

Monitoring Link Faults in Nonlinear Diffusively Coupled Networks

Miel Sharf , Member, IEEE, and Daniel Zelazo , Senior Member, IEEE

Abstract—Fault detection and isolation is an area of engineering dealing with designing online protocols for systems that allow one to identify the existence of faults, pinpoint their exact location, and overcome them. We consider multi-agent systems, where faults correspond to the disappearance of links in the underlying graph, simulating communication failures between agents. We study the case in which the agents and controllers are maximal equilibrium-independent passive, and use the known connection between steady states of these multi-agent systems and network optimization theory. We first study asymptotic methods of differentiating the faultless system from its faulty versions by studying their steady-state outputs. We explain how to apply the asymptotic differentiation to detect and isolate communication faults, with graph-theoretic guarantees on the number of faults that can be isolated, assuming the existence of a “convergence assertion protocol,” a data-driven method of asserting that a multi-agent system converges to a conjectured limit. We then construct two data-based convergence assertion protocols. We demonstrate our results by a case study.

Index Terms—Fault tolerant Control, fault detection and isolation, multi-agent systems, networked control systems, nonlinear control systems, passivity.

I. INTRODUCTION

MULTI-AGENT systems (MASs) have been widely studied in recent years, presenting both a variety of applications and a deep theoretical framework [1]. One of the deepest concerns when considering applications of MASs is communication failures, which can drive the agents to act poorly or fail their task altogether. These communication failures, which we term *network failures*, can either be accidental or planned by an adversary. There is a need of detecting network faults and dealing with them in real time for the network to be secure.

Fault detection and isolation (FDI) for MASs usually deals with faults in one of the agents (see, e.g., [3] and references

Manuscript received July 14, 2020; revised December 6, 2020 and February 18, 2021; accepted June 20, 2021. Date of publication July 7, 2021; date of current version May 31, 2022. Recommended by Associate Editor P. Rapisarda. (Corresponding author: Miel Sharf.)

Miel Sharf is with the Division of Decision and Control Systems, KTH Royal Institute of Technology and Digital Futures, 114 28 Stockholm, Sweden (e-mail: msharf@tx.technion.ac.il).

Daniel Zelazo is with the Faculty of Aerospace Engineering, Technion—Israel Institute of Technology, Haifa 3200003, Israel (e-mail: dzelazo@technion.ac.il).

Color versions of one or more figures in this article are available at <https://doi.org/10.1109/TAC.2021.3095258>.

Digital Object Identifier 10.1109/TAC.2021.3095258

therein). The possibility of faults in the communication links was first studied in [4] using the notion of structural controllability, which was later used in [5] to solve the problem of leader localization. The problem of network FDI, i.e., detection and isolation of network faults, was studied primarily for linear time-invariant (LTI) systems with a known model. In [6] and [7], Rahimian and Preciado used jump discontinuities in the derivative of the output to detect topological changes in the network. Tools from switching systems theory, namely, mode observability, was used in [8] for network FDI. Combinatorial tools were used in [9] and [10] to solve the FDI problem for consensus-seeking networks. Recently, Zhang *et al.* [11] have proposed a network FDI method, which allows an uncertainty in the model, but is restricted for networks with LTI systems. A compressed-sensing-based approach was presented in [12], providing an FDI algorithm for nonlinear power networks with measurement noise. A related problem in which one tries to distinguish between MASs with identical agents but different communication graphs was studied in [13]–[15], from which only [14] also deals with nonlinear agents.

We aim at a network FDI scheme applicable also for nonlinear systems by relying on another concept widespread in MASs, namely, passivity. Passivity was first used to address faults by Yang *et al.* [16] for control-affine systems, although only fault tolerance is addressed, and no synthesis procedures are suggested. Later works addressed FDI for a single nonlinear agent [16]–[18], [37]. To the extent of our knowledge, passivity has not been previously used for network FDI schemes, and no other works consider networks with nonlinear components.

Passivity theory is a cornerstone of the theoretical framework of networks of dynamical systems [19]. It allows for the analysis of MASs to be decoupled into two separate layers: the dynamic system layer and the information exchange layer. Passivity theory was first used to study the convergence properties of network systems in [20]. Many variations and extensions of passivity have been applied in different aspects of MASs. For example, the related concepts of incremental passivity or relaxed co-coercivity and have been used to study synchronization problems [2], [38], port-Hamiltonian systems [21], and robotic systems [22].

One prominent variant is maximal equilibrium-independent passivity (MEIP), which was applied in [23] in order to reinterpret the analysis problem for an MAS as a pair of network optimization problems. Network optimization is a branch of optimization theory dealing with optimization of functions defined over graphs [24]. The main result of [23] showed that

the asymptotic behavior of these MASs is (*inverse*) *optimal* with respect to a family of network optimization problems. In fact, the steady-state input–output signals of both the agents and the controllers comprising the MAS can be associated with the optimization variables of either an *optimal flow* or an *optimal potential* problem; these are the two canonical dual network optimization problems described in [24]. The results of [23] were used in [25] and [26] to solve the synthesis problem for MASs and were further used in [14] to solve the network identification problem.

We aim to use the network optimization framework of [23], [25], and [26] for analysis and synthesis of MASs in order to provide a strategy for detecting and isolating network faults. We also consider adversarial games regarding communication faults. We strive to give graph-theoretic-based results, showing that network FDI can be done for any MEIP MAS, so long that the graph \mathcal{G} satisfies certain conditions. We show that if the graph \mathcal{G} is “connected enough,” we can solve the network FDI problem. Namely, if \mathcal{G} is 2-connected, then detecting the existence of any number of faults is possible, and if \mathcal{G} is k -connected with $k > 2$, we can isolate up to $k - 2$ faults.

The rest of this article is organized as follows. Section II surveys the relevant parts of the network optimization framework. Section III presents the problem formulation of this work and states the assumptions used throughout this article. Section IV presents the first technical tool used for building the network fault detection schemes, namely, edge-indication vectors, and shows how to construct them. Section V uses edge-indication vectors to design network FDI schemes, as well as strategies for adversarial games, assuming the existence of a “convergence assertion protocol,” a data-driven method of asserting that a given MAS converges to a conjectured limit. Section VI studies these convergence assertion protocols, prescribing two approaches for constructing them. Finally, we present simulations demonstrating the constructed algorithms.

Notations: We use basic notions from algebraic graph theory [27]. An undirected graph $\mathcal{G} = (\mathbb{V}, \mathbb{E})$ consists of a finite set of vertices \mathbb{V} and edges $\mathbb{E} \subset \mathbb{V} \times \mathbb{V}$. We denote by $e = \{i, j\} \in \mathbb{E}$ the edge that has ends i and j in \mathbb{V} . For each edge e , we pick an arbitrary orientation and denote $e = (i, j)$. The incidence matrix of \mathcal{G} , denoted $\mathcal{E}_{\mathcal{G}} \in \mathbb{R}^{|\mathbb{E}| \times |\mathbb{V}|}$, is defined such that for edge $e = (i, j) \in \mathbb{E}$, $[\mathcal{E}_{\mathcal{G}}]_{ie} = +1$, $[\mathcal{E}_{\mathcal{G}}]_{je} = -1$, and $[\mathcal{E}_{\mathcal{G}}]_{\ell e} = 0$ for $\ell \neq i, j$. We also use simple notions from graph theory [28]. A path is a sequence of distinct nodes v_1, v_2, \dots, v_n such that $\{v_i, v_{i+1}\} \in \mathbb{E}$ for all i . A cycle is a path v_1, \dots, v_n for which $v_1 = v_n$. A cycle $v_1, v_2, \dots, v_{n-1}, v_1$ is called simple if $v_i \neq v_j$ for all $i \neq j$. A collection of paths is called vertex disjoint if no two share a node, except possibly for their first and last nodes. Furthermore, for a linear map $T : U \rightarrow V$ between vector spaces, we denote the kernel of T by $\ker T$.

II. NETWORK OPTIMIZATION AND MEIP MULTI-AGENT SYSTEMS

This section briefs over the main results we need about the connection between MASs and network optimization, originally presented in [23], [25] and [26].

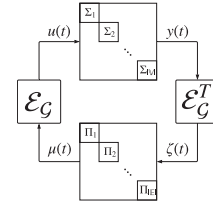


Fig. 1. Block diagram of the closed loop.

A. Diffusively Coupled Networks and Their Steady States

Consider a collection of single-input single-output (SISO) agents interacting over a network $\mathcal{G} = (\mathbb{V}, \mathbb{E})$. The agents $\{\Sigma_i\}_{i \in \mathbb{V}}$ and the controllers $\{\Pi_e\}_{e \in \mathbb{E}}$ are governed by:

$$\Sigma_i : \begin{cases} \dot{x}_i = f_i(x_i, u_i) \\ y_i = h_i(x_i) \end{cases}, \quad \Pi_e : \begin{cases} \dot{\eta}_e = \phi_e(\eta_e, \zeta_e) \\ \mu_e = \psi_e(\eta_e, \zeta_e). \end{cases} \quad (1)$$

We consider stacked vectors of the form $u = [u_1^\top, \dots, u_{|\mathbb{V}|}^\top]^\top$ and similarly for x, y, ζ, η , and μ . The agents and controllers are coupled by defining the controller input as $\zeta = \mathcal{E}_{\mathcal{G}}^\top y$ and the control input as $u = -\mathcal{E}_{\mathcal{G}} \mu$. This closed-loop system is called a *diffusively coupled network* and is denoted by $(\mathcal{G}, \Sigma, \Pi)$. Its structure is illustrated in Fig. 1. We wish to study the steady states of the closed loop. Suppose that (u, y, ζ, μ) is a steady state of $(\mathcal{G}, \Sigma, \Pi)$. For every $i \in \mathbb{V}, e \in \mathbb{E}$, (u_i, y_i) is a steady-state input–output pair of the i th agent, and (ζ_e, μ_e) is a steady-state pair of the e th controller. This motivates the following definition, originally from [23].

Definition 1: The steady-state input–output relation k of a dynamical system is the collection of all steady-state input–output pairs of the system. Given a steady-state input u and a steady-state output y , we define

$$k(u) = \{y : (u, y) \in k\}, \quad k^{-1}(y) = \{u : (u, y) \in k\}.$$

Let k_i be the steady-state input–output relation for the i th agent, γ_e be the steady-state input–output relation for the e th controller, and k and γ be their stacked versions. The closed-loop steady-state (u, y, ζ, μ) has to satisfy $y \in k(u)$, $\zeta = \mathcal{E}_{\mathcal{G}}^\top y$, $\mu \in \gamma(\zeta)$, and $u = -\mathcal{E}_{\mathcal{G}} \mu$. By a simple manipulation, one can show that y is a closed-loop steady state for the agents’ output if and only if $0 \in k^{-1}(y) + \mathcal{E}_{\mathcal{G}} \gamma(\mathcal{E}_{\mathcal{G}}^\top y)$ [26].

B. MEIP Systems and Closed-Loop Convergence

The convergence of the diffusively coupled network $(\mathcal{G}, \Sigma, \Pi)$ can be assured using passivity. We first recall the classic definition of (shifted) passivity.

Definition 2 (Passivity [29]): Let Υ be a SISO system with input $u(t)$, output $y(t)$, and state $x(t)$, and let (u, y) be a steady-state input–output pair of the system. For a differentiable function $S = S(x)$ and a number $\rho > 0$, we consider the inequality $\frac{d}{dt} S(x) \leq -\rho \|y(t) - y\|^2 + (y(t) - y)(u(t) - u)$. We say Υ is passive (w.r.t. (u, y)) if there exists a semidefinite storage function $S(x)$ and $\rho \geq 0$ such that the inequality holds for any trajectory. Also, we say the system is output-strictly passive

(w.r.t. (u, y)) if the same condition holds for some $\rho > 0$. The largest number ρ for which the condition holds is called the (output) passivity index w.r.t. (u, y) .

Passivity was first used for diffusively coupled networks in [20]. It is known that if (u, y, ζ, μ) is an equilibrium of the network, and the agents and controllers are passive with respect to (u_i, y_i) and (ζ_e, μ_e) , then the network converges to said equilibrium. The existence of an equilibrium for the closed-loop network can be proved using network optimization tools under certain monotonicity assumptions on the steady-state input–output relation of the agents and controllers [23], [26], namely, under the following variant of passivity.

Definition 3 (MEIP [23]): Consider the SISO dynamical system Υ governed by $\dot{x} = f(x, u), y = h(x, u)$ with input–output relation r . The system Υ is (output-strictly) MEIP if the following conditions hold.

- i) The system Υ is (output-strictly) passive with respect to any steady-state pair $(u, y) \in r$.
- ii) The steady-state input–output relation r is maximally monotone. That is, if $(u_1, y_1), (u_2, y_2) \in r$, then $(u_1 - u_2)(y_1 - y_2) \geq 0$, and r is not contained in any larger monotone relation [24].

The passivity index of the system Υ is defined as $\min_{y \in r(u)} \rho_{u, y}$, where $\rho_{u, y}$ is the passivity index with respect to (u, y) .

Such systems include single integrators, gradient systems, port-Hamiltonian systems on graphs, and others (see [23] and [26] for more examples). In this work, we often consider networks of control-affine systems. We provide a sufficient condition for a control-affine system to be MEIP.

Theorem 1: Let Σ be the SISO system of the form $\dot{x} = -f(x) + q(x)u, y = h(x)$. Suppose that $q(x)$ is positive for all x , that h is strictly monotone C^1 ascending, and that f/q is C^1 and monotone ascending.

- 1) A pair (u, y) is a steady-state input–output pair for Σ if and only if there exists some $x \in \mathbb{R}$ such that $u = f(x)/q(x)$ and $y = h(x)$.
- 2) For any $x \in \mathbb{R}$, the function $S(x) = \int_x^x \frac{h(\sigma) - h(x)}{q(\sigma)} d\sigma$ is a storage function for the steady-state input–output pair $u = f(x)/q(x)$ and $y = h(x)$.
- 3) The function $S(x)$ proves that Σ is passive w.r.t. (u, y) with passivity index $\rho = \inf_{x \in \mathbb{R}} \frac{\frac{f(x) - f(x)}{q(x)} - \frac{f(x)}{q(x)}}{h(x) - h(x)} \geq 0$.
- 4) If either $\lim_{|t| \rightarrow \infty} |f(t)/q(t)| = \infty$ or $\lim_{|t| \rightarrow \infty} |h(t)| = \infty$, then the system is MEIP.
- 5) If the derivative of h is always positive, then the inverse steady-state relation k^{-1} is differentiable.

Proof: The first, second, and fourth parts are proved in [30, Proposition 1]. As for the third, we note that

$$\begin{aligned} \dot{S} &= \frac{(h(x) - h(x))}{q(x)} \dot{x} = \frac{(h(x) - h(x))}{q(x)} (-f(x) + q(x)u) = \\ &= (h(x) - y)(u - u) - (h(x) - h(x)) \left(\frac{f(x)}{q(x)} - \frac{f(x)}{q(x)} \right) \\ &\leq (y - y)(u - u) - \rho(y - y)^2. \end{aligned}$$

We note that $\rho \geq 0$ as $(h(x) - h(x)) \left(\frac{f(x)}{q(x)} - \frac{f(x)}{q(x)} \right) \geq 0$ by monotonicity. Moreover, $\int_x^x \frac{h(\sigma) - h(x)}{q(\sigma)} d\sigma \geq 0$, with strict inequality whenever $x \neq x$, as h is strictly monotone and $q(x) > 0$. Thus, S is a C^1 storage function, and hence the system is passive with passivity index $\rho \geq 0$ with respect to the steady-state input–output pair (u, y) . Finally, as h is strictly monotone, the inverse h^{-1} can be defined. Thus, the inverse steady-state relation k^{-1} is given by $k^{-1}(y) = \frac{f(h^{-1}(y))}{q(h^{-1}(y))}$, which is differentiable by the inverse function theorem. ■

As we previously remarked, MEIP can be used to prove existence of a closed-loop equilibrium for networks.

Theorem 2 (see [23] and [25]): Consider the network $(\mathcal{G}, \Sigma, \Pi)$. Assume the agents Σ_i are MEIP, and the controllers Π_e are output-strictly MEIP (or vice versa). Then, the signals $u(t), y(t), \zeta(t)$, and $\mu(t)$ of the closed-loop converge to steady-state values u, y, ζ , and μ , where $0 \in k^{-1}(y) + \mathcal{E}_{\mathcal{G}} \gamma(\mathcal{E}_{\mathcal{G}}^\top y)$.

C. Synthesis Problem for MEIP Multi-Agent Systems

The synthesis problem of MAS with MEIP agents has been studied in [25] and [26]. The problem deals with synthesizing controllers $\{\Pi_e\}$ forcing the closed-loop network to converge to some desired steady-state output y^* , when the agents Σ and the graph \mathcal{G} are known. We cite the following results from [26].

Theorem 3 (see [26]): Let Σ be any MEIP agents and let \mathcal{G} be any graph. Let $y^* \in \mathbb{R}^{|\mathcal{V}|}$ be any desired steady-state output. Then, there exists a solution to the synthesis problem (i.e., a realization of the controllers Π) with desired output y^* for which the controllers are output-strictly MEIP.

Remark 1: The paper [26] depicts many possible solutions to the synthesis problem with output-strictly MEIP controllers. It is shown that one can always solve the problem using affine controllers. Another suggested solution is an augmentation of any preferred output-strictly MEIP controller using a constant exogenous input. In practice, we will usually opt for the augmentation procedure when using the theorem as a tool for synthesis, as many real-world networks are already equipped with some given controllers.

III. PROBLEM FORMULATION

This section presents the problem we aim to solve and states the assumptions we make to tackle it. We consider a diffusively coupled network of the form $\mathcal{N}_{\mathcal{G}} = (\mathcal{G}, \{\Sigma_i\}_{i \in \mathcal{V}}, \{\Pi_e\}_{e \in \mathbb{E}_{\mathcal{G}}})$, where $\mathcal{G} = (\mathcal{V}, \mathbb{E}_{\mathcal{G}})$ is the interaction graph, Σ_i are the agents, and Π_e are the edge controllers. For any subgraph $\mathcal{H} = (\mathcal{V}, \mathbb{E}_{\mathcal{H}})$ of \mathcal{G} , we can consider another diffusively coupled network $\mathcal{N}_{\mathcal{H}} = (\mathcal{H}, \{\Sigma_i\}_{i \in \mathcal{V}}, \{\Pi_e\}_{e \in \mathbb{E}_{\mathcal{H}}})$. We can think of $\mathcal{N}_{\mathcal{H}}$ as a faulty version of $\mathcal{N}_{\mathcal{G}}$, in which the edge controllers corresponding to the edges $\mathbb{E}_{\mathcal{G}} \setminus \mathbb{E}_{\mathcal{H}}$ have malfunctioned and stopped working. Edges can fault mid-run, but we assume that once an edge has malfunctioned, it remains faulty for the remainder to the run. If we let \mathfrak{G} be the collection of all nonempty subgraphs of \mathcal{G} , then one can think of the closed-loop diffusively coupled network as a switched system, where the switching signal $\varsigma : [0, \infty) \rightarrow \mathfrak{G}$

designates the functioning edges at each time instant. The assumption that faulty edges remain faulty throughout the run can be described using the switching signal ς . Namely, we require that the switching signal ς is *nonincreasing*, in the sense that for all times $t_1 < t_2$, $\varsigma(t_2)$ is a subgraph of $\varsigma(t_1)$. We denote the number of faulty edges at time t by $\zeta(t)$.

Consider a collection of agents $\{\Sigma_i\}$ and a graph \mathcal{G} . Fix a constant vector $y^* \in \mathbb{R}^{|\mathcal{V}|}$. Our goal is to design a control scheme, for which the closed-loop network will converge to the steady-state output y^* . In the absence of faults, we can solve the synthesis problem as in Theorem 3. However, designing controllers while ignoring faults might prevent the system from achieving the control goal. Thus, we also seek for a *fault monitoring system*, monitoring the agents and controllers, attempting to identify aberrant behavior. When it does, it declares a fault [31]. We now formulate the problems of network FDI:

Problem 1 (Network fault detection): Let $\{\Sigma_i\}_{i \in \mathcal{V}}$ be a set of agents, \mathcal{G} be a graph, y^* be any desired steady-state output, and let $\varsigma(t)$ be any nonincreasing switching signal. Find edge controllers $\{\Pi_e\}_{e \in \mathbb{E}_{\mathcal{G}}}$ and a fault monitoring system such that:

- i) if no faults occur, i.e., $\varsigma(t) = \mathcal{G}$, $\forall t$, then the closed-loop network converges to the steady-state output y^* , and the fault monitoring system never declares a fault;
- ii) if faults do occur, i.e. $\exists t, \varsigma(t) \neq \mathcal{G}$, then the fault monitoring system declares a fault.

Problem 2 (Network FDI): Let $\{\Sigma_i\}_{i \in \mathcal{V}}$ be a set of agents, \mathcal{G} be a graph, and y^* be any desired steady-state output. Given some $r < |\mathbb{E}_{\mathcal{G}}|$, find a synthesis for the edge controllers such that for any monotone nonincreasing switching signal ς such that $\zeta(t) \leq r$, $\forall t$, the closed-loop diffusively coupled network converges to the steady-state output y^* , i.e., the effect of up to r faults can be isolated from the network.¹

A. Assumptions

We now state the assumptions used throughout the work. For the remainder of this work, we fix the agents $\{\Sigma_i\}$ and make the following assumption.

Assumption 1: The agent dynamics $\{\Sigma_i\}$ are MEIP, and the chosen controller dynamics $\{\Pi_e\}$ are output-strictly MEIP (or vice versa). Moreover, the relations k_i^{-1} and γ_e are C^1 functions. Furthermore, the derivative $\frac{dk_i^{-1}}{dy_i}$ is positive at any $y_i \in \mathbb{R}$.

The passivity assumption assures that all the systems $\mathcal{N}_{\mathcal{H}}$ will globally asymptotically converge to some limit. The added smoothness assumptions, together with the positive derivative assumption, are technical assumptions we need to apply tools from manifold theory used later. The passivity assumption allows the consideration of, e.g., port-Hamiltonian systems and gradient-descent systems [23]. Moreover, if a system satisfies any dissipation inequality with respect to all equilibria, one can use output feedback and input feedthrough to force MEIP [32]. Theorem 1 shows the smoothness assumption holds for many control-affine systems. Moreover, it can be shown by definition that if Σ_i is output-strictly MEIP with passivity index ρ , then $\frac{dk_i^{-1}}{dy_i} \geq \rho > 0$ whenever k_i^{-1} is differentiable. Furthermore, the

smoothness assumption can be relaxed by allowing k_i^{-1} and γ_e to not be differentiable at finitely many points. The arguments presented below still hold, but require heavier tools from measure theory, so we avoid them for clarity of presentation.

In some cases, we need to sense the state of the system, including the state of the controllers. Sometimes, the control model is such that the controller state has a physical meaning that can be measured even for nonconnected agents. For example, in the traffic control model in [33], the state η_e is the relative position between two vehicles. However, the controller state of some systems might not have a physical meaning. For example, consider a collection of robots trying to synchronize their positions, where the output $y(t)$ is the position of each robot and the edge controllers are proportional-integral controllers. The controller state $\eta(t)$ has no physical meaning and, thus, cannot be defined for nonconnected agents. Some of the techniques developed later require us to be able to sense the state of the system, including the controllers' states, and we will then make the following assumption.

Assumption 2: The controllers Π_e are static nonlinearities given by the functions g_e , i.e., $\mu_e(t) = g_e(\zeta_e(t))$ for all t . In this case, the steady-state relation γ_e is equal to the function g_e , and the closed-loop system is $\dot{x} = f(x, -\mathcal{E}_{\mathcal{G}}g(\mathcal{E}_{\mathcal{G}}^T h(x)))$, or equivalently, $\dot{x}_i = f_i(x_i, \sum_{e=\{i,j\}} g_e(h_j(x_j) - h_i(x_i)))$.

In one of the methods below, we will want to have a clear relationship between the measurements $h_i(x_i)$ and the storage functions $S_i(x_i)$. To achieve this, we follow Theorem 1 and assume that the agents are control-affine.

Assumption 3: Assumption 2 holds, and the agents have the form $\dot{x}_i = -f_i(x_i) + q_i(x_i)u_i$; $y_i = h_i(x_i)$. Thus, the closed-loop system is governed by

$$\dot{x}_i = -f_i(x_i) + q_i(x_i) \sum_{e=\{i,j\}} g_e(h_j(x_j) - h_i(x_i)). \quad (2)$$

It should be noted that the MEIP property for the static controllers g_e reduces to monotonicity of the functions g_e .

In the next section, we start heading toward a solution to Problems 1 and 2 by exhibiting a method for asymptotically differentiating between the nominal dynamical system $\mathcal{N}_{\mathcal{G}}$ and its faulty versions $\mathcal{N}_{\mathcal{H}}$. Later, we show how this asymptotic differentiation can induce a finite-time differentiation of the systems.

IV. ASYMPTOTIC DIFFERENTIATION BETWEEN NETWORKS

In this section, we develop the notion of edge-indication vectors, which will be used for network fault detection later. In [14], the notion of indication vectors was first developed. These are constant exogenous inputs used to drive the closed-loop system, chosen appropriately to give different steady-state limits for systems with identical agents and controllers, but different underlying graphs. The idea of using constant exogenous inputs to drive the system into favorable steady-state outputs was also used in [34] to give a network reconstruction algorithm with optimal time complexity, although it considers sets of multiple constant exogenous inputs applied in succession. Here, we opt for a slightly different strategy.

¹Some authors refer to fault isolation in this case as identifying the faulty links, which is achieved by the algorithms in Section V-C as a side effect.

In [14] and [34], the problem of network reconstruction was considered, in which we cannot affect the agents, controllers, or the underlying graph. In network FDI, we are doing synthesis, so we can manipulate the controllers and (in most cases) the underlying network. For that reason, we opt for a slightly different idea, in which we add a constant exogenous signal to the *output of the controllers*, that is, we consider $u(t) = -\mathcal{E}_G(\mu(t) + w)$. A system implementing this control law is said to have the interaction protocol (Π, w) . Analogously to the notion of indication vectors, we desire that networks with identical agents and controllers, but different underlying graphs, will be forced to converge to different steady-state outputs. We can then monitor the output y of the system and use it to detect changes in the underlying graph, i.e., network faults. For that, we first compute the steady-state limit for these systems $(\mathcal{G}, \Sigma, (\Pi, w))$.

Proposition 1: Consider a diffusively coupled network $\mathcal{N}_{\mathcal{H}} = (\mathcal{H}, \Sigma, \Pi)$ satisfying Assumption 1. Suppose that $w \in \mathbb{R}^{|\mathbb{E}_{\mathcal{H}}|}$ is any constant signal added to the controller output, i.e., the loop is closed as $u(t) = -\mathcal{E}_{\mathcal{H}}(\mu(t) + w)$. Then, y is a steady-state output of the closed-loop system if and only if

$$k^{-1}(y) + \mathcal{E}_{\mathcal{H}}\gamma(\mathcal{E}_{\mathcal{H}}^{\top}y) = -\mathcal{E}_{\mathcal{H}}w. \quad (3)$$

Proof: The proof follows from the discussion after Definition 1, as the new steady-state relation for the controllers is given as $\tilde{\gamma}(\zeta) = \gamma(\zeta) + w$. ■

In our case, the constant signal w will be in $\mathbb{R}^{|\mathbb{E}_{\mathcal{G}}|}$, as we determine the exogenous controller output on each edge of \mathcal{G} . If one then considers the system $\mathcal{N}_{\mathcal{H}}$ for some $\mathcal{H} \in \mathfrak{G}$, then the exogenous controller output will be different from w , as it will only have entries of w corresponding to edges in \mathcal{H} . To formulate this, take any graph $\mathcal{H} \in \mathfrak{G}$, and let $P_{\mathcal{H}}$ be the linear map $\mathbb{R}^{|\mathbb{E}_{\mathcal{G}}|} \rightarrow \mathbb{R}^{|\mathbb{E}_{\mathcal{H}}|}$ removing entries corresponding to edges absent from \mathcal{H} . In other words, this is an $\mathbb{R}^{|\mathbb{E}_{\mathcal{H}}| \times |\mathbb{E}_{\mathcal{G}}|}$ matrix with entries in $\{0, 1\}$, whose rows are the rows of the identity matrix $\text{Id} \in \mathbb{R}^{|\mathbb{E}_{\mathcal{G}}| \times |\mathbb{E}_{\mathcal{G}}|}$ corresponding to the edges of \mathcal{H} .

We can now define the notion of edge-indication vectors.

Definition 4: Let $(\mathcal{G}, \Sigma, \Pi)$ be a network satisfying Assumption 1. Let $w \in \mathbb{R}^{|\mathbb{E}_{\mathcal{G}}|}$ by any vector, and for any graph $\mathcal{H} \in \mathfrak{G}$, we denote the solution of (3) with underlying graph \mathcal{H} and exogenous input $P_{\mathcal{H}}w$ by $y_{\mathcal{H}}$.

- 1) The vector w is called a $(\mathfrak{G}, \mathcal{H})$ -edge-indication vector if for any $\mathcal{H}' \in \mathfrak{G}$ such that $\mathcal{H}' \neq \mathcal{H}$, we have $y_{\mathcal{H}} \neq y_{\mathcal{H}'}$.
- 2) The vector w is called a \mathfrak{G} -edge-indication vector if for any two graphs $\mathcal{H}_1 \neq \mathcal{H}_2$ in \mathfrak{G} , $y_{\mathcal{H}_1} \neq y_{\mathcal{H}_2}$.

Note 1: An edge-indication vector is a bias chosen on each edge in \mathcal{G} . This bias can be programmed into the controllers and nodes and need not be changed nor computed online. In this light, for any $w \in \mathbb{R}^{|\mathbb{E}_{\mathcal{G}}|}$, (3) transforms into

$$k^{-1}(y) + \mathcal{E}_{\mathcal{H}}\gamma(\mathcal{E}_{\mathcal{H}}^{\top}y) = -\mathcal{E}_{\mathcal{H}}P_{\mathcal{H}}w. \quad (4)$$

We wish to find a \mathfrak{G} -edge-indication vector for given agents and controllers or at least a $(\mathfrak{G}, \mathcal{G})$ -edge-indication vector. As in [14], we use randomization. We claim that random vectors are \mathfrak{G} -edge-indication vectors with probability 1.

Theorem 4: Let \mathbb{P} be any absolutely continuous² probability measure on $\mathbb{R}^{|\mathbb{E}_{\mathcal{G}}|}$. Let w be a vector sampled according to \mathbb{P} . Then, $\mathbb{P}(w \text{ is a } \mathfrak{G}\text{-edge-indication vector}) = 1$.

Proof: By definition, w is not a \mathfrak{G} -edge-indication vector if and only if there are $\mathcal{G}_1, \mathcal{G}_2 \in \mathfrak{G}$ such that the same vector y solves (4) for both graphs. We show that for any $\mathcal{G}_1, \mathcal{G}_2 \in \mathfrak{G}$, the probability that the two equations share a solution is zero.

Let n be the number of vertices in \mathcal{G} . For each graph $\mathcal{H} \in \mathfrak{G}$, define a function $F_{\mathcal{H}}: \mathbb{R}^n \times \mathbb{R}^{|\mathbb{E}_{\mathcal{G}}|} \rightarrow \mathbb{R}^n$ by $F_{\mathcal{H}}(y, w) = k^{-1}(y) + \mathcal{E}_{\mathcal{H}}\gamma(\mathcal{E}_{\mathcal{H}}^{\top}y) + \mathcal{E}_{\mathcal{H}}P_{\mathcal{H}}w$. The set of steady-state exogenous input and output pairs for the system $\mathcal{N}_{\mathcal{H}}$ is given by the set $\mathcal{A}_{\mathcal{H}} = \{(y, w) : F_{\mathcal{H}}(y, w) = 0\}$. We note that the differential $dF_{\mathcal{H}}$ always has rank n . Indeed, it can be written as $[\nabla k^{-1}(y) + \mathcal{E}_{\mathcal{H}}\nabla\gamma(\mathcal{E}_{\mathcal{H}}^{\top}y)\mathcal{E}_{\mathcal{H}}^{\top}, \mathcal{E}_{\mathcal{H}}P_{\mathcal{H}}]$, where $\nabla\gamma(\mathcal{E}_{\mathcal{H}}^{\top}y) \in \mathbb{R}^{|\mathbb{E}_{\mathcal{H}}| \times n}$. By Assumption 1, the first matrix, of size $n \times n$, is positive definite as a sum of a positive-definite matrix and a positive-semidefinite matrix, hence invertible. Thus, by the implicit function theorem, $\mathcal{A}_{\mathcal{H}}$ is a manifold of dimension $|\mathbb{E}_{\mathcal{G}}|$. Moreover, by Assumption 1, for any w , there is a unique y such that (3) is satisfied. Thus, \mathbb{P} gives rise to an absolutely continuous³ probability measure on each manifold $\mathcal{A}_{\mathcal{H}}$.⁴ Hence, it suffices to show that for any $\mathcal{G}_1 \neq \mathcal{G}_2$, the intersection $\mathcal{A}_{\mathcal{G}_1} \cap \mathcal{A}_{\mathcal{G}_2}$ has dimension $\leq |\mathbb{E}_{\mathcal{G}}| - 1$.

To show this, we take any point $(y, w) \in \mathcal{A}_{\mathcal{G}_1} \cap \mathcal{A}_{\mathcal{G}_2}$. As both $\mathcal{A}_{\mathcal{G}_1}$ and $\mathcal{A}_{\mathcal{G}_2}$ are of dimension $|\mathbb{E}_{\mathcal{G}}|$, it suffices to show they do not have the same tangent space at (y, w) . The tangent space of the manifold $\mathcal{A}_{\mathcal{H}}$ is given by the kernel of the differential $dF_{\mathcal{H}}(y, w)$, so we show that if $\mathcal{G}_1 \neq \mathcal{G}_2$, the kernels $\ker dF_{\mathcal{G}_1}$ and $\ker dF_{\mathcal{G}_2}$ are different at (y, w) . As $\mathcal{G}_1 \neq \mathcal{G}_2$, we can find an edge existing in one of the graphs and not the other. Assume the edge e exists in \mathcal{G}_1 but not in \mathcal{G}_2 , and let $v = (0, \mathbb{1}_e)$, where $\mathbb{1}_e$ is the vector in $\mathbb{R}^{|\mathbb{E}_{\mathcal{G}}|}$ with all entries zero, except for the e th entry, which is equal to 1. Then, $v \in \ker dF_{\mathcal{H}}$ if and only if $\mathbb{1}_e \in \ker(\mathcal{E}_{\mathcal{H}}P_{\mathcal{H}})$. Note that $\mathbb{1}_e \notin \ker(\mathcal{E}_{\mathcal{G}_1}P_{\mathcal{G}_1})$, as $P_{\mathcal{G}_1}\mathbb{1}_e = \mathbb{1}_e$, and thus, $\mathcal{E}_{\mathcal{G}_1}P_{\mathcal{G}_1}\mathbb{1}_e = \mathcal{E}_{\mathcal{G}_1}\mathbb{1}_e \neq 0$. Moreover, $\mathbb{1}_e \in \ker(\mathcal{E}_{\mathcal{G}_2}P_{\mathcal{G}_2})$, as $P_{\mathcal{G}_2}\mathbb{1}_e = 0$, so $\ker dF_{\mathcal{G}_1} \neq \ker dF_{\mathcal{G}_2}$ at (y, w) . Thus, $\mathcal{A}_{\mathcal{G}_1} \cap \mathcal{A}_{\mathcal{G}_2}$ is of dimension $\leq |\mathbb{E}_{\mathcal{G}}| - 1$, and is a zero-measure set inside both $\mathcal{A}_{\mathcal{G}_1}, \mathcal{A}_{\mathcal{G}_2}$. ■

Theorem 4 presents a way to choose a \mathfrak{G} -edge-indication vector, but does not deal with the control goal. One could satisfy the control goal by using Theorem 3 to solve the synthesis problem for the original graph \mathcal{G} , but we cannot assure we get an edge-indication vector. Note that any $w \in \ker \mathcal{E}_{\mathcal{G}}P_{\mathcal{G}}$ gives a solution of (4) identical to the solution for $w = 0$, so choosing an exogenous control input in $\ker \mathcal{E}_{\mathcal{G}}P_{\mathcal{G}}$ does not change the steady-state output of the system $\mathcal{N}_{\mathcal{G}}$. However, it does change the steady-state output of all other systems $\mathcal{N}_{\mathcal{H}}$. This suggests to search for an edge-indication vector in $\ker \mathcal{E}_{\mathcal{G}}P_{\mathcal{G}}$, which is possible if \mathcal{G} is “sufficiently connected,” defined below in an exact manner.

Proposition 2 (Menger’s theorem [28]): Let \mathcal{G} be any connected graph. The following conditions are equivalent.

²Unless stated otherwise, absolute continuity is with respect to the Lebesgue measure.

³With respect to the $|\mathbb{E}_{\mathcal{G}}|$ -dimensional Hausdorff measure, or equivalently, with respect to the standard Riemannian volume form on $\mathcal{A}_{\mathcal{H}}$.

⁴As the push-forward measure of \mathbb{P} under the map $w \mapsto (\varphi(w), w)$, where φ is the local map $w \mapsto y$ given by the implicit function theorem.

- 1) Between every two nodes, there are k vertex-disjoint simple paths.
- 2) For any $k - 1$ vertices $v_1, \dots, v_{k-1} \in \mathbb{V}$, the graph $\mathcal{G} - \{v_1, \dots, v_{k-1}\}$ is connected.

A graph satisfying these conditions is called k -connected.

We will take special interest in 2-connected graphs. Specifically, we can state the following theorem about edge-indication vectors in $\ker \mathcal{E}_{\mathcal{G}} P_{\mathcal{G}}$.

Theorem 5: Let \mathbb{P} be any absolutely continuous probability distribution on $\ker \mathcal{E}_{\mathcal{H}} P_{\mathcal{H}}$, where \mathcal{H} is 2-connected. Suppose furthermore that w is a vector sampled according to \mathbb{P} . Then, $\mathbb{P}(w \text{ is a } (\mathcal{G}, \mathcal{H})\text{-edge-indication vector}) = 1$.

We first need to state and prove a lemma.

Lemma 1: Let \mathcal{H} be a 2-connected graph. Suppose that we color the edges of \mathcal{H} in two colors: red and blue. If not all edges have the same color, then there is a simple cycle in \mathcal{H} with both red and blue edges.

Proof: Suppose, heading toward contradiction, that any simple cycle in \mathcal{H} is monochromatic. We claim that for each vertex x , all the edges touching x have the same color. Indeed, take any vertex x , and suppose that there are two neighbors v_1 and v_2 of x such that the edge $\{x, v_1\}$ is blue and the edge $\{x, v_2\}$ is red. We note that $v_1 \rightarrow x \rightarrow v_2$ is a path from v_1 to v_2 , meaning that there is another path from v_1 to v_2 , which does not pass through x . Adding both edges to the path yields a simple cycle with edges of both colors, as $\{x, v_1\}$ is blue and $\{x, v_2\}$ is red. Thus, every node touches edges of a single color.

Let \mathbb{V}_{red} be the set of nodes touching red edges, and \mathbb{V}_{blue} be the set of nodes touching blue edges. We know that \mathbb{V}_{red} and \mathbb{V}_{blue} do not intersect. Moreover, if we had an edge between \mathbb{V}_{red} and \mathbb{V}_{blue} , it had a color. Assume, without loss of generality, that it is blue, so some vertex in \mathbb{V}_{red} touches a blue edge, which is impossible. Thus, there are no edges between \mathbb{V}_{red} and \mathbb{V}_{blue} . By assumption, there is at least one edge of each color in the graph, meaning that both sets are nonempty. This is a contradiction to the fact \mathcal{H} is connected, completing the proof. ■

We can now prove Theorem 5.

Proof: We denote $m_1 = \dim \ker \mathcal{E}_{\mathcal{H}} P_{\mathcal{H}}$. The proof is similar to the proof of Theorem 4. We again define functions $F_{\mathcal{G}_1}$ for graphs $\mathcal{G}_1 \in \mathcal{G}$ as $F_{\mathcal{G}_1}(y, w) = k^{-1}(y) + \mathcal{E}_{\mathcal{G}_1} g(\mathcal{E}_{\mathcal{G}_1}^\top y) + \mathcal{E}_{\mathcal{G}_1} P_{\mathcal{G}_1} w$, but the function $F_{\mathcal{G}_1}$ is now defined on $\ker \mathcal{E}_{\mathcal{H}} P_{\mathcal{H}} \subset \mathbb{R}^{|\mathbb{E}_{\mathcal{G}}|}$. As before, we define $\mathcal{A}_{\mathcal{G}_1} = \{(y, w) : F_{\mathcal{G}_1}(y, w) = 0\}$ and use the implicit function theorem to show that $\mathcal{A}_{\mathcal{G}_1}$ are all manifolds, but their dimension is now $m_1 = \dim \ker \mathcal{E}_{\mathcal{H}} P_{\mathcal{H}}$. This time, we want to show that if $\mathcal{H} \neq \mathcal{G}_1$, then $\mathcal{A}_{\mathcal{H}} \cap \mathcal{A}_{\mathcal{G}_1}$ is an embedded submanifold of dimension $\leq m_1 - 1$, as we want to show that (with probability 1), the solutions (4) for the graphs \mathcal{G}_1 and \mathcal{H} are different. As before, it suffices to show that if $(y, w) \in \mathcal{A}_{\mathcal{G}_1} \cap \mathcal{A}_{\mathcal{H}}$, then $\ker dF_{\mathcal{G}_1}$ and $\ker dF_{\mathcal{H}}$ are different at (y, w) . For any graph \mathcal{G}_1 ,

$$dF_{\mathcal{G}_1} = [\nabla k^{-1}(y) + \mathcal{E}_{\mathcal{G}_1} \nabla \gamma(\mathcal{E}_{\mathcal{G}_1}^\top y) \mathcal{E}_{\mathcal{G}_1}^\top, (\mathcal{E}_{\mathcal{G}_1} P_{\mathcal{G}_1})|_{\ker \mathcal{E}_{\mathcal{H}} P_{\mathcal{H}}}]$$

where $\cdot|_{\ker \mathcal{E}_{\mathcal{H}} P_{\mathcal{H}}}$ is the restriction of the matrix to $\ker \mathcal{E}_{\mathcal{H}} P_{\mathcal{H}}$. Thus, if $\mathcal{G}_1 \in \mathcal{G}$ is not a subgraph of \mathcal{H} , it contains an edge e absent from \mathcal{H} . As in the proof of Theorem 4, we note that $1_e \in \ker \mathcal{E}_{\mathcal{H}} P_{\mathcal{H}}$ and conclude that $\ker dF_{\mathcal{G}_1} \neq \ker dF_{\mathcal{H}}$ at (y, w) . Thus, we only consider nonempty subgraphs \mathcal{G}_1 of \mathcal{H} .

For any collection $E \subset \mathbb{E}_{\mathcal{H}}$, we consider $v = (0, \mathbb{1}_E)$, where $\mathbb{1}_E = \sum_{e \in E} \mathbb{1}_e$. If E corresponds to a cycle in \mathcal{H} , then $v \in \ker dF_{\mathcal{H}}(y, w)$. We show there exists cycle in \mathcal{H} such that $v \notin \ker dF_{\mathcal{G}_1}(y, w)$, completing the proof.

The graph \mathcal{G}_1 defines a coloring of the graph \mathcal{H} —edges in \mathcal{G}_1 are colored in blue, whereas edges absent from \mathcal{G}_1 are colored in red. Because \mathcal{G}_1 is a nonempty proper subgraph of \mathcal{H} , this coloring contains both red and blue edges. By the lemma, there is a simple cycle in \mathcal{H} having both red and blue edges. Let E be the set of the edges traversed by the cycle. We claim that $\mathcal{E}_{\mathcal{G}_1} P_{\mathcal{G}_1} \mathbb{1}_E \neq 0$, which will complete the proof of the theorem. Indeed, because the simple cycle contains both red and blue edges, we can find a vertex touching both a red edge in the cycle and a blue edge in the cycle. We let v be the vertex, and let e_1 and e_2 be the corresponding blue and red edges. Recalling that the cycle is simple, these are the only cycle edges touching v . However, by the coloring, e_1 is in \mathcal{G}_1 , but e_2 is not. Thus,

$$(\mathcal{E}_{\mathcal{G}_1} P_{\mathcal{G}_1} \mathbb{1}_E)_v = (\mathcal{E}_{\mathcal{G}_1})_{ve_1} = \pm 1 \neq 0$$

and in particular, $\mathbb{1}_E \notin \ker \mathcal{E}_{\mathcal{G}_1} P_{\mathcal{G}_1}$. ■

V. MONITORING NETWORK FAULTS

In this section, we consider two applications of the developed framework, namely, network FDI and defense strategies for adversarial games over networks. We first present a simple algorithm for network fault detection. Then, we discuss defense strategies for adversarial games over networks, which will require a bit more effort. Finally, we exhibit a network fault isolation protocol by combining the previous two algorithms, which will be demonstrated in a case study in Section VII. In order to apply the framework of edge-indication vectors, we need an algorithm elevating the asymptotic differentiation achieved in the previous section to an online differentiation scheme. Thus, we make the following assumption.

Assumption 4: There exists an algorithm \mathcal{A} that receives a model for a network $(\mathcal{G}, \Sigma, \Pi)$ and a conjectured limit y^* as input and takes measurements of the network in-run. The algorithm stops and declares “no” if the network does not converge to y^* and otherwise runs indefinitely. The algorithm \mathcal{A} is called a convergence assertion algorithm.

In the language of Section III, this is a fault-monitoring system that gives neither false positives nor false negatives. For now, we assume that such algorithm exists. We will discuss this assumption in Section VI.

A. Detecting Network Faults

We first focus on Problem 1. To tackle the problem, we use the notion of edge-indication vectors from Section IV. Suppose that we have MEIP agents $\{\Sigma_i\}$. We first take any output-strictly MEIP controllers $\{\Pi_e\}$ solving the classical synthesis problem, i.e., forcing the closed loop system to converge to y^* (see Theorem 3). As we noted, if $w \in \mathbb{R}^{|\mathbb{E}_{\mathcal{G}}|}$ lies in the kernel of $\mathcal{E}_{\mathcal{G}} P_{\mathcal{G}}$, then the solution of the following equations is the same:

$$k^{-1}(y) + \mathcal{E}_{\mathcal{G}} \gamma(\mathcal{E}_{\mathcal{G}}^\top y) = -\mathcal{E}_{\mathcal{G}} P_{\mathcal{G}} w, \quad k^{-1}(y) + \mathcal{E}_{\mathcal{G}} \gamma(\mathcal{E}_{\mathcal{G}}^\top y) = 0.$$

Algorithm 1: Network Fault Detection in MEIP MAS.

- 1: Find edge controllers $\{\Pi_e\}_{e \in \mathbb{E}_G}$ solving the synthesis problem with graph \mathcal{G} , agents Σ , and control goal y^* (see Theorem 3 and Remark 1).
- 2: Find a basis $\{b_1, \dots, b_k\}$ for the linear space $\ker \mathcal{E}_G P_G$
- 3: Pick a Gaussian vector $\alpha \in \mathbb{R}^k$ and let $w = \sum_i \alpha_i b_i$
- 4: Define the interaction protocol as (Π, w) .
- 5: Run the system with the chosen interaction protocol.
- 6: Implement the algorithm \mathcal{A} for the system $(\mathcal{G}, \Sigma, (\Pi, w))$ with limit y^* . Declare a fault in the network if \mathcal{A} declares that the system does not converge to the prescribed value.

Thus, if w lies in $\ker(\mathcal{E}_G P_G)$, running the interaction protocol (Π, w) does not change the steady-state output of the system. However, by Theorem 5, a random vector in $\ker(\mathcal{E}_G P_G)$ gives a $(\mathfrak{G}, \mathcal{G})$ -edge-indication vector, as long as \mathcal{G} is 2-connected. Thus, using the interaction protocol (Π, w) , where $w \in \ker \mathcal{E}_G P_G$ is chosen randomly, will guarantee that all faulty systems would converge to a steady state different than y^* , with probability 1. Thus, applying the algorithm \mathcal{A} allows an online finite-time distinction between the nominal faultless system and its faulty versions. We explicitly write the prescribed algorithm as follows.

Theorem 6 (Network fault detection): Consider an underlying graph \mathcal{G} and n agents $\{\Sigma_i\}$ satisfying Assumption 1. Assume \mathcal{G} is 2-connected. Then, with probability 1, Algorithm 1 synthesizes an interaction protocol (Π, w) solving Problem 1, i.e., the algorithm satisfies the following properties.

- i) If no faults occur in the network, the output of the closed-loop system converges to y^* .
- ii) If any number of faults occur in the network, the algorithm detects their existence.

Proof: The proof follows from the discussion preceding Algorithm 1. Namely, Theorem 5 assures that w is a $(\mathfrak{G}, \mathcal{G})$ -edge-indication vector, so long that \mathcal{G} is 2-connected. In other words, the output of the closed-loop system with graph \mathcal{G} converges to y^* , and for any graph $\mathcal{G} \neq \mathcal{H} \in \mathfrak{G}$, the output of the closed-loop system with graph \mathcal{H} converges to a value different from y^* .

It remains to show that the algorithm declares a fault if and only if a fault occurs. If no faults occur, \mathcal{A} never declares a fault, and the same is true for Algorithm 1. On the contrary, assume that any number of faults occur in the network, and let \mathcal{H} be the current underlying graph. The output of the closed loop system converges to $y \neq y^*$, so \mathcal{A} eventually stops and declares a fault. Thus, Algorithm 1 declares a fault. ■

B. Synthesis in the Presence of an Adversary

Consider the following two-player game. Both players are given the same n SISO agents $\Sigma_1, \dots, \Sigma_n$, the same graph \mathcal{G} on n vertices and m edges, and the same vector $y^* \in \mathbb{R}^n$. There is also a server that can measure the state of the agents at certain intervals and broadcast a single message to all agents once. The planner acts first and designs a control scheme for the network and the server. The adversary acts second, removing at most r edges from \mathcal{G} . The system is then run. The planner wins if

Algorithm 2: Planner Strategy in Adversarial Multi-Agent Synthesis Game With MEIP Agents—Synthesis.

- 1: Define $N = \sum_{\ell=0}^r \binom{m}{\ell}$. Let Graphs be an array with N entries, and let $j = 1$;
- 2: **for** $\ell = 0, \dots, r$ **do**
- 3: **for** $1 \leq i_1 < i_2 < \dots < i_\ell \leq m$ **do**
- 4: Insert the graph $\mathcal{H} = \mathcal{G} - \{e_{i_1}, \dots, e_{i_\ell}\}$ to the j th entry of Graphs. Advance j by 1.
- 5: Define two arrays Controllers, SSLimits of length N .
- 6: Choose w as Gaussian random vector of length m .
- 7: **for** $j = 1, \dots, N$ **do**
- 8: Take any edge controllers $\{\Pi_e\}_{e \in \mathbb{E}_G}$ satisfying Assumption 1.
- 9: Compute the steady-state limit of the network $(\text{Graphs}(j), \Sigma, (\Pi, w))$. Insert the result into $\text{SSLimits}(j)$
- 10: Solve the synthesis problem for agents Σ and underlying graph $\text{Graphs}(j)$ as in Theorem 3 and Remark 1. Insert the result into $\text{Controllers}(j)$

Algorithm 3: Planner Strategy-In-Run Protocol for Agents.

- 1: Run the interaction protocol (Π, w) .
- 2: When a message j is received, run the interaction protocol described by $\text{Controllers}(j)$.

Algorithm 4: Planner Strategy-In-Run Protocol for Server.

- 1: Define HasFaulted as an array of zeros of size N .
- 2: **while** HasFaulted has at least two null entries **do**
- 3: Run N instances of the algorithm \mathcal{A} simultaneously, with conjectured steady-state limits from SSLimits .
- 4: **for** $j = 1$ to N **do**
- 5: **if** The j th instance declared “no” **then**
- 6: Change the value of $\text{HasFaulted}(j)$ to 1.
- 7: Find the index j such that $\text{HasFaulted}(j) = 0$. Broadcast the message j to the agents.

the closed-loop system converges to y^* , and the adversary wins otherwise. We show that the planner can always win by using a strategy stemming from edge-indication vectors, assuming that the agents are MEIP.

Namely, consider the following strategy. Take all possible $\sum_{\ell=0}^r \binom{m}{\ell}$ underlying graphs. For each graph, the planner solves the synthesis problem as in Theorem 3. If the planner finds out the adversary changed the underlying graph to \mathcal{H} , he could alert the agents (through the server) and ask them to run the protocol solving the synthesis problem for \mathcal{H} . Thus, the planner needs to find a way to identify the underlying graph after the adversary acted, without using the server’s broadcast. This can be done by running the system with a \mathfrak{G} -edge-indication vector and using the server to identify the network’s steady state. Namely, consider Algorithms 2–4, detailing the synthesis procedure and in-run protocol for the planner. We prove:

Theorem 7: Consider the game above. With probability 1, Algorithms 2–4 describe a winning strategy for the planner. Moreover, if r is independent of n (i.e., $r = O(1)$), the synthesis algorithm has polynomial time complexity. Otherwise, the time complexity is $O(n^{c_1 r + c_2})$ for some universal constants $c_1, c_2 > 0$. Furthermore, the size of the message broadcasted by the server is $O(r \log n)$.

Proof: Suppose the adversary changed the underlying graph to \mathcal{H} , which has entry j in `Graphs`. We note that Assumption 4 assures that \mathcal{A} never declares a fault if and only if the closed-loop system converges to the conjectured steady state, and that w is a \mathfrak{G} -edge-indication vector by Theorem 4. Thus, the j th instance of convergence assertion protocol can never return a fault, and all other instances must eventually declare a fault. Thus, the server correctly guesses the underlying graph. It then broadcasts the index j to the agents, allowing them to change the interaction protocol and run the solution of the synthesis problem with desired output y^* and underlying graph \mathcal{H} . Thus, the network will converge to y^* , and the planner wins.

We now move to time complexity. Note that $N = O(m^r)$. The first for-loop has N iterations; each having $O(mn)$ actions (where a graph is saved in memory by its incidence matrix, which is of size $\leq m \times n$). Thus, the first for-loop takes a total of $O(m^{r+1}n)$ time. The second for-loop solves the synthesis problem for \mathcal{H} by solving an equation of the form $\mathcal{E}_{\mathcal{H}}v = v_0$ for a known vector v_0 and an unknown v . This can be done using least squares, taking $O(\max\{m, n\}^3)$ time. As for finding the steady state, this can be done by solving a convex network optimization problem [23, Problem (OPP)], which takes a polynomial amount of time in n and m (e.g., via gradient descent). Recalling that $m \leq \binom{n}{2} = O(n^2)$, we conclude that if r is bounded, the total time used is polynomial in n . Moreover, if r is unbounded, the bottleneck is the first for-loop which takes $O(m^{r+1}n)$ time. Plugging $m \leq n^2$ gives a bound on the time complexity of the form $O(n^{2r+3})$. As for the communication complexity, the server broadcasts a number between 1 and N , so a total of $O(\log_2 N)$ bits are needed to transmit the message. Plugging in $N = O(m^r)$ gives that $O(r \log_2 m) = O(r \log n)$ bits are needed.

C. Detection and Isolation of Network Faults

We now consider Problem 2, in which faults occur throughout the run, and we want to detect their existence and overcome them. This problem can be thought of as a tougher hybrid of the previous two problems—in Section V-A, the faults could appear throughout the run, but we only needed to find out they exist. In Section V-B, all of the faults occur before the run, but we had to overcome them. Motivated by this view, we offer a hybrid solution. Ideally, the interaction protocol will have two disjoint phases—a first “stable” phase in which the underlying graph is known and no extra faults have been found, and a second “exploratory” phase in which extra faults have been found, and the current underlying graph is not yet known. The first phase can be solved by using the network fault detection Algorithm 1, as long as the current underlying graph is 2-connected. The

Algorithm 5: Synthesis for Network Fault Isolation.

- 1: Define $N = \sum_{\ell=0}^r \binom{m}{\ell}$. Let `Graphs` be an array with N entries, and let $j = 1$;
 - 2: **for** $\ell = 0, \dots, r$ **do**
 - 3: **for** $1 \leq i_1 < i_2 < \dots < i_\ell < m$ **do**
 - 4: Insert the graph $\mathcal{H} = \mathcal{G} - \{e_{i_1}, \dots, e_{i_\ell}\}$ to the j th entry of `Graphs`. Advance j by 1.
 - 5: Define two arrays `IP`, `SSLimits` of length N .
 - 6: Choose w as a Gaussian random vector of length m .
 - 7: Choose controllers $\{\Pi_e\}_{e \in \mathbb{E}}$ satisfying Assumption 1.
 - 8: **for** $j = 1, \dots, N$ **do**
 - 9: Run steps 1–1 of Algorithm 1. Insert the resulting interaction protocol into `IP(j)`.
 - 10: Compute the steady-state limit of the closed-loop system with the interaction protocol (Π, w) . Insert the result into `SSLimits(j)`
-

second phase can be solved by the pre-broadcast stage of the planner strategy described in Algorithms 2–4.

The main issue that remains is what happens if the underlying graph changes again in the exploratory phase, i.e., we entered the exploratory phase with underlying graph \mathcal{H}_1 , but it changed to \mathcal{H}_2 before we identified that graph. In the exploratory phase, we run an instance of \mathcal{A} on all of the possible graphs simultaneously, until all but one instance declared a fault. If the instance related to \mathcal{H}_2 has not declared a fault yet, it will not declare a fault from now on, unless another fault occurs before the end of the exploratory phase. If the same instance has already declared a fault, we have a problem—all other instances will eventually also declare a fault, and there are two options in this case.

The first option is that one instance will declare a fault last, i.e., there is a time in which all but one instance have declared a fault. In this case, we identify the graph as some \mathcal{H}_3 . When we return to the stable phase and run the interaction protocol (Π, w) corresponding to \mathcal{H}_3 , a fault will be declared, and we will return to the exploratory phase. This is because w was synthesized as a $(\mathfrak{G}, \mathcal{H}_3)$ -edge-indication vector, meaning that the de-facto steady-state limit (with graph \mathcal{H}_2) will be different than the conjectured steady-state limit (with graph \mathcal{H}_3). The second option is that the last few instances of \mathcal{A} declare a fault simultaneously, which is dealt with by restarting the exploratory phase. We get the synthesis algorithm and the in-run protocol presented in Algorithms 5 and 6, respectively. We claim these solve Problem 2.

Theorem 8: Let $\Sigma_1, \dots, \Sigma_m$ be agents satisfying Assumption 1, and let \mathcal{G} be a k -connected graph for $k \geq 3$ on n vertices and m edges. Then, with probability 1, Algorithms 5 and 6, run with $r = k - 2$, solve Problem 2 for up to r faults.

Proof: We refer to steps 2 and 3 of Algorithm 6 as the stable phase of the algorithm, and to steps 4–13 as the exploratory phase. As the number of faults is no bigger than $r = k - 2$, the underlying graph remains 2-connected throughout the run. We claim that the theorem follows from the following claims:

Algorithm 6: In-Run Protocol for Network Fault Isolation.

- 1: Find the index j for which $\text{Graphs}(j) = \mathcal{G}$.
- 2: Command the agents to change their interaction protocol to the one described in $\text{IP}(j)$. Define $\mathcal{H} = \text{Graphs}(j)$.
- 3: Run \mathcal{A} for the closed-loop system with graph \mathcal{H} and interaction protocol $\text{IP}(j)$. Only if the algorithm declares a fault, continue to step 4.
- 4: Define HasFaulted as an array of zeros of size N .
- 5: Change the agents' interaction protocol to (Π, w) .
- 6: **while** HasFaulted has at least two null entries **do**
- 7: Run N instances of \mathcal{A} simultaneously, with conjectured limits from SSLimits .
- 8: **for** $j = 1$ to N **do**
- 9: **if** The j th instance has declared a fault **then**
- 10: Change the value of $\text{HasFaulted}(j)$ to 1.
- 11: **if** HasFaulted has no entries equal to zero **then**
- 12: Go to step 4.
- 13: Find the index j such that $\text{HasFaulted}(j) = 0$. Set $\mathcal{H} = \text{Graphs}(j)$. Go to step 2.

- 1) If we are in the stable phase, the current underlying graph is $\mathcal{H} = \text{Graphs}(j)$, and no more faults occur throughout the run, the closed-loop system converges to y^* .
- 2) If we are in the stable phase, but the assumed graph $\mathcal{H} = \text{Graphs}(j)$ is not the current underlying graph, we will eventually move to the exploratory phase.
- 3) Each instance of the exploratory phase eventually ends.
- 4) If an instance of the exploratory phase is executed, and no more faults occur throughout the run, it correctly identifies the current underlying graph.

We first explain why the claims hold with probability 1. Claims 1 and 2 follow from Theorem 6 with probability 1, as the underlying graph \mathcal{G} is always 2-connected. Claim 4 holds with probability 1, as follows from Theorem 7. As for Claim 3, if no faults occur throughout the instance of the exploratory phase, then it eventually ends by Claim 4. If faults do occur throughout the run, the discussion above shows that at some time, all (except possibly one) instances of the convergence assertion protocol have declared a fault. If all of them declared a fault, we start another instance of the exploratory phase, and otherwise, we move to the stable phase. In either case, the instance of the exploratory phase ends, and Claim 3 is proved.

We now explain how the theorem follows from these claims. Suppose that a total of $\ell \leq r$ faults occur throughout the run. Let $T < \infty$ be the time at which the last fault occurs. We look at the phase of the algorithm at times $t > T$ and show that in either case, the system must converge to y^* .

- i) If we arrive at the stable phase and the conjectured graph \mathcal{H} is the true underlying graph, then the system converges to y^* (Claim 1).
- ii) If we start an instance of the exploratory phase, it eventually ends (Claim 3) and the stable phase starts, in which the conjectured graph \mathcal{H} is the true underlying graph (Claim 4). By i), the system converges to y^* .

- iii) If we are in the stable phase, but the conjectured graph \mathcal{H} is incorrect, we eventually start an exploratory phase (Claim 2). Thus, the system converges to y^* by ii).
- iv) Finally, we could be in the middle of an instance of the exploratory phase. In that case, the instance eventually ends (Claim 3), after which we either apply a new instance of the exploratory phase, or the stable phase. In both cases, the system must converge to y^* by i), ii) or iii). ■

Remark 2: We can use a similar protocol to isolate more complex faults. For example, we can consider a case in which each agent communicates with all other agents by a single transceiver, and if it faults, all edges touching the corresponding vertex are removed from the graph. We can even use a hybrid fault model, in which faults correspond to the removal of certain subsets of edges touching a common vertex. For example, suppose that there are two distant groups of agents. Agents in the same group are close and communicate using Bluetooth communication. Agents in different groups are farther and communicate using Wi-Fi. When an agent's Bluetooth transceiver faults, all inter-group edges are removed, and when the Wi-Fi transceiver faults, all intragroup edges are removed.

VI. ONLINE ASSERTION OF NETWORK CONVERGENCE

In the previous section, we used the notion of edge-indication vectors, together with Assumption 4, to suggest algorithms for network FDI. In this section, we propose algorithms \mathcal{A} satisfying Assumption 4 by using convergence estimates, relying on passivity. We revisit a result from [23].

Proposition 3 (see [23]): Let (u, y, ζ, μ) be a steady state of $(\mathcal{G}, \Sigma, \Pi)$ of the form (1). Suppose that the agents Σ_i , with state x_i , are passive with respect to (u_i, y_i) with passivity index $\rho_i \geq 0$, and that the controllers Π_e , with state η_e , are passive with respect to (ζ_e, μ_e) , with passivity index $\nu_e \geq 0$. Let S_i and W_e be the agents' and the controllers' storage functions. Then, $S(x, \eta) = \sum_{i \in \mathbb{V}} S_i(x_i) + \sum_{e \in \mathbb{E}} W_e(\eta_e)$ is a positive-definite C^1 -function, which nulls only at the steady states (x, η) corresponding to (u_i, y_i) and (ζ_e, μ_e) , and satisfies the inequality

$$\frac{dS}{dt} \leq - \sum_{i \in \mathbb{V}} \rho_i (y_i(t) - y_i)^2 - \sum_{e \in \mathbb{E}} \nu_e (\mu_e(t) - \mu_e)^2. \quad (5)$$

Proof: The proof follows immediately from S_i, W_e being positive-definite C^1 -functions nulling only at x_i, η_e , by summing the following inequalities:

$$\begin{aligned} \frac{dS_i}{dt} &\leq (u_i(t) - u_i)(y_i(t) - y_i) - \rho_i (y_i(t) - y_i)^2 \\ \frac{dW_e}{dt} &\leq (\mu_e(t) - \mu_e)(\zeta_e(t) - \zeta_e) - \nu_e (\mu_e(t) - \mu_e)^2 \end{aligned}$$

and using the equality $(u(t) - u)^\top (y(t) - y) = -(\mu(t) - \mu)^\top \mathcal{E}_G^\top (y(t) - y) = -(\mu(t) - \mu)(\zeta(t) - \zeta)$. ■

The inequality (5) can be thought of as a way to check that the system is functioning properly. Indeed, we can monitor x, y, η , and μ and check that the inequality holds. If it does not, there must have been a fault in the system. This idea has a few drawbacks, linked to one another. First, as we commented in Section III-A, in some networks, the controller state $\eta_e(t)$ can

be defined only for existing edges, so using $\eta(t)$ requires us to know the functioning edges, which is absurd. Thus, in some cases, we must use Assumption 2. Second, in practice, even if we have access to x , we cannot measure it continuously. Instead, we measure it at certain time intervals. One can adapt (5) to an equivalent integral form

$$S(x(t_{k+1}), \eta(t_{k+1})) - S(x(t_k), \eta(t_k)) \leq - \int_{t_k}^{t_{k+1}} \left(\sum_{i \in \mathbb{V}} \rho_i \Delta y_i(t)^2 + \sum_{e \in \mathbb{E}} \nu_e \Delta \mu_e(t)^2 \right) dt \quad (6)$$

where $\Delta y_i = y_i(t) - y_i$ and $\Delta \mu_e = \mu_e(t) - \mu_e$. However, this gives rise to the third problem—unlike the function S , we cannot assure that the functions $(y_i(t) - y_i)^2$ and $(\mu_e(t) - \mu_e)^2$ (or their sum) are monotone. Thus, we cannot easily bound the integral appearing on the right-hand side of the inequality.

We present two approaches to address this problem. First, we estimate the integral by linearizing the right-hand side of (6) and bounding the error. Second, we bound the right-hand side as a function of S , resulting in an inequality of the form $\dot{S} \leq -\mathcal{F}(S)$, providing a convergence estimate.

A. Asserting Convergence Using High-Rate Sampling

Consider the inequality (6), and assume $t_{k+1} - t_k = \Delta t_k$ is very small, so the functions $y_i(t) - y_i$ and $\mu_e(t) - \mu_e$ are roughly constant in the time period $[t_k, t_{k+1}]$. More precisely, recalling that $y = h(x)$ and $\mu = \phi(\eta, \mathcal{E}_G^\top y)$, and assuming these functions are differentiable near $x(t_k)$ and $\eta(t_k)$, we expand the right-hand side of (6) to a Taylor series

$$\int_{t_k}^{t_{k+1}} \left(\sum_{i \in \mathbb{V}} \rho_i \Delta y_i(t)^2 + \sum_{e \in \mathbb{E}} \nu_e \Delta \mu_e(t)^2 \right) dt = \left(\sum_{i \in \mathbb{V}} \rho_i \Delta y_i(t_k)^2 + \sum_{e \in \mathbb{E}} \nu_e \Delta \mu_e(t_k)^2 \right) \Delta t_k + O(\Delta t_k^2). \quad (7)$$

We wish to give a more explicit bound on the $O(\Delta t_k^2)$ term. We consider the following function G , defined on $[t_k, t_{k+1}]$:

$$G(t) = \sum_i \rho_i (y_i(t) - y_i)^2 + \sum_e \nu_e (\mu_e(t) - \mu_e)^2. \quad (8)$$

Equation (7) arises from the approximation $G(t) = G(t_k) + O(|t - t_k|)$, which is true for differentiable functions. Using Lagrange's mean value theorem for $t \in [t_k, t_{k+1}]$, we find a point $s \in (t, t_{k+1})$ such that $G(t) = G(t_k) + \frac{dG}{dt}(s)(t - t_k)$. If we would bound the derivative $\frac{dG}{dt}$ in the interval $[t_k, t_{k+1}]$, we would find a computational way to assert convergence. By the chain rule, the derivative of G is given by

$$\frac{dG}{dt} = \sum_{i \in \mathbb{V}} \rho_i (y_i(t) - y_i) \dot{y}_i + \sum_{e \in \mathbb{E}} \nu_e (\mu_e(t) - \mu_e) \dot{\mu}_e. \quad (9)$$

In order to compute the time derivative of y_i and μ_i , we recall that both are functions of x and η , namely, $y = h(x)$ and $\mu =$

$\phi(\eta, \mathcal{E}_G^\top y) = \phi(\eta, \mathcal{E}_G^\top h(x))$. Thus, we have that

$$\begin{cases} \dot{y} &= \nabla_x h(x(t)) \dot{x} \\ \dot{\mu} &= \nabla_\eta \phi(\eta(t), \zeta(t)) \dot{\eta} + \nabla_x \phi(\eta(t), \zeta(t)) \mathcal{E}_G^\top \nabla h(x(t)) \dot{x} \end{cases} \quad (10)$$

where $\zeta(t) = \mathcal{E}_G^\top h(x(t))$, $\dot{x} = f(x, u) = f(x, -\mathcal{E}_G \phi(\eta, \zeta))$, and $\dot{\eta} = \psi(\eta, \zeta) = \psi(\eta, \mathcal{E}_G^\top h(x))$. Thus, we can write the time derivative of G as a continuous function of $x(t)$, $\eta(t)$ by plugging (10) into (9). However, we do not know the value of $x(t)$, $\eta(t)$ between measurements. To tackle this problem, notice that we have some information on where $x(t)$ and $\eta(t)$ can lie. Namely, (5) shows that $S(x(t), \eta(t))$ is a monotone descending function, so $(x(t), \eta(t))$ lie in the set $\mathbb{B} = \{(x, \eta) : S(x, \eta) \leq S(x(t_k), \eta(t_k))\}$. More precisely, we have:

Proposition 4: Assume that the functions h_i , f_i , ϕ_e , and ψ_e are all continuously differentiable. Then, for any time $t \in [t_k, t_{k+1}]$, the following inequality holds:

$$\left| \frac{dG}{dt}(t) \right| \leq (\rho_* M_{\Delta y} M_{\dot{y}} + \nu_* M_{\Delta \mu} M_{\dot{\mu}, x}) M_{\dot{x}} + \nu_* M_{\Delta \mu} M_{\dot{\mu}, \eta} M_{\dot{\eta}}$$

where $M_{\dot{x}} = \max_{(x, \eta) \in \mathbb{B}} \|f(x, -\mathcal{E}_G \phi(\eta, \mathcal{E}_G^\top h(x)))\|$, $M_{\dot{\eta}} = \max_{(x, \eta) \in \mathbb{B}} \|\psi(\eta, \mathcal{E}_G^\top h(x))\|$, $M_{\dot{y}} = \max_{(x, \eta) \in \mathbb{B}} \|\nabla_x h(x)\|$, $M_{\dot{\mu}, x} = \max_{(x, \eta) \in \mathbb{B}} \|\nabla_\zeta \phi(\eta, \mathcal{E}_G^\top h(x)) \mathcal{E}_G^\top \nabla_x h(x)\|$, $M_{\dot{\mu}, \eta} = \max_{(x, \eta) \in \mathbb{B}} \|\nabla_\eta \phi(\eta, \mathcal{E}_G^\top h(x))\|$, $M_{\delta y} = \max_{(x, \eta) \in \mathbb{B}} \|h(x) - h(x)\|$, $M_{\delta \mu} = \max_{(x, \eta) \in \mathbb{B}} \|\psi((\eta, \mathcal{E}_G^\top h(x)) - \mu)\|$, $\rho_* = \max_i \rho_i$, $\nu_* = \max_e \nu_e$, and $\mathbb{B} = \{(x, \eta) : S(x, \eta) \leq S(x(t_k), \eta(t_k))\}$.

Proof: We fix some $t \in [t_k, t_{k+1}]$, so that $(x(t), \eta(t)) \in \mathbb{B}$. We use the expressions for \dot{x} , $\dot{\eta}$, \dot{y} , and $\dot{\mu}$ found in (10). First, the conditions $\|\dot{x}\| \leq M_{\dot{x}}$ and $\|\dot{\eta}\| \leq M_{\dot{\eta}}$ are obvious. Equation (10) shows that $\|\dot{y}\| \leq M_{\dot{y}} M_{\dot{x}}$ and $\|\dot{\mu}\| \leq M_{\dot{\mu}, x} M_{\dot{x}} + M_{\dot{\mu}, \eta} M_{\dot{\eta}}$. By using the Cauchy-Schwarz inequality on (9), we obtain $\left| \frac{dG}{dt} \right| \leq \rho_* M_{\delta y} \|\dot{y}\| + \nu_* M_{\delta \mu} \|\dot{\mu}\|$, concluding the proof. ■

Corollary 1: Fix any two times $t_k < t_{k+1}$, and consider the notation of Proposition 4. Then, the following inequality holds:

$$S(x(t_{k+1})) - S(x(t_k)) \leq - \left(\sum_i \rho_i \Delta y_i(t_k)^2 + \sum_{e \in \mathbb{E}} \nu_e \Delta \mu_e(t_k)^2 \right) \Delta t_k + \frac{M}{2} \Delta t_k^2, \quad (11)$$

where

$$M = (\rho_* M_{\Delta y} M_{\dot{y}} + \nu_* M_{\Delta \mu} M_{\dot{\mu}, x}) M_{\dot{x}} + \nu_* M_{\Delta \mu} M_{\dot{\mu}, \eta} M_{\dot{\eta}}.$$

Proof: Recall that $G(t) = \sum_{i \in \mathbb{V}} \rho_i (y_i(t) - y_i)^2 + \sum_{e \in \mathbb{E}} \nu_e (\mu_e(t) - \mu_e)^2$. By Proposition 4, for every $t \in [t_k, t_{k+1}]$ we have $G(t) \leq G(t_{k+1}) + M|t - t_{k+1}|$. Thus, (6) implies that (11) holds. ■

The corollary proposes a mathematically sound method for asserting convergence of the output $y(t)$ to y^* . One samples $y(t)$, $x(t)$, $\eta(t)$, and $\mu(t)$ at times t_1, t_2, t_3, \dots . At every time instance t_{k+1} , one checks that the inequality (11) holds. We show that when $\Delta t_k \rightarrow 0$, this method asserts that the output of the system converges to the said value. In other words, assuming that we

sample the system at a high-enough rate, we can assert that it converges very closely to the supposed steady-state output:

Proposition 5: Let t_1, t_2, \dots , be any monotone sequence of times such that $t_k \rightarrow \infty$, and suppose that the inequality (11) holds for any k . Then, for any $\varepsilon > 0$, there are infinitely many $N > 0$ such that $\sum_{i \in \mathbb{V}} \rho_i \Delta y_i(t_N)^2 + \sum_{e \in \mathbb{E}} \nu_e \Delta \mu_e(t_N)^2 < \frac{M}{2} \Delta t_N + \varepsilon$. More precisely, for any two times $t_{N_1} \leq t_{N_2}$, if $t_{N_2} \geq t_{N_1} + \varepsilon^{-1} S(x(t_{N_1}), \eta(t_{N_1}))$, then there exists some $k \in \{N_1, N_1 + 1, \dots, N_2\}$ such that $\sum_{i \in \mathbb{V}} \rho_i \Delta y_i(t_k)^2 + \sum_{e \in \mathbb{E}} \nu_e \Delta \mu_e(t_k)^2 < \frac{M}{2} \Delta t_k + \varepsilon$.

The proposition can be thought of as a close-convergence estimate. The left-hand side, viewed as a function of x, η , is a nonnegative smooth function, which nulls only at the steady state (x, η) . Thus, it is small only when $x(t)$ and $\eta(t)$ are close to (x, η) , and because $S(x(t), \eta(t))$ is monotone descending, once the trajectory arrives near (x, η) , it must remain near (x, η) . One might ask why “infinitely many times” is more useful in this case. It does not add any more information if the time intervals Δt_k are taken as constant (i.e., we sample the system at a constant rate). However, we can measure the system at an ever-increasing rate, at least theoretically. Taking $\Delta t_k \rightarrow 0$ (while still having $t_k \rightarrow \infty$, e.g., $\Delta t_k = 1/k$), we see that we must have $x(t) \rightarrow x$ and $\eta(t) \rightarrow \eta$, hence we can use the proposition to assert convergence. We now prove the proposition.

Proof: It is enough to show that for each $\varepsilon > 0$ and any $N_1 > 0$, there is some $N > N_1$ such that $\sum_{i \in \mathbb{V}} \rho_i \Delta y_i(t_N)^2 + \sum_{e \in \mathbb{E}} \nu_e \Delta \mu_e(t_N)^2 < \frac{M}{2} \Delta t_N + \varepsilon$. Indeed, suppose this is not the case. Then, for any $k > N_1$, the right-hand side of (11) is upper-bounded by $-\varepsilon \Delta t_k$. We sum the telescopic series and conclude that for any $k > N_1$,

$$S(x(t_k)) - S(x(t_{N_1})) \leq - \sum_{j=N_1+1}^k \varepsilon \Delta t_j = -\varepsilon(t_k - t_{N_1}), \quad (12)$$

so $t_{N_2} \geq t_{N_1} + \varepsilon^{-1} S(x(t_{N_1}), \eta(t_{N_1}))$ implies that $S(x(t_k), \eta(t_k)) < 0$. This is absurd, as $S \geq 0$. Thus, there must exist some $N \in \{N_1, N_1 + 1, \dots, N_2\}$ such that $\sum_{i \in \mathbb{V}} \rho_i \Delta y_i(t_N)^2 + \sum_{e \in \mathbb{E}} \nu_e \Delta \mu_e(t_N)^2 < \frac{M}{2} \Delta t_N + \varepsilon$. ■

Proposition 5 can be used for convergence assertion. We can consider the following scheme—begin at time t_0 and state x_0, η_0 . We want to show that $S(x(t), \eta(t)) \rightarrow 0$. We instead show that $G(t)$, defined in (8), gets arbitrarily close to 0. As we said, this is enough as $G(t)$ is only small when $x(t), \eta(t)$ are close to the steady state (x, η) . We prove the following.

Theorem 9: Consider the algorithm \mathcal{A} , defined in the following form. Sample the system at times t_1, t_2, \dots and check whether the inequality (11) holds. If it does, continue, and if does not, then stop and declare “no.” Then, there exists a sequence t_1, t_2, \dots , depending on the system and the initial conditions, such that \mathcal{A} satisfies Assumption 4.

Proof: We present the following method of choosing t_1, t_2, \dots . We first choose $t_0 = 0$, an arbitrary $\delta_1 > 0$, compute M as in Proposition 4, and choose $\Delta_1 t = \frac{\delta_1}{M}$ and $\varepsilon = \frac{\delta_1}{2}$. Sample the system at rate $\Delta_1 t$ until time $t_{N_1} > t_0 + \varepsilon^{-1} S(x_0, \eta_0)$. The process is then reiterated with $\delta_{k+1} = \delta_1/2^k$ for $k = 1, 2, \dots$, giving rates $\Delta_k t$ and times t_{N_k} . We claim that \mathcal{A} ,

with this choice of sample times, satisfies Assumption 4. If the network $(\mathcal{G}, \Sigma, \Pi)$ converges to (x, η) , then Corollary 1 implies that the algorithm never stops, as required. We show that if the algorithm never stops, the network converges to the conjectured limit. By the discussion above, and the fact that $S(x(t), \eta(t))$ is a monotone descending function, it suffices to show that $\liminf_{k \rightarrow \infty} G(t_k) = 0$. By choice of $\Delta_1 t$, if (11) holds at each time, then when we reach time t_{N_1} , we know that at some point, we had $G(t) \leq \frac{M}{2} \Delta_1 t + \varepsilon = \delta_1$. Reiterating shows that at some times t_k^* , $G(t_k^*) \leq \delta_k$, where $\delta_{k+1} = \frac{\delta_k}{2}$, so $\liminf_{k \rightarrow \infty} G(t_k) = 0$. ■

The term “high-rate sampling” comes from the fact that if M is not updated when we reiterate with smaller δ , then eventually, $t_{k+1} - t_k \rightarrow 0$, which is impractical in real-world cases. However, we note that the number M decreases as $S(x(t), \eta(t))$ decreases, as shown in Proposition 4. Thus, if M is updated between iterations, we might have $\Delta t \not\rightarrow 0$.

Remark 3: There is a tradeoff between the time steps Δt and the time it takes to find a point in which $G(t) < \frac{M}{2} \Delta t + \varepsilon$, given by $t = \frac{S(x(0), \eta(0))}{\varepsilon}$. On one hand, we want larger time steps (to avoid high-rate sampling) and shorter overall times; however, increasing both Δt and ε creates a worse eventual bound on $G(t)$. We can choose both by maximizing an appropriate cost function $C(\Delta t, \varepsilon)$, monotone in both Δt and ε , subject to $\frac{M}{2} \Delta t + \varepsilon = \delta_1, \varepsilon \geq 0, \Delta t \geq 0$. Choosing $C(\Delta t, \varepsilon)$ as linear is inadvisable, as maximizing a linear function with linear constraints always leads to the optimizer being on the boundary, which means either $\Delta t = 0$ or $\varepsilon = 0$. The choice $\Delta t = \frac{\delta_1}{M}$ and $\varepsilon = \frac{\delta_1}{2}$ mentioned above corresponds to the geometric average cost function $C(\Delta t, \varepsilon) = \sqrt{\Delta t \varepsilon}$. Other possible choices of C include the Harmonic mean $\frac{2\Delta t \varepsilon}{\Delta t + \varepsilon}$, or weighted versions of these means, where the weights express a preference between lowering ε and lowering Δt .

B. Asserting Convergence Using Convergence Profiles

For this subsection, we now assume that Assumption 3 holds, and that the agents are output-strictly MEIP, i.e., that $\rho_i > 0$. Consider (5) and suppose that there is a nonnegative monotone function \mathcal{F} such that for any t , the right-hand side of (5) is bounded from above by $-\mathcal{F}(S)$. In that case, we get an estimate of the form $\dot{S} \leq -\mathcal{F}(S)$. This is a weaker estimate than (5), but it has a more appealing discrete-time form

$$\begin{aligned} S(x(t_{k+1})) - S(x(t_k)) &\leq - \int_{t_k}^{t_{k+1}} \mathcal{F}(S(x(t))) dt \\ &\leq - \mathcal{F}(S(x(t_{k+1}))) \cdot (t_{k+1} - t_k) \end{aligned} \quad (13)$$

where we use the monotonicity of \mathcal{F} and the fact that $S(x(t))$ is monotone nonascending. Due to Assumption 3, we focus on the elements of the right-hand side of (5) corresponding to the agents and neglect the ones corresponding to controllers, as S is now the sum of the functions $S_i(x_i)$. Because controllers are passive, we have $\nu_e \geq 0$, so removing the said term does not change the inequality’s validity.

In order to find \mathcal{F} , it is natural to look for functions Ω_i satisfying $\Omega_i(S_i) \leq (y_i(t) - y_i)^2$. We define the existence of the functions Ω_i properly in the following definition.

Definition 5: Let $\Omega : [0, \infty) \rightarrow [0, \infty)$ be any function on the nonnegative real numbers. We say that an autonomous system has the *convergence profile* (ρ, Ω) with respect to the steady state (u, y) if there exists a C^1 storage function $S(x)$ such that the following inequalities hold.

- i) $\frac{dS(x(t))}{dt} \leq (u(t) - u)(y(t) - y) - \rho(y(t) - y)^2$;
- ii) $\Omega(S(x(t))) \leq (y(t) - y)^2$.

Example 1: Consider the SISO system Σ defined by $\dot{x} = -x + u$, $y = x$, and consider the steady-state input–output pair $(0, 0)$. The storage function $S(x(t)) = \frac{1}{2}x(t)^2$ satisfies

$$\dot{S}(x(t)) = x(t)\dot{x}(t) = (u(t) - 0)(y(t) - 0) - (y(t) - 0)^2.$$

Thus, Σ has convergence profile $(1, \Omega)$ for $\Omega(\theta) = 2\theta$.

More generally, when considering an LTI system with no input feedthrough, both functions $S(x)$ and $(y(t) - y)^2$ are quadratic in x . Thus, there is a monotone linear function Ω such that the inequality $\Omega(S(x(t))) \leq (y(t) - y)^2$ holds. In particular, a function Ω exists in this case. We show that functions Ω exist for rather general systems.

Theorem 10: Let Σ be the SISO system of the form $\dot{x} = -f(x) + q(x)u$, $y = h(x)$. Suppose that q is a positive continuous function, that f/q is C^1 and monotone ascending, and that h is C^1 and strictly monotone ascending. Let $(u = f(x)/q(x), y = h(x))$ be any steady-state input–output pair of the system. Then, we have the following.

- i) Using the storage function $S(x) = \int_x^x \frac{h(\sigma) - h(x)}{q(\sigma)} d\sigma$, the system Σ has the convergence profile (ρ, Ω) for a strictly ascending function Ω and $\rho = \inf_x \frac{f(x) - f(x)}{h(x) - h(x)} \geq 0$.
- ii) Suppose that there exists some $\alpha > 0$ such that the limit $\lim_{x \rightarrow x} \frac{|h(x) - h(x)|}{|x - x|^\alpha}$ exists and is finite. Then, the limit $\lim_{\theta \rightarrow 0} \frac{\Omega(\theta)}{\theta^\beta}$ exists and is finite, where $\beta = \frac{2\alpha}{\alpha + 1}$. In other words, if h behaves like a power law near x , then Ω behaves like a power law near 0.

Proof: The passivation inequality follows from Theorem 1, so we focus on constructing the function Ω . For every $\theta \geq 0$, we define the set $\mathbb{A}_\theta = \{x \in \mathbb{R} : (h(x) - h(x))^2 \leq \theta\}$. We want that $x \in \mathbb{A}_\theta$ would imply that $\Omega(S(x)) \leq \theta$. Because h is continuous and monotone, it is clear that \mathbb{A}_θ is an interval containing x . Now, let ω be the function on $[0, \infty)$ defined as $\omega(\theta) = \sup_{x \in \mathbb{A}_\theta} S(x)$. We note that ω can take infinite values (e.g., when h is bounded, but S is not). However, we show that the restriction of ω on $\{\theta : \omega(\theta) < \infty\}$ is strictly monotone. If we show that this claim is true, then ω has an inverse function, which is also strictly monotone. Define $\Omega = \omega^{-1}$ as the strictly monotone inverse function. By definition, for any $x \in \mathbb{R}$, we have that $x \in \mathbb{A}_\theta$ for $\theta = (h(x) - h(x))^2$, so $S(x) \leq \omega(\theta)$. Thus, $\Omega(S(x)) \leq \Omega(\omega(\theta)) = \theta = (h(x) - h(x))^2$, concluding the first part of the proof.

We now prove that the restriction of ω on $\{\theta : \omega(\theta) < \infty\}$ is strictly monotone. Note that if $0 \leq \theta_1 < \theta_2$, the interval $\{x : (h(x) - h(x))^2 \leq \theta_1\} = \mathbb{A}_{\theta_1}$ is strictly contained in \mathbb{A}_{θ_2} , as h is strictly monotone. Moreover, S is strictly ascending in $[x, \infty)$

and strictly descending in $(-\infty, x]$, as the function $\frac{h(x) - h(x)}{g(x)}$ is positive on (x, ∞) and negative on $(-\infty, x)$. Thus, we have $\omega(\theta_1) < \omega(\theta_2)$, unless $\omega(\theta_1) = \infty$, which is what we wanted to prove.

We now move to the second part of theorem, i.e., that if h behaves like a power law near x , then Ω behaves like a power law near 0. We use big- O notation (in the limit $x \rightarrow x$). By assumption and strict monotonicity of h , we have

$$h(x) - h(x) = C \operatorname{sgn}(x - x) |x - x|^\alpha + o(|x - x|^\alpha) \quad (14)$$

for some constant $C > 0$, implying $(h(x) - h(x))^2 = C^2 |x - x|^{2\alpha} + o(|x - x|^{2\alpha})$. By definition, for $\theta > 0$ small enough, \mathbb{A}_θ is an interval centered at x of radius $\theta^{1/2\alpha} / C^{1/\alpha} + o(\theta^{1/2\alpha})$. Writing $S(x) = \int_x^x \frac{h(\sigma) - h(x)}{q(\sigma)} d\sigma$ and $q(x) = q(x) + o(1)$, (14) implies that

$$\frac{h(\sigma) - h(x)}{q(\sigma)} = \frac{1}{q(x)} (C \operatorname{sgn}(x - x) |x - x|^\alpha + o(|x - x|^\alpha)).$$

Thus, $S(x) = \frac{C}{q(x)(\alpha + 1)} |x - x|^{\alpha + 1} + o(|x - x|^{\alpha + 1})$. We now compute $\omega(\theta)$ by definition, using our characterization of \mathbb{A}_θ :

$$\begin{aligned} \omega(\theta) &= \max_{x \in \mathbb{A}_\theta} S(x) = \max_{x \in \mathbb{A}_\theta} \left(\frac{C |x - x|^{\alpha + 1}}{q(x)(\alpha + 1)} + o(|x - x|^{\alpha + 1}) \right) \\ &= \frac{C}{q(x)(\alpha + 1)} \left(\frac{\theta^{1/2\alpha}}{C^{1/\alpha}} \right)^{\alpha + 1} + o\left(\left(\frac{\theta^{1/2\alpha}}{C^{1/\alpha}} \right)^{\alpha + 1} \right) \\ &= (D + o(1)) \theta^{\frac{\alpha + 1}{2\alpha}} \end{aligned}$$

for $D = \frac{1}{q(x)(\alpha + 1)C^{1/\alpha}} > 0$. Thus, the inverse function $\Omega(\theta)$ is given as $\Omega(\theta) = (D^{-\frac{2\alpha}{\alpha + 1}} - o(1)) \theta^{\frac{2\alpha}{\alpha + 1}}$, as desired. ■

Example 2: Consider a system with $q(x) = 1$, $h(x) = \sqrt[3]{x}$ and a steady state $u = x = y = 0$. $h(x)$ behaves like a power law with power $\alpha = \frac{1}{3}$. Part ii) of Theorem 10 implies that Ω also behaves like a power law, with power $\beta = \frac{2\alpha}{\alpha + 1} = \frac{1}{2}$. We exemplify the computation of Ω and show that it behaves like a power law with $\beta = \frac{1}{2}$, as forecasted by the theorem. Indeed, $S(x) = \int_0^x \sqrt[3]{\sigma} d\sigma = \frac{3}{4}x^{4/3}$, and $(h(x) - h(x))^2 = x^{2/3}$.

For every $\theta \geq 0$, $\mathbb{A}_\theta = \{x : x^{2/3} \leq \theta\} = [-\theta^{1.5}, \theta^{1.5}]$. Thus, $\omega(\theta) = \sup_{x \in \mathbb{A}_\theta} S(x) = \sup_{|x| \leq \theta^{1.5}} \frac{3}{4}x^{4/3} = \frac{3}{4}\theta^2$, implying that Ω , the inverse function of ω , is given by $\sqrt{\frac{4}{3}\theta}$, and one observes that actually $(h(x) - h(x))^2 = \Omega(S(x))$.

Remark 4: Theorem 10 gives a prescription to design the function Ω . However, some steps, namely, the inversion of ω , are computationally hard. For example, if $h(x) = 1 - e^{-x}$ and $q(x) = 1$, then $\omega(\theta) = \log_e \frac{1}{1 - \sqrt{\theta}}$ for $\theta < 1$ and $\omega(\theta) = \infty$ for $\theta \geq 1$, which is almost impossible to invert analytically. To solve this problem, we can either precompute the different values of Ω numerically and store them in a table, or approximate them online using the bisection method. The strength of Theorem 10 is that it shows that a function Ω can always be found, implying that this method is always applicable.

So far, we transformed (6) into the equation $\frac{dS}{dt} \leq \sum_i -\rho_i \Omega_i(S_i)$, for nonnegative monotone functions Ω_i . This is closer to an inequality of the form $\dot{S} \leq -\mathcal{F}(S)$, but we still

cannot use it without high-rate sampling, as we cannot assume that $S_i(x_i(t))$ are monotone decreasing. We want to transform the right-hand side into a function of S . We note that $\Omega_i(\theta_i) = 0$ only at $\theta_i = 0$, as $S_i = 0$ happens only at x_i . We claim the following.

Proposition 6: Let ρ_1, \dots, ρ_n be any positive numbers, and let $\Omega_1, \dots, \Omega_n : [0, \infty) \rightarrow [0, \infty)$ be any C^1 strictly monotone functions such that $\Omega_i(\theta_i) = 0$ only at $\theta_i = 0$. Suppose further that for any i , there exists some $\beta_i > 0$ such that the limit $\lim_{\theta_i \rightarrow 0} \frac{\Omega_i(\theta_i)}{\theta_i^{\beta_i}}$ exists and is positive. Define $\Omega_* : [0, \infty) \rightarrow [0, \infty)$ as $\Omega_*(\theta) = \min_i \Omega_i(\theta)$. Then, for every $D > 0$, there exists some constant $C > 0$ such that for all $D \geq \theta_1, \dots, \theta_n \geq 0$, we have $\sum_{i=1}^n \rho_i \Omega_i(\theta_i) \geq C \cdot \Omega_*(\sum_{i=1}^n \theta_i)$.

The proof of the proposition can be found in the Appendix.

Corollary 2: Let S_1, \dots, S_n be the storage functions of the agents, let $S = \sum_i S_i$, and let $\Omega_1, \dots, \Omega_n$ be C^1 strictly monotone functions such that $\Omega_i(\theta_i) = 0$ only at $\theta_i = 0$. Suppose that for any i , there exists some $\beta_i > 0$ such that the limit $\lim_{\theta_i \rightarrow 0} \frac{\Omega_i(\theta_i)}{\theta_i^{\beta_i}}$ exists and is positive. Moreover, suppose that $\dot{S} \leq \sum_i \rho_i \Omega_i(S_i)$. Then, for every bounded set $B \subset \mathbb{R}^n$, there exists a constant $C > 0$ such that for any trajectory of the closed-loop system with initial condition in B , the inequality $\dot{S} \leq -C \cdot \Omega_*(S)$ holds, where $\Omega_*(\theta) = \min_i \Omega_i(\theta)$.

Proof: Use $\theta_i = S_i$ and $D = S(x(0))$ in Proposition 6. ■

Proposition 6 and Corollary 2 show that an inequality of the form (13) can be achieved, so long the functions Ω_i from Theorem 10 “behave nicely” around 0, namely, do not grow faster nor slower than a power law. This condition is very general and only excludes pathologies as $\Omega(\theta) = \frac{1}{\log(1/\theta)}$, growing faster than any power law, and $\Omega(\theta) = \exp(-1/\theta^2)$, growing slower than any power law. Theorem 10 shows that if h behaves like a power law near x , then so does Ω , so pathological functions Ω_* can only come from pathological measurement functions h_i . We show that it is enough to check the discretized equation (13) to assert convergence.

Proposition 7: Let $\Omega_* : [0, \infty) \rightarrow [0, \infty)$ be any continuous function such that $\Omega_*(\theta) = 0$ only at $\theta = 0$. Let $\tilde{S}(t)$ be any time-dependent monotone decreasing function $\tilde{S} : [0, \infty) \rightarrow [0, \infty)$. Let t_1, t_2, t_3, \dots be any unbounded sequence of times such that $\liminf_{k \rightarrow \infty} (t_{k+1} - t_k) > 0$, and suppose that for every k , the inequality $\tilde{S}(t_{k+1}) - \tilde{S}(t_k) \leq -\Omega_*(\tilde{S}(t_{k+1}))(t_{k+1} - t_k)$ holds. Then, $\tilde{S}(t) \rightarrow 0$ as $t \rightarrow \infty$.

The proof of the proposition can be found in the Appendix. We want to use $\tilde{S}(t) = S(x(t))$. The results above suggest an algorithm for convergence assertion.

Theorem 11: Algorithm 7, taking the system $(\mathcal{G}, \Sigma, \Pi)$, the initial state $x(0)$, and the conjectured steady state $\hat{x} = h^{-1}(y)$ as input, satisfies Assumption 4.

Proof: We denote the true limit of the system $(\mathcal{G}, \Sigma, \Pi)$ by x . We first assume that the algorithm never stops and show that $\hat{x} = x$. We show that $S(x(t_k)) \leq \delta_k$, which would suffice as $\delta_k \rightarrow 0$ and $S(t) \rightarrow 0$ implies that $x(t) \rightarrow \hat{x}$, and thus, $x = \hat{x}$. Suppose, heading toward contradiction, that $S(x(t_k)) \not\leq \delta_k$. Then, $\Omega(S(x(t_k))) \geq M$, meaning that the right-hand side of

Algorithm 7: Convergence Assertion Using Convergence Profile.

- Input:** A diffusively coupled network $(\mathcal{G}, \Sigma, \Pi)$, an initial condition $x(0)$ and a conjectured steady state \hat{x} ;
- 1: Let $S_i(x_i) = \int_{\hat{x}_i}^{x_i} \frac{h_i(\sigma_i) - h_i(\hat{x}_i)}{q_i(\sigma_i)} d\sigma_i$,
 $S(x) = \sum_i S_i(x_i)$;
 - 2: Use Theorem 10, Proposition 6, and Corollary 2 to find a function \mathcal{F} such that $\dot{S} \leq -\mathcal{F}(S)$ for all times t , with initial condition $x(0)$;
 - 3: Choose $\delta_0 = S(x(0))$ and $t_0 = 0$;
 - 4: **for** $k = 0, 1, 2, 3, \dots$ **do**
 - 5: Define $\delta_{k+1} = \delta_k/2$;
 - 6: Define $M = \min_{x: S(x) \geq \delta_{k+1}} \mathcal{F}(S(x))$;
 - 7: Sample the system at time some $t_{k+1} > t_k + \frac{S(x_0)}{M}$.
 - 8: **if** if (13) does not hold **then**
 - 9: Stop and return “no”;
-

(13) is larger than $-S(x(t_k))$. Hence, if the inequality holds, then $S(x(t_{k+1})) < 0$, which is absurd. Thus, $S(x(t_k)) \leq \delta_k$, and $\hat{x} = x$. On the contrary, if the conjectured limit \hat{x} is the true limit of the network, then Theorem 10, Proposition 6, and Corollary 2 show that (13) always holds, so the algorithm never stops, as expected. ■

Remark 5: Proposition 7 shows that we can take any sample times t_k such that $\liminf_{k \rightarrow \infty} (t_k - t_{k-1}) > 0$ and still get a valid convergence assertion algorithm. The suggested algorithm gives extra information, as it also bounds the distance of $x(t_k)$ from x . Another way to choose t_k is to use the solution of the ordinary differential equation (ODE) $\dot{\tilde{S}} = -\Omega(\tilde{S})$ with $\tilde{S}(t_0) = S(x_0)$. Let t_k be the earliest time in which $\tilde{S}(t_k) \leq \delta_k$. The inequality $\dot{\tilde{S}} \leq -\Omega(\tilde{S})$ assures that $S(x(t_k)) \leq \delta_k$. This method is more demanding, as the minima M in the previous section can be precomputed, but the solution to the ODE must be computed online.

Remark 6: Although we can prove convergence with this method using very seldom measurements, we should still sample the system at a reasonable rate. This is because we want to detect faults as soon as possible. If we sample the system in too large intervals, we will not be able to sense a fault until a large amount of time has passed.

We conclude this section with a short discussion about the perks and drawbacks of the two presented convergence assertion methods. The convergence profile method allows the designer to sample the system at any desired rate, allowing one to prove convergence using very seldom measurements. Moreover, it gives rate of convergence guarantees before running the system. On the contrary, the high-rate sampling method can require a long time to assert convergence to a δ -ball around the desired steady state, unless one is willing to increase the sampling rate, perhaps arbitrarily. However, its main upshot is that Assumption 3 need not hold, and that the method is computationally easier, as one can avoid function inversion needed to compute Ω .

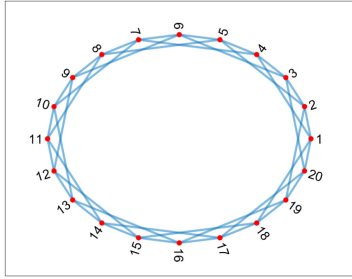


Fig. 2. Faultless underlying graph in the case study.

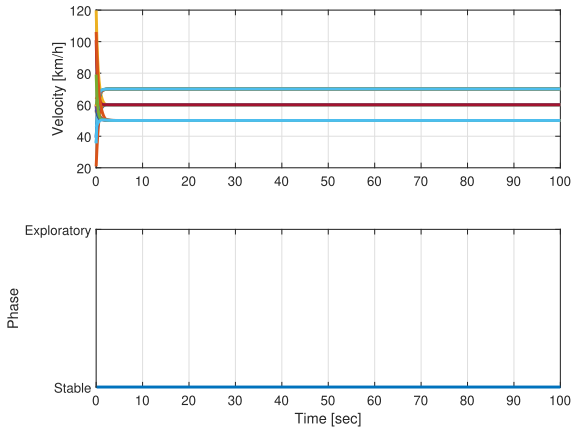


Fig. 3. Results of the first scenario.

VII. CASE STUDY: VELOCITY COORDINATION IN VEHICLES WITH DRAG

Consider a collection of $n = 20$ vehicles trying to coordinate their velocity. Each vehicle is modeled as a double integrator $G(s) = 1/s^2$ in vacuum, but is subject to drag in the real world. The drag model of a vehicle is usually chosen as a force against the direction of the vehicle's motion, which is quadratic in the size of the velocity [35, p. 164]. Thus, each vehicle is governed by the equations: $\dot{x}_i = -C_i x_i |x_i| + u_i$, $y_i = x_i$ where x_i is the velocity of the i th vehicle, and C_i is a constant aggregating different parameters affecting the drag, e.g., density of the air, viscosity of the air, the geometry of the vehicle, and the mass of the vehicle. The vehicles are trying to coordinate their velocity—Agents #1–#7 want to travel at 60 km/h, agents #8–#13 want to travel at 70 km/h, and agents #14–#20 want to travel at 50 km/h. The edge controllers are static nonlinearities given by sigmoid functions of the form $\mu_e = \tanh(\zeta_e)$. This diffusively coupled network satisfies both Assumptions 1 and 3, and we note that both the agents and the controllers are nonlinear. We choose an interaction graph \mathcal{G} , as seen in Fig. 2. One can check that \mathcal{G} is 4-connected, either using Menger's theorem or using other known algorithms [36].

The parameters C_i were chosen log-uniformly between 0.01 and 0.1. The initial velocity of the agents was chosen to be Gaussian with mean $\mu = 70$ km/h and standard deviation $\sigma = 20$ km/h. We solve the synthesis problem, forcing the network to converge to $y^* = [60 \cdot \mathbb{1}_7^T, 70 \cdot \mathbb{1}_6^T, 50 \cdot \mathbb{1}_7^T]^T$ km/h, where $\mathbb{1}_m$

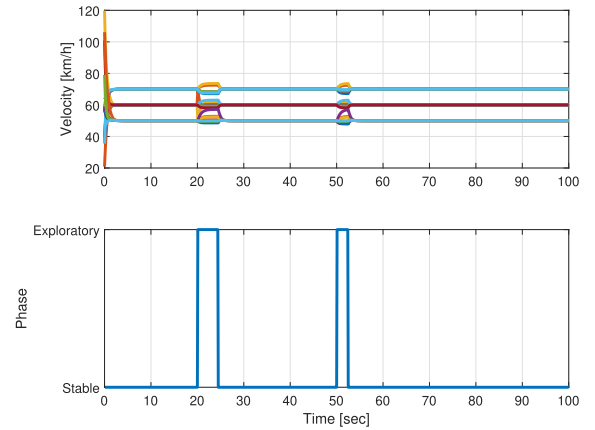


Fig. 4. Results of the second scenario.

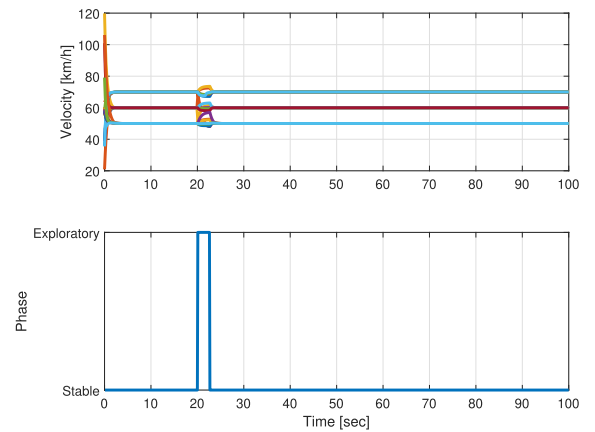


Fig. 5. Results of the third scenario.

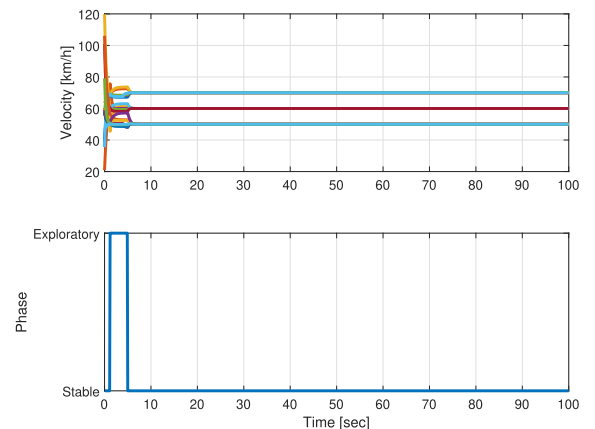


Fig. 6. Results of the fourth scenario.

is the all-one vector of size m , allowing up to two edges to fault. We run our FDI protocol, implementing the profile-based convergence assertion scheme, sampling the system at 10 Hz (i.e., a modified version of Algorithm 7). We consider four different scenarios, each lasting 100 s:

- 1) a faultless scenario;

- 2) at time $t = 20$ s, the edge $\{2, 3\}$ faults, and at time $t = 50$ s, the edge $\{13, 14\}$ faults;
- 3) at time $t = 20$ s, the edge $\{2, 3\}$ faults, and at time $t = 21$ s, the edge $\{13, 14\}$ faults;
- 4) at time $t = 0.5$ s, the edge $\{2, 3\}$ faults, and at time $t = 4$ s, the edge $\{13, 14\}$ faults.

The first scenario tests the nominal behavior of the protocol. The second tests its ability to handle with one single fault at a time. The third tests its ability to handle more than one fault at a time. The last tests its ability to deal with faults before the network converged. The results are available in Figs. 3–6. It can be seen that we achieve the control goal in all four scenarios. Moreover, in all scenarios and at all times, the velocities of the agents are not too far from the values found in y^* , meaning that this protocol cannot harm the agents by demanding them to have very wild states. In the second and third scenarios, the exploratory phase begins at the first measurement after the fault occurred. On the contrary, in the fourth scenario, the exploratory phase begins only at $t = 1.3$ s, a little under a second after the first fault. This is because the steady states of the faulty and nominal closed-loop systems are relatively close, meaning that it takes a little extra time to find that a fault exists.

VIII. CONCLUSION

We considered multi-agent networks prone to communication faults, in which the agents are output-strictly MEIP and the controllers are MEIP. We exhibited a protocol, in which a constant bias w is added to the controller output, and showed that if w is chosen randomly, no matter what the underlying graph \mathcal{G} is, we can asymptotically differentiate between any two versions (faulty or faultless) of the system. We also showed that if w is chosen randomly within a certain set, we asymptotically differentiate the faultless version of the system from its faulty version, while also solving the synthesis problem for the faultless version, assuming that \mathcal{G} was connected enough, i.e., 2-connected. These results were used to describe algorithms for network FDI protocols for general MEIP MASs, where the number of isolable faults is given by a graph-theoretic characteristic of \mathcal{G} , while no extra information on the agents and controllers but MEIP is used. These were achieved by assuming the existence of an online algorithm asserting that a given network converges to a conjectured steady state, allowing us to move from asymptotic differentiation to online differentiation. Later, two such algorithms were built using passivity of the agents and controllers, and their correctness was proved. We demonstrated our protocols by a case study, in which we successfully detect communication faults in a nonlinear network. We emphasize that the proposed method is proved to work so long the agents and controllers are MEIP, and the graph \mathcal{G} is connected enough. In particular, there is no assumption on the scale of the network.

APPENDIX

This appendix includes the proof of technical propositions from Section VI-B. We start with Proposition 6.

Proof: Without loss of generality, we assume that $\Omega_i = \Omega_*$ for all i , as proving that $\sum_{i=1}^n \rho_i \Omega_*(\theta_i) \geq C \Omega_*(\sum_{i=1}^n \theta_i)$

would imply the desired inequality. We also assume that $\rho_i = 1$ for all i , as proving that $\sum_i \Omega_i(\theta_i) \geq C \cdot \Omega_*(\sum_i \theta_i)$ would give $\sum_i \rho_i \Omega_i(\theta_i) \geq C \min_i \rho_i \cdot \Omega_*(\sum_i \theta_i)$. Define $F : [0, D]^n \setminus \{0\} \rightarrow \mathbb{R}$ as $F(\theta_1, \dots, \theta_n) = \frac{\sum_{i=1}^n \Omega_*(\theta_i)}{\Omega_*(\sum_{i=1}^n \theta_i)}$, where the claim is equivalent to F being bounded from below. For any $r > 0$, F is continuous on the compact set $[0, D]^n \setminus \{x : \|x\| < r\}$, so its minimum is obtained at some point. As F does not vanish on the set, the minimum is positive, so F is bounded from below on that set by a constant greater than zero. It remains to show that $\liminf_{\theta_1, \dots, \theta_n \rightarrow 0} F(\theta_1, \dots, \theta_n) > 0$. Let $\beta = \max_i \beta_i$, so that $\lim_{\theta \rightarrow 0} \frac{\Omega_*(\theta)}{\theta^\beta} > 0$. Then, $F(\theta_1, \dots, \theta_n) = \frac{\sum_{i=1}^n \Omega_*(\theta_i)}{\Omega_*(\sum_{i=1}^n \theta_i)} = \frac{\sum_{i=1}^n \Omega_*(\theta_i)}{(\sum_{i=1}^n \theta_i)^\beta} \cdot \frac{(\sum_{i=1}^n \theta_i)^\beta}{\Omega_*(\sum_{i=1}^n \theta_i)}$. We want to bound both factors from below when $\theta_1, \dots, \theta_n \rightarrow 0$. It is clear that the second factor is equal to $(\lim_{\theta \rightarrow 0} \frac{\Omega_*(\theta)}{\theta^\beta})^{-1}$, which is a positive real number by assumption. As for the first factor, we can bound it as $\lim_{\theta_1, \dots, \theta_n \rightarrow 0} \frac{\sum_{i=1}^n \Omega_*(\theta_i)}{(\sum_{i=1}^n \theta_i)^\beta} \geq \lim_{\theta_1, \dots, \theta_n \rightarrow 0} \frac{\Omega_*(\max_i \theta_i)}{(n \max_i \theta_i)^\beta} > 0$ as $\sum_{i=1}^n \theta_i \leq n \max_i \theta_i$ and $\sum_i \Omega_*(\theta_i) \geq \Omega_*(\max_i \theta_i)$. ■

We now prove Proposition 7.

Proof: By assumption, $\tilde{S}(t_k)$ is monotone decreasing and bounded from below, as $\tilde{S}(t_k) \geq 0$. Thus, it converges to some value, denoted \tilde{S}_∞ . Using $\tilde{S}(t_{k+1}) - \tilde{S}(t_k) \leq -\Omega_*(\tilde{S}(t_{k+1}))(t_{k+1} - t_k)$ and taking $k \rightarrow \infty$ gives that $0 \leq -\Omega_*(\tilde{S}_\infty)$. However, Ω_* is nonnegative, so $\Omega_*(\tilde{S}_\infty) = 0$, and thus, $\tilde{S}_\infty = 0$, meaning that $\tilde{S}(t_k) \rightarrow 0$. By monotonicity of \tilde{S} , we conclude that $\tilde{S}(t) \rightarrow 0$ as $t \rightarrow \infty$. ■

REFERENCES

- [1] K.-K. Oh, M.-C. Park, and H.-S. Ahn, "A survey of multi-agent formation control," *Automatica*, vol. 53, pp. 424–440, 2015.
- [2] L. Scardovi, M. Arcak, and E. D. Sontag, "Synchronization of interconnected systems with applications to biochemical networks: An input-output approach," *IEEE Trans. Autom. Control*, vol. 55, no. 6, pp. 1367–1379, Jun. 2010.
- [3] A. Teixeira, I. Shames, H. Sandberg, and K. H. Johansson, "Distributed fault detection and isolation resilient to network model uncertainties," *IEEE Trans. Cybern.*, vol. 44, no. 11, pp. 2024–2037, Nov. 2014.
- [4] S. Jafari, A. Ajorlou, A. G. Aghdam, and S. Tafazoli, "On the structural controllability of multi-agent systems subject to failure: A graph-theoretic approach," in *Proc. 49th IEEE Conf. Decis. Control*, 2010, pp. 4565–4570.
- [5] S. Jafari, A. Ajorlou, and A. G. Aghdam, "Leader localization in multi-agent systems subject to failure: A graph-theoretic approach," *Automatica*, vol. 47, no. 8, pp. 1744–1750, 2011.
- [6] M. A. Rahimian and V. M. Preciado, "Detection and isolation of failures in directed networks of LTI systems," *IEEE Control Netw. Syst.*, vol. 2, no. 2, pp. 183–192, Jun. 2015.
- [7] M. A. Rahimian and V. M. Preciado, "Failure detection and isolation in integrator networks," in *Proc. Amer. Control Conf.*, 2015, pp. 677–682.
- [8] G. Battistelli and P. Tesi, "Detecting topology variations in dynamical networks," in *Proc. 54th IEEE Conf. Decis. Control*, 2015, pp. 3349–3354.
- [9] M. A. Rahimian, A. Ajorlou, and A. G. Aghdam, "Detectability of multiple link failures in multi-agent systems under the agreement protocol," in *Proc. IEEE 51st IEEE Conf. Decis. Control*, 2012, pp. 118–123.
- [10] M. E. Valcher and G. Parlangeli, "On the effects of communication failures in a multi-agent consensus network," in *Proc. 23rd Int. Conf. Syst. Theory, Control Comput.*, 2019, pp. 709–720.
- [11] Y. Zhang, Y. Xia, J. Zhang, and J. Shang, "Generic detectability and isolability of topology failures in networked linear systems," *IEEE Trans. Control Netw. Syst.*, vol. 8, no. 1, pp. 500–512, Mar. 2021.

- [12] W. Pan, Y. Yuan, H. Sandberg, J. Gonçalves, and G.-B. Stan, "Online fault diagnosis for nonlinear power systems," *Automatica*, vol. 55, pp. 27–36, 2015.
- [13] M. A. Rahimian, A. Ajorlou, and A. G. Aghdam, "Digraphs with distinguishable dynamics under the multi-agent agreement protocol," *Asian J. Control*, vol. 16, no. 5, pp. 1300–1311, 2014.
- [14] M. Sharf and D. Zelazo, "Network identification: A passivity and network optimization approach," in *Proc. IEEE 57th Annu. Conf. Decis. Control*, 2018, pp. 2107–2113.
- [15] D. Patil, P. Tesi, and S. Trenn, "Indiscernible topological variations in DAE networks," *Automatica*, vol. 101, pp. 280–289, 2019.
- [16] H. Yang, V. Cocquempot, and B. Jiang, "Fault tolerance analysis for switched systems via global passivity," *IEEE Trans. Circuits Syst. II: Exp. Briefs*, vol. 55, no. 12, pp. 1279–1283, Dec. 2008.
- [17] W. Chen, S. X. Ding, A. Q. Khan, and M. Abid, "Energy based fault detection for dissipative systems," in *Proc. Conf. Control Fault-Tolerant Syst.*, Oct. 2010, pp. 517–521.
- [18] Q. Lei, R. Wang, and J. Bao, "Fault diagnosis based on dissipativity property," *Comput. Chem. Eng.*, vol. 108, pp. 360–371, 2018.
- [19] H. Bai, M. Arcak, and J. Wen, "Cooperative control design: A systematic, passivity-based approach," in *Communications and Control Engineering*. New York, NY, USA: Springer, 2011.
- [20] M. Arcak, "Passivity as a design tool for group coordination," *IEEE Trans. Autom. Control*, vol. 52, no. 8, pp. 1380–1390, Aug. 2007.
- [21] A. van der Schaft and B. Maschke, "Port-Hamiltonian systems on graphs," *SIAM J. Control Optim.*, vol. 51, no. 2, pp. 906–937, 2013.
- [22] N. Chopra and M. Spong, "Passivity-based control of multi-agent systems," in *Advances in Robot Control, From Everyday Physics to Human-Like Movements*. New York, NY, USA: Springer, 2006, pp. 107–134.
- [23] 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.
- [24] R. T. Rockafellar, *Network Flows and Monotropic Optimization*. Belmont, MA, USA: Athena Scientific, 1998.
- [25] M. Sharf and D. Zelazo, "A network optimization approach to cooperative control synthesis," *IEEE Control Syst. Lett.*, vol. 1, no. 1, pp. 86–91, Jul. 2017.
- [26] M. Sharf and D. Zelazo, "Analysis and synthesis of MIMO multi-agent systems using network optimization," *IEEE Trans. Autom. Control*, vol. 64, no. 11, pp. 4512–4524, Nov. 2019.
- [27] C. Godsil and G. Royle, *Algebraic Graph Theory (Graduate Texts in Mathematics)*, 1st ed. New York, NY, USA: Springer, 2001.
- [28] J. Bondy and U. Murty, *Graph Theory With Applications*. New York, NY, USA: Macmillan, 1977.
- [29] H. K. Khalil, *Nonlinear Systems*, 3rd ed. London, U.K.: Pearson, 2001.
- [30] M. Sharf, A. Romer, D. Zelazo, and F. Allgöwer, "Model-free practical cooperative control for diffusively coupled systems," *IEEE Trans. Autom. Control*, to be published, doi: [10.1109/TAC.2021.3056582.1](https://doi.org/10.1109/TAC.2021.3056582.1).
- [31] A. S. Willsky, "A survey of design methods for failure detection in dynamic systems," *Automatica*, vol. 12, no. 6, pp. 601–611, 1976.
- [32] M. Sharf, A. Jain, and D. Zelazo, "A Geometric method for passivation and cooperative control of equilibrium-independent passivity-short systems," *IEEE Trans. Autom. Control*, to be published, doi: [10.1109/TAC.2020.3043390](https://doi.org/10.1109/TAC.2020.3043390).
- [33] M. Bando, K. Hasebe, A. Nakayama, A. Shibata, and Y. Sugiyama, "Dynamical model of traffic congestion and numerical simulation," *Phys. Rev. E*, vol. 51, pp. 1035–1042, Feb. 1995.
- [34] M. Sharf and D. Zelazo, "Network identification for diffusively-coupled systems with minimal time complexity," Mar. 2019, *arXiv:1903.04923*.
- [35] R. A. Serway and J. W. Jewett, *Physics for Scientists and Engineers With Modern Physics*. Boston, MA, USA: Cengage Learning, 2018.
- [36] Z. Galil, "Finding the vertex connectivity of graphs," *SIAM J. Comput.*, vol. 9, no. 1, pp. 197–199, 1980.
- [37] L. Marton, "Energetic approach to deal with faults in robot actuators," *20th Mediterranean Conf. Control Autom. (MED)*, pp. 85–90, Jul. 2012.
- [38] G.-B. Stan and R. Sepulchre, "Analysis of interconnected oscillators by dissipativity theory," *IEEE Trans. Autom. Control*, vol. 52, no. 2, pp. 256–270, 2007.



Miel Sharf (Member, IEEE) received the B.Sc. and M.Sc. degrees in mathematics from the Technion—Israel Institute of Technology, Haifa, Israel, in 2013 and 2016, respectively, and the Ph.D. degree from the Department of the Aerospace Engineering, Technion—Israel Institute of Technology, in 2020.

Since September 2020, he has been a Postdoctoral Researcher with the Division of Decision and Control Systems, KTH Royal Institute of Technology, Stockholm, Sweden. His research interests include the relation between graph theory to multi-agent

systems, nonlinear control, data-driven control, security in networked systems, and formal verification of control systems.

Dr. Sharf is a recipient of the Springer Thesis Award. He was selected as a part of the 2021 class of Forbes Israel "30 under 30."



Daniel Zelazo (Senior Member, IEEE) received the B.Sc. and M.Eng. degrees in electrical engineering from the Massachusetts Institute of Technology, Cambridge, MA, USA, in 1999 and 2001, respectively, and the Ph.D. degree from the Department of Aeronautics and Astronautics, University of Washington, Seattle, WA, USA, in 2009.

He is currently an Associate Professor with the Department of Aerospace Engineering, Technion—Israel Institute of Technology, Haifa,

Israel. From 2010 to 2012, he was a Postdoctoral Research Associate and Lecturer with the Institute for Systems Theory and Automatic Control, University of Stuttgart, Stuttgart, Germany. His research interests include topics related to multi-agent systems, optimization, and graph theory.