SLS-BRD: A System-Level Approach to Seeking Generalized Feedback Nash Equilibria

Otacilio B. L. Neto, *Student Member, IEEE*, Michela Mulas, and Francesco Corona

***Abstract*—This work proposes a policy learning algorithm for seeking generalized feedback Nash equilibria in N_P-player noncooperative dynamic games. We consider linear-quadratic games with stochastic dynamics and design a best-response dynamics in which players update and broadcast a parametrization of their state-feedback policies. Our approach leverages the system level synthesis framework to formulate each player's update rule as the solution to a robust optimization problem. Under certain conditions, rates of convergence to a feedback Nash equilibrium can be established. The algorithm is showcased in exemplary problems ranging from the decentralized control of unstable systems to competition in oligopolistic markets.**

***Index Terms*—Best-response dynamics (BRD), feedback Nash equilibrium (NE), noncooperative games, system level synthesis (SLS).**

I. Introduction

MODERN cyber-physical systems are often comprised of interacting subsystems operated locally by noncooperative decision-making agents. Ideally, each agent operate its subsystem according to a feedback policy that optimize local objectives, while satisfying global constraints and being robust to their rivals' interference. However, the large-scale, decentralized, and multiobjective nature of such applications hinders most traditional approaches to policy design. The dynamic game theory provides an alternative framework through the concept of noncooperative equilibria (e.g., the generalized Nash equilibrium (GNE) [1], [2]) describing efficient, yet strategically stable, strategies for each agent. Under a dynamic game setting, decision-making agents must design feedback policies (i.e., strategies) to operate their subsystems while aware that their choice affects (and is affected by) the other agents' choice. An equilibrium would then correspond to a set of policies that satisfy the global constraints and is agreeable to all agents given their objectives. A policy design based on GNE seeking thus presents a promising venue for the decentralized control of cyber-physical systems.

In general, solving a noncooperative game has distinct goals: For the agents, to design efficient and robust policies in competitive environments, and, for the game designer, to examine (and potentially control) the local and global behavior of rational agents. Regardless, obtaining a Nash equilibrium (NE) is a notoriously difficult task [3]. The algorithmic game theory thus emerges as the field concerned with designing methods to bridge this computational gap [4]. An important class of algorithms, denoted *equilibrium-seeking* methods, places the task of searching for an equilibrium on the players: These include best-response dynamics (BRD, [5], [6], [7]), no-regret learning [8], [9], [10], and operator splitting [11], [12] methods. Broadly, these are fixed-point methods designed for (repeated) static games centered on the idea of players improving their strategies using only the information available to them. Aside from enabling the computation of Nash equilibria, these routines have the advantage of mimicking how noncooperative players would learn their policies in reality. In particular, BRD stands out as a simple, yet fundamental, model of policy learning for uncoordinated but communicating players. This method has become an important tool in economics and engineering, with applications in networked systems [13], [14], robotics [15], [16], [17], and resource management [18], [19].

We are interested in NE seeking algorithms for dynamic games in which the underlying system has linear, stochastic, and potentially unstable dynamics. The focus is on *closed-loop perfect state* information structures, as they yield equilibrium policies that are less sensitive to modeling and decision errors than their *open-loop* counterparts [1]. Specifically, we consider the relevant task of players learning a generalized feedback Nash equilibrium (GFNE) of state-feedback policies, which stabilize the system against disturbances, satisfy constraints on the control and state signals, and incorporate some specific structure (e.g., encoding communication delays). In the current practice, feedback Nash equilibria is obtained by either applying dynamic programming principles [1], [20], [21], [22], [23], [24], deriving equivalent complementary problems [25], [26], or using iterative heuristics [27], [28]. Recently, Laine et al. [29] extended the scope to provide a systematic method to compute (approximate) GFNE. While remarkable, these solutions cannot enforce either closed-loop stability, operational constraints, or design of the policy structure, in practice. Moreover, most of them still require some central coordinator to solve the game; the players themselves do not perform equilibrium seeking. The available

Received 26 February 2025; accepted 27 April 2025. Date of publication 9 May 2025; date of current version 27 October 2025. Recommended by Associate Editor. *(Corresponding author: Otacilio B. L. Neto.)*

Otacilio B. L. Neto and Francesco Corona are with the Department of Chemical and Metallurgical Engineering, Aalto University, 02150 Espoo, Finland (e-mail: otacilio.neto@aalto.fi; franciscu.corona@gmail.com).

Michela Mulas is with the Department of Teleinformatics Engineering, Federal University of Ceará, Fortaleza, CE 60455-760, Brazil (e-mail: michela.mulas@ufc.br).

Digital Object Identifier 10.1109/TAC.2025.3568560

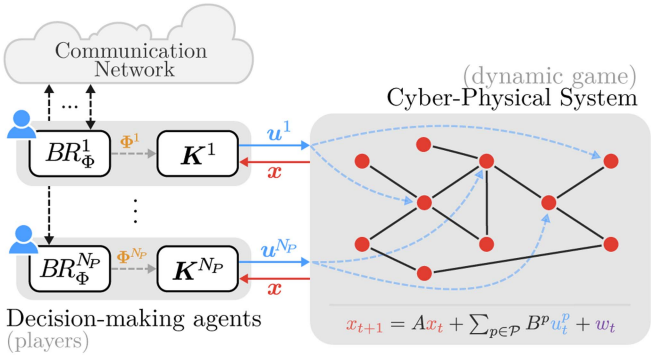

Fig. 1. SLS-BRD: Control architecture and the learning dynamics. A set of players $(1,\dots,N_P)$ control a decentralized dynamical system by measuring its state $(\boldsymbol{x})$, and then, applying actions $(\boldsymbol{u}^1,\dots,\boldsymbol{u}^{N_P})$ based on state-feedback policies $(\boldsymbol{K}^1,\dots,\boldsymbol{K}^{N_P})$ whose parameterizations $(\boldsymbol{\Phi}^1,\dots,\boldsymbol{\Phi}^{N_P})$ are chosen as best responses to each others' strategies, simultaneously, and broadcasted via a communication network.

literature on decentralized policy learning is concentrated on the field of reinforcement learning, where the focus is on finite-duration games described by Markov decision processes [30], [31], [32], [33], [34]. To the best of our knowledge, there are no equilibrium-seeking solutions to the aforementioned class of GFNE problems.

In this work, we propose a GFNE seeking algorithm to bridge this research gap. Leveraging the system level synthesis (SLS, [35]) framework, we first recast each player's policy synthesis problem as the search for an optimal *system level response* to disturbances. The system level response can be used to recover a corresponding optimal feedback policy, thus serving as its parametrization. Under this setup, each player's best-response strategy is the solution to a tractable optimization problem, which enforces closed-loop stability, operational constraints (on the control and state signals), and structural constraints (on the policy's parameters), already at the synthesis level. We then design a best-response routine in which players update the parametrization of their policies, in parallel, and then, announce them through some communication network (see Fig. 1). The algorithm does not depend on the state and actions applied to the underlying system and thus can be executed simultaneously with its operation. In summary, our contributions are as follows.

1) A realization of the associated best-response mappings in GFNE problems as robust convex optimization problems that are amenable to numerical solutions.
2) A system-level best-response dynamics (SLS-BRD) algorithm for GFNE seeking in dynamic games. In this equilibrium-seeking routine, players converge to a GFNE by iteratively updating the parametrization of their policies as the best response to their rivals' current policies.
3) In the absence of shared constraints, a formal analysis of the convergence properties of the SLS-BRD algorithm.

This policy learning algorithm is demonstrated in simulated experiments on the decentralized control of an unstable network and price management in a competitive oligopolistic market.

The rest of this article is organized as follows. Section II overviews the classes of (generalized) static and dynamic games, and the BRD algorithm. In Section III, we provide a system level parametrization for linear-quadratic games, then design a BRD for GFNE seeking. Section IV illustrates this approach in exemplary problems, and finally, Section V concludes this article. Toward a concise presentation, only essential results are given in the main text.

A. Notation

We use Latin letters to denote vectors and functions, and boldface to distinguish signals, operators, and their respective spaces. Sets are in calligraphic font; exceptions are the usual $\mathbb{R}$ and $\mathbb{N}$, and the sets of $(N\times N)$ symmetric $(\mathbb{S}^N)$, positive semidefinite $(\mathbb{S}^N_+)$, and positive definite matrices $(\mathbb{S}^N_{++})$. In particular, sequences are written as $\boldsymbol{x}=(x_t)_{t\in\mathcal{I}}$ for a countable set $\mathcal{I}\subseteq\mathbb{N}$, or $\boldsymbol{x}=(x_t)_{t=0}^T$ if $\mathcal{I}=\{0,\dots,T\}$. For $p\in(0,\infty)$, we define the space of N_x-dimensional vector-sequences $\ell_p^{N_x}(\mathcal{I})=\{\boldsymbol{x}:\|\boldsymbol{x}\|_{\ell_p}=(\sum_{t\in\mathcal{I}}\|x_t\|^p)^{1/p}<\infty\}$, with $\ell_\infty^{N_x}$ the space of all bounded sequences. The set $\boldsymbol{\mathcal{Y}}^{\boldsymbol{\mathcal{X}}}$ denotes all relations $\boldsymbol{A}:\boldsymbol{\mathcal{X}}\to\boldsymbol{\mathcal{Y}}$ with $\mathcal{L}(\boldsymbol{\mathcal{X}},\boldsymbol{\mathcal{Y}})\subseteq\boldsymbol{\mathcal{Y}}^{\boldsymbol{\mathcal{X}}}$ being the set of bounded linear operators. We sometimes write transformed signals as $\boldsymbol{Ax}=(Ax_t)_{t\in\mathcal{I}}$. We use the standard definition of Hardy spaces $\mathcal{H}_\infty$ and $\mathcal{RH}_\infty$, and write $\frac{1}{z}\mathcal{RH}_\infty$ for the set of real-rational strictly proper transfer functions. Finally, signals and operators used in this article include the following: The impulse signal $\boldsymbol{\delta}=(\delta_t)_{t\in\mathcal{I}}$, the identity operator I and matrix I_{N_x}, the matrices $\mathbf{1}_{n\times m}$ and $\mathbf{0}_{n\times m}$ of all 1's and 0's, respectively, the shift operator $\boldsymbol{S}_+:(x_0,x_1,\dots)\mapsto(0,x_0,\dots)$, and the Kronecker and Hadamard products $\otimes$ and $\odot$, respectively.

We distinguish set-valued mappings from ordinary functions using the notation $F:\mathcal{X}\rightrightarrows\mathcal{Y}$. A mapping is L_F-Lipschitz if $\|a-b\|\le L_F\|x-y\|$ for all $x,y\in\mathcal{X}$, $a\in F(x)$, $b\in F(y)$ and appropriate norm $\|\cdot\|$. F is said to be nonexpansive if $L_F=1$ and contractive if $L_F<1$. For any tuple $s=(s^p)_{p\in\mathcal{P}}\in\mathcal{S}$, we frequently write $s=(s^p,s^{-p})$ to highlight the element s^p; this should not be interpreted as a reordering. Similarly, if $\mathcal{S}=\prod_{p\in\mathcal{P}}\mathcal{S}^p$, we define the product $\mathcal{S}^{-p}=\prod_{\tilde{p}\in\mathcal{P}\setminus\{p\}}\mathcal{S}^{\tilde{p}}$.

II. Noncooperative Games and BRD

A (static) N_P-player game, denoted by a tuple

$$\mathcal{G}:=(\mathcal{P},\{\mathcal{S}^p\}_{p\in\mathcal{P}},\{L^p\}_{p\in\mathcal{P}}) \tag{1}$$

defines the problem in which *players* $p\in\mathcal{P}=\{1,\dots,N_P\}$ each decides on a strategy $s^p\in S^p(s^{-p})\subseteq\mathcal{S}^p$ to minimize an objective function $L^p:\mathcal{S}^1\times\cdots\times\mathcal{S}^{N_P}\to\mathbb{R}$. The strategy spaces S^p $(\forall p\in\mathcal{P})$ determine the actions available to the players, with the mappings $S^p:\mathcal{S}^{-p}\rightrightarrows\mathcal{S}^p$ restricting this choice based on the actions from their opponents. As such, both the players' objectives and feasible strategies depend explicitly on their competitors' actions. Finally, the players are assumed to be rational, noncooperative, and acting simultaneously.

A solution to the game $\mathcal{G}$ is understood as a strategy profile $s=(s^1,\dots,s^{N_P})\in\mathcal{S},\mathcal{S}=\mathcal{S}^1\times\cdots\times\mathcal{S}^{N_P}$, having some specified property that makes it agreeable to all players if they act rationally. In noncooperative settings, a widely accepted solution

Algorithm 1: Prototypical learning dynamics.

Input: Game $\mathcal{G} \coloneqq (\mathcal{P}, \{\mathcal{S}^p\}_{p\in\mathcal{P}}, \{L^p\}_{p\in\mathcal{P}})$
Output: GNE $s^\star = (s^{1^\star}, \dots, s^{N_P^\star})$

Initialize $s_0 \coloneqq (s_0^1, \dots, s_0^{N_P})$ and $k \coloneqq 0$;
for $k = 0, 1, 2, \dots$ **do**
  if $s_k \in \Omega_\mathcal{G}$ **then return** s_k;
  for $p \in \mathcal{P}$ **do**
    Update $s_{k+1}^p \in T^p(s_k^p, R^p(s_k) \mid L^p)$;

Algorithm 2: Best-Response Dynamics (BRD).

Input: Game $\mathcal{G} \coloneqq (\mathcal{P}, \{\mathcal{S}^p\}_{p\in\mathcal{P}}, \{L^p\}_{p\in\mathcal{P}})$
Output: GNE $s^\star = (s^{1^\star}, \dots, s^{N_P^\star})$

Initialize $s_0 \coloneqq (s_0^1, \dots, s_0^{N_P})$ and $k \coloneqq 0$;
for $k = 0, 1, 2, \dots$ **do**
  if $s_k \in BR(s_k)$ **then return** s_k;
  for $p \in \mathcal{P}$ **do**
    Update $s_{k+1}^p \in (1-\eta)s_k^p + \eta BR^p(s_k^{-p})$;

concept is that of a GNE: The game is *solved* when no player can improve its objective by unilaterally deviating from the agreed strategy profile. Formally, the following.

Definition 1: A strategy profile $s^\star = (s^{1^\star}, \dots, s^{N_P^\star}) \in \mathcal{S}$ is a GNE for the game $\mathcal{G}$ if

$$L^p(s^{p^\star}, s^{-p^\star}) \leq \min_{s^p \in S^p(s^{-p^\star})} L^p(s^p, s^{-p^\star}) \tag{2}$$

holds for every player $p \in \mathcal{P}$.

In general, the set of GNEs that solve a game $\mathcal{G}$

$$\Omega_\mathcal{G} := \{s^\star \in \mathcal{S} : s^\star \text{ satisfies } (2)\}$$

is not a singleton and can include strategy profiles that favor a specific subset of players (that is, *nonadmissible* GNEs [1]). The game might also not admit any GNE (i.e., $\Omega_\mathcal{G} = \emptyset$), then being characterized as unsolvable. Hereafter, we ensure that the problems being discussed are well-posed by considering the following conditions on their primitives.

Assumption 1: For each player $p \in \mathcal{P}$,

1) the objective $L^p : \mathcal{S}^1 \times \cdots \mathcal{S}^{N_P} \to \mathbb{R}$ is jointly continuous in all of its arguments and convex in the pth argument, $s^p \in S^p(s^{-p})$, for every $s^{-p} \in \mathcal{S}^{-p}$;
2) the mapping $S^p : \mathcal{S}^p \rightrightarrows \mathcal{S}^{-p}$ takes the form

$$S^p(s^{-p}) := \{s^p \in \mathcal{S}^p : (s^p, s^{-p}) \in \mathcal{S}_\mathcal{G}\}$$

where $\mathcal{S}_\mathcal{G}$ is some global constraint set shared by all players. Moreover, $\mathcal{S}^p$ and $\mathcal{S}_\mathcal{G}$ are both compact convex sets and they satisfy $\mathcal{S}_\mathcal{G} \cap (\mathcal{S}^1 \times \cdots \times \mathcal{S}^{N_P}) \neq \emptyset$.

Under Assumption 1, a generalization of the Kakutani fixed-point theorem ensures that $\mathcal{G}$ has a GNE, that is, $\Omega_\mathcal{G} \neq \emptyset$ [36]. In practice, these conditions consider each objective to have a unique optimal value, while imposing the feasible set of strategies to be nonempty and coupled only through a common constraint. Although restrictive, these assumptions still cover a broad class of problems of practical relevance.

We investigate algorithms for solving $\mathcal{G}$. A direct computation of a GNE is equivalent to solving N_P optimization problems simultaneously, as implied by Definition 1. Such an approach would require players to be coordinated and their objectives to be public. Conversely, we consider adaptive procedures in which the players learn their GNE strategies independently. In this direction, consider that $\mathcal{G}$ admits episodic repetitions, and let $s_k := (s_k^1, \dots, s_k^{N_P})$ be the strategy profile taken by players $\mathcal{P}$ at the kth episode. A prototypical learning routine for equilibrium seeking is outlined in Algorithm 1, with

1) $T^p : \mathcal{S}^p \times \mathcal{S}^{-p} \rightrightarrows \mathcal{S}^p$ describing how the pth player updates its strategy, based on a prediction of its opponents' next actions, given its individual objective;
2) $R^p : \mathcal{S} \rightrightarrows \mathcal{S}^{-p}$ describing how the pth player predicts its opponents' strategies for the next episode, based on the strategy profile currently being played.

Algorithm 1 belongs to the class of fixed-point methods: Its termination implies that $s^\star$ is a fixed-point of both T^p and R^p, that is, $s^\star \in T^p(s^{p^\star}, R^p(s^\star)) \subseteq T^p(s^{p^\star}, s^{-p^\star})$. In its general form, it is difficult to establish the conditions (and convergence rates) for these learning dynamics to approach an equilibrium. In this work, we build upon a specific yet fundamental instance of this algorithm: The BRD. This routine is overviewed in the following.

Best-response dynamics (BRD): Let the map $\mathrm{BR}^p : \mathcal{S}^{-p} \rightrightarrows \mathcal{S}^p$

$$\mathrm{BR}^p(s^{-p}) := \operatorname*{arg\,min}_{s^p \in S^p(s^{-p})} L^p(s^p, s^{-p}) \tag{3}$$

denote the best-response of $p \in \mathcal{P}$ to other players' strategies. Collectively, $\mathrm{BR}(s) := \mathrm{BR}^1(s^{-1}) \times \cdots \times \mathrm{BR}^{N_P}(s^{-N_P}) \subseteq \mathcal{S}$ is the joint best-response to any given profile $s \in \mathcal{S}$. From Definition 1, a strategy profile $s^\star = (s^{1^\star}, \dots, s^{N_P^\star}) \in \mathcal{S}$ is a GNE for $\mathcal{G}$ whenever $s^\star \in \mathrm{BR}(s^\star)$ or, equivalently

$$s^{p^\star} \in \mathrm{BR}^p(s^{-p^\star}) \quad \forall p \in \mathcal{P}. \tag{4}$$

The task of computing an NE can thus be translated into the search for a fixed-point of the set-valued mapping $\mathrm{BR} : \mathcal{S} \rightrightarrows \mathcal{S}$ [7]. The set of GNE solutions for $\mathcal{G}$ is the set of all such fixed points, $\Omega_\mathcal{G} := \{s^\star \in \mathcal{S} : s^\star \in \mathrm{BR}(s^\star)\}$. A natural procedure for GNE seeking consists of players adapting their strategies toward best responses to their rivals' strategies, which they assume will remain constant. Formally

$$T^p(s_k^p, R^p(s_k)) := (1-\eta)s_k^p + \eta \mathrm{BR}^p(R^p(s_k)) \tag{5}$$

given $R^p(s_k) = s_k^{-p}$ and a learning rate factor of $\eta \in (0, 1)$. This learning dynamics, summarized in Algorithm 2, is known as (discrete-time) BRD.

After each episode, the strategy profile is updated to

$$s_{k+1} = T(s_k) = (1-\eta)s_k + \eta \mathrm{BR}(s_k) \tag{6}$$

given the global update rule $T = (1-\eta)I + \eta\mathrm{BR}$. Notably, the mappings T and BR share the same set of fixed-points: The GNEs $\Omega_\mathcal{G}$. We can then establish the following result.

Lemma 1: Let $\mathrm{BR} : \mathcal{S} \rightrightarrows \mathcal{S}$ be a nonexpansive mapping. Then, $s_{k+1} = T(s_k)$ converge monotonically to a GNE solution

$s^\star \in \Omega_\mathcal{G}$, that is, $\lim_{k\to\infty} \inf_{s^\star\in\Omega_\mathcal{G}} \|T(s_k) - s^\star\| = 0$ given any appropriate norm $\|\cdot\|$ for $\mathcal{S}$.

This convergence result stems from the fixed-point theory, where the BRD algorithm is interpreted as belonging to the class of averaged (or Krasnosel'skii-Mann) iteration methods (see [37] for a formal proof). Moreover, if the best-response mapping BR is a contraction, then so is T and the BRD must converge geometrically to a GNE $s^\star \in \Omega_\mathcal{G}$, which is unique [37].

Lemma 2: Let $\mathrm{BR} : \mathcal{S} \rightrightarrows \mathcal{S}$ be L_{BR}-Lipschitz, $L_{\mathrm{BR}} < 1$. Then, from any feasible $s_0 \in \mathcal{S}$, the BRD $s_{k+1} = T(s_k)$ converge to the unique GNE $s^\star \in \Omega_\mathcal{G}$ with rate

$$\frac{\|s_k - s^\star\|}{\|s_0 - s^\star\|} \leq \left((1-\eta) + \eta L_{\mathrm{BR}}\right)^k \tag{7}$$

given any appropriate norm $\|\cdot\|$ for $\mathcal{S}$.

Importantly, these results might not hold in practice whenever the best-response maps, $\{\mathrm{BR}^p\}_{p\in\mathcal{P}}$, are only approximated (e.g., by solving (3) numerically). However, such *inexact* averaged operators are still known to converge under reasonable assumptions on the accuracy of this approximation [38]. The learning rate η plays a central role in the numerical stability of the BRD algorithm: A careful choice is required to ensure that strategy updates do not escape the feasible set, that is, to ensure that $T(s_k) \in \mathcal{S}_\mathcal{G} \cap \mathcal{S}$ for all $s_k \in \mathcal{S}_\mathcal{G} \cap \mathcal{S}$. For nongeneralized games, T trivially satisfy the constraints for any $\eta \in (0, 1)$, and thus, a careful design of the learning rate might not be necessary. In such cases, $\eta \to 1$ is the optimal choice if BR is a contraction and the convergence rate in (7) simplifies to L_{BR}^k.

Finally, we note that the stopping criteria in Algorithm 2 can be modified to allow for earlier termination. In this case, interrupting the BRD at some episode $k_f > 0$ will produce a strategy profile $s_{k_f} \in \mathcal{S}$ for which

$$L^p(s^p_{k_f}, s^{-p}_{k_f}) \leq \min_{s^p \in S^p(s^{-p}_{k_f})} L^p(s^p, s^{-p}_{k_f}) + \varepsilon \tag{8}$$

holds for every player $p \in \mathcal{P}$ with an "equilibrium gap" $\varepsilon > 0$. This profile characterises an ε-GNE: No player can improve its cost more than ε by unilaterally changing its strategy. The set of all ε-GNEs is denoted $\Omega^\varepsilon_\mathcal{G} = \{s^\varepsilon \in \mathcal{S} : s^\varepsilon \text{ satisfies (8)}\}$.

A. Infinite-Horizon Dynamic Games

A dynamic N_P-player game, denoted by a tuple

$$\mathcal{G}_\infty := (\mathcal{P}, \boldsymbol{\mathcal{X}}, \{\boldsymbol{\mathcal{U}}^p\}_{p\in\mathcal{P}}, \boldsymbol{\mathcal{W}}, \{J^p\}_{p\in\mathcal{P}}) \tag{9}$$

is defined by the stochastic linear dynamics

$$x_{t+1} = Ax_t + \sum_{p\in\mathcal{P}} B^p u^p_t + w_t, \quad x_0 \text{ given} \tag{10}$$

describing how the state of the game, $\boldsymbol{x} = (x_t)_{t\in\mathbb{N}} \in \boldsymbol{\mathcal{X}}$, evolves in response to the players' actions $\boldsymbol{u}^p = (u^p_t)_{t\in\mathbb{N}} \in \boldsymbol{\mathcal{U}}^p$ $(\forall p \in \mathcal{P})$ and the additive random noise $\boldsymbol{w} = (w_t)_{t\in\mathbb{N}} \in \boldsymbol{\mathcal{W}}$. For each realization $\boldsymbol{w} \in \boldsymbol{\mathcal{W}}$ and initial x_0, the state is explicitly expressed as $\boldsymbol{x} = \boldsymbol{F_w u}$ via the causal affine operator

$$\boldsymbol{F_w} : \boldsymbol{u} \mapsto (I - \boldsymbol{S_+ A})^{-1} \left(\sum_{p\in\mathcal{P}} \boldsymbol{S_+ B^p u^p} + \boldsymbol{S_+ w} + \boldsymbol{\delta} x_0 \right) \tag{11}$$

with $\boldsymbol{A} : \boldsymbol{x} \mapsto (Ax_t)_{t\in\mathbb{N}}$ and $\boldsymbol{B}^p : \boldsymbol{u}^p \mapsto (B^p u^p_t)_{t\in\mathbb{N}}$. Because known, the dependence on x_0 is omitted to simplify notation. Moreover, we assume $\mathrm{E} w_t = 0$ and $\mathrm{E}(w_{t+\tau} w_t^\mathsf{T}) = \delta_\tau \Sigma_w$, given a covariance matrix $\Sigma_w \in \mathbb{S}^{N_x}_{++}$, for every $t, \tau \in \mathbb{N}$. Finally, the sets $\boldsymbol{\mathcal{X}}, \boldsymbol{\mathcal{U}}^p$ $(\forall p \in \mathcal{P})$, and $\boldsymbol{\mathcal{W}}$ define all permissible state, action, and noise sequences; they take the form

$$\boldsymbol{\mathcal{X}} := \{\boldsymbol{x} \in \ell^{N_x}_\infty(\mathbb{N}) : x_t \in \mathcal{X},\ t \in \mathbb{N}\}$$
$$\boldsymbol{\mathcal{U}}^p := \{\boldsymbol{u}^p \in \ell^{N^p_u}_\infty(\mathbb{N}) : u^p_t \in \mathcal{U}^p,\ t \in \mathbb{N}\}$$
$$\boldsymbol{\mathcal{W}} := \{\boldsymbol{w} \in \ell^{N_x}_\infty(\mathbb{N}) : w_t \in \mathcal{W},\ t \in \mathbb{N}\}$$

given sets $\mathcal{X} \subseteq \mathbb{R}^{N_x}$, $\mathcal{U}^p \subseteq \mathbb{R}^{N^p_u}$ $(\forall p \in \mathcal{P})$, and $\mathcal{W} \subseteq \mathbb{R}^{N_x}$.

In infinite-horizon games, each player chooses a plan of action $\boldsymbol{u}^p \in U^p(\boldsymbol{u}^{-p})$ to minimize its objective functional

$$J^p(\boldsymbol{u}^p, \boldsymbol{u}^{-p}) := \mathrm{E}\left[\sum_{t=0}^{\infty} L^p(x_t, u^p_t, u^{-p}_t)\right] \tag{12}$$

defined by cost function $L^p : \mathcal{X} \times \mathcal{U}^1 \times \cdots \times \mathcal{U}^{N_P} \to \mathbb{R}$. The mappings $U^p : \boldsymbol{\mathcal{U}}^{-p} \rightrightarrows \boldsymbol{\mathcal{U}}^p$ restrict the permissible actions for each player based on its rivals' strategies. Under this setup, the dynamic game $\mathcal{G}_\infty$ is stationary and can be interpreted as a static game on the appropriate functional spaces. A plan of action $\boldsymbol{u} = (\boldsymbol{u}^1, \ldots, \boldsymbol{u}^{N_P}) \in \boldsymbol{\mathcal{U}}$ can then be characterized as a GNE solution to $\mathcal{G}_\infty$ when no player can improve its objective by unilaterally deviating from this profile. Formally, the following.

Definition 2: A strategy profile $\boldsymbol{u}^\star = (\boldsymbol{u}^{1^\star}, \ldots, \boldsymbol{u}^{N^\star_P}) \in \boldsymbol{\mathcal{U}}$ is a GNE for the game $\mathcal{G}_\infty$ if

$$J^p(\boldsymbol{u}^{p^\star}, \boldsymbol{u}^{-p^\star}) \leq \min_{\boldsymbol{u}^p \in U^p(\boldsymbol{u}^{-p^\star})} J^p(\boldsymbol{u}^p, \boldsymbol{u}^{-p^\star}) \tag{13}$$

holds for every player $p \in \mathcal{P}$.

As before, the set of GNEs that solve $\mathcal{G}_\infty$

$$\Omega_{\mathcal{G}_\infty} := \{\boldsymbol{u}^\star \in \boldsymbol{\mathcal{U}} : \boldsymbol{u}^\star \text{ satisfies (13)}\}$$

is not necessarily a singleton and the game is considered unsolvable if $\Omega_{\mathcal{G}_\infty} = \emptyset$. The following assumptions are taken.

Assumption 2: For each player $p \in \mathcal{P}$ and noise $\boldsymbol{w} \in \boldsymbol{\mathcal{W}}$

1) the cost functional $J^p : \boldsymbol{\mathcal{U}}^1 \times \cdots \boldsymbol{\mathcal{U}}^{N_P} \to \mathbb{R}$ is jointly continuous in all of its arguments and convex in the pth argument, $\boldsymbol{u}^p \in U^p(\boldsymbol{u}^{-p})$, for every sequence $\boldsymbol{u}^{-p} \in \boldsymbol{\mathcal{U}}^{-p}$;
2) the mapping $U^p : \boldsymbol{\mathcal{U}}^{-p} \rightrightarrows \boldsymbol{\mathcal{U}}^p$ takes the form

$$U^p(\boldsymbol{u}^{-p}) := \{\boldsymbol{u}^p \in \boldsymbol{\mathcal{U}}^p : (\boldsymbol{u}^p, \boldsymbol{u}^{-p}) \in \boldsymbol{\mathcal{U}}_\mathcal{G}\}$$

given the global constraint set

$$\boldsymbol{\mathcal{U}}_\mathcal{G} = \{\boldsymbol{u} \in \ell^{N_u}_\infty(\mathbb{N}) : (F_w u)_t \in \mathcal{X},\ u_t \in \mathcal{U}_\mathcal{G},\ t \in \mathbb{N}\}.$$

The sets $\mathcal{U}^p$, $\mathcal{U}_\mathcal{G}$, and $\mathcal{X}$ are all nonempty, compact, and convex. Finally, $\boldsymbol{\mathcal{U}}_\mathcal{G} \cap (\boldsymbol{\mathcal{U}}^1 \times \cdots \times \boldsymbol{\mathcal{U}}^{N_P}) \neq \emptyset$.

These conditions are analogous to those of Assumption 1: They aim to ensure the existence of GNE solutions to $\mathcal{G}_\infty$, that is, $\Omega_{\mathcal{G}_\infty} \neq \emptyset$. Here, the shared constraints $\boldsymbol{\mathcal{U}}_\mathcal{G}$ also require

the players to ensure that state trajectories lie in a feasible set ($x \in \boldsymbol{\mathcal{X}}$) against all realizations of the noise. These constraints describe operational desiderata and/or limitations in the game.

The equilibria $\boldsymbol{u}^\star \in \Omega_{\mathcal{G}_\infty}$ have an open-loop information pattern: Actions $u_t^\star = (u_t^{1^\star}, \dots, u_t^{N_P^\star})$ depend explicitly only on initial state $x_0 \in \mathcal{X}$ and stage-index $t \in \mathbb{N}$. A plan of action with such representation is undesirable, as players become sensible to noise disturbances and decision errors. Conversely, state-feedback policies $\boldsymbol{u}^\star = K(\boldsymbol{x})$, for some $K : \boldsymbol{\mathcal{X}} \to \boldsymbol{\mathcal{U}}$, can detect such errors and provide corrective actions. A feedback (respectively, open-loop) representation of $\boldsymbol{u}^\star \in \Omega_{\mathcal{G}_\infty}$ is thus said to be strongly (weakly) time consistent [1]. In this work, we investigate state-feedback solutions to the game $\mathcal{G}_\infty$.

We consider a closed-loop information pattern and assume that each pth player's actions are represented as

$$\boldsymbol{u}^p := \boldsymbol{K}^p \boldsymbol{x}, \quad \boldsymbol{K}^p : \boldsymbol{x} \mapsto \boldsymbol{\Phi}^{\boldsymbol{p}} * \boldsymbol{x} \tag{14}$$

given a linear causal operator $\boldsymbol{K}^p \in \boldsymbol{\mathcal{C}}^p \subseteq \mathcal{L}(\ell_\infty^{N_x}, \ell_\infty^{N_u^p})$ defined by its convolution kernel $\boldsymbol{\Phi}^p = (\Phi_n^p)_{n\in\mathbb{N}} \in \ell_1(\mathbb{N})$. The sets $\{\boldsymbol{\mathcal{C}}^p\}_{p\in\mathcal{P}}$ describe the operators that satisfy some $\mathcal{G}_\infty$-related restrictions (e.g., information patterns incurred by communication, actuation, and sensing delays). In this setup, players do not plan their actions explicitly but rather by designing a state-feedback policy profile $\boldsymbol{K} := (\boldsymbol{K}^1, \dots, \boldsymbol{K}^{N_P}) \in \boldsymbol{\mathcal{C}}$, with $\boldsymbol{\mathcal{C}} = \boldsymbol{\mathcal{C}}^1 \times \cdots \times \boldsymbol{\mathcal{C}}^{N_P}$. The solution concept that naturally arises is that of a GFNE.

Definition 3: A policy profile $\boldsymbol{K}^\star = (\boldsymbol{K}^{1^\star}, \dots, \boldsymbol{K}^{N_P^\star})$ is a GFNE for $\mathcal{G}_\infty$ if

$$J^p(\boldsymbol{u}^{p^\star}, \boldsymbol{u}^{-p^\star}) \leq \min_{\boldsymbol{u}^p \in U^p(\boldsymbol{u}^{-p^\star})} J^p(\boldsymbol{u}^p, \boldsymbol{u}^{-p^\star}) \tag{15}$$

where $\boldsymbol{u}^\star \in \operatorname{Ker}(I - \boldsymbol{K}^\star \boldsymbol{F}_{\boldsymbol{w}})$, holds for every $p \in \mathcal{P}$.

The set of GFNE that solve $\mathcal{G}_\infty$ is defined as

$$\Omega_{\mathcal{G}_\infty}^{\boldsymbol{K}} := \{\boldsymbol{K}^\star \in \boldsymbol{\mathcal{C}} : \boldsymbol{u}^\star = \boldsymbol{K}^\star \boldsymbol{x}^\star \text{ satisfies } (15)\}.$$

We consider $\boldsymbol{K}^\star \in \Omega_{\mathcal{G}_\infty}^{\boldsymbol{K}}$ to be admissible only if it renders the game stable, that is, if the closed-loop evolution

$$\boldsymbol{x}^\star = \left(I - \boldsymbol{S}_+(\boldsymbol{A} - \sum_{p\in\mathcal{P}} \boldsymbol{B}^p \boldsymbol{K}^{p^\star})\right)^{-1} (\boldsymbol{S}_+ \boldsymbol{w} + \boldsymbol{\delta} x_0)$$

is bounded ($\boldsymbol{x}^\star \in \ell_\infty^{N_x}$) for all bounded noise ($\boldsymbol{w} \in \ell_\infty^{N_x}$). A policy satisfying this requirement is said to be *stabilizing*. From Assumption 2, we have that $\Omega_{\mathcal{G}_\infty}^{\boldsymbol{K}} \neq \emptyset$ when $\boldsymbol{\mathcal{C}} = \boldsymbol{\mathcal{U}}^{\boldsymbol{\mathcal{X}}}$. In the case of $\boldsymbol{\mathcal{C}} \subseteq \mathcal{L}(\ell_\infty^{N_x}, \ell_\infty^{N_u})$, establishing the existence (and, especially, uniqueness) of a solution is demanding [39]. In practice, the set $\Omega_{\mathcal{G}_\infty}^{\boldsymbol{K}}$ could be constructed from open-loop equilibria $\boldsymbol{u}^\star \in \Omega_{\mathcal{G}_\infty}$ by parameterizing the set of all possible trajectories $\{\boldsymbol{x}_{\boldsymbol{w}}^\star = \boldsymbol{F}_{\boldsymbol{w}} \boldsymbol{u}^\star\}_{\boldsymbol{w}\in\boldsymbol{\mathcal{W}}}$, and then, identifying policies $(\boldsymbol{K}^{1^\star}, \dots, \boldsymbol{K}^{N_P^\star})$ that satisfy $\{\boldsymbol{u}^{p^\star} = \boldsymbol{K}^{p^\star} \boldsymbol{x}_{\boldsymbol{w}}^\star\}_{\boldsymbol{w}\in\boldsymbol{\mathcal{W}}}$, $p \in \mathcal{P}$. Highlighting this equivalence, we refer to such $\boldsymbol{u}^\star$ as an *open-loop realization* of the *closed-loop policy* $\boldsymbol{K}^\star$, and vice versa.

BRD for GFNE seeking: The mapping

$$\mathrm{BR}^p(\boldsymbol{u}^{-p}) := \operatorname*{arg\,min}_{\boldsymbol{u}^p \in U^p(\boldsymbol{u}^{-p})} J^p(\boldsymbol{u}^p, \boldsymbol{u}^{-p}) \tag{16}$$

is the best-response of $p \in \mathcal{P}$ to other players' plan of action. Under its feedback representation, $\boldsymbol{u}^p = \boldsymbol{K}^p \boldsymbol{x} \in \mathrm{BR}^p(\boldsymbol{u}^{-p})$ is a solution to the infinite-horizon stochastic control problem

$$\operatorname*{minimize}_{\boldsymbol{u}^p := \boldsymbol{K}^p \boldsymbol{x}} \ \mathrm{E}\left[\sum_{t=0}^{\infty} L^p(x_t, u_t^p, u_t^{-p})\right] \tag{17a}$$

$$\text{subject to} \ \underset{\forall t\in\mathbb{N}}{} \ x_{t+1} = Ax_t + \sum_{\tilde{p}\in\mathcal{P}} B^{\tilde{p}} u_t^{\tilde{p}} + w_t \tag{17b}$$

$$x_t \in \mathcal{X}, \ u_t^p \in \mathcal{U}^p, \ (u_t^p, u_t^{-p}) \in \mathcal{U}_{\mathcal{G}} \tag{17c}$$

$$\boldsymbol{K}^p \in \boldsymbol{\mathcal{C}}^p \tag{17d}$$

$$(x_0 \text{ given}). \tag{17e}$$

While posed in terms of action signals ($\boldsymbol{u}^p$, $p \in \mathcal{P}$), Problem (17) should be interpreted as the direct search for a best-response policy $\boldsymbol{K}^p$ against the (fixed) plan of action from other players, $\boldsymbol{u}^{-p} := (\boldsymbol{u}^{\tilde{p}})_{\tilde{p}\in\mathcal{P}\setminus\{p\}}$. We slightly abuse notation and let $\mathrm{BR}^p(\boldsymbol{K}^{-p})$ be its solutions when parameterized by $\boldsymbol{u}^{-p} := \boldsymbol{K}^{-p}\boldsymbol{x} = (\boldsymbol{K}^{\tilde{p}}\boldsymbol{x})_{\tilde{p}\in\mathcal{P}\setminus\{p\}}$. The mapping $\mathrm{BR} : \boldsymbol{\mathcal{C}} \rightrightarrows \boldsymbol{\mathcal{C}}$, defined by $\mathrm{BR}(\boldsymbol{K}) = \mathrm{BR}^1(\boldsymbol{K}^{-1}) \times \cdots \times \mathrm{BR}^{N_P}(\boldsymbol{K}^{-N_P})$, is the joint best-response to a strategy profile $\boldsymbol{K} \in \boldsymbol{\mathcal{C}}$. The GFNE of $\mathcal{G}_\infty$ thus correspond to the fixed points of this mapping: That is, $\Omega_{\mathcal{G}_\infty}^{\boldsymbol{K}} = \{\boldsymbol{K}^\star \in \boldsymbol{\mathcal{C}} : \boldsymbol{K}^\star \in \mathrm{BR}(\boldsymbol{K}^\star)\}$. Due to constraints $(\mathcal{X}, \mathcal{U}^p, \mathcal{U}_{\mathcal{G}})$ and $\boldsymbol{\mathcal{C}}^p$, an analytical solution to Problem (17) does not exist. Moreover, because infinite-dimensional, its numerical approximation cannot be obtained.

Algorithm 3: BRD for GFNE seeking (BRD-GFNE).

Input: Game $\mathcal{G}_\infty := (\mathcal{P}, \boldsymbol{\mathcal{X}}, \{\boldsymbol{\mathcal{U}}^p\}_{p\in\mathcal{P}}, \boldsymbol{\mathcal{W}}, \{J^p\}_{p\in\mathcal{P}})$
Output: GFNE $\boldsymbol{K}^\star = (\boldsymbol{K}^{1^\star}, \dots, \boldsymbol{K}^{N_P^\star})$

```
Initialize K_0 := (K_0^1, ..., K_0^{N_P}) and k := 0;
for t = 0, 1, 2, ... do
    /* Players apply actions {u^p_{k,t} = K^p_k x_{k,t}}_{p∈P} */
    if K_k ∈ BR(K_k) then return K_k;
    if t = (k+1)ΔT then
        for p ∈ P do
            Update K^p_{k+1} ∈ (1−η)K^p_k + ηBR^p(K^{−p}_k) ;
        k := k + 1;
```

A BRD for GFNE seeking is outlined in Algorithm 3. As $\mathcal{G}_\infty$ is dynamic and stationary, the procedure does not require episodic repetitions of the game. Instead, the learning dynamics occurs simultaneously with the game's execution: Players learn and announce their new policies at stages $t \in \{(k+1)\Delta T\}_{k\in\mathbb{N}}$. $\boldsymbol{K}_k := (\boldsymbol{K}_k^1, \dots, \boldsymbol{K}_k^{N_P})$ denotes the strategy profile after $k \in \mathbb{N}$ updates. The period $\Delta T \geq 1$ defines the rate at which policies are updated, reflecting some communication structure (e.g., the time needed for each $p \in \mathcal{P}$ to collect $\{\boldsymbol{K}_k^{\tilde{p}}\}_{\tilde{p}\in\mathcal{P}\setminus\{p\}}$).

A verbal execution of Algorithm 3 yields the following.

1) The players $p \in \mathcal{P}$ act on $\mathcal{G}_\infty$ according to the policies

$$\boldsymbol{u}_k^p = \boldsymbol{K}_k^p \boldsymbol{x}_k, \quad k \in \mathbb{N}$$

where $\boldsymbol{u}_k^p = (u_t^p)_{t\in\mathcal{T}_k}$ and $\boldsymbol{x}_k = (x_t)_{t\in\mathcal{T}_k}$ are the signals restricted to the interval $\mathcal{T}_k = [k\Delta T, (k+1)\Delta T)$.

2) At $t = (k+1)\Delta T$, every pth player updates its policy

$$\boldsymbol{K}_{k+1}^p \in (1-\eta)\boldsymbol{K}_k^p + \eta \mathrm{BR}^p(\boldsymbol{K}_k^{-p})$$

which is then announced to the other players.

The BRD-GFNE induces an operator $T = (1-\eta)I + \eta\mathrm{BR}$, which is equivalent to the update rule of its static counterpart. Thus, it possesses the same properties: The learning dynamics converge if BR is nonexpansive and the convergence rate is geometric if BR is also a contraction (Lemmas 1 and 2). These properties can also be stated in terms of stage indices $t \in \mathbb{N}$ by replacing $k = \lfloor t/\Delta T \rfloor$. As in the static case, a careful choice of the learning rate $\eta \in (0,1)$ is required to ensure that this fixed-point iteration is well-defined. Finally, we stress that the BRD-GFNE can be interrupted at any $k_f > 1$, thus producing an ϵ-GFNE policy $\boldsymbol{K}_{k_f}$ with associated equilibrium gap $\epsilon > 0$.

III. BRD Via SLS

In this section, we present an approach for GFNE seeking in (stationary) stochastic dynamic games. First, we introduce the system level parameterization of the players' feedback policies ($\boldsymbol{K}^p$, $p \in \mathcal{P}$) and reformulate their best-response mappings (BR^p, $p \in \mathcal{P}$) through finite-dimensional robust optimization problems. Then, a modified BRD-GFNE procedure is proposed and its convergence properties are investigated.

We focus on N_P-player linear-quadratic stochastic games $\mathcal{G}_\infty^{\mathrm{LQ}} = (\mathcal{P}, \boldsymbol{\mathcal{X}}, \{\boldsymbol{\mathcal{U}}^p\}_{p\in\mathcal{P}}, \boldsymbol{\mathcal{W}}, \{J^p\}_{p\in\mathcal{P}})$ with dynamics

$$x_{t+1} = Ax_t + \sum_{p\in\mathcal{P}} B^p u_t^p + w_t, \quad x_0 \text{ given} \tag{18}$$

and objective functionals

$$J^p(\boldsymbol{u}^p, \boldsymbol{u}^{-p}) = \mathrm{E}\left[\sum_{t=0}^{\infty}\left(\|C^p x_t\|_2^2 + \|\textstyle\sum_{\tilde{p}\in\mathcal{P}} D^{p\tilde{p}} u_t^{\tilde{p}}\|_2^2\right)\right] \tag{19}$$

defined by matrices $C^p \in \mathbb{R}^{N_z \times N_x}$ and $D^{p\tilde{p}} \in \mathbb{R}^{N_z \times N_u^{\tilde{p}}}$ with dimension $N_z \geq N_x + N_u$. The following assumptions ensure that stabilising GFNE solutions to $\mathcal{G}_\infty^{\mathrm{LQ}}$ exist.

Assumption 3: For each player $p \in \mathcal{P}$,

1) the pair (A, B^p) is stabilizable;
2) the pair (C^p, A) is detectable;
3) the matrix D^{pp} is full column rank, i.e., $D^{pp\intercal} D^{pp} \in \mathbb{S}_{++}^{N_u^p}$. Moreover, $D^{p\tilde{p}\intercal} C^p = 0 = C^{p\intercal} D^{p\tilde{p}}$ for all $\tilde{p} \in \mathcal{P}$.

Finally, the sets $\mathcal{X}$, $\mathcal{U}^p$ ($\forall p$), and $\mathcal{U}_\mathcal{G}$, are convex polyhedra satisfying $0 \in \mathbf{relint}\,\mathcal{X}$, $0 \in \mathbf{relint}\,\mathcal{U}^p$, and $0 \in \mathbf{relint}\,\mathcal{U}_\mathcal{G}$.

The class $\mathcal{G}_\infty^{\mathrm{LQ}}$ describe problems in which N_P noncooperative agents have to agree on stationary policies that jointly stabilize a global system, robustly to the noise process, while penalizing state- and input-deviations differently. While representative of many practically relevant problems, this choice is not restrictive. Our derivations should follow similarly for any collection of cost functions $\{L^p\}_{p\in\mathcal{P}}$ satisfying Assumption 2.

A. System-Level Best-Response Mappings

SLS [35] is a novel methodology for controller design centered on the equivalent representation of control policies in terms of the closed-loop responses that they achieve. Unlike similar approaches, such as the Youla [40] and input–output [41] parameterizations, SLS allows the synthesis of state-feedback policies to be posed as the solution to convex optimization problems, even when subjected to constraints on the state- and input-signals, and on the structure of the policy itself. In this section, we present a system-level parameterization for the best-response mappings in $\mathcal{G}_\infty^{\mathrm{LQ}}$.

We start by assuming a stabilizing profile $(\boldsymbol{K}^1, \ldots, \boldsymbol{K}^{N_P})$, guaranteed by Assumption 3. Each policy is associated with a transfer matrix $\hat{\boldsymbol{K}}^p \in \mathcal{RH}_\infty$, $\hat{\boldsymbol{K}}^p = \sum_{n=0}^{\infty} \frac{1}{z^n}\Phi_n^p$, which defines the state-feedback $\hat{\boldsymbol{u}}^p = \hat{\boldsymbol{K}}^p \hat{\boldsymbol{x}}$ in the frequency domain. Considering the linear dynamics (18)

$$z\hat{\boldsymbol{x}} = A\hat{\boldsymbol{x}} + \sum_{p\in\mathcal{P}} B^p \hat{\boldsymbol{u}}^p + \hat{\boldsymbol{w}} \tag{20a}$$

$$\hat{\boldsymbol{u}}^p = \hat{\boldsymbol{K}}^p \hat{\boldsymbol{x}}, \quad (\forall p \in \mathcal{P}) \tag{20b}$$

the signals $(\hat{\boldsymbol{x}}, \hat{\boldsymbol{u}}^1, \ldots, \hat{\boldsymbol{u}}^{N_P})$ can be expressed in terms of $\hat{\boldsymbol{w}}$

$$\begin{bmatrix} \hat{\boldsymbol{x}} \\ \hat{\boldsymbol{u}}^1 \\ \vdots \\ \hat{\boldsymbol{u}}^{N_P} \end{bmatrix} = \begin{bmatrix} \hat{\boldsymbol{\Phi}}_{\boldsymbol{x}} \\ \hat{\boldsymbol{\Phi}}_{\boldsymbol{u}}^1 \\ \vdots \\ \hat{\boldsymbol{\Phi}}_{\boldsymbol{u}}^{N_P} \end{bmatrix} \hat{\boldsymbol{w}} \tag{21}$$

where $\hat{\boldsymbol{\Phi}}_{\boldsymbol{x}} = (zI - A - \sum_{p\in\mathcal{P}} B^p \hat{\boldsymbol{K}}^p)^{-1}$ and $\hat{\boldsymbol{\Phi}}_{\boldsymbol{u}}^p = \hat{\boldsymbol{K}}^p \hat{\boldsymbol{\Phi}}_{\boldsymbol{x}}$ ($p \in \mathcal{P}$). The introduced transfer matrices $(\hat{\boldsymbol{\Phi}}_{\boldsymbol{x}}, \hat{\boldsymbol{\Phi}}_{\boldsymbol{u}}^1, \ldots, \hat{\boldsymbol{\Phi}}_{\boldsymbol{u}}^{N_P})$ are referred to as *system level responses* or *closed-loop maps*. Under this representation, the following result holds.

Theorem 1 (System level parameterization): Consider the dynamics (20) under state-feedback $\hat{\boldsymbol{u}}^p = \hat{\boldsymbol{K}}^p \hat{\boldsymbol{x}}$ ($\forall p \in \mathcal{P}$). The following statements are true.

1) The affine space

$$\begin{bmatrix} zI - A & -B^1 & \cdots & -B^{N_P} \end{bmatrix} \begin{bmatrix} \hat{\boldsymbol{\Phi}}_{\boldsymbol{x}} \\ \hat{\boldsymbol{\Phi}}_{\boldsymbol{u}}^1 \\ \vdots \\ \hat{\boldsymbol{\Phi}}_{\boldsymbol{u}}^{N_P} \end{bmatrix} = I \tag{22}$$

with $\hat{\boldsymbol{\Phi}}_{\boldsymbol{x}}, \hat{\boldsymbol{\Phi}}_{\boldsymbol{u}}^1, \ldots, \hat{\boldsymbol{\Phi}}_{\boldsymbol{u}}^{N_P} \in \frac{1}{z}\mathcal{RH}_\infty$, parameterizes all system responses from $\hat{\boldsymbol{w}}$ to $(\hat{\boldsymbol{x}}, \hat{\boldsymbol{u}}^1, \ldots, \hat{\boldsymbol{u}}^{N_P})$ achievable by internally stabilising policies $(\hat{\boldsymbol{K}}^1, \ldots, \hat{\boldsymbol{K}}^{N_P})$.

2) Any response $(\hat{\boldsymbol{\Phi}}_{\boldsymbol{x}}, \hat{\boldsymbol{\Phi}}_{\boldsymbol{u}}^1, \ldots, \hat{\boldsymbol{\Phi}}_{\boldsymbol{u}}^{N_P})$ satisfying (22) is achieved by the policies $\hat{\boldsymbol{K}}^p = \hat{\boldsymbol{\Phi}}_{\boldsymbol{u}}^p \hat{\boldsymbol{\Phi}}_{\boldsymbol{x}}^{-1}$ ($\forall p \in \mathcal{P}$), which are internally stabilising and can be implemented as

$$z\hat{\boldsymbol{\xi}} = \tilde{\boldsymbol{\Phi}}_{\boldsymbol{x}} \hat{\boldsymbol{\xi}} + \hat{\boldsymbol{x}} \tag{23a}$$

$$\hat{\boldsymbol{u}}^p = \tilde{\boldsymbol{\Phi}}_{\boldsymbol{u}}^p \hat{\boldsymbol{\xi}} \tag{23b}$$

with $\tilde{\boldsymbol{\Phi}}_{\boldsymbol{x}} = z(I - \hat{\boldsymbol{\Phi}}_{\boldsymbol{x}})$ and $\tilde{\boldsymbol{\Phi}}_{\boldsymbol{u}}^p = z\hat{\boldsymbol{\Phi}}_{\boldsymbol{u}}^p$ (see Fig. 2).

Proof: Defining $B := [B^1 \;\; B^2 \;\; \cdots \;\; B^{N_P}]$, and transfer matrices $\hat{\boldsymbol{\Phi}}_{\boldsymbol{u}} := \mathtt{col}(\hat{\boldsymbol{\Phi}}_{\boldsymbol{u}}^1, \ldots, \hat{\boldsymbol{\Phi}}_{\boldsymbol{u}}^{N_P})$ and $\hat{\boldsymbol{K}} := \mathtt{col}(\hat{\boldsymbol{K}}^1, \ldots, \hat{\boldsymbol{K}}^{N_P})$, the proof is as in [35, Thm. 4.1]. In the second statement, we consider an alternative representation $\hat{\boldsymbol{K}} = \tilde{\boldsymbol{\Phi}}_{\boldsymbol{u}}(zI - \tilde{\boldsymbol{\Phi}}_{\boldsymbol{x}})^{-1}\tilde{\boldsymbol{\Phi}}_{\boldsymbol{y}}$ with $\tilde{\boldsymbol{\Phi}}_{\boldsymbol{x}} = z(I - \hat{\boldsymbol{\Phi}}_{\boldsymbol{x}})$, $\tilde{\boldsymbol{\Phi}}_{\boldsymbol{u}} = z\hat{\boldsymbol{\Phi}}_{\boldsymbol{u}}$,

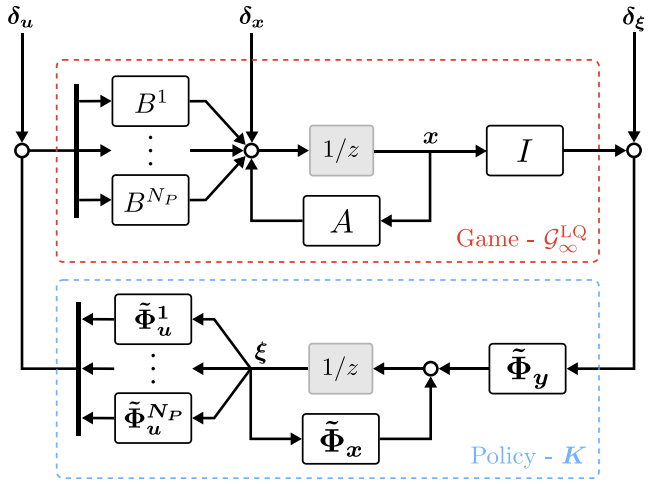

Fig. 2. Feedback structure for the policy $\hat{K} = \tilde{\Phi}_u(zI - \tilde{\Phi}_x)^{-1}\tilde{\Phi}_y = \hat{\Phi}_u^p \hat{\Phi}_x^{-1}$, equivalent to the internal representation in (23).

and $\tilde{\Phi}_y = I$. This leads to the transfer matrices from $(\hat{\delta}_x, \hat{\delta}_u, \hat{\delta}_\xi)$ to $(\hat{x}, \hat{u}, \hat{\xi})$

$$\begin{bmatrix} \hat{x} \\ \hat{u} \\ \hat{\xi} \end{bmatrix} = \begin{bmatrix} \hat{\Phi}_x & \hat{\Phi}_x B & \hat{\Phi}_x(zI - A) \\ \hat{\Phi}_u & I + \hat{\Phi}_u B & \hat{\Phi}_u(zI - A) \\ \frac{1}{z} I & \frac{1}{z} B & \frac{1}{z}(zI - A) \end{bmatrix} \begin{bmatrix} \hat{\delta}_x \\ \hat{\delta}_u \\ \hat{\delta}_\xi \end{bmatrix} \tag{24}$$

which are all stable due to $\hat{\Phi}_x, \hat{\Phi}_u \in \frac{1}{z}\mathcal{RH}_\infty$. Thus, the policy $\hat{K} = (\hat{K}^1, \hat{K}^2, \dots, \hat{K}^{N_P})$ is internally stabilising. □

We refer to the system level responses through their kernels, $\Phi_x = (\Phi_{x,n})_{n\in\mathbb{N}} \in \ell_2(\mathbb{N})$ and $\Phi_u^p = (\Phi_{u,n}^p)_{n\in\mathbb{N}} \in \ell_2(\mathbb{N})$, for all $p \in \mathcal{P}$. Due to strict causality, $\Phi_{x,0} = 0$ and $\Phi_{u,0}^p = 0$. From Theorem 1, the operators $\{K^p \in \mathcal{C}^p\}_{p\in\mathcal{P}}$ and the transfer matrices $\{\hat{K}^p \in \mathcal{RH}_\infty\}_{p\in\mathcal{P}}$ are equivalent representations of the feedback policies. Hence, provided there is no confusion, we use exclusively the first notation. In particular, $K^p = \Phi_u^p \Phi_x^{-1}$ $(p \in \mathcal{P})$ denotes the policy parameterized by $(\hat{\Phi}_x, \hat{\Phi}_u^p)$ and $K = (\Phi_u^1, \dots, \Phi_u^{N_P})\Phi_x^{-1}$ denotes the corresponding profile. A time-domain characterization of K is given in the following.

Corollary 1.1: A policy $K^p = \Phi_u^p \Phi_x^{-1}$ $(p \in \mathcal{P})$ is defined by the kernel $\Phi^p = \Phi_u^p * \Phi_x^{-1}$, and can be implemented as

$$\xi_t = -\sum_{\tau=1}^{t} \Phi_{x,\tau+1}\xi_{t-\tau} + x_t \tag{25a}$$

$$u_t^p = \sum_{\tau=0}^{t} \Phi_{u,\tau+1}^p \xi_{t-\tau} \tag{25b}$$

using an auxiliary *internal state* $\xi = (\xi_n)_{n\in\mathbb{N}}$ with $\xi_0 = x_0$.

Proof: The statement $\Phi^p = \Phi_u^p * \Phi_x^{-1}$ follows directly from the inverse $\mathcal{Z}$-Transform of $K^p = \Phi_u^p \Phi_x^{-1}$ and $\Phi^p = (\Phi_n^p)_{n\in\mathbb{N}}$ being the kernel of K^p. The operations (25) are obtained as the inverse $\mathcal{Z}$-Transform of (23) and the fact that $\mathcal{Z}^{-1}[z(I - \hat{\Phi}_x)\hat{\xi}] = \xi_{t+1} - \xi_t - \sum_{\tau=2}^{t+1} \Phi_{x,\tau}\xi_{t+1-\tau}$. □

The system level parameterization enables a methodology for policy synthesis consisting of searching the space of stabilising policies (in)directly through (Φ_x, Φ_u^p), $p \in \mathcal{P}$. In particular, this parameterization can be leveraged to reformulate the best-response mappings in $\mathcal{G}_\infty^{\mathrm{LQ}}$ as numerically tractable problems. In this direction, consider that players design stabilising policies $K = (K^1, \dots, K^{N_P})$ by choosing their desired system level responses $\Phi_u = (\Phi_u^1, \dots, \Phi_u^{N_P})$, simultaneously. From the affine space (22), the signal Φ_x, common to all players, satisfies the (deterministic) linear dynamics

$$\Phi_{x,n+1} = A\Phi_{x,n} + \sum_{p\in\mathcal{P}} B^p \Phi_{u,n}^p, \quad \Phi_{x,1} = I_{N_x} \tag{26}$$

or $\Phi_x = F_\Phi \Phi_u$ given the causal affine operator

$$F_\Phi : \Phi_u \mapsto (I - S_+ A)^{-1} \left(\sum_{p\in\mathcal{P}} S_+ B^p \Phi_u^p + \delta I_{N_x} \right). \tag{27}$$

Using the system level responses, the objective functionals of $\mathcal{G}_\infty^{\mathrm{LQ}}$ can be shown as equivalent to the functional

$$J^p(\Phi_u^p, \Phi_u^{-p}) = \sum_{n=1}^{\infty} \left(\|C^p \Phi_{x,n} \Sigma_w^{\frac{1}{2}}\|_F^2 + \| \sum_{\tilde{p}\in\mathcal{P}} D^{p\tilde{p}} \Phi_{u,n}^{\tilde{p}} \Sigma_w^{\frac{1}{2}}\|_F^2 \right).$$

The game $\mathcal{G}_\infty^{\mathrm{LQ}}$ thus induces a *system-level* dynamic game

$$\mathcal{G}_\infty^\Phi := (\mathcal{P}, \mathcal{C}_x, \{\mathcal{C}_u^p\}_{p\in\mathcal{P}}, \mathcal{W}, \{J^p\}_{p\in\mathcal{P}}) \tag{28}$$

defining the problem in which players $p \in \mathcal{P}$ each plans a closed-loop response $\Phi_u^p \in U_\Phi^p(\Phi_u^{-p}) \subseteq \mathcal{C}_u^p$ to minimize its individual cost functional $J^p : \mathcal{C}_u^1 \times \cdots \times \mathcal{C}_u^{N_P} \to \mathbb{R}$. Here, the set-valued mappings $U_\Phi^p : \mathcal{C}_u^{-p} \rightrightarrows \mathcal{C}_u^p$ are defined as

$$U_\Phi^p(\Phi_u^{-p}) := \{\Phi_u^p \in \mathcal{C}_u^p : F_\Phi \Phi_u \in \mathcal{C}_x, \; \Phi_u^p * w \in U^p(\Phi_u^{-p} * w)\}$$

which incorporate the constraints U^p from the original $\mathcal{G}_\infty^{\mathrm{LQ}}$. The sets $(\mathcal{C}_x, \mathcal{C}_u^p)$ are designed to enforce the policy constraints $K^p \in \mathcal{C}^p$ directly through the kernels (Φ_x, Φ_u^p): They are related as $\mathcal{C}_u^p = \{K^p \mathcal{C}_x : K^p \in \mathcal{C}^p\}$. We refer to a joint response $\Phi_u = (\Phi_u^1, \dots, \Phi_u^{N_P}) \in \mathcal{C}_u, \mathcal{C}_u = \mathcal{C}_u^1 \times \cdots \times \mathcal{C}_u^{N_P}$, as a system-level strategy profile. Finally, the set of (open-loop) system-level GNEs for this game is denoted as $\Omega_{\mathcal{G}_\infty^\Phi}$.

The best-response mappings for $\mathcal{G}_\infty^\Phi$ take the form

$$\mathrm{BR}_\Phi^p(\Phi_u^{-p}) := \arg\min_{\Phi_u^p \in U_\Phi^p(\Phi_u^{-p})} J^p(\Phi_u^p, \Phi_u^{-p})$$

consisting of the set of closed-loop maps Φ_u^p, which are best responses to the maps of other players, $\Phi_u^{-p} = (\Phi_u^{\tilde{p}})_{\tilde{p}\in\mathcal{P}\setminus\{p\}}$. They are solutions to the SLS problem

$$\underset{\Phi_u^p}{\text{minimize}} \quad \sum_{n=1}^{\infty} \left(\|C^p \Phi_{x,n} \Sigma_w^{\frac{1}{2}}\|_F^2 + \Big\| \sum_{\tilde{p}\in\mathcal{P}} D^{p\tilde{p}} \Phi_{u,n}^{\tilde{p}} \Sigma_w^{\frac{1}{2}} \Big\|_F^2 \right) \tag{29a}$$

$$\text{subject to} \quad \Phi_{x,n+1} = A\Phi_{x,n} + \textstyle\sum_{\tilde{p}\in\mathcal{P}} B^{\tilde{p}} \Phi_{u,n}^{\tilde{p}} \tag{29b}$$

$$\forall n \in \mathbb{N}^+ \quad (\Phi_x * w)_n \in \mathcal{X}, \; (\Phi_u^p * w)_n \in \mathcal{U}^p$$

$$\left((\Phi_u^p * w)_n, (\Phi_u^{-p} * w)_n\right) \in \mathcal{U}_\mathcal{G} \tag{29c}$$

$$\Phi_x \in \mathcal{C}_x, \quad \Phi_u^p \in \mathcal{C}_u^p \tag{29d}$$

$$\Phi_{x,1} = I_{N_x}. \tag{29e}$$

Collectively, the mapping $\mathrm{BR}_\Phi(\boldsymbol{\Phi_u}) = \mathrm{BR}^1_\Phi(\boldsymbol{\Phi}^{-1}_{\boldsymbol{u}}) \times \cdots \times \mathrm{BR}^{N_P}_\Phi(\boldsymbol{\Phi}^{-N_P}_{\boldsymbol{u}})$, is the joint best-response to a system-level strategy profile $\boldsymbol{\Phi_u}$. The GNEs of $\mathcal{G}^\Phi_\infty$ are equivalent to the fixed points of this map, $\Omega_{\mathcal{G}^\Phi_\infty} = \{\boldsymbol{\Phi}^\star_{\boldsymbol{u}} \in \boldsymbol{\mathcal{C}}_u : \boldsymbol{\Phi}^\star_{\boldsymbol{u}} \in \mathrm{BR}_\Phi(\boldsymbol{\Phi}^\star_{\boldsymbol{u}})\}$. Considering how $\mathcal{G}^{\mathrm{LQ}}_\infty$ induces $\mathcal{G}^\Phi_\infty$, a relationship can be established between the best-responses BR and BR_Φ, and consequently, between their fixed-points, $\Omega^{\boldsymbol{K}}_{\mathcal{G}^{\mathrm{LQ}}_\infty}$ and $\Omega_{\mathcal{G}^\Phi_\infty}$. In Section III-B, we formalize this relationship and propose a learning dynamics for GFNE seeking based on the system-level best-response mappings.

The best-responses $\{\mathrm{BR}^p_\Phi\}_{p\in\mathcal{P}}$ are still intractable: They are defined by infinite-dimensional problems with no general solution and that require full knowledge of the noise process $(w_n)_{n\in\mathbb{N}}$ to formulate the constraints U^p_Φ.

In the following, we tackle both issues and provide a class of finite-dimensional robust optimization problems that approximate Problem (29). We conclude the section by presenting a class of system level constraints that enforce a richer feedback information pattern.

1) Finite-Horizon Approximation: The programs in $\{\mathrm{BR}^p_\Phi\}_{p\in\mathcal{P}}$ can be made finite dimensional by restricting the system level responses to the set of finite-impulse responses (FIR)

$$\boldsymbol{\mathcal{C}}_x = \{\boldsymbol{\Phi_x} \in \ell_2[0,N] : \Phi_{x,n} \in \mathcal{C}_{x,n},\ n\in[0,N),\ \Phi_{x,N} = 0\}$$

$$\boldsymbol{\mathcal{C}}^p_u = \{\boldsymbol{\Phi}^p_{\boldsymbol{u}} \in \ell_2[0,N) : \Phi^p_{u,n} \in \mathcal{C}^p_{u,n},\ n\in[0,N)\}$$

given horizon $N \in (1,\infty)$. We enforce $\boldsymbol{\Phi_x} \in \boldsymbol{\mathcal{C}}_x$ and $\boldsymbol{\Phi}^p_{\boldsymbol{u}} \in \boldsymbol{\mathcal{C}}^p_u$ in Problem (29) by adding a terminal constraint $\Phi_{x,N} = 0$, then letting $\boldsymbol{\Phi_x} = (\Phi_{x,n} \in \mathcal{C}_{x,n})_{n=1}^N$ and $\boldsymbol{\Phi}^p_{\boldsymbol{u}} = (\Phi^p_{u,n} \in \mathcal{C}^p_{u,n})_{n=1}^{N-1}$. The sets $\mathcal{C}_{x,n} \subseteq \mathbb{R}^{N_x\times N_x}$ and $\mathcal{C}^p_{u,n} \subseteq \mathbb{R}^{N^p_u\times N_x}$ are required only to be compact convex sets. Under this condition, Problem (29) is finite dimensional with $(N-1)N_xN^p_u$ decision variables (the entries of $\Phi^p_{u,1},\dots,\Phi^p_{u,N-1} \in \mathbb{R}^{N^p_u\times N_x}$), and thus, can be solved numerically. We remark that the policies $\boldsymbol{K}^p = \boldsymbol{\Phi}^p_{\boldsymbol{u}}\boldsymbol{\Phi}^{-1}_{\boldsymbol{x}}$ ($p\in\mathcal{P}$) are still solutions to the infinite-horizon Problem (17) regardless of the closed-loop maps $\{\boldsymbol{\Phi_x}, \boldsymbol{\Phi}^p_{\boldsymbol{u}}\}$ being FIR.

Although realizing Problem (29) into a tractable program, the constraint $\Phi_{x,N} = 0$ is only feasible when the pair (A, B^p) is full-state controllable. This is a difficult requirement in multiagent settings, as often $N^p_u \ll N_x$ for all $p\in\mathcal{P}$, leading to overdetermined problems. Furthermore, enforcing FIR constraints is known to result in *deadbeat* policies: Control actions are excessively large in magnitude for small $N < \infty$. Alternatively, we restrict the system level responses to the sets

$$\boldsymbol{\mathcal{C}}_x = \{\boldsymbol{\Phi_x}\in\ell_2[0,N] : \Phi_{x,n}\in\mathcal{C}_{x,n},\ n\in[0,N),\ \|\Phi_{x,N}\|_F^2 \le \gamma\}$$

$$\boldsymbol{\mathcal{C}}^p_u = \{\boldsymbol{\Phi}^p_{\boldsymbol{u}}\in\ell_2[0,N) : \Phi^p_{u,n}\in\mathcal{C}^p_{u,n},\ n\in[0,N)\}$$

with $\|\Phi_{x,N}\|_F^2 = \sum_i \sigma_i(\Phi_{x,N})^2 \le \gamma$ for some factor $\gamma\in(0,1)$ and $\sigma_i(\cdot)$ denoting the ith largest singular value of a matrix. These are denoted as the set of (soft) FIR for $N > 0$. The pth player's best-response map thus corresponds to the problem

$$\underset{\boldsymbol{\Phi}^p_{\boldsymbol{u}}}{\text{minimize}} \quad \sum_{n=1}^{N-1}\left(\|C^p\Phi_{x,n}\Sigma_w^{\frac12}\|_F^2 + \Big\|\sum_{\tilde p\in\mathcal{P}} D^{p\tilde p}\Phi^{\tilde p}_{u,n}\Sigma_w^{\frac12}\Big\|_F^2\right) + \|C^p\Phi_{x,N}\Sigma_w^{\frac12}\|_F^2 \tag{30a}$$

$$\text{subject to}\ \underset{\forall n\in[1,N)}{} \quad \Phi_{x,n+1} = A\Phi_{x,n} + \textstyle\sum_{\tilde p\in\mathcal{P}} B^{\tilde p}\Phi^{\tilde p}_{u,n} \tag{30b}$$

$$(\Phi_x * w)_n \in \mathcal{X},\ (\Phi^p_u * w)_n \in \mathcal{U}^p,\ \big((\Phi^p_u * w)_n, (\Phi^{-p}_u * w)_n\big) \in \mathcal{U}_{\mathcal{G}} \tag{30c}$$

$$\Phi_{x,n} \in \mathcal{C}_{x,n},\ \Phi^p_{u,n} \in \mathcal{C}^p_{u,n} \tag{30d}$$

$$\Phi_{x,1} = I_{N_x},\ \|\Phi_{x,N}\|_F^2 \le \gamma. \tag{30e}$$

The solutions to Problem (30) approximate those of the infinite-horizon Problem (29): With respect to N, the performance of the former converges to that achieved by the latter [35]. In this case, feasibility only requires (A, B^p) stabilizable and a sufficiently large horizon N to ensure that $\|\Phi_{x,N}\|_F^2 \le \gamma$ is achievable for some $\boldsymbol{\Phi}^p_{\boldsymbol{u}} \in U^p_\Phi(\boldsymbol{\Phi}^{-p}_{\boldsymbol{u}})$. Computationally, this is still a finite-dimensional convex problem, which can be solved numerically. Here, we let $\widehat{\mathrm{BR}}^p_\Phi : \boldsymbol{\mathcal{C}}^{-p}_u \rightrightarrows \boldsymbol{\mathcal{C}}^p_u$ be the solutions of Problem (30) parameterized by $\boldsymbol{\Phi}^{-p}_{\boldsymbol{u}}$. The map $\widehat{\mathrm{BR}}_\Phi : \boldsymbol{\mathcal{C}}_u \rightrightarrows \boldsymbol{\mathcal{C}}_u$, $\widehat{\mathrm{BR}}_\Phi(\boldsymbol{\Phi_u}) = \widehat{\mathrm{BR}}^1_\Phi(\boldsymbol{\Phi}^{-1}_{\boldsymbol{u}}) \times \cdots \times \widehat{\mathrm{BR}}^{N_P}_\Phi(\boldsymbol{\Phi}^{-N_P}_{\boldsymbol{u}})$, is the joint *(approximately)best-response* to the system-level profile $\boldsymbol{\Phi_u}$.

The complexity of Problem (30) is agnostic to the number of players: The effect of other players' strategies can always be condensed as affine terms (e.g., $z^{-p}_n = \sum_{\tilde p\in\mathcal{P}\setminus\{p\}} B^{\tilde p}\Phi^{\tilde p}_{u,n}$). Conversely, it scales quickly with the FIR horizon N and model dimensions N_x and N^p_u. We remark, however, that this problem is still convex and can be solved efficiently by exploiting its structure using techniques from real-time optimization of linear-quadratic control problems [42]. Moreover, if the constraints (30c)–(30e) are *column-separable* (see [35]), then Problem (30) can be decomposed into smaller subproblems to be solved in parallel, substantially reducing its computational costs.

Conversely to BR_Φ, the fixed points of $\widehat{\mathrm{BR}}_\Phi$ do not coincide with the set of GNEs $\Omega_{\mathcal{G}^\Phi_\infty}$, but are rather contained in the set of ε-GNEs $\Omega^\varepsilon_{\mathcal{G}^\Phi_\infty}$ for some equilibrium gap $\epsilon > 0$. This is clear from the fact that the original Problem (29) and the approximation Problem (30) have different optimal values. Under certain conditions, this fact can be shown explicitly.

Theorem 2: Consider a fixed point $\boldsymbol{\Phi}^\varepsilon_{\boldsymbol{u}} \in \widehat{\mathrm{BR}}_\Phi(\boldsymbol{\Phi}^\varepsilon_{\boldsymbol{u}})$ and assume that $\|\Phi^\star_{x,N}\|_F^2 \le \gamma$ for $\boldsymbol{\Phi}^\star_{\boldsymbol{x}} = \boldsymbol{F_\Phi}\boldsymbol{\Phi}^\star_{\boldsymbol{u}}$ obtained from the original best response $\boldsymbol{\Phi}^\star_{\boldsymbol{u}} \in \mathrm{BR}_\Phi(\boldsymbol{\Phi}^\varepsilon_{\boldsymbol{u}})$. Then, the profile $\boldsymbol{\Phi}^\varepsilon_{\boldsymbol{u}} = (\boldsymbol{\Phi}^{1^\varepsilon}_{\boldsymbol{u}},\dots,\boldsymbol{\Phi}^{N^\varepsilon_P}_{\boldsymbol{u}})$ is an ε-GNE of $\mathcal{G}^\Phi_\infty$ satisfying

$$J^p\left(\boldsymbol{\Phi}^{p^\varepsilon}_{\boldsymbol{u}}, \boldsymbol{\Phi}^{-p^\varepsilon}_{\boldsymbol{u}}\right) \le \min_{\boldsymbol{\Phi}^p_{\boldsymbol{u}}\in U^p_\Phi(\boldsymbol{\Phi}^{-p^\varepsilon}_{\boldsymbol{u}})} J^p\left(\boldsymbol{\Phi}^p_{\boldsymbol{u}}, \boldsymbol{\Phi}^{-p^\varepsilon}_{\boldsymbol{u}}\right) + \varepsilon \tag{31}$$

with $\varepsilon = \max_{p\in\mathcal{P}} \gamma J^p(\boldsymbol{\Phi}^{p^\varepsilon}_{\boldsymbol{u}}, \boldsymbol{\Phi}^{-p^\varepsilon}_{\boldsymbol{u}})$ for every player $p\in\mathcal{P}$.

Proof: Let $\boldsymbol{\Phi}^\star_{\boldsymbol{u}} \in \mathrm{BR}_\Phi(\boldsymbol{\Phi}^\varepsilon_{\boldsymbol{u}})$ and consider a candidate fixed-point $\tilde{\boldsymbol{\Phi}}^\varepsilon_{\boldsymbol{u}} = (\boldsymbol{\Phi}^\star_{u,n})_{n=1}^{N-1}$. This profile satisfies the constraints in $\widehat{\mathrm{BR}}^p_\Phi$ by construction and $(\Phi^\star_{x,n})_{n=1}^N = \tilde{\boldsymbol{\Phi}}^\varepsilon_{\boldsymbol{x}} = F_{\boldsymbol{\Phi}}\tilde{\boldsymbol{\Phi}}^\varepsilon_{\boldsymbol{u}}$ satisfies $\|\Phi^\star_{x,N}\|_F^2 \le \gamma$ by our assumption. From optimality, $J^p(\boldsymbol{\Phi}^{p^\varepsilon}_{\boldsymbol{u}}, \boldsymbol{\Phi}^{-p^\varepsilon}_{\boldsymbol{u}}) \le J^p(\tilde{\boldsymbol{\Phi}}^{p^\varepsilon}_{\boldsymbol{u}}, \boldsymbol{\Phi}^{-p^\varepsilon}_{\boldsymbol{u}}) \le \frac{1}{1-\gamma} J^p(\boldsymbol{\Phi}^\star_{\boldsymbol{u}}, \boldsymbol{\Phi}^{-p^\varepsilon}_{\boldsymbol{u}})$, with the second inequality resulting from the quadratic objective being larger for the infinite-horizon response $\Phi^\star_{\boldsymbol{u}}$ and $\frac{1}{1-\gamma} > 1$. Finally, $\boldsymbol{\Phi}^{p^\star}_{\boldsymbol{u}} = \arg\min_{\boldsymbol{\Phi}^p_{\boldsymbol{u}}\in U^p_\Phi(\boldsymbol{\Phi}^{-p^\varepsilon}_{\boldsymbol{u}})} J^p(\boldsymbol{\Phi}^p_{\boldsymbol{u}}, \boldsymbol{\Phi}^{-p^\varepsilon}_{\boldsymbol{u}})$ by definition and thus (31) follows by letting $\varepsilon = \max_{p\in\mathcal{P}} \gamma J^p(\boldsymbol{\Phi}^{p^\varepsilon}_{\boldsymbol{u}}, \boldsymbol{\Phi}^{-p^\varepsilon}_{\boldsymbol{u}})$. □

The equilibrium gap associated with $\boldsymbol{\Phi}^\varepsilon_{\boldsymbol{u}} \in \widehat{\mathrm{BR}}_\Phi(\boldsymbol{\Phi}^\varepsilon_{\boldsymbol{u}})$ thus depends on the choice of parameter γ, assuming that the terminal

constraint (30e) also holds for solutions to the original Problem (29). Hereafter, $U_\Phi^p : \boldsymbol{\mathcal{C}}_u^{-p} \rightrightarrows \boldsymbol{\mathcal{C}}_u^p$ $(\forall p)$ are assumed to incorporate the sets $(\boldsymbol{\mathcal{C}}_x, \boldsymbol{\mathcal{C}}_u^p)$ of (soft-constrained) FIR approximations as defined above for a $N > 1$.

2) Robust Operational Constraints: From Assumption 3, the sets $\mathcal{X}$, $\mathcal{U}^p$ $(p \in \mathcal{P})$, and $\mathcal{U}_\mathcal{G}$, can be expressed by linear inequalities

$$\mathcal{X} = \{x_t \in \mathbb{R}^{N_x} : G_x x_t \preceq \mathbf{1}_{N_\mathcal{X}}\}$$
$$\mathcal{U}^p = \{u_t^p \in \mathbb{R}^{N_u^p} : G_u^p u_t^p \preceq \mathbf{1}_{N_{\mathcal{U}^p}}\}$$
$$\mathcal{U}_\mathcal{G} = \{u_t \in \textstyle\prod_{p\in\mathcal{P}} \mathbb{R}^{N_u^p} : G_\mathcal{G} u_t \preceq \mathbf{1}_{N_{\mathcal{U}_\mathcal{G}}}\}$$

given some matrices $G_x \in \mathbb{R}^{N_\mathcal{X}\times N_x}$, $G_u^p \in \mathbb{R}^{N_{\mathcal{U}^p}\times N_u^p}$, and $G_\mathcal{G} \in \mathbb{R}^{N_{\mathcal{U}_\mathcal{G}}\times N_u}$. The map $U^p(u^{-p})$ is then equivalent to the actions $\boldsymbol{u}^p = \boldsymbol{\Phi}_{\boldsymbol{u}}^p * \boldsymbol{w}$ whose associated response $\boldsymbol{\Phi}_{\boldsymbol{u}}^p$ satisfy

$$[G_x]_i(\Phi_x * w)_n \le 1, \quad i = 1, \dots, N_\mathcal{X} \tag{32}$$
$$[G_u^p]_j(\Phi_u^p * w)_n \le 1, \quad j = 1, \dots, N_{\mathcal{U}^p} \tag{33}$$
$$[G_\mathcal{G}]_l(\Phi_u * w)_n \le 1, \quad l = 1, \dots, N_{\mathcal{U}_\mathcal{G}} \tag{34}$$

with $([G_x]_i, [G_u^p]_j, [G_\mathcal{G}]_l)$ being the ith, jth, and lth rows of the corresponding matrices. In this work, players are assumed to synthesize policies satisfying these constraints for any $\boldsymbol{w} \in \boldsymbol{\mathcal{W}}$. Equivalently, we cast Problem (30) as a robust optimization problem by considering the worst-case realization of the noise. Specifically, we reformulate the constraint (32) as

$$\sup_{\bar{w}\in\mathcal{W}^N} \left\{\textstyle\sum_{n'=0}^N [G_x]_i \Phi_{x,n'} \bar{w}_{n'}\right\} \le 1, \quad i = 1, \dots, N_\mathcal{X}$$

and similarly for (33) and (34). Since $(\boldsymbol{\Phi}_{\boldsymbol{x}}, \boldsymbol{\Phi}_{\boldsymbol{u}}^p)$ are FIR maps, it suffices to consider noise sequences of length N, $\bar{w} \in \mathcal{W}^N = \prod_{n'=1}^N \mathcal{W}$, then exploit our knowledge of $\mathcal{W}$ to analytically solve this supremum. A common instance of $\mathcal{G}_\infty^{\mathrm{LQ}}$ consists on the problems in which $(w_n)_{n\in\mathbb{N}}$ is known to satisfy

$$w_n \in \mathcal{W} = \{P\zeta \in \mathbb{R}^{N_x} : \|\zeta\|_q \le 1\} \quad \forall n \in \mathbb{N}$$

given a full-column-rank matrix $P \in \mathbb{R}^{N_x\times N_\zeta}$ and the ℓ_q-norm $\|\cdot\|_q$. Using standard results from linear algebra [43], the robust counterpart of constraints (32)–(34) are the constraints

$$N^{1/q}\left\|(I_N \otimes P^\mathsf{T})([G_x]_i \bar{\Phi}_x)^\mathsf{T}\right\|_q^* \le 1 \quad (\forall i) \tag{35}$$
$$(N-1)^{1/q}\left\|(I_{N-1} \otimes P^\mathsf{T})([G_u^p]_j \bar{\Phi}_u^p)^\mathsf{T}\right\|_q^* \le 1 \quad (\forall j) \tag{36}$$
$$(N-1)^{1/q}\left\|(I_{N-1} \otimes P^\mathsf{T})([G_\mathcal{G}]_l \bar{\Phi}_u)^\mathsf{T}\right\|_q^* \le 1 \quad (\forall l) \tag{37}$$

in which $\bar{\Phi}_x = [\Phi_{x,1} \cdots \Phi_{x,N}]$, $\bar{\Phi}_u^p = [\Phi_{u,1}^p \cdots \Phi_{u,N-1}^p]$, and $\bar{\Phi}_u = [\Phi_{u,1} \cdots \Phi_{u,N-1}]$. The Problem (30) is thus rendered robust to the uncertainty in $\boldsymbol{w} \in \boldsymbol{\mathcal{W}}$ by incorporating the worst-case constraints (35)–(37) in place of (30c).

Remark 1: If $\mathcal{X} = \mathbb{R}^{N_x}$, $\mathcal{U}^p = \mathbb{R}^{N_u^p}$, or $\mathcal{U}_\mathcal{G} = \prod_{p\in\mathcal{P}} \mathbb{R}^{N_u^p}$, then the corresponding constraints are trivially satisfied for any $w \in \boldsymbol{\mathcal{W}}$ and can be removed from Problem (29). Conversely, if $\mathcal{W} = \mathbb{R}^{N_x}$ when either $\mathcal{X}$, $\mathcal{U}^p$, or $\mathcal{U}_\mathcal{G}$ is bounded, then no stabilising policy can enforce those constraints for all $\boldsymbol{w}$.

Remark 2: The constraints (35)–(37) are equivalent to those with $(-[G_x]_i, -[G_u^p]_j, -[G_\mathcal{G}]_l)$. As such, matrices in the form $\widetilde{G} = \mathtt{col}(G, -G)$ (i.e., enforcing $-\mathbf{1}_{N_z} \preceq Gz \preceq \mathbf{1}_{N_z}$) can be replaced by G $(Gz \preceq \mathbf{1}_{N_z})$ and yield the same results.

3) Structural Constraints: The sets $(\mathcal{C}_{x,n}, \mathcal{C}_{u,n}^p)_{n\in\mathbb{N}^+}$ $(\forall p \in \mathcal{P})$ are designed to impose some structure directly on the policy $\boldsymbol{K}^p$ (Corollary 1.1), often in the form of sparsity constraints. We consider the class of structural constraints that encode information patterns incurred by actuation and communication delays: Let $\boldsymbol{K}^p \in \boldsymbol{\mathcal{C}}^p$ $(\forall p)$ satisfy the restrictions

$$\boldsymbol{\mathcal{C}}^p = \{\boldsymbol{K}^p \in \mathcal{L}(\ell_\infty^{N_x}, \ell_\infty^{N_u^p}) : \text{The state } [x_t]_i \text{ is propagated to and affected by } [B^p K^p x_t]_i \text{ with delays } d_c, d_a > 0, \text{ respectively}\}$$

and consider the operators $\boldsymbol{S}_x : \boldsymbol{\Phi}_{\boldsymbol{x}} \mapsto (S_{x,n} \odot \Phi_{x,n})_{n\in\mathbb{N}_+}$ and $\boldsymbol{S}_u^p : \boldsymbol{\Phi}_{\boldsymbol{u}}^p \mapsto (S_{u,n}^p \odot \Phi_{u,n}^p)_{n\in\mathbb{N}_+}$ $(\forall p \in \mathcal{P})$, given the signals

$$(S_{x,n})_{n\in\mathbb{N}_+} = \left(\mathtt{Sp}\left(A^{\max(0,\lfloor\frac{n-d_a}{d_c}\rfloor)}\right)\right)_{n\in\mathbb{N}_+}$$
$$(S_{u,n}^p)_{n\in\mathbb{N}_+} = \left(\mathtt{Sp}\left(B^{p\mathsf{T}} A^{\max(0,\lfloor\frac{n-d_a}{d_c}\rfloor)}\right)\right)_{n\in\mathbb{N}_+}$$

with $\mathtt{Sp}(\cdot)$ denoting the sparsity pattern of a matrix, that is, $[\mathtt{Sp}(X)]_{i,j} = 1$ if $[X]_{i,j} \neq 0$ and $[\mathtt{Sp}(X)]_{i,j} = 0$ otherwise, for any matrix X. It can be shown that $\boldsymbol{K}^p = \boldsymbol{\Phi}_{\boldsymbol{u}}^p \boldsymbol{\Phi}_{\boldsymbol{x}}^{-1} \in \boldsymbol{\mathcal{C}}^p$ if $(\Phi_{x,n} \in \mathcal{C}_{x,n})_{n\in\mathbb{N}^+}$ and $(\Phi_{u,n}^p \in \mathcal{C}_{u,n}^p)_{n\in\mathbb{N}^+}$ with convex sets

$$\mathcal{C}_{x,n} = \{\Phi_{x,n} \in \mathbb{R}^{N_x\times N_x} : \Phi_{x,n} = S_{x,n} \odot \Phi_{x,n}\} \tag{38a}$$
$$\mathcal{C}_{u,n}^p = \{\Phi_{u,n}^p \in \mathbb{R}^{N_u^p\times N_x} : \Phi_{u,n}^p = S_{u,n}^p \odot \Phi_{u,n}^p\}. \tag{38b}$$

The constraints (38) enforce that the closed-loop response to the noise obeys an information pattern induced by the dynamics of the game $\mathcal{G}_\infty^{\mathrm{LQ}}$. Specifically, $[S_{x,n}]_{i,j} = 0$ (resp., $[S_{u,n}^p]_{i,j} = 0$) implies that disturbances to the jth component of the state, $[x_t]_j$, should not affect the state $[x_{t+n}]_i$ (the action $[u_{t+n}^p]_i$) when the policies $\boldsymbol{K}^p = \boldsymbol{\Phi}_{\boldsymbol{u}}^p \boldsymbol{\Phi}_{\boldsymbol{x}}^{-1}$ are employed. Enforcing sparsity also benefits the tractability of the best-response maps, as this reduces the effective number of decision variables.

The ability to impose a desired policy structure using convex constraints is a central feature of the SLS framework. In the context of dynamic games, it allows for describing, and most importantly, solving problems where players have asymmetric information patterns; a major challenge for feedback NE problems [39], [44]. Considering $(\mathcal{C}_{x,n}, \mathcal{C}_{u,n}^p)_{n\in\mathbb{N}^+}$ from (38), the feasibility of the best-response maps BR^p $(\forall p)$ become dependent on the parameters d_a and d_c: The larger their values the more restrictive the search space of matrices $\{\Phi_{x,n}, \Phi_{u,n}^p\}_{n=1}^N$. As such, the actuation and communication delays associated with the game directly affect the existence of GFNE (that is, $\Omega_{\mathcal{G}_\infty^{\mathrm{LQ}}}^{\boldsymbol{K}} \neq \emptyset$). In practice, whether these structural constraints are consistent or not can be assessed by solving the feasibility problem associated with Problem (30) [43]. A formal analysis of the requirements for d_a and d_c to ensure $\Omega_{\mathcal{G}_\infty^{\mathrm{LQ}}}^{\boldsymbol{K}} \neq \emptyset$ is beyond the scope of this article.

Algorithm 4: System-level BRD (SLS-BRD).

Input: Game $\mathcal{G}_\infty^{\mathrm{LQ}} := (\mathcal{P}, \boldsymbol{\mathcal{X}}, \{\boldsymbol{\mathcal{U}}^{\boldsymbol{p}}\}_{p\in\mathcal{P}}, \boldsymbol{\mathcal{W}}, \{J^p\}_{p\in\mathcal{P}})$
Output: GFNE $\boldsymbol{K}^\star = (\boldsymbol{K}^{1^\star}, \dots, \boldsymbol{K}^{N_P^\star})$

Initialize $\boldsymbol{K}_0 := (\boldsymbol{K}_0^1, \dots, \boldsymbol{K}_0^{N_P})$ and $k := 0$;
for $t = 0, 1, 2, \dots$ **do**
  /* Players apply actions $\{u_{k,t}^p = K_k^p x_{k,t}\}_{p\in\mathcal{P}}$ */
  if $\boldsymbol{\Phi_{u,k}} \in BR_\Phi(\boldsymbol{\Phi_{u,k}})$ **then return** $\boldsymbol{K}_k$;
  if $t = (k+1)\Delta T$ **then**
    for $p \in \mathcal{P}$ **do**
      Update $\boldsymbol{K}_{k+1}^p$ by computing the kernels
      $\boldsymbol{\Phi}_{\boldsymbol{u},\boldsymbol{k}+1}^p := (1-\eta)\boldsymbol{\Phi}_{\boldsymbol{u},\boldsymbol{k}}^p + \eta BR_\Phi^p(\boldsymbol{\Phi}_{\boldsymbol{u},\boldsymbol{k}}^{-p})$,
      $\boldsymbol{\Phi}_{\boldsymbol{x},\boldsymbol{k}+1}^p := \boldsymbol{F}_{\boldsymbol{\Phi}}(\boldsymbol{\Phi}_{\boldsymbol{u},\boldsymbol{k}+1}^p, \boldsymbol{\Phi}_{\boldsymbol{u},\boldsymbol{k}}^{-p})$
    $k := k+1$;

B. System-Level BRD

A learning dynamics based on the system-level best responses $\{\mathrm{BR}_\Phi^p\}_{p\in\mathcal{P}}$ relies on the following central result.

Theorem 3: A policy profile $\boldsymbol{K}^\star = (\boldsymbol{\Phi}_{\boldsymbol{u}}^{1^\star}, \dots, \boldsymbol{\Phi}_{\boldsymbol{u}}^{N_P^\star})\boldsymbol{\Phi}_{\boldsymbol{x}}^{\star^{-1}} \in \boldsymbol{\mathcal{C}}$, is a GFNE of $\mathcal{G}_\infty^{\mathrm{LQ}}$ if $\boldsymbol{\Phi}_{\boldsymbol{u}}^\star \in \mathrm{BR}_\Phi(\boldsymbol{\Phi}_{\boldsymbol{u}}^\star)$, or, equivalently

$$\boldsymbol{\Phi}_{\boldsymbol{u}}^{p^\star} \in \mathrm{BR}_\Phi^p(\boldsymbol{\Phi}_{\boldsymbol{u}}^{-p^\star}) \quad \forall p \in \mathcal{P}. \tag{39}$$

Proof: Consider an arbitrary fixed-point $\boldsymbol{\Phi}_{\boldsymbol{u}}^\star \in \mathrm{BR}_\Phi^p(\boldsymbol{\Phi}_{\boldsymbol{u}}^\star)$. From Theorem 1, we have $\boldsymbol{\Phi}_{\boldsymbol{x}}^\star = \boldsymbol{F}_{\boldsymbol{\Phi}}\boldsymbol{\Phi}_{\boldsymbol{u}}^\star$. Now, consider policies $\boldsymbol{K}^{p^\star} = \boldsymbol{\Phi}_{\boldsymbol{u}}^{p^\star}(\boldsymbol{\Phi}_{\boldsymbol{x}}^\star)^{-1}, p \in \mathcal{P}$. Clearly, $\boldsymbol{\Phi}_{\boldsymbol{u}}^{p^\star} = \boldsymbol{K}^{p^\star}\boldsymbol{\Phi}_{\boldsymbol{x}}^\star$. As a consequence, for any $\boldsymbol{w} \in \boldsymbol{\mathcal{W}}$

$$\boldsymbol{\Phi}_{\boldsymbol{u}}^{p^\star}\boldsymbol{w} = \boldsymbol{K}^{p^\star}\boldsymbol{\Phi}_{\boldsymbol{x}}^\star\boldsymbol{w} \iff \boldsymbol{u}^{p^\star} = \boldsymbol{K}^{p^\star}\boldsymbol{x}^\star$$

and by definition, $\boldsymbol{\Phi}_{\boldsymbol{u}}^{p^\star} \in U_\Phi^p(\boldsymbol{\Phi}_{\boldsymbol{u}}^{-p^\star})$ imply $\boldsymbol{u}^{p^\star} \in U^p(\boldsymbol{u}^{-p^\star})$. Thus, $\boldsymbol{u}^\star$ is an open-loop realization of the policy $\boldsymbol{K}^\star$. Finally, since $J^p(\boldsymbol{u}^{p^\star}, \boldsymbol{u}^{-p^\star}) \cong J^p(\boldsymbol{\Phi}_{\boldsymbol{u}}^{p^\star}, \boldsymbol{\Phi}_{\boldsymbol{u}}^{-p^\star})$ and $\boldsymbol{\Phi}_{\boldsymbol{u}}{}^\star \in \Omega_{\mathcal{G}_\infty^\Phi}$, we conclude that no player can obtain an admissible policy that unilaterally improves its cost, that is, $\boldsymbol{K}^\star \in \Omega_{\mathcal{G}_\infty^{\mathrm{LQ}}}^{\boldsymbol{K}}$. □

The relationship between BR and BR_Φ implies that a GFNE of $\mathcal{G}_\infty^{\mathrm{LQ}}$ can be obtained analytically from a GNE of $\mathcal{G}_\infty^\Phi$. This allows us to adapt the BRD-GFNE procedure from Algorithm 3 into a procedure for GFNE seeking in constrained infinite-horizon dynamic games based on the mappings $\{\mathrm{BR}_\Phi^p\}_{p\in\mathcal{P}}$. This SLS-BRD approach is given in Algorithm 4. We remark on some technical aspects

1) The pair $(\boldsymbol{\Phi}_{\boldsymbol{x},\boldsymbol{k}}^p, \boldsymbol{\Phi}_{\boldsymbol{u},\boldsymbol{k}}^p)$ is the parameterization of pth player's policy after $k \in \mathbb{N}$ updates, that is, the signals $\boldsymbol{\Phi}_{\boldsymbol{x},\boldsymbol{k}}^p = (\Phi_{x,k,t}^p)_{t\in\mathbb{N}}$ and $\boldsymbol{\Phi}_{\boldsymbol{u},\boldsymbol{k}}^p = (\Phi_{u,k,t}^p)_{t\in\mathbb{N}}$. The index $k \in \mathbb{N}$ should not be mistaken for the stage index $t \in \mathbb{N}$.
2) Updating the policy to $\boldsymbol{K}_{k+1}^p$ consists of employing the Corollary 1.1 with the updated maps $(\boldsymbol{\Phi}_{\boldsymbol{x},\boldsymbol{k}+1}^p, \boldsymbol{\Phi}_{\boldsymbol{u},\boldsymbol{k}+1}^p)$.
3) Responses $\{\boldsymbol{\Phi}_{\boldsymbol{x},\boldsymbol{k}}^p\}_{p\in\mathcal{P}}$ are most likely distinct at $k < \infty$, that is, $\boldsymbol{\Phi}_{\boldsymbol{x},\boldsymbol{k}}^p \neq \boldsymbol{\Phi}_{\boldsymbol{x},\boldsymbol{k}}^{\tilde{p}}$ for $p \neq \tilde{p}$. Consequently, the system level parameterization (22) might not hold for the profile $\boldsymbol{K}_k = (\boldsymbol{\Phi}_{\boldsymbol{u},\boldsymbol{k}}^1(\boldsymbol{\Phi}_{\boldsymbol{x},\boldsymbol{k}}^1)^{-1}, \dots, \boldsymbol{\Phi}_{\boldsymbol{u},\boldsymbol{k}}^{N_P}(\boldsymbol{\Phi}_{\boldsymbol{x},\boldsymbol{k}}^{N_P})^{-1})$ and this could lead to stability issues. However, this policy still satisfies a robust variant of Theorem 1 when the distances $\|\boldsymbol{\Phi}_{\boldsymbol{x},\boldsymbol{k}}^p - \boldsymbol{F}_{\boldsymbol{\Phi}}\boldsymbol{\Phi}_{\boldsymbol{u},\boldsymbol{k}}\|$ $(\forall p \in \mathcal{P})$ are sufficiently small [35].

The SLS-BRD induces an operator $T_\Phi = (1-\eta)I + \eta\mathrm{BR}_\Phi$, which defines the global update $\boldsymbol{\Phi}_{\boldsymbol{u},\boldsymbol{k}+1}$ from $\boldsymbol{\Phi}_{\boldsymbol{u},\boldsymbol{k}}$. As before, T_Φ and BR_Φ share the same fixed points: The GNEs $\Omega_{\mathcal{G}_\infty^\Phi}$. From Theorem 3, convergence to a response $\boldsymbol{\Phi}_{\boldsymbol{u}}^\star \in T_\Phi(\boldsymbol{\Phi}_{\boldsymbol{u}}^\star)$ then implies convergence to a policy $\boldsymbol{K}^\star \in \Omega_{\mathcal{G}_\infty}^{\boldsymbol{K}}$. Hence, this learning dynamics is a formal procedure for GFNE seeking.

In this general form, Algorithm 4 is still unpractical due to $\{\mathrm{BR}_\Phi^p\}_{p\in\mathcal{P}}$ being intractable. The SLS-BRD can be adapted to consider instead $\{\widehat{\mathrm{BR}}_\Phi^p\}_{p\in\mathcal{P}}$: The players' updates become

$$\boldsymbol{\Phi}_{\boldsymbol{u},\boldsymbol{k}+1}^p := (1-\eta)\,\boldsymbol{\Phi}_{\boldsymbol{u},\boldsymbol{k}}^p + \eta\widehat{\mathrm{BR}}_\Phi^p(\boldsymbol{\Phi}_{\boldsymbol{u},\boldsymbol{k}}^{-p}).$$

The global update rule induced by this modified algorithm is $\widehat{T}_\Phi = (1-\eta)I + \eta\widehat{\mathrm{BR}}_\Phi$. In this case, the fixed-points of $\widehat{T}_\Phi$ coincide with those of $\widehat{\mathrm{BR}}_\Phi$. As such, this (approximately)BRD is a procedure for ϵ-GFNE seeking, $\varepsilon > 0$.

1) Convergence of the SLS-BRD: The convergence of the Algorithm 4 depends on BR_Φ (or $\widehat{\mathrm{BR}}_\Phi$, for its tractable version) being at least nonexpansive (Lemmas 1–2). Formally, the following.

Corollary 3.1: Let $\widehat{\mathrm{BR}}_\Phi : \boldsymbol{\mathcal{C}}_u \rightrightarrows \boldsymbol{\mathcal{C}}_u$ be $L_{\widehat{\mathrm{BR}}_\Phi}$-Lipschitz, with $L_{\widehat{\mathrm{BR}}_\Phi} < 1$. Then, the SLS-BRD $\boldsymbol{\Phi}_{\boldsymbol{u},\boldsymbol{k}+1} = \widehat{T}_\Phi(\boldsymbol{\Phi}_{\boldsymbol{u},\boldsymbol{k}})$ converge to the unique ε-GNE $\boldsymbol{\Phi}_{\boldsymbol{u}}^\star \in \Omega_{\mathcal{G}_\infty^\Phi}^\varepsilon$ with rate

$$\frac{\|\boldsymbol{\Phi}_{\boldsymbol{u},\boldsymbol{k}} - \boldsymbol{\Phi}_{\boldsymbol{u}}^\star\|_{\ell_2}}{\|\boldsymbol{\Phi}_{\boldsymbol{u},\boldsymbol{0}} - \boldsymbol{\Phi}_{\boldsymbol{u}}^\star\|_{\ell_2}} \leq \left((1-\eta) - \eta L_{\widehat{\mathrm{BR}}_\Phi}\right)^k \tag{40}$$

from any feasible initial $\boldsymbol{\Phi}_{\boldsymbol{u},\boldsymbol{0}}$.

In general, determining a Lipschitz constant for such mappings is challenging. However, for linear-quadratic games $\mathcal{G}_\infty^{\mathrm{LQ}}$ where $\mathcal{W}$ is a polyhedron, best-responses are piecewise-affine operators and their Lipschitz properties are straightforward. In the following, we use this fact to establish some conditions for convergence of the SLS-BRD for a specific class of games.

Consider the (approximately)best-response maps $\{\widehat{\mathrm{BR}}_\Phi^p\}_{p\in\mathcal{P}}$ and assume N sufficiently large to ensure that $\|\Phi_{x,N}\|_F^2 < \gamma$ is strictly satisfied. For notational convenience, let us introduce the operators $\boldsymbol{F}_{\boldsymbol{\Phi}}^p = (I - \boldsymbol{S}_+\boldsymbol{A})^{-1}\boldsymbol{S}_+\boldsymbol{B}^p$ $(\forall p \in \mathcal{P})$ and signal $\boldsymbol{F}_{\boldsymbol{\Phi}}^0 = (I - \boldsymbol{S}_+\boldsymbol{A})^{-1}\boldsymbol{\delta}I_{N_x}$, and the objective-related mappings $\boldsymbol{C}^p : \boldsymbol{\Phi}_{\boldsymbol{x}} \mapsto (C^p\Phi_{x,n})_{n\in\mathbb{N}}$ and $\boldsymbol{D}^{p\tilde{p}} : \boldsymbol{\Phi}_{\boldsymbol{u}}{}^{\tilde{p}} \mapsto (D^{p\tilde{p}}\Phi_{u,n}^{\tilde{p}})_{n\in\mathbb{N}}$, with $\boldsymbol{D}^{p0} = 0$. Moreover, for all $p, \tilde{p} \in \mathcal{P}$, define the operators

$$\boldsymbol{H}^{p\tilde{p}} = (\boldsymbol{C}^p\boldsymbol{F}_{\boldsymbol{\Phi}}^p + \boldsymbol{D}^{pp})^{\mathsf{T}}\left(\boldsymbol{C}^p\boldsymbol{F}_{\boldsymbol{\Phi}}^{\tilde{p}} + \boldsymbol{D}^{p\tilde{p}}\right) \tag{41}$$

and $\boldsymbol{H}^{p,-p} = (\boldsymbol{H}^{p,\tilde{p}})_{\tilde{p}\in\mathcal{P}}$. Because $\{\boldsymbol{\Phi}_{\boldsymbol{u}}^p\}_{p\in\mathcal{P}}$ are FIR, all the elements defined previously have equivalent matrix representations. Using this notation, Problem (30) can be reformulated as

$$\begin{aligned}\underset{\boldsymbol{\Phi}_{\boldsymbol{u}}^p \in U_\Phi^p(\boldsymbol{\Phi}_{\boldsymbol{u}}^{-p})}{\text{minimize}} \quad \mathrm{Tr}\Bigg[&\boldsymbol{\Phi}_{\boldsymbol{u}}^{p\,\mathsf{T}}\boldsymbol{H}^{pp}\boldsymbol{\Phi}_{\boldsymbol{u}}^p \\ &+ 2\left(\sum_{\tilde{p}\in\mathcal{P}\setminus\{p\}} \boldsymbol{H}^{p\tilde{p}}\boldsymbol{\Phi}_{\boldsymbol{u}}^{\tilde{p}} + \boldsymbol{H}^{p0}\right)^{\mathsf{T}}\boldsymbol{\Phi}_{\boldsymbol{u}}^p\Bigg].\end{aligned} \tag{42}$$

The maps $\mathrm{BR}^p(\boldsymbol{\Phi}_{\boldsymbol{u}}^{-p})$, $p \in \mathcal{P}$, thus correspond to the solution of quadratic programs with convex constraints $U_\Phi^p(\boldsymbol{\Phi}_{\boldsymbol{u}}^{-p})$, for $\boldsymbol{\Phi}_{\boldsymbol{u}}^{-p} \in \boldsymbol{\mathcal{C}}_u^{-p}$. In the case of $\mathcal{W} = \{w_t \in \mathbb{R}^{N_x} : \|w_t\|_\infty \leq 1\}$ and

no structural constraints ($\boldsymbol{S}_x = \boldsymbol{S}_u^p = I$), we have

$$\begin{aligned} U_\Phi^p(\boldsymbol{\Phi}_{\boldsymbol{u}}^{-p}) = \{&\boldsymbol{\Phi}_{\boldsymbol{u}}^p \in \ell_2[0,N) : \\ &\Phi_{x,n+1} = A\Phi_{x,n} + \textstyle\sum_{\tilde{p}\in\mathcal{P}} B^{\tilde{p}}\Phi_{u,n}^{\tilde{p}},\ \Phi_{x,1} = I_{N_x}, \\ &\left\|([G_x]_i\bar{\Phi}_x)^{\mathsf{T}}\right\|_1 \le 1,\ i = 1,\dots,N_{\mathcal{X}}; \\ &\left\|([G_u^p]_j\bar{\Phi}_u^p)^{\mathsf{T}}\right\|_1 \le 1,\ j = 1,\dots,N_{\mathcal{U}^p}; \\ &\left\|([G_{\mathcal{G}}]_l\bar{\Phi}_u)^{\mathsf{T}}\right\|_1 \le 1,\ l = 1,\dots,N_{\mathcal{U}_{\mathcal{G}}}\}. \end{aligned} \tag{43}$$

Using standard techniques from optimization, the constraints (43) can be incorporated into Problem (42) as linear inequalities. The best-responses mappings $\{\widehat{\mathrm{BR}}_\Phi^p\}_{p\in\mathcal{P}}$ are thus piecewise affine in $\boldsymbol{\Phi}_{\boldsymbol{u}}^{-p} \in \boldsymbol{\mathcal{C}}_u^{-p}$ [45]. Consequently, also $\widehat{\mathrm{BR}}_\Phi$ must be piecewise affine: A local Lipschitz constant can then be derived for each region of $\boldsymbol{\mathcal{C}}_u^{-p}$ that leads to a subset of the operational constraints being active. For nongeneralized games, this fact can be exploited to derive a global Lipschitz constant for $\widehat{\mathrm{BR}}_\Phi$.

Theorem 4: Consider the mapping $U_\Phi^p(\boldsymbol{\Phi}_{\boldsymbol{u}}^{-p})$ from (43) with G_u^p full-row-rank and $G_x = G_{\mathcal{G}} = 0$. Then, the best-response map $\widehat{\mathrm{BR}}_\Phi$ is $L_{\widehat{\mathrm{BR}}_\Phi}$-Lipschitz with

$$(L_{\widehat{\mathrm{BR}}_\Phi})^2 \le \sum_{p\in\mathcal{P}} \left(L_{\widehat{\mathrm{BR}}_\Phi}^p\right)^2 \tag{44}$$

given the player-specific $L_{\widehat{\mathrm{BR}}_\Phi}^p = \sigma_{\max}(\boldsymbol{H}^{p,-p})/\sigma_{\min}(\boldsymbol{H}^{pp})$.

Proof: See the Appendix. □

We remark that this Lipschitz constant is not tight and the condition that $L_{\widehat{\mathrm{BR}}_\Phi} < 1$ using (44) is only a sufficient (but not necessary) condition for Corollary 3.1 to hold in practice. Regardless, it highlights some intuitive, but nontrivial, facts about the convergence of the SLS-BRD.

1) The condition $L_{\widehat{\mathrm{BR}}_\Phi} < 1$ is satisfied if the block operator $\boldsymbol{H} = [H^{p\tilde{p}}]_{p,\tilde{p}\in\mathcal{P}}$ is diagonally dominant. This highlights the relationship between the SLS-BRD and classical Jacobi iterative methods, where diagonal dominance of the linear system being solved is a known requirement.
2) The convergence rates of the SLS-BRD depend directly on $\|D^{pp}\|_2$ $(\forall p)$ through $\sigma_{\min}(\boldsymbol{H}^{pp}) = \sigma_{\min}^2(\boldsymbol{C}^p\boldsymbol{F}_{\boldsymbol{\Phi}}^p + \boldsymbol{D}^{pp})$. As such, faster convergence is expected for games in which players apply strong penalties to their own actions.
3) The convergence rates of the SLS-BRD depend on the number of players since $L_{\widehat{\mathrm{BR}}_\Phi}^p > 0$ for all $p \in \mathcal{P}$. In large-scale games, players might need to apply strong penalties to their actions (through D^{pp}) to ensure convergence.
4) The convergence rate of the SLS-BRD can be dominated by the slowest player: Whenever there exists a $p \in \mathcal{P}$ such that $L_{\widehat{\mathrm{BR}}_\Phi}^p \gg L_{\widehat{\mathrm{BR}}_\Phi}^{\tilde{p}}$ for all $\tilde{p} \in \mathcal{P}$, then $L_{\widehat{\mathrm{BR}}_\Phi} \approx L_{\widehat{\mathrm{BR}}_\Phi}^p$.

At the cost of interpretability, similar Lipschitz constants as in (44) can also be obtained for general $(\mathcal{X},\mathcal{U}_{\mathcal{G}},\mathcal{U}^p)$ and P, and when structural constraints are present. We stress that Theorem 4 is obtained under the assumption that $\|\Phi_{x,N}\|_F^2 < \gamma$ holds strictly. The best-response maps $\{\widehat{\mathrm{BR}}_\Phi^p\}_{p\in\mathcal{P}}$ when this constraint is active, or when $\mathcal{W}$ is an ellipsoid, are solutions to quadratically constrained quadratic programs and their Lipschitz properties are less intuitive.

IV. Examples

In the following, we demonstrate the SLS-BRD algorithm in two exemplary problems: decentralized control of an unstable network and management of a competitive market. In both problems, we set the initial profile $\boldsymbol{K}_0 = (\boldsymbol{\Phi}_{\boldsymbol{u},\boldsymbol{0}}^1,\dots,\boldsymbol{\Phi}_{\boldsymbol{u},\boldsymbol{0}}^{N_P})\boldsymbol{\Phi}_{\boldsymbol{x},\boldsymbol{0}}^{-1}$ by projecting the zero-response $\hat{\boldsymbol{\Phi}}_{\boldsymbol{u},\boldsymbol{0}} = \boldsymbol{0}$ into the feasible set. An SLS-BRD routine is then simulated by having players update their policies through best responses to the parameterization of their rivals' policies, assumed to be available. Due to numerical limitations, we interrupt the updates whenever the condition $\|\boldsymbol{\Phi}_{\boldsymbol{u},\boldsymbol{k}}^p - \boldsymbol{\Phi}_{\boldsymbol{u},\boldsymbol{k-1}}^p\|_{\ell_2}/\|\boldsymbol{\Phi}_{\boldsymbol{u},\boldsymbol{k}}^p\|_{\ell_2} \le 10^{-8}$ $(\forall p \in \mathcal{P})$ is satisfied.

A. Stabilization of a Bidirectional Chain Network

Consider a game $\mathcal{G}_\infty^{\mathrm{LQ}} = \{\mathcal{P},\boldsymbol{\mathcal{X}},\{\boldsymbol{\mathcal{U}}^p\}_{p\in\mathcal{P}},\boldsymbol{\mathcal{W}},\{J^p\}_{p\in\mathcal{P}}\}$ with players $\mathcal{P} = \{1,2,3\}$ operating a chain network of $N_x = 14$ single-state nodes whose dynamics are described by

$$A = \begin{bmatrix} 1 & 0.2 & & \\ -0.2 & \ddots & \ddots & \\ & \ddots & \ddots & 0.2 \\ & & -0.2 & 1 \end{bmatrix};\ \left\{ B^p = \begin{bmatrix} \boldsymbol{0}_{6(p-1)\times 2} \\ I_2 \\ \boldsymbol{0}_{6(3-p)\times 2} \end{bmatrix} \right\}_{p\in\mathcal{P}}$$

where $A \in \mathbb{R}^{N_x\times N_x}$ and $B^p \in \mathbb{R}^{N_x\times N_u^p}$ with $N_u^p = 2$ $(\forall p)$. This game has unstable dynamics, since $\|A\|_2 = 1.073 > 1$, however it is stabilizable for each (A,B^p). Moreover, the game is subjected to a noise process described by $w_t \sim \mathrm{Uniform}(\mathcal{W})$, $t \in \mathbb{N}$, defined over $\mathcal{W} = \{w_t \in \mathbb{R}^{N_x} : \|w_t\|_\infty \le 1\}$. In this problem, players are interested in stabilizing the game $\mathcal{G}_\infty^{\mathrm{LQ}}$, while minimizing their individual objective functionals

$$J^p(\boldsymbol{u}^p,\boldsymbol{u}^{-p}) = \mathrm{E}\left[\sum_{t=0}^{\infty}\left(\|x_t\|_2^2 + \beta^p\|u_t^p\|_2^2\right)\right]$$

equivalent to (19) after setting $C^p = [I_{N_x}\ \boldsymbol{0}_{N_x\times N_u}]^{\mathsf{T}}$, $D^{pp} = [\boldsymbol{0}_{N_u\times N_x}\ \sqrt{\beta^p}I_{Nu}]^{\mathsf{T}}$, and $D^{p\tilde{p}} = 0$ for all $\tilde{p} \in \mathcal{P}\backslash\{p\}$. The players' actions are subjected to operational constraints, $\boldsymbol{u}^p \in U^p(\boldsymbol{u}^{-p})$, defined by the constraint sets

$$\begin{aligned} \mathcal{X} &= \mathbb{R}^{N_x}; \\ \mathcal{U}^p &= \left\{u_t^p \in \mathbb{R}^{N_u^p} : \left[\tfrac{1}{12}I_{N_u^p}\right] u_t^p \preceq \boldsymbol{1}_{N_u^p}\right\}; \\ \mathcal{U}_{\mathcal{G}} &= \prod_{p\in\mathcal{P}} \mathbb{R}^{N_u^p}, \end{aligned}$$

enforcing $\|u_t^p\|_\infty \le 12$ for each $t \in \mathbb{N}$ and $p \in \mathcal{P}$ (see Remark 2). We assume that players design their state-feedback policies, $\boldsymbol{K}^p = \boldsymbol{\Phi}_{\boldsymbol{u}}^p\boldsymbol{\Phi}_{\boldsymbol{x}}^{-1} \in \boldsymbol{\mathcal{C}}^p$, considering an FIR horizon of $N = 50$ and the constraints $(\Phi_{x,n} \in \mathcal{C}_{x,n})_{n=1}^N$ and $(\Phi_{u,n}^p \in \mathcal{C}_{u,n}^p)_{n=1}^N$,

$$\mathcal{C}_{x,n} = \{\Phi_{x,n} \in \mathbb{R}^{N_x\times N_x} : \Phi_{x,n} = \mathtt{Sp}(A^{n-1}) \odot \Phi_{x,n}\}$$

$$\mathcal{C}_{u,n}^p = \{\Phi_{u,n}^p \in \mathbb{R}^{N_u^p\times N_x} : \Phi_{u,n}^p = \mathtt{Sp}(B^{p\mathsf{T}}A^{n-1}) \odot \Phi_{u,n}^p\}$$

defined by setting the delay parameters $d_a = d_c = 1$. Under this setup, $\mathcal{G}_\infty^{\mathrm{LQ}}$ belongs to the class of dynamic potential games (DPG, [46]) and the (unique) GFNE can be obtained in advance by solving a centralized optimization problem.

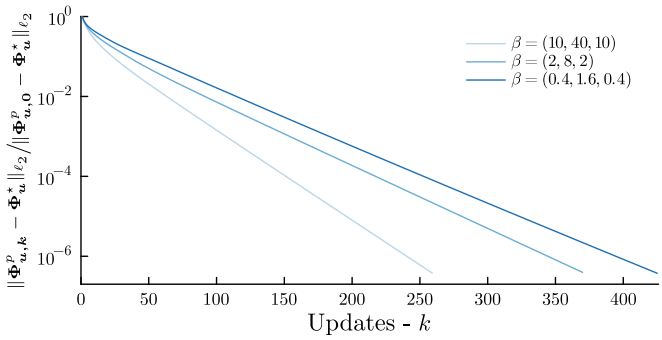

Fig. 3. N_P-chain game: Convergence of the SLS-BRD routine.

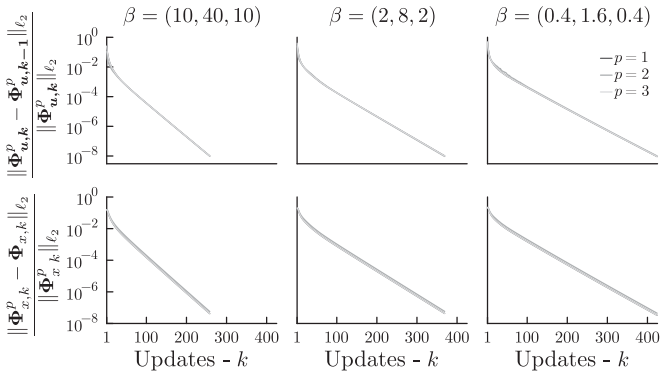

Fig. 4. N_P-chain game: Relative distance between the local updates $(\Phi^p_{u,k}, \Phi^p_{u,k-1})$, top row, and responses $(\Phi^p_{x,k}, \Phi_{x,k})$, bottom row.

In this experiment, we simulate a GFNE seeking procedure for each value $\beta \in \{(10, 40, 10), (2, 8, 2), (0.4, 1.6, 0.4)\}$. The game $\mathcal{G}^{\mathrm{LQ}}_\infty$ is executed alongside the updating of players' policies according to some learning dynamics. The players are assumed to seek ε-GFNE policies by adhering to the SLS-BRD routine (Algorithm 4) using (approximately)best-response maps, $\{\widehat{\mathrm{BR}}^p_\Phi\}_{p\in\mathcal{P}}$, with $\gamma = 0.95$. The policies are updated simultaneously every $\Delta T = 1$ stage with learning rate $\eta = 1/2$.

The convergence of the SLS-BRD routine to the fixed-point $\boldsymbol{K}^\star = \boldsymbol{\Phi}^\star_{\boldsymbol{u}}\boldsymbol{\Phi}^{-1}_{\boldsymbol{x}} = (\boldsymbol{\Phi}^{1^\star}_{\boldsymbol{u}}, \ldots, \boldsymbol{\Phi}^{N^\star_P}_{\boldsymbol{u}})\boldsymbol{\Phi}^{-1}_{\boldsymbol{x}}$ is shown in Fig. 3. The results demonstrate that the players' policies are sufficiently close to the ϵ-GFNE profile after 260, 370, and 425 iterations for each respective weighting configuration. In each case, the (soft) FIR constraints are satisfied strictly after the initial profile (that is, $\|\Phi^p_{x,k,N}\|^2_F < 0.95,\ k > 2\ \forall p \in \mathcal{P}$). Moreover, this fixed-point iteration seems to follow the behavior discussed in Section III-B: The convergence improves when players apply stronger penalties to their actions ($\beta = (10, 40, 10)$) when compared to that obtained by weaker control penalties ($\beta = (0.4, 1.6, 0.4)$). However, we stress that the Lipschitz constants from Theorem 4 do not consider structural constraints ($\boldsymbol{S}_x$, $\boldsymbol{S}^p_u$), and thus, cannot be applied to this experiment. Regardless, the convergence is shown to be geometric, indicating that the best-response maps $\widehat{\mathrm{BR}}_\Phi$ are contractive. If not interrupted, and disregarding numerical limitations, the SLS-BRD should continue to approach the fixed-point $\boldsymbol{K}^\star$ at this rate.

In Fig. 4(top), we show the relative distance between the individual updates $\{\boldsymbol{\Phi}^p_{\boldsymbol{u},\boldsymbol{k}}, \boldsymbol{\Phi}^p_{\boldsymbol{u},\boldsymbol{k-1}}\}_{p\in\mathcal{P}}$ from each player. The updates are of similar magnitude for all players $p \in \mathcal{P}$, in each scenario, except for player $p = 3$, which shows a slightly faster convergence. In general, these local changes become numerically negligible at a faster rate than the global convergence in Fig. 3. Specifically, when the relative distance between updates approaches the aforementioned threshold of 10^{-8}, the corresponding policy profile has become closer to the ϵ-GFNE by a factor of 10^{-6}. Finally, we consider the relative distances $\|\boldsymbol{\Phi}^p_{\boldsymbol{x},\boldsymbol{k}} - \boldsymbol{\Phi}_{\boldsymbol{x},\boldsymbol{k}}\|_{\ell_2}/\|\boldsymbol{\Phi}^p_{\boldsymbol{x},\boldsymbol{k}}\|_{\ell_2}$, $k \in \mathbb{N}_+$, between the responses $\{\boldsymbol{\Phi}^p_{\boldsymbol{x},\boldsymbol{k}}\}_{p\in\mathcal{P}}$ and $\boldsymbol{\Phi}_{\boldsymbol{x},\boldsymbol{k}} = \boldsymbol{F}_{\boldsymbol{\Phi}}\boldsymbol{\Phi}_{\boldsymbol{u},\boldsymbol{k}}$ obtained from the system-level parameterization associated with $\boldsymbol{\Phi}_{\boldsymbol{u},\boldsymbol{k}}$ (see Theorem 1). As shown in Fig. 4(bottom), these distances decrease at a similar rate and are relatively small since the initial stages of the game. The policies $\boldsymbol{K}_k = (\boldsymbol{\Phi}^1_{\boldsymbol{u},\boldsymbol{k}}(\boldsymbol{\Phi}^1_{\boldsymbol{x},\boldsymbol{k}})^{-1}, \ldots, \boldsymbol{\Phi}^{N_P}_{\boldsymbol{u},\boldsymbol{k}}(\boldsymbol{\Phi}^{N_P}_{\boldsymbol{x},\boldsymbol{k}})^{-1})$ thus approximate $(\boldsymbol{\Phi}^1_{\boldsymbol{u},\boldsymbol{k}}, \ldots, \boldsymbol{\Phi}^{N_P}_{\boldsymbol{u},\boldsymbol{k}})\boldsymbol{\Phi}^{-1}_{\boldsymbol{x},\boldsymbol{k}}$ as $k \to \infty$ and are thus expected to stabilize the system during this learning dynamics.

The evolution of the game given is displayed in Fig. 5 for the period $t \in (550, 600]$, after policy updates are interrupted. The results demonstrate that the policies obtained by the SLS-BRD routine are capable of jointly stabilising the networked system, robustly against the random noise. These policies are also shown to satisfy the operational constraints: Each player's actions satisfy $\|u^p_t\|_\infty \leq 5$, $t \in \mathbb{N}$, in all scenarios. The enforcement of this strategy highlights the conservativeness resulting from the robust operational constraints. Finally, note that the state- and control-trajectories from operating the system through these policies are similar for all β configurations; however, the policies with $\beta = (2, 8, 2)$ and $\beta = (0.4, 1.6, 0.4)$ achieve a slightly better noise rejection at the expense of more aggressive control actions. In this experiment, the choice $\beta = (10, 40, 10)$ seems to be preferable as it converges faster while still achieving satisfactory performance.

B. Price Competition in Oligopolistic Markets

Consider a game $\mathcal{G}^{\mathrm{LQ}}_\infty = \{\mathcal{P}, \boldsymbol{\mathcal{X}}, \{\boldsymbol{\mathcal{U}}^p\}_{p\in\mathcal{P}}, \boldsymbol{\mathcal{W}}, \{J^p\}_{p\in\mathcal{P}}\}$ consisting of a set of companies $\mathcal{P} = \{1, 2, 3, 4\}$ participating in a single-product market. Assume that these companies have equivalent production capacities and are able to satisfy the demand for their products. The product offered by company $p \in \mathcal{P}$ has a daily local demand $\boldsymbol{d^p} = (d^p(t))_{t\in\mathbb{R}_{\geq 0}}$, which evolves according to the continuous-time dynamics

$$\tau \dot{d}^p(t) = \underbrace{d^p_{\mathrm{base}}(t) - \textstyle\sum_{\tilde{p}\in\mathcal{P}} \widetilde{B}^{p\tilde{p}} u^{\tilde{p}}(t)}_{\text{linear price-demand curve}} - d^p(t),$$

where $\boldsymbol{u}^{\tilde{p}} = (u^{\tilde{p}}(t))_{t\in\mathbb{R}_{\geq 0}}$ are price changes around the value at which $\tilde{p} \in \mathcal{P}$ sell its products and $\boldsymbol{d}^p_{\mathrm{base}} = (d^p_{\mathrm{base}}(t))_{t\in\mathbb{R}_{\geq 0}}$ is some fluctuating baseline demand. Specifically, the baseline demands are of the form $d^p_{\mathrm{base}}(t) = \bar{d}_{\mathrm{base}} + v^p(t)$ for all $p \in \mathcal{P}$, given fixed $\bar{d}_{\mathrm{base}} = 10$ and noise process $\boldsymbol{v^p} = (v^p(t))_{t\in\mathbb{R}_{\geq 0}}$. The market parameters $\tau \in \mathbb{R}_{\geq 0}$ and $\widetilde{B}^{p\tilde{p}} \in \mathbb{R}_{\geq 0}$ $(\forall p, \tilde{p} \in \mathcal{P})$ describe how local demands respond to price changes: We set $\tau = 1.2$ and sample $B^{p\tilde{p}} \sim \mathrm{Uniform}(0.5, 1.5)$ for all $p, \tilde{p} \in \mathcal{P}$. In this problem, companies aim at devising pricing policies to

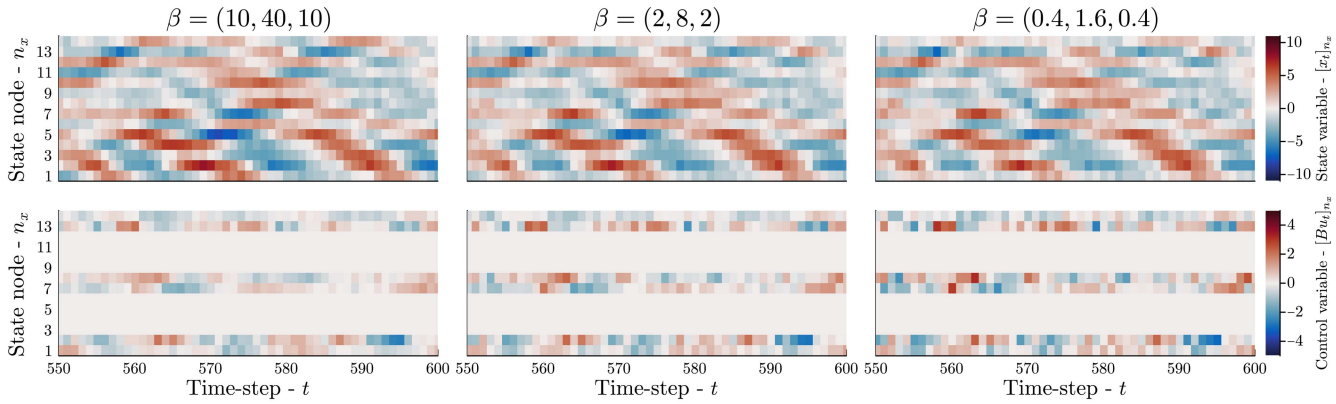

Fig. 5. N_P-chain game, $t \in (550, 600]$: State $\boldsymbol{x}$ (top panels) and applied control $\boldsymbol{Bu}$ (bottom panels) trajectories for each execution of $\mathcal{G}_\infty^{\mathrm{LQ}}$ with $\beta \in \{(10, 40, 10), (2, 8, 2), (0.4, 1.6, 0.4)\}$. The vertical axis represents each node in the chain-networked system.

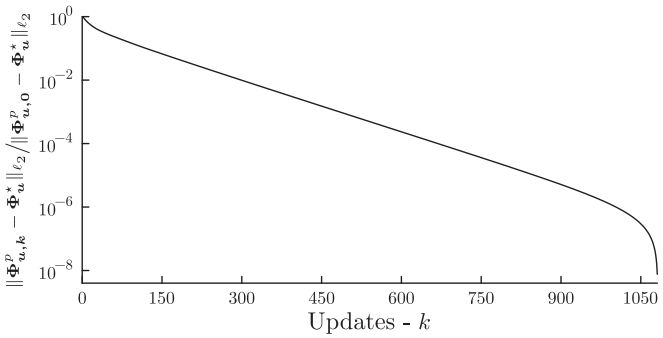

Fig. 6. Market game: Convergence of the SLS-BRD routine.

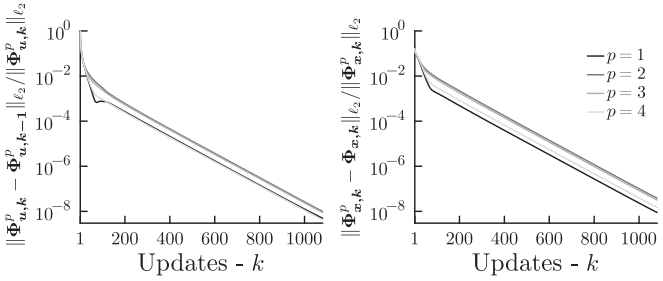

Fig. 7. Market game: Relative distance between the local updates $(\boldsymbol{\Phi}^p_{\boldsymbol{u},\boldsymbol{k}}, \boldsymbol{\Phi}^p_{\boldsymbol{u},\boldsymbol{k-1}})$, left row, and responses $(\boldsymbol{\Phi}^p_{\boldsymbol{x},\boldsymbol{k}}, \boldsymbol{\Phi}_{\boldsymbol{x},\boldsymbol{k}})$, right row.

stabilize their demands around $\bar{d}_{\text{base}}$, which provides a stable profit margin, while satisfying a price-cap regulation enforcing

$$-\bar{u}_{\text{avg}} \leq \tfrac{1}{N_P} \sum_{p \in \mathcal{P}} u^p(t) \leq \bar{u}_{\text{avg}}, \quad \bar{u}_{\text{avg}} = 0.5.$$

Define $\boldsymbol{x} = (\boldsymbol{x}^1, \ldots, \boldsymbol{x}^{N_P})$, with $\boldsymbol{x}^p = (d^p(t) - \bar{d}_{\text{base}})_{t \in \mathbb{R}_{\geq 0}}$, and $\widetilde{B}^p = [\widetilde{B}^{p1} \cdots \widetilde{B}^{pN_P}]$ for all $p \in \mathcal{P}$. Considering a zero-order hold of inputs with period $\Delta t = 1/4$ [days], the game $\mathcal{G}_\infty^{\mathrm{LQ}}$ can be described by the discrete-time dynamics[1]

$$x_{t+1} = Ax_t + \sum_{p \in \mathcal{P}} B^p u_t^p + w_t$$

[1] With a slight abuse of notation, we use t to index both the continuous-time $(x(t), t \in \mathbb{R}_{\geq 0})$ and discrete-time $(x_t, t \in \mathbb{N})$ signals in this example.

with $A = \exp(-\tau \Delta t) I_{N_x}$ and $\{B^p = -\frac{1}{\tau}(A - I_{N_x})\widetilde{B}^p\}_{p \in \mathcal{P}}$, and the noise process $\boldsymbol{w} = (-\frac{1}{\tau}(A - I_{N_x})v_t)_{t \in \mathbb{N}}$ [47]. The companies assume that the baseline demand fluctuations satisfy $w_t \in \mathcal{W} = \{w_t \in \mathbb{R}^{N_x} : \|w_t\|_\infty \leq 1\}$ for all $t \in \mathbb{N}$. Under this representation, each player's objective is formulated as solving for a policy which minimizes the functional

$$J^p(\boldsymbol{u}^p, \boldsymbol{u}^{-p}) = \mathrm{E}\left[\sum_{t=0}^{\infty} \left(\alpha^p \|x_t\|_2^2 + \beta^p \|u_t^p\|_2^2\right)\right]$$

given weights $\alpha^p, \beta^p \in \mathbb{R}_{\geq 0}$. We sample $\alpha^p \sim \text{Uniform}(5, 15)$ and $\beta^p \sim \text{Uniform}(0.3, 0.6)$ for each $p \in \mathcal{P}$. The operational constraints, $\boldsymbol{u}^p \in U^p(\boldsymbol{u}^{-p})$, are defined by the constraint sets

$$\mathcal{X} = \mathbb{R}^{N_x}$$
$$\mathcal{U}^p = \mathbb{R}^{N_u^p}$$
$$\mathcal{U}_{\mathcal{G}} = \left\{u_t \in \textstyle\prod_{p \in \mathcal{P}} \mathbb{R}^{N_u^p} : \left[\frac{1}{N_P \cdot \bar{u}_{\text{avg}}} \mathbf{1}_{N_u}^{\mathsf{T}}\right] u_t \leq 1\right\}.$$

We consider that players design their state-feedback policies, $\boldsymbol{K}^p = \boldsymbol{\Phi}_{\boldsymbol{u}}^p \boldsymbol{\Phi}_{\boldsymbol{x}}^{-1} \in \boldsymbol{\mathcal{C}}^p$, with an FIR horizon of $N = 16$ and no structural constraints, that is, $\boldsymbol{S}_x = I$ and $\boldsymbol{S}_u^p = I$ $(\forall p \in \mathcal{P})$. In practice, this implies that companies have perfect information of any demand $\boldsymbol{x}^p$ and price $\boldsymbol{u}^p$ changes in the market.

As in Section IV-A, we simulate an instance of the game $\mathcal{G}_\infty^{\mathrm{LQ}}$ alongside the SLS-BRD routine (Algorithm 4) with players using (approximately)best-response maps, $\{\widehat{\mathrm{BR}}_\Phi^p\}_{p \in \mathcal{P}}$, given $\gamma = 0.95$. Policies are updated simultaneously every $\Delta T = 1$ stage with learning rate $\eta = 1/4$. Under the aforementioned setup, the game $\mathcal{G}_\infty^{\mathrm{LQ}}$ is not a potential game and thus a GFNE cannot be easily computed in advance. Finally, we remark that solutions to this problem are not unique: Different initial profiles may result in convergence to different equilibrium profiles.

The convergence of the SLS-BRD routine to a fixed-point $\boldsymbol{K}^\star = \boldsymbol{\Phi}_{\boldsymbol{u}}^\star \boldsymbol{\Phi}_{\boldsymbol{x}}^{-1} = (\boldsymbol{\Phi}_{\boldsymbol{u}}^{1^\star}, \ldots, \boldsymbol{\Phi}_{\boldsymbol{u}}^{N_P^\star}) \boldsymbol{\Phi}_{\boldsymbol{x}}^{-1}$ is shown in Fig. 6. Since the exact fixed-point to which the routine will converge is not known for this problem, we let $\boldsymbol{\Phi}_{\boldsymbol{u}}^\star \approx \boldsymbol{\Phi}_{\boldsymbol{u},\boldsymbol{k_f}}$ for $k_f = 1080$ (when the policy updates are interrupted) and analyze the convergence with respect to this point. In this case, the iterates approach the fixed-point mostly at a geometric rate with the

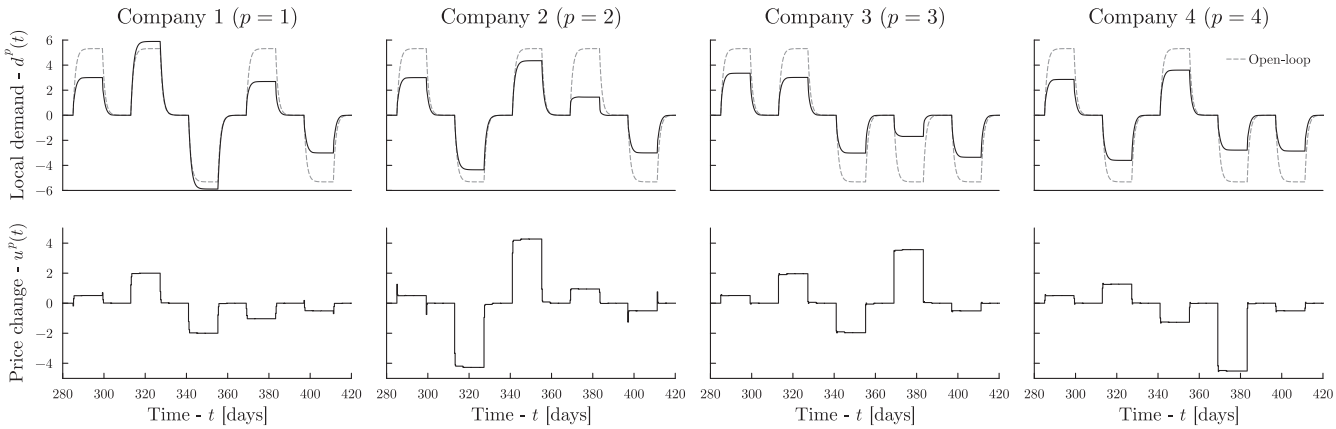

Fig. 8. Market game, $t \in (280, 420]$ days: State $\boldsymbol{x}$ (top panels) and applied control $\boldsymbol{Bu}$ (bottom panels) trajectories for an execution of the game $\mathcal{G}_\infty^{\mathsf{LQ}}$. The dashed lines refer to the state trajectories resulting from an open-loop operation of the market with $u(t) = 0$ for all $t \in \mathbb{R}_{\geq 0}$.

policies requiring roughly 1000 updates (or 250 days in-game) before changes become numerically negligible. We stress that the best-responses $\{\widehat{\mathrm{BR}}_\Phi^p\}_{p\in\mathcal{P}}$ cannot be (globally) contractive in this case, as the set of fixed points $\Omega^{\varepsilon}_{\mathcal{G}_\infty^\Phi}$ is not a singleton. Finally, this experiment imply that multiagent markets might require a half-year of learning dynamics (if policies are updated every $\Delta t = 1/4$ [days]) before converging to an equilibrium.

In Fig. 7(left), the relative distances between the individual updates $\{\boldsymbol{\Phi}^p_{\boldsymbol{u},\boldsymbol{k}}, \boldsymbol{\Phi}^p_{\boldsymbol{u},\boldsymbol{k-1}}\}_{p\in\mathcal{P}}$ are displayed. As in the previous example, the updates are of similar magnitude for all players $p \in \mathcal{P}$ and they become numerically negligible at a faster rate than the global convergence in Fig. 6. The convergence of the distance between updates is shown to slow considerably, especially for $p \in \{1, 4\}$, around $k = 50$. Thereafter, these relative distances continue to decrease at a steady rate. In Fig. 7(right), the relative distances $\|\boldsymbol{\Phi}_{\boldsymbol{x},\boldsymbol{k}}{}^p - \boldsymbol{\Phi}_{\boldsymbol{x},\boldsymbol{k}}\|_{\ell_2} / \|\boldsymbol{\Phi}_{\boldsymbol{x},\boldsymbol{k}}{}^p\|_{\ell_2}$, $k \in \mathbb{N}_+$, between responses $\{\boldsymbol{\Phi}^p_{\boldsymbol{x},\boldsymbol{k}}\}_{p\in\mathcal{P}}$ and $\boldsymbol{\Phi}_{\boldsymbol{x},\boldsymbol{k}} = \boldsymbol{F}_{\boldsymbol{\Phi}}\boldsymbol{\Phi}_{\boldsymbol{u},\boldsymbol{k}}$ are shown. As before, these distances decrease at a similar rate and are relatively small since the initial stages of the game.

The evolution of the game given each player's actions is displayed Fig. 8 for the *in-game* period $t \in (280, 420]$ days, after the policy updates have been interrupted. We compare the performance of the policy profile $\boldsymbol{K}^\star = \boldsymbol{\Phi}^\star_{\boldsymbol{u}}\boldsymbol{\Phi}^{-1}_{\boldsymbol{x}}$ with the evolution obtained by the open-loop operation $u(t) = 0$. During this period, we simulate a worst-case fluctuation on the baseline demand for the companies' product by defining

$$w(t) = \begin{cases} (\ 1, \ \ 1, \ \ 1, \ \ 1) & \text{for } t \in [285, 299], \\ (\ 1, -1, \ \ 1, -1) & \text{for } t \in [313, 327], \\ (-1, \ \ 1, -1, \ \ 1) & \text{for } t \in [341, 355], \\ (\ 1, \ \ 1, -1, -1) & \text{for } t \in [369, 383], \\ (-1, -1, -1, -1) & \text{for } t \in [397, 411], \end{cases}$$

and $w(t) = 0$ otherwise. The results show that the policy profile obtained by the SLS-BRD routine allows the companies to efficiently respond to changes in their local demands. In general, the players coordinate price changes to alleviate the deviations from the baseline demand caused by the disturbances, while still satisfying the price-cap constraint $\frac{1}{N_P}|\sum_{p\in\mathcal{P}} u^p(t)| \leq 0.5$. The best performance is observed for the companies $p \in \{3, 4\}$, whereas $p = 1$ behaves noticeably worse than all players (especially for $w(t) = (1, -1, 1, -1)$ and $w(t) = (-1, 1, -1, 1)$, when the open-loop operation attains better results). This highlights the fact that fixed-points obtained through the SLS-BRD routine are not necessarily *admissible* GFNE, and might be unfavorable for a subset of players. Finally, we note that these policies are still not able to completely reject the effect of the noise: Achieving zero-offset requires incorporating integral action into the feedback policies.

V. Conclusion

This work presents the SLS-BRD, an algorithm for GFNE seeking in N_P-player noncooperative games. The method is based on the BRD class of algorithms for NE seeking and consists of players updating and announcing a parameterization of their policies until converging to a fixed point. Our approach leverages the SLS framework to formulate each player's best-response map as the solution to robust finite-horizon optimal control problems. Because not updating control actions explicitly, this learning dynamics can be performed alongside the execution of the game. This framework also benefits the SLS-BRD by allowing richer information patterns to be enforced directly at the synthesis level. Using results from the operator theory, we then established sufficient conditions for this procedure to converge to a GFNE. After the theoretical aspects are discussed, the algorithm is showcased on exemplary problems on the noncooperative control of multiagent systems.

Appendix
Proof of Theorem 4

In the following, for ease of notation and without loss of generality, we treat linear operators as matrices.

The assumptions in Theorem 4 imply that the mapping $\widehat{\mathrm{BR}}^p_\phi$ (Problem (42)) is the solution to a convex quadratic problem with linear inequality constraints [43]. The optimiser can be obtained

from the Karush-Kuhn-Tucker system of equations

$$\begin{bmatrix} \boldsymbol{H}^{pp} & (\boldsymbol{G}^p_{\mathcal{A}})^{\mathsf{T}} \\ \boldsymbol{G}^p_{\mathcal{A}} & \end{bmatrix} \begin{bmatrix} \boldsymbol{\Phi}^p_{\boldsymbol{u}} \\ \nu \end{bmatrix} = \begin{bmatrix} \boldsymbol{H}^{p,-p}\boldsymbol{\Phi}^{-p}_{\boldsymbol{u}} + \boldsymbol{H}^{p0} \\ \mathbf{1}_{N_{\mathcal{A}}} \end{bmatrix} \tag{45}$$

with $(\boldsymbol{G}^p_{\mathcal{A}}, \mathbf{1}_{N_{\mathcal{A}}})$ defining the inequality constraints that are active at the solution. Using block-elimination [43, §C.4.3], the best-response is thus obtained as

$$\widehat{\mathrm{BR}}^{\,p}_{\phi}(\boldsymbol{\Phi}^{-p}_{\boldsymbol{u}}) = \boldsymbol{Q}^p \boldsymbol{H}^{p,-p}\boldsymbol{\Phi}^{-p}_{\boldsymbol{u}} + (\text{affine terms}),$$

with $\boldsymbol{Q}^p = (\boldsymbol{H}^{pp})^{-1} + (\boldsymbol{H}^{pp})^{-1}(\boldsymbol{G}^p_{\mathcal{A}})^{\mathsf{T}}(\boldsymbol{S}^p)^{-1}\boldsymbol{G}^p_{\mathcal{A}}(\boldsymbol{H}^{pp})^{-1}$ and $\boldsymbol{S}^p$ being the Schur complement of the matrix in Eq. (45). Using standard results from operator theory [37, §2.2.3] and linear algebra [48], we have that

$$\begin{aligned} L^p_{\widehat{\mathrm{BR}}_\phi} = \|\boldsymbol{Q}^p \boldsymbol{H}^{p,-p}\|_2 &= \sigma_{\max}(\boldsymbol{Q}^p)\sigma_{\max}(\boldsymbol{H}^{p,-p}) \\ &\le \sigma_{\max}((\boldsymbol{H}^{pp})^{-1})\sigma_{\max}(\boldsymbol{H}^{p,-p}) \\ &= \sigma_{\max}(\boldsymbol{H}^{p,-p})/\sigma_{\min}(\boldsymbol{H}^{pp}) \end{aligned}$$

since $(\boldsymbol{H}^{pp})^{-1}$ and $(\boldsymbol{H}^{pp})^{-1}(\boldsymbol{G}^p_{\mathcal{A}})^{\mathsf{T}}(\boldsymbol{S}^p)^{-1}\boldsymbol{G}^p_{\mathcal{A}}(\boldsymbol{H}^{pp})^{-1}$ are positive and negative definite, respectively. Finally, we have

$$\begin{aligned} (L_{\widehat{\mathrm{BR}}_\phi})^2 &= \left\|\mathrm{col}(\boldsymbol{Q}^p \boldsymbol{H}^{p,-p})_{p\in\mathcal{P}}\right\|_2^2 \\ &= \left\|\sum\nolimits_{p\in\mathcal{P}} (\boldsymbol{Q}^p \boldsymbol{H}^{p,-p})^{\mathsf{T}} \boldsymbol{Q}^p \boldsymbol{H}^{p,-p}\right\|_2 \\ &\le \sum\nolimits_{p\in\mathcal{P}} \|\boldsymbol{Q}^p \boldsymbol{H}^{p,-p}\|_2^2. \end{aligned}$$

which then implies that $(L_{\widehat{\mathrm{BR}}_\phi})^2 \le \sum_{p\in\mathcal{P}} (L^p_{\widehat{\mathrm{BR}}_\phi})^2$.

REFERENCES

[1] T. Başar and G. J. Olsder, *Dynamic Noncooperative Game Theory*. Philadelphia, PA, USA: SIAM, 1998.

[2] F. Facchinei and C. Kanzow, "Generalized Nash equilibrium problems," *Ann. Operations Res.*, vol. 175, no. 1, pp. 177–211, 2010.

[3] C. Daskalakis, P. W. Goldberg, and C. H. Papadimitriou, "The complexity of computing a Nash equilibrium," *Commun. ACM*, vol. 52, no. 2, pp. 89–97, 2009.

[4] N. Nisan, T. Roughgarden, E. Tardos, and V. V. Vazirani, *Algorithmic Game Theory*. Cambridge, U.K.: Cambridge Univ. Press, 2011.

[5] A. Cortés and S. Martínez, "Self-triggered best-response dynamics for continuous games," *IEEE Trans. Autom. Control*, vol. 60, no. 4, pp. 1115–1120, Apr. 2015.

[6] S. Grammatico, "Dynamic control of agents playing aggregative games with coupling constraints," *IEEE Trans. Autom. Control*, vol. 62, no. 9, pp. 4537–4548, Sep. 2017.

[7] B. Swenson, R. Murray, and S. Kar, "On best-response dynamics in potential games," *SIAM J. Control Optim.*, vol. 56, no. 4, pp. 2734–2767, 2018.

[8] N. Cesa-Bianchi and G. Lugosi, *Prediction, Learning, and Games*. Cambridge, U.K.: Cambridge Univ. Press, 2006.

[9] G. J. Gordon, A. Greenwald, and C. Marks, "No-regret learning in convex games," in *Proc. 25th Int. Conf. Mach. Learn.*, 2008, pp. 360–367.

[10] A. Celli, A. Marchesi, G. Farina, and N. Gatti, "No-regret learning dynamics for extensive-form correlated equilibrium," in *Proc. Int. Conf. Adv. Neural Inf. Process. Syst.*, 2020, vol. 33, pp. 7722–7732.

[11] P. Yi and L. Pavel, "An operator splitting approach for distributed generalized Nash equilibria computation," *Automatica*, vol. 102, pp. 111–121, 2019.

[12] G. Belgioioso, P. Yi, S. Grammatico, and L. Pavel, "Distributed generalized Nash equilibrium seeking: An operator-theoretic perspective," *IEEE Control Syst. Mag.*, vol. 42, no. 4, pp. 87–102, Aug. 2022.

[13] W. Saad, Z. Han, H. V. Poor, and T. Basar, "Game-theoretic methods for the smart grid: An overview of microgrid systems, demand-side management, and smart grid communications," *IEEE Signal Process. Mag.*, vol. 29, no. 5, pp. 86–105, Sep. 2012.

[14] S. Maghsudi and S. Stańczak, "Channel selection for network-assisted D2D communication via no-regret bandit learning with calibrated forecasting," *IEEE Trans. Wireless Commun.*, vol. 14, no. 3, pp. 1309–1322, Mar. 2015.

[15] Z. Wang, R. Spica, and M. Schwager, "Game theoretic motion planning for multi-robot racing," in *Proc. 14th Int. Symp. Distrib. Auton. Robot. Syst.*, 2019, pp. 225–238.

[16] J. F. Fisac, E. Bronstein, E. Stefansson, D. Sadigh, S. S. Sastry, and A. D. Dragan, "Hierarchical game-theoretic planning for autonomous vehicles," in *Proc. IEEE Int. Conf. Robot. Automat.*, 2019, pp. 9590–9596.

[17] R. Spica, E. Cristofalo, Z. Wang, E. Montijano, and M. Schwager, "A real-time game theoretic planner for autonomous two-player drone racing," *IEEE Trans. Robot.*, vol. 36, no. 5, pp. 1389–1403, Oct. 2020.

[18] M. Braverman, J. Mao, J. Schneider, and M. Weinberg, "Selling to a no-regret buyer," in *Proc. ACM Conf. Econ. Comput.*, 2018, pp. 523–538.

[19] D. Aussel, L. Brotcorne, S. Lepaul, and L. von Niederhäusern, "A trilevel model for best response in energy demand-side management," *Eur. J. Oper. Res.*, vol. 281, no. 2, pp. 299–315, 2020.

[20] J. Engwerda, W. Van Den Broek, and J. M. Schumacher, "Feedback Nash equilibria in uncertain infinite time horizon differential games," in *Proc. 14th Int. Symp. Math. Theory Netw. Syst.*, 2000, pp. 1–6.

[21] D. Vrabie and F. Lewis, "Integral reinforcement learning for online computation of feedback Nash strategies of nonzero-sum differential games," in *Proc. IEEE 49th Conf. Decis. Control*, 2010, pp. 3066–3071.

[22] G. Kossioris, M. Plexousakis, A. Xepapadeas, A. de Zeeuw, and K.-G. Mäler, "Feedback Nash equilibria for non-linear differential games in pollution control," *J. Econ. Dyn. Control*, vol. 32, no. 4, pp. 1312–1331, 2008.

[23] R. Kamalapurkar, J. R. Klotz, and W. E. Dixon, "Concurrent learning-based approximate feedback-Nash equilibrium solution of N-player nonzero-sum differential games," *IEEE/CAA J. Autom. Sinica*, vol. 1, no. 3, pp. 239–247, Jul. 2014.

[24] A. Tanwani and Q. Zhu, "Feedback Nash equilibrium for randomly switching differential–algebraic games," *IEEE Trans. Autom. Control*, vol. 65, no. 8, pp. 3286–3301, Aug. 2020.

[25] P. V. Reddy and G. Zaccour, "Feedback Nash equilibria in linear-quadratic difference games with constraints," *IEEE Trans. Autom. Control*, vol. 62, no. 2, pp. 590–604, Feb. 2017.

[26] P. V. Reddy and G. Zaccour, "Open-loop and feedback Nash equilibria in constrained linear–quadratic dynamic games played over event trees," *Automatica*, vol. 107, pp. 162–174, 2019.

[27] D. Fridovich-Keil, E. Ratner, L. Peters, A. D. Dragan, and C. J. Tomlin, "Efficient iterative linear-quadratic approximations for nonlinear multi-player general-sum differential games," in *Proc. IEEE Int. Conf. Robot. Automat.*, 2020, pp. 1475–1481.

[28] D. Fridovich-Keil, V. Rubies-Royo, and C. J. Tomlin, "An iterative quadratic method for general-sum differential games with feedback linearizable dynamics," in *Proc. IEEE Int. Conf. Robot. Automat.*, 2020, pp. 2216–2222.

[29] F. Laine, D. Fridovich-Keil, C.-Y. Chiu, and C. Tomlin, "The computation of approximate generalized feedback Nash equilibria," *SIAM J. Optim.*, vol. 33, no. 1, pp. 294–318, 2023.

[30] G. Arslan and S. Yüksel, "Decentralized Q-learning for stochastic teams and games," *IEEE Trans. Autom. Control*, vol. 62, no. 4, pp. 1545–1558, Apr. 2017.

[31] Z. Gao, Q. Ma, T. Başar, and J. R. Birge, "Finite-sample analysis of decentralized Q-learning for stochastic games," 2021, *arXiv:2112.07859*.

[32] K. Zhang, Z. Yang, and T. Başar, "Multi-agent reinforcement learning: A selective overview of theories and algorithms," in *Handbook of Reinforcement Learning and Control*. Berlin, Germany: Springer, 2021, pp. 321–384.

[33] B. Yongacoglu, G. Arslan, and S. Yüksel, "Decentralized learning for optimality in stochastic dynamic teams and games with local control and global state information," *IEEE Trans. Autom. Control*, vol. 67, no. 10, pp. 5230–5245, Oct. 2022.

[34] B. Yongacoglu, G. Arslan, and S. Yüksel, "Satisficing paths and independent multiagent reinforcement learning in stochastic games," *SIAM J. Math. Data Sci.*, vol. 5, no. 3, pp. 745–773, 2023.

[35] J. Anderson, J. C. Doyle, S. H. Low, and N. Matni, "System level synthesis," *Annu. Rev. Control*, vol. 47, pp. 364–393, 2019.

[36] I. L. Glicksberg, "A further generalization of the Kakutani fixed theorem, with application to Nash equilibrium points," *Proc. Amer. Math. Soc.*, vol. 3, no. 1, pp. 170–174, 1952.

[37] E. K. Ryu and W. Yin, *Large-Scale Convex Optimization: Algorithms & Analyses via Monotone Operators*. Cambridge, U.K.: Cambridge Univ. Press, 2022.

[38] J. Liang, J. Fadili, and G. Peyré, "Convergence rates with inexact nonexpansive operators," *Math. Program.*, vol. 159, pp. 403–434, 2016.
[39] T. Başar, "Stochastic differential games and intricacy of information structures," in *Dynamic Games in Economics*. Berlin, Germany: Springer, 2014, pp. 23–49.
[40] M. Rotkowitz and S. Lall, "A characterization of convex problems in decentralized control," *IEEE Trans. Autom. Control*, vol. 50, no. 12, pp. 1984–1996, Feb. 2006.
[41] L. Furieri, Y. Zheng, A. Papachristodoulou, and M. Kamgarpour, "An input–output parametrization of stabilizing controllers: Amidst Youla and system level synthesis," *IEEE Contr. Syst. Lett.*, vol. 3, no. 4, pp. 1014–1019, Oct. 2019.
[42] Y. Wang and S. Boyd, "Fast model predictive control using online optimization," *IEEE Trans. Control Syst. Technol.*, vol. 18, no. 2, pp. 267–278, Mar. 2010.
[43] S. P. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, U.K.: Cambridge Univ. Press, 2004.
[44] A. Nayyar and T. Başar, "Dynamic stochastic games with asymmetric information," in *Proc. IEEE 51st Conf. Decis. Control*, 2012, pp. 7145–7150.
[45] F. Borrelli, A. Bemporad, and M. Morari, *Predictive Control for Linear and Hybrid Systems*. Cambridge, U.K.: Cambridge Univ. Press, 2017.
[46] S. Zazo, S. Valcarcel Macua, M. Sánchez-Fernández, and J. Zazo, "Dynamic potential games with constraints: Fundamentals and applications in communications," *IEEE Trans. Signal Process.*, vol. 64, no. 14, pp. 3806–3821, Jul. 2016.
[47] K. J. Åström and B. Wittenmark, *Computer-Controlled Systems: Theory and Design*. New York, NY, USA: Dover, 2013.
[48] R. A. Hom and C. R. Johnson, *Matrix Analysis*. Cambridge Univ. Press, 2 ed. 2012.

Otacilio B. L. Neto (Student Member, IEEE) received the B.Eng. degree in computer engineering and the M.Sc. degree in teleinformatics engineering from the Federal University of Ceará, Fortaleza, Brazil, in 2019 and 2021, respectively.

He is currently a Doctoral Researcher with the Process Control and Automation Group, Aalto University, Espoo, Finland. His research interests include automatic control, optimization, and dynamic game theory, with applications to (bio)chemical systems.

Michela Mulas received the Laurea degree in chemical engineering and the Ph.D. degree in industrial engineering from the University of Cagliari, Cagliari, Italy, in 2001 and 2006, respectively.

She has been a Researcher with the Laboratory of Process Control and Automation, Helsinki University of Technology, Espoo, Finland, and a Research Fellow with the Water Engineering Research Group, Aalto University. She is currently a Professor with the Department of Teleinformatics Engineering, Federal University of Ceará, Fortaleza, Brazil. Her research interests include predictive control, soft sensors, and fault diagnosis theory with special focus on their application on wastewater treatment plants.

Francesco Corona received the M.Sc. degree in chemical engineering and the Ph.D. degree in industrial engineering from the University of Cagliari, Cagliari, Italy, in 2002 and 2007, respectively.

He is an Associate Professor in process control and automation and a Docent in computer science with the Aalto University, Espoo, Finland. His research interests include automatic control, statistical machine learning, and information visualization, with applications to environmental process systems.