SLS-BRD: A system-level approach to seeking generalised feedback Nash equilibria

Otacilio B. L. Neto, Michela Mulas, and Francesco Corona

***Abstract*—This work proposes a policy learning algorithm for generalised feedback Nash equilibrium seeking in N_P-players non-cooperative dynamic games. We consider linear-quadratic games with stochastic dynamics and design a best-response dynamics in which players update and communicate a parametrisation of their state-feedback policies. Our approach leverages the System Level Synthesis (SLS) framework to formulate each player's update rule as the solution of a tractable robust optimisation problem. Under certain conditions, the conditions and rates of convergence can be established. The algorithm is showcased for an exemplary problem from decentralised control of multi-agent systems.**

***Index Terms*—Non-cooperative games, Feedback Nash equilibrium, Best-response dynamics, System level synthesis**

I. Introduction

MODERN cyber-physical systems are often comprised of interacting subsystems operated locally by selfish decision-making agents. Ideally, agents operate these systems according to feedback policies that optimise local objectives, while satisfying global requirements and being robust to their rivals' interference. However, the large-scale and decentralised nature of most applications, alongside the usual lack of coordination between agents, hinders most traditional approaches to policy design. Dynamic game theory provides an alternative framework based on the concept of competitive equilibria (e.g., the Nash equilibrium [1, 2]), which describes locally optimal, yet strategically stable, operating conditions for each non-cooperative agent. A policy design methodology based on Nash equilibrium seeking thus presents a promising venue for enabling a decentralised control of cyber-physical systems.

In general, solving a non-cooperative game can concern distinct goals: *i*) For the unbiased observer, to examine the behaviour of rational agents from a local to a global scale; *ii*) for the game designer, to actuate the self-interested players towards system-wide objectives; and *iii*) for the player, to determine locally optimal and robust policies in competitive environments. Regardless, computing a Nash equilibrium (NE) solution is a notoriously difficult task [3]. Algorithmic game theory thus emerges as the field concerned with designing procedures to bridge this computational gap [4]. An important class of algorithms are those which place to the players the task of converging towards a competitive equilibrium: These include best-response dynamics (BRD, [5]–[10]) and no-regret learning (NRL, [11]–[15]). Recently, online feedback optimisation has also been employed for equilibrium seeking problems [16]. In general, these are iterative optimisation-based methods centred on players updating their strategies, simultaneously and independently, using the information available to them. In particular, best-response dynamics stands out as a simple, yet fundamental, model of policy learning for uncoordinated but communicating players. These algorithms have become important tools in economics and engineering, with applications including the design and control of networked systems [17]–[19], robotics [20, 21], and resource management [22]–[24].

For dynamic games, any solution concept depends qualitatively on a description of the data available to each player; a notion termed *information pattern* [1]. Under *open-loop* information patterns, when players only access the initial state of the game, the analysis and solution of NE problems are well-understood. Conversely, games under *closed-loop* (or *feedback*) information patterns, when players monitor the state of the game, are still under active research. A feedback Nash equilibrium (FNE) solution is often more attractive than its open-loop counterpart: They render players robust to disturbances and decision errors. However, designing routines for computing FNE is demanding for all but simple cases. A notably challenging class of problems consists of generalised feedback Nash equilibrium (GFNE, [2]) seeking: When players' policies are required to satisfy coupled constraints (e.g., restrictions on the state of the game). To the best of our knowledge, methods for FNE seeking exist only for linear-quadratic games under restrictive assumptions [25]–[28], unconstrained nonlinear and control-affine games [29]–[32], and application-specific problems [33, 34]. Recently, [35] expanded these results and presented a systematic (approximate) solution for GFNE problems. Despite remarkable, these efforts are restricted to finite-duration games and thus exclude decision-making processes in which agents are continuously operating the underlying system. Finally, a systematic solution is also lacking for GFNE games associated with stochastic dynamics.

In this work, we investigate algorithms for GFNE seeking in (stationary) difference games with linear-stochastic dynamics. Leveraging the System Level Synthesis (SLS, [36]) framework, we propose a best-response approach based on a parametrisation of all stabilising policies for the game. The algorithm consists of players iteratively, and simultaneously, updating the parameters of their individual policies, then announcing these changes through some communication network (Figure 1). Under our approach, each player's update rule is formulated as the solution to a tractable robust optimisation problem. Moreover, these problems allow for constraints to be imposed directly on the policies' structure, thus having the ability to encode information patterns at the synthesis level. Finally, this proposed learning

Otacilio B. L. Neto and Francesco Corona are with the Department of Chemical and Metallurgical Engineering, Aalto University, Finland (e-mails: {otacilio.neto, francesco.corona}@aalto.fi).

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

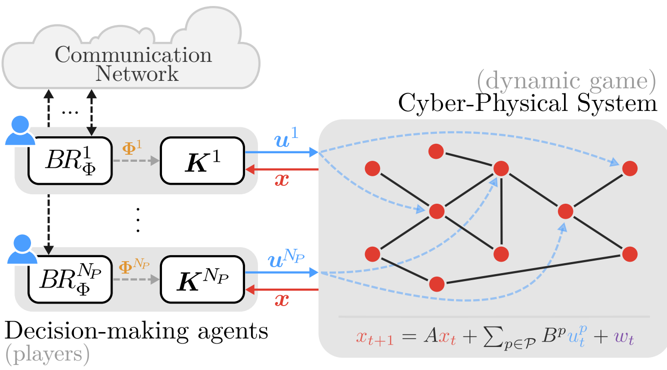

Fig. 1. SLS-BRD: Control architecture and the learning dynamics.

dynamics do not depend on the state and actions applied to the system, and thus can be performed simultaneously with the game's execution. In summary, our contributions are:

(i) A system-level best-response dynamics (SLS-BRD) algorithm for GFNE seeking in linear-stochastic dynamic games. Specifically, we design a class of best-response maps which compute an optimal system level parametrisation of a player's policy in response to its rivals' choices.

(ii) A realisation of the best-response mappings as robust finite-dimensional programs amenable to numerical solutions. Considering these best-responses, the SLS-BRD algorithm is formally an ε-GFNE seeking algorithm for an equilibrium gap $\varepsilon > 0$ which can be explicitly established.

(iii) For a specific but important class of dynamic games, an analysis of the conditions and rates of convergence of the SLS-BRD to feedback Nash equilibria.

This policy learning algorithm is showcased in a simulated experiment on the decentralised control of an unstable network.

The paper is organised as follows: Section II overviews the classes of (generalised) static and dynamic games, and the best-response dynamics algorithm. In Section III, we provide a system level parametrisation for linear-stochastic games, then design a best-response dynamics for GFNE seeking. Finally, Section IV illustrates this learning dynamics in simulated examples, and Section V provides some concluding remarks. Towards a concise presentation, only some essential theorems are proved in the main text: The remaining are in the Appendix.

A. Notation

We use Latin letters to denote vectors (lowercase) and mappings (uppercase), and use 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 denoted $\boldsymbol{x} = (x_t)_{t\in\mathcal{I}}$ 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 use 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 $\ell_\infty^{N_x}$ and $\ell_{\infty,e}^{N_x}$ the space of all bounded sequences and all sequences, respectively. $\mathcal{L}(\boldsymbol{\mathcal{X}}, \boldsymbol{\mathcal{Y}})$ is the set of all bounded linear operators $\boldsymbol{A} : \boldsymbol{\mathcal{X}} \to \boldsymbol{\mathcal{Y}}$ and we sometimes denote transformed signals by $\boldsymbol{A}\boldsymbol{x} = (Ax_t)_{t\in\mathcal{I}}$. Whenever $(\boldsymbol{\mathcal{X}}, \boldsymbol{\mathcal{Y}})$ are finite-dimensional, we let M_A refer to a matrix representation of $\boldsymbol{A}$. We use the standard definition of Hardy spaces $\mathcal{H}_\infty$ and $\mathcal{RH}_\infty$, and write $\frac{1}{z}\mathcal{RH}_\infty$ to refer to the set of real-rational strictly proper transfer functions. Finally, some standard signals and operators used in this paper are: The impulse signal $\boldsymbol{\delta} = (\delta_t)_{t\in\mathcal{I}}$, the identity operator I and the identity matrix I_{N_x}, and the shift operator $\boldsymbol{S}_+ : (x_0, x_1, \dots) \mapsto (0, x_0, \dots)$.

We distinguish set-valued mappings from ordinary functions using the notation $F : \mathcal{X} \rightrightarrows \mathcal{Y}$. 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 re-ordering. 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. Non-cooperative games and best-response dynamics

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 minimise an objective function $L^p : \mathcal{S}^1 \times \dots \times \mathcal{S}^{N_P} \to \mathbb{R}$. 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 rivals' strategies. Finally, the players are assumed to be rational, non-cooperative, and acting simultaneously.

A solution to the game $\mathcal{G}$ is understood as a strategy profile $s = (s^1, \dots, s^{N_P}) \in \mathcal{S}$, $\mathcal{S} = \mathcal{S}^1 \times \dots \times \mathcal{S}^{N_P}$, having some specified property that makes it agreeable to all players if they act rationally. In non-cooperative settings, a widely accepted solution concept is that of a generalized Nash equilibrium: The game is *solved* when no player can improve its objective by unilaterally deviating from the agreed strategy profile. Formally,

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

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

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

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

$$\Omega_\mathcal{G} := \{s^* \in \mathcal{S} : s^* \text{ satisfies Eq. (2)}\},$$

is not a singleton. As such, players might favour different solutions based on their individual objectives. We thus seek to characterise which solutions are "acceptable", in the sense that no other equilibrium can improve the objective of all players simultaneously. Consider the partial ordering of $\Omega_\mathcal{G}$ defined by

$$s^* \prec \tilde{s}^* \iff L^p(s^{p^*}) \leq L^p(\tilde{s}^{p^*}), \quad \forall p \in \mathcal{P}, \tag{3}$$

with at least one inequality being strict, for any pair $s^*, \tilde{s}^* \in \Omega_\mathcal{G}$. A solution $s^* \in \Omega_\mathcal{G}$ is then characterised as an *admissible* GNE if there are no other equilibrium $\tilde{s}^* \in \Omega_\mathcal{G}$ such that $\tilde{s}^* \prec s^*$. Note that an admissible GNE can still be favourable to only a subset of players (that is, $\tilde{s}^* \not\prec s^* \not\prec \tilde{s}^*$), and thus it is possible that $\mathcal{G}$ has no solution that is "fair" to all players. The game might also not admit any GNE (i.e., $\Omega_{\mathcal{G}} = \emptyset$), then being characterised 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 $p \in \mathcal{P}$,*

a) 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 p-th argument, $s^p \in S^p(s^{-p})$, for every $s^{-p} \in \mathcal{S}^{-p}$.

b) the mapping $S^p : \mathcal{S}^p \rightrightarrows \mathcal{S}^{-p}$ takes the form

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

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 generalisation of the Kakutani fixed-point theorem ensures that $\mathcal{G}$ has a GNE, that is, $\Omega_{\mathcal{G}} \neq \emptyset$ [37]. 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 optimisation 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 $s_k \coloneqq (s_k^1, \ldots, s_k^{N_P})$ be the strategy profile taken by players $\mathcal{P}$ at the k-th episode. A prototypical learning procedure for equilibrium seeking is outlined in Algorithm 1, where

- $T^p : \mathcal{S}^p \times \mathcal{S}^{-p} \rightrightarrows \mathcal{S}^p$ describes how the p-th player updates its strategy, based on its opponents' predicted next actions and given its individual objective;
- $R^p : \mathcal{S} \rightrightarrows \mathcal{S}^{-p}$ describes how the p-th player's predicts its opponents' strategies for the next episode, based on the strategy profile currently being played.

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 $s^* = (s^{1^*}, \ldots, s^{N_P^*})$

```
Initialize s_0 = (s_0^1, ..., s_0^{N_P}) and k = 0;
for k = 1, 2, ... do
    if s_k ∈ T(s_k) then return s_k;
    for p ∈ P do
        Update s_{k+1}^p ∈ T^p(s_k^p, R^p(s_k) | L^p);
```

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

Best-response dynamics: Let the mapping $BR^p : \mathcal{S}^{-p} \rightrightarrows \mathcal{S}^p$,

$$BR^p(s^{-p}) \coloneqq \underset{s^p \in S^p(s^{-p})}{\arg\min} \; L^p(s^p, s^{-p}) \tag{4}$$

denote the best-response of $p \in \mathcal{P}$ to other players' strategies. Collectively, $BR(s) \coloneqq BR^1(s^{-1}) \times \cdots \times BR^{N_P}(s^{-N_P}) \subseteq \mathcal{S}$ is the joint best-response to any given profile $s \in \mathcal{S}$. Then,

Theorem 1. *A strategy profile $s^* = (s^{1^*}, \ldots, s^{N_P^*}) \in \mathcal{S}$ is a GNE for $\mathcal{G}$ if and only if $s^* \in BR(s^*)$ or, equivalently,*

$$s^{p^*} \in BR^p(s^{-p^*}), \quad \forall p \in \mathcal{P}. \tag{5}$$

The task of computing a Nash equilibrium can thus be translated into searching for a fixed-point of the set-valued mapping $BR : \mathcal{S} \rightrightarrows \mathcal{S}$. The set of GNE solutions for $\mathcal{G}$ is the set of all such fixed-points, $\Omega_{\mathcal{G}} \coloneqq \{s^* \in \mathcal{S} : s^* \in BR(s^*)\}$. A natural procedure for GNE seeking thus consists of players adapting their strategies towards best-responses to their rivals' strategies, which they assume will remain constant. Formally,

$$T^p(s_k^p, R^p(s_k)) \coloneqq (1-\eta)s_k^p + \eta BR^p(R^p(s_k)), \tag{6}$$

given $R^p(s_k) = s_k^{-p}$ and a learning rate factor of $\eta \in (0, 1)$. This learning dynamics, summarised in Algorithm 2, is known as best-response dynamics or fictitious play.

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 $s^* = (s^{1^*}, \ldots, s^{N_P^*})$

```
Initialize s_0 = (s_0^1, ..., s_0^{N_P}) and k = 0;
for k = 0, 1, 2, ... do
    if s_k ∈ BR(s_k) then return s_k;
    for p ∈ P do
        Update s_{k+1}^p ∈ (1−η)s_k^p + ηBR^p(s_k^{−p}) ;
```

After each episode, the strategy profile gets updated as

$$s_{k+1} = \underbrace{(1-\eta)s_k + \eta BR(s_k)}_{T(s_k)}, \tag{7}$$

given the global update rule $T = (1-\eta)I + \eta BR$. 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.

Theorem 2. *Let $BR : \mathcal{S} \rightrightarrows \mathcal{S}$ be a non-expansive mapping. Then, the best-response dynamics $s_{k+1} = T(s_k)$ converge monotonically to a GNE solution $s^* \in \Omega_{\mathcal{G}}$, that is,*

$$\lim_{k\to\infty} \inf_{s^* \in \Omega_{\mathcal{G}}} \|T(s_k) - s^*\| = 0 \tag{8}$$

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

This convergence result stems from fixed-point theory, where the BRD algorithm is interpreted as belonging to the class of averaged (or Mann-Krasnosel'skii) iteration methods [38, 39]. For non-generalised quadratic problems, we show in Section III-B that the best-response mapping BR is non-expansive under specific conditions on $L^1, \dots, L^{N_P}$. Importantly, if the best-response mapping is a contraction, then the BRD converge geometrically to a GNE solution $s^* \in \Omega_{\mathcal{G}}$, which is unique.

Theorem 3. *Let $BR : \mathcal{S} \rightrightarrows \mathcal{S}$ be L_{BR}-Lipschitz, $L_{BR} < 1$. Then, from any feasible $s_0 \in \mathcal{S}$, the best-response dynamics $s_{k+1} = T(s_k)$ converge to the unique GNE $s^* \in \Omega_{\mathcal{G}}$ with rate*

$$\frac{\|s_k - s^*\|}{\|s_0 - s^*\|} \leq \big((1-\eta) + \eta L_{BR}\big)^k \tag{9}$$

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

The learning rate η plays an important role in the 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 $s_k \in \mathcal{S}_{\mathcal{G}} \cap \mathcal{S}$. Otherwise, the learning dynamics become undefined and the aforementioned convergence results do not apply. For the choice $\eta = 1/2$, the update rule T is *firmly non-expansive* whenever BR is non-expansive and thus carries better convergence properties [39]. For non-generalised games, the updates trivially satisfy the constraints for any $\eta \in (0, 1)$ and thus a careful design of the learning rate might not be necessary. In this case, $\eta \to 1$ is the optimal choice if BR is a contraction and the convergence rate in Eq. (9) can be simplified to L_{BR}^k. However, in practice, choosing a small η can still improve the stability of the BRD when the mappings $\{BR_p\}_{p\in\mathcal{P}}$ are solved numerically.

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

$$L^p(s^p_{k_f}, s^{-p}_{k_f}) \leq \min_{s^p \in S^p(s^{-p}_{k_f})} L^p(s^p, s^{-p}_{k_f}) + \varepsilon \tag{10}$$

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 his strategy. The set of all ε-GNEs is denoted $\Omega^\varepsilon_{\mathcal{G}} = \{s^\varepsilon \in \mathcal{S} : s^\varepsilon \text{ satisfies Eq. (10)}\}$.

A. Infinite-horizon dynamic games

An infinite-horizon dynamic N_P-player game, denoted by tuple

$$\mathcal{G}_\infty := (\mathcal{P}, \boldsymbol{\mathcal{X}}, \{\boldsymbol{\mathcal{U}}^p\}_{p\in\mathcal{P}}, \boldsymbol{\mathcal{W}}, \{J^p\}_{p\in\mathcal{P}}), \tag{11}$$

is defined by the stochastic linear dynamics

$$x_{t+1} = Ax_t + \sum_{p\in\mathcal{P}} B^p u^p_t + w_t, \quad x_0 \text{ given}, \tag{12}$$

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 realisation $\boldsymbol{w} \in \boldsymbol{\mathcal{W}}$ and initial x_0, the state is explicitly expressed as $\boldsymbol{x} = \boldsymbol{F_w}\boldsymbol{u}$ via the causal affine operator

$$\boldsymbol{F_w} : \boldsymbol{u} \mapsto (I - \boldsymbol{S}_+\boldsymbol{A})^{-1}\Big(\sum_{p\in\mathcal{P}} \boldsymbol{S}_+\boldsymbol{B}^p\boldsymbol{u}^p + \boldsymbol{S}_+\boldsymbol{w} + \boldsymbol{\delta}x_0\Big), \tag{13}$$

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 dependency on x_0 is omitted to simplify notation. Moreover, we assume $\mathrm{E}w_t = 0$ and $\mathrm{E}(w_t w_{t'}^\mathsf{T}) = \delta_{t-t'}\Sigma_w$, given a covariance matrix $\Sigma_w \in \mathbb{S}^{N_x}_{++}$, for every $t, t' \in \mathbb{N}$. 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

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

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 dynamic stochastic games, each player decides on a plan of action $\boldsymbol{u}^p \in U^p(\boldsymbol{u}^{-p})$ to minimize its objective functional

$$J^p(\boldsymbol{u}^p, \boldsymbol{u}^{-p}) := \mathrm{E}\left[\sum_{t=0}^{\infty} L^p(x_t, u^p_t, u^{-p}_t)\right], \tag{14}$$

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 strategies for each player based on their rivals' strategies. Under this formulation, the dynamic game $\mathcal{G}_\infty$ is stationary and can be interpreted as a static game defined on the appropriate functional spaces. A plan of action $\boldsymbol{u} = (\boldsymbol{u}^1, \dots, \boldsymbol{u}^{N_P}) \in \boldsymbol{\mathcal{U}}$ can then be characterised as a generalised Nash Equilibrium solution to $\mathcal{G}_\infty$ when no player can improve its objective by unilaterally deviating from this agreed profile. Formally,

Definition 2. A strategy profile $\boldsymbol{u}^* = (\boldsymbol{u}^{1^*}, \dots, \boldsymbol{u}^{N^*_P}) \in \boldsymbol{\mathcal{U}}$ is a generalized Nash equilibrium (GNE) for the game $\mathcal{G}_\infty$ if

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

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

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

$$\Omega_{\mathcal{G}_\infty} := \{\boldsymbol{u}^* \in \boldsymbol{\mathcal{U}} : \boldsymbol{u}^* \text{ satisfies Eq. (15)}\},$$

is not necessarily a singleton and can include non-admissible equilibria. The game is considered unsolvable if $\Omega_{\mathcal{G}_\infty} = \emptyset$. The following assumptions are taken for this class of games:

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

a) 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 p-th argument, $\boldsymbol{u}^p \in U^p(\boldsymbol{u}^{-p})$, for every sequence $\boldsymbol{u}^{-p} \in \boldsymbol{\mathcal{U}}^{-p}$.

b) the mapping $U^p : \boldsymbol{\mathcal{U}}^{-p} \rightrightarrows \boldsymbol{\mathcal{U}}^p$ takes the form

$$U^p(\boldsymbol{u}^{-p}) := \{\boldsymbol{u}^p \in \boldsymbol{\mathcal{U}}^p : (\boldsymbol{u}^p, \boldsymbol{u}^{-p}) \in \boldsymbol{F}^\dagger_{\boldsymbol{w}}(\boldsymbol{\mathcal{X}}) \cap \boldsymbol{\mathcal{U}}_{\mathcal{G}}\},$$

where $\boldsymbol{F}^\dagger_{\boldsymbol{w}}$ is a left-inverse of the operator in Eq. (13) and

$$\boldsymbol{\mathcal{U}}_{\mathcal{G}} = \{\boldsymbol{u} \in \textstyle\prod_{p\in\mathcal{P}} \ell^{N^p_u}_\infty(\mathbb{N}) : (u^p_t, u^{-p}_t) \in \mathcal{U}_{\mathcal{G}},\ t \in \mathbb{N}\}$$

is a global constraint set shared by all players. The sets $\mathcal{U}^p$, $\mathcal{U}_{\mathcal{G}}$, and $\mathcal{X}$ are all non-empty, compact, and convex. Finally, we have that $\boldsymbol{F}^\dagger_{\boldsymbol{w}}(\boldsymbol{\mathcal{X}}) \cap \boldsymbol{\mathcal{U}}_{\mathcal{G}} \cap (\boldsymbol{\mathcal{U}}^1 \times \cdots \times \boldsymbol{\mathcal{U}}^{N_P}) \neq \emptyset$.

These conditions are analogous to those of Assumption 1: They ensure the existence of GNE solutions to $\mathcal{G}_\infty$, that is, $\Omega_{\mathcal{G}_\infty} \neq \emptyset$. In this case, the players not only share a global constraint $\boldsymbol{\mathcal{U}}_{\mathcal{G}}$, but are also required to ensure that state trajectories lie in a feasible set, $x \in \boldsymbol{\mathcal{X}}$, against all possible noise realisations. These constraints often describe operational desiderata and/or limitations concerning the game $\mathcal{G}_\infty$. Moreover, note that Assumption 2 implies that the operator $\boldsymbol{F_w}$ is injective, that is, that distinct control signals cannot produce the same state trajectory. A sufficient condition for this property consists on the matrix $B = [B^1 \; B^2 \; \cdots \; B^{N_P}]$ being full-rank.

Any equilibrium $\boldsymbol{u}^* \in \Omega_{\mathcal{G}_\infty}$ incorporates an open-loop information pattern: Actions $u_t^* = (u_t^{1^*}, \ldots, u_t^{N_P^*})$ depend explicitly only on initial $x_0 \in \mathcal{X}$ and stage $t \in \mathbb{N}$. A plan of action with such representation is undesirable, as players become sensible to noise disturbances and decision errors. Conversely, a state-feedback decision policy, $\boldsymbol{u}^* = K(\boldsymbol{x})$ for some $K : \boldsymbol{\mathcal{X}} \to \boldsymbol{\mathcal{U}}$, often carry desirable robustness properties. In this sense, a feedback (respectively, open-loop) representation of $\boldsymbol{u}^* \in \Omega_{\mathcal{G}_\infty}$ is 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 p-th player's actions are represented as

$$\boldsymbol{u}^p := \boldsymbol{K}^p \boldsymbol{x}, \quad \boldsymbol{K}^p : \boldsymbol{x} \mapsto \boldsymbol{\Phi^p} * \boldsymbol{x}, \tag{16}$$

given a causal operator $\boldsymbol{K}^p \in \boldsymbol{\mathcal{C}}^p \subseteq \mathcal{L}(\boldsymbol{\mathcal{X}}, \boldsymbol{\mathcal{U}}^p)$ 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 setting, players do not devise their actions explicitly but rather by designing a state-feedback policy profile $\boldsymbol{K} := (\boldsymbol{K}^1, \ldots, \boldsymbol{K}^{N_P}) \in \boldsymbol{\mathcal{C}}$, 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 generalized feedback Nash equilibrium.

Definition 3. A policy profile $\boldsymbol{K}^* = (\boldsymbol{K}^{1^*}, \ldots, \boldsymbol{K}^{N_P^*})$ is a generalized feedback Nash equilibrium (GFNE) for $\mathcal{G}_\infty$ if

$$J^p(\boldsymbol{u}^{p^*}, \boldsymbol{u}^{-p^*}) \leq \min_{\boldsymbol{u}^p \in U^p(\boldsymbol{u}^{-p^*})} J^p(\boldsymbol{u}^p, \boldsymbol{u}^{-p^*}), \tag{17}$$

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

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

$$\Omega_{\mathcal{G}_\infty}^{\boldsymbol{K}} := \{\boldsymbol{K}^* \in \boldsymbol{\mathcal{C}} : \boldsymbol{u}^* = \boldsymbol{K}^* \boldsymbol{x}^* \text{ satisfies Eq. (17)}\}.$$

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

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

is bounded ($\boldsymbol{x}^* \in \ell_\infty^{N_x}$) for all bounded noise ($\boldsymbol{w} \in \ell_\infty^{N_x}$). A policy satisfying this requirement is said to be *stabilising*. From Assumption 2, we have $\Omega_{\mathcal{G}_\infty}^{\boldsymbol{K}} \neq \emptyset$ when $\boldsymbol{\mathcal{C}} = \mathcal{L}(\boldsymbol{\mathcal{X}}, \boldsymbol{\mathcal{U}})$. For a more specific $\boldsymbol{\mathcal{C}} \subset \mathcal{L}(\boldsymbol{\mathcal{X}}, \boldsymbol{\mathcal{U}})$, establishing the existence (and, especially, uniqueness) of a solution is demanding [40]. In practice, the set $\Omega_{\mathcal{G}_\infty}^{\boldsymbol{K}}$ can be constructed from open-loop equilibria $\boldsymbol{u}^* \in \Omega_{\mathcal{G}_\infty}$ by *i)* parametrising the set of all possible trajectories $\{\boldsymbol{x}_{\boldsymbol{w}}^* = \boldsymbol{F_w} \boldsymbol{u}^*\}_{\boldsymbol{w} \in \boldsymbol{\mathcal{W}}}$, then *ii)* identifying policies $(\boldsymbol{K}^{1^*}, \ldots, \boldsymbol{K}^{N_P^*})$ that satisfy $\{\boldsymbol{u}^{p^*} = \boldsymbol{K}^{p^*} \boldsymbol{x}_{\boldsymbol{w}}^*\}_{\boldsymbol{w} \in \boldsymbol{\mathcal{W}}}$, $p \in \mathcal{P}$. Highlighting this equivalence, we refer to such $\boldsymbol{u}^*$ as the *open-loop realisation* of the *closed-loop policy* $\boldsymbol{K}^*$, and vice-versa.

Best-response dynamics for GFNE seeking: The mapping

$$BR^p(\boldsymbol{u}^{-p}) := \operatorname*{arg\,min}_{\boldsymbol{u}^p \in U^p(\boldsymbol{u}^{-p})} J^p(\boldsymbol{u}^p, \boldsymbol{u}^{-p}) \tag{18}$$

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

$$\begin{aligned}
\underset{\boldsymbol{u}^p := \boldsymbol{K}^p \boldsymbol{x}}{\text{minimize}} \quad & \mathrm{E}\left[\sum_{t=0}^{\infty} L^p(x_t, u_t^p, u_t^{-p})\right] && \text{(19a)}\\
\underset{\forall t \in \mathbb{N}}{\text{subject to}} \quad & x_{t+1} = A x_t + \textstyle\sum_{\tilde{p} \in \mathcal{P}} B^{\tilde{p}} u_t^{\tilde{p}} + w_t, && \text{(19b)}\\
& x_t \in \mathcal{X}, \quad u_t^p \in \mathcal{U}^p, \quad (u_t^p, u_t^{-p}) \in \mathcal{U}_{\mathcal{G}}, && \text{(19c)}\\
& \boldsymbol{K}^p \in \boldsymbol{\mathcal{C}}^p, && \text{(19d)}\\
& (x_0 \text{ given}). && \text{(19e)}
\end{aligned}$$

While posed in terms of action signals ($\boldsymbol{u}^p$, $p \in \mathcal{P}$), Problem (19) 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 $BR^p(\boldsymbol{K}^{-p})$ be its solutions, when the problem is parametrised by $\boldsymbol{u}^{-p} := \boldsymbol{K}^{-p}\boldsymbol{x} = (\boldsymbol{K}^{\tilde{p}}\boldsymbol{x})_{\tilde{p} \in \mathcal{P} \setminus \{p\}}$. The mapping $BR : \boldsymbol{\mathcal{C}} \rightrightarrows \boldsymbol{\mathcal{C}}$, defined by $BR(\boldsymbol{K}) = BR^1(\boldsymbol{K}^{-1}) \times \cdots \times BR^{N_P}(\boldsymbol{K}^{-N_P})$, is the joint best-response to a strategy profile $\boldsymbol{K}$. The GFNEs of $\mathcal{G}_\infty$ are thus the fixed-points of this mapping: That is, $\Omega_{\mathcal{G}_\infty}^{\boldsymbol{K}} = \{\boldsymbol{K}^* \in \boldsymbol{\mathcal{C}} : \boldsymbol{K}^* \in BR(\boldsymbol{K}^*)\}$. Due to constraints ($\mathcal{X}, \mathcal{U}^p, \mathcal{U}_{\mathcal{G}}$) and $\boldsymbol{\mathcal{C}}^p$, an analytical solution to Problem (19) does not exist. Moreover, because infinite-dimensional, its numerical approximation cannot be obtained.

The BRD method for GFNE seeking is given in Algorithm 3. As $\mathcal{G}_\infty$ is dynamic and stationary, the procedure needs not 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 $t \in \{(k{+}1)\Delta T\}_{k \in \mathbb{N}}$. $\boldsymbol{K}_k := (\boldsymbol{K}_k^1, \ldots, \boldsymbol{K}_k^{N_P})$ 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 $\{\boldsymbol{K}_k^{\tilde{p}}\}_{\tilde{p} \in \mathcal{P} \setminus \{p\}}$).

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 $\boldsymbol{K}^* = (\boldsymbol{K}^{1^*}, \ldots, \boldsymbol{K}^{N_P^*})$

```
Initialize K_0 = (K_0^1, ..., K_0^{N_P}) and k = 0;
for t = 0, 1, 2, ... do
    /* Players apply actions {u_{k,t}^p = K_k^p 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_{k+1}^p ∈ (1−η)K_k^p + η BR^p(K_k^{−p}) ;
        k := k + 1;
```

If we verbally execute Algorithm 3, we have the following:

- 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 $\mathcal{T}_k = [k\Delta T, (k+1)\Delta T)$.

- At $t = (k+1)\Delta T$, every p-th player updates its policy,
$$\boldsymbol{K}_{k+1}^p \in (1-\eta)\boldsymbol{K}_k^p + \eta BR^p(\boldsymbol{K}_k^{-p}),$$
which is then announced to the other players.

The BRD-GFNE induces an operator $T = (1-\eta)I + \eta BR$ which is equivalent to the update rule of its static counterpart. Thus, it possesses the same properties: The learning dynamics converge if BR is non-expansive and the convergence rate is geometric if BR is also a contraction (Theorems 2–3). These properties can also be stated in terms of stage indices $t \in \mathbb{N}$ by substituting $k = \lfloor t/\Delta T \rfloor$. 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. Moreover, the BRD-GFNE can be interrupted at any episode $k_f > 1$ (producing an ϵ-GFNE, $\epsilon > 0$) and, if necessary, a central coordinator assigned to project the resulting strategy profile $\boldsymbol{K}_{k_f}$ onto the feasible set $\boldsymbol{F}_{\boldsymbol{w}}^{\dagger}(\boldsymbol{\mathcal{X}}) \cap \boldsymbol{\mathcal{U}}_{\mathcal{G}} \cap (\boldsymbol{\mathcal{U}}^1 \times \cdots \times \boldsymbol{\mathcal{U}}^{N_P})$. Finally, we note that such considerations can be disregarded when the game is not generalised, that is, when $\mathcal{X} = \mathbb{R}^{N_x}$ and $\mathcal{U}_{\mathcal{G}} = \prod_{p\in\mathcal{P}} \mathbb{R}^{N_u^p}$.

III. Best-response dynamics via system level synthesis

In this section, we present an approach for GFNE seeking in (stationary) stochastic dynamic games. Firstly, we introduce the system level parametrisation 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 optimisation problems. Then, a modified BRD-GFNE procedure is proposed and its convergence properties are investigated.

We focus on N_P-players linear-quadratic stochastic games $\mathcal{G}_\infty^{\text{LQ}} = (\mathcal{P}, \boldsymbol{\mathcal{X}}, \{\boldsymbol{\mathcal{U}}^p\}_{p\in\mathcal{P}}, \boldsymbol{\mathcal{W}}, \{J^p\}_{p\in\mathcal{P}})$ with dynamics
$$x_{t+1} = Ax_t + \sum_{p\in\mathcal{P}} B^p u_t^p + w_t, \quad x_0 \text{ given}, \tag{20}$$
and objective functionals
$$J^p(\boldsymbol{u}^p, \boldsymbol{u}^{-p}) = \mathrm{E}\left[\sum_{t=0}^{\infty}\left(\|C^p x_t\|_2^2 + \|\textstyle\sum_{\tilde{p}\in\mathcal{P}} D^{p\tilde{p}} u_t^{\tilde{p}}\|_2^2\right)\right], \tag{21}$$
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 GFNE of $\mathcal{G}_\infty^{\text{LQ}}$ are admissible (that is, stabilising):

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

a) The pair (A, B^p) is stabilisable;

b) The pair (C^p, A) is detectable;

c) The matrix D^{pp} is full column rank, i.e., $D^{pp^{\mathsf{T}}} D^{pp} \in \mathbb{S}_{++}^{N_u^p}$. Moreover, $D^{p\tilde{p}^{\mathsf{T}}} C^p = 0 = C^{p^{\mathsf{T}}} D^{p\tilde{p}}$ for all $\tilde{p} \in \mathcal{P}$.

Finally, the constraint sets $\mathcal{X}$ and $\mathcal{U}^p$ ($\forall p$) are convex polyhedra and contain their respective origins: $0 \in \mathcal{X}$ and $0 \in \mathcal{U}^p$.

The class $\mathcal{G}_\infty^{\text{LQ}}$ describe problems in which N_P non-cooperative agents have to agree on stationary policies that jointly stabilise a global system, robustly to the noise process, while penalising 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

System level synthesis (SLS, [36]) is a novel methodology for controller design which lifts the search for optimal controllers into the search of optimal closed-loop responses to disturbances. Under this approach, the synthesis of stationary control policies can be equivalently posed as the solution to scalable numerical problems, even when operational constraints on the state and control signals need to be enforced. Additionally, the method allows for constraints to be enforced directly on the structure of the control policies. In this section, we present a system-level parametrisation for the best-response mappings in $\mathcal{G}_\infty^{\text{LQ}}$.

We start by assuming a stabilising 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 Eq. (20),
$$z\hat{\boldsymbol{x}} = A\hat{\boldsymbol{x}} + \textstyle\sum_{p\in\mathcal{P}} B^p\hat{\boldsymbol{u}}^p + \hat{\boldsymbol{w}}; \tag{22a}$$
$$\hat{\boldsymbol{u}}^p = \hat{\boldsymbol{K}}^p\hat{\boldsymbol{x}}, \quad (\forall p \in \mathcal{P}), \tag{22b}$$
the signals $(\hat{\boldsymbol{x}}, \hat{\boldsymbol{u}}^1, \ldots, \hat{\boldsymbol{u}}^{N_P})$ can be expressed in terms of $\hat{\boldsymbol{w}}$,
$$\begin{aligned}\begin{bmatrix}\hat{\boldsymbol{x}}\\ \hat{\boldsymbol{u}}^1\\ \vdots\\ \hat{\boldsymbol{u}}^{N_P}\end{bmatrix} &= \begin{bmatrix}(zI - A - \sum_{p\in\mathcal{P}} B^p\hat{\boldsymbol{K}}^p)^{-1}\\ \hat{\boldsymbol{K}}^1(zI - A - \sum_{p\in\mathcal{P}} B^p\hat{\boldsymbol{K}}^p)^{-1}\\ \vdots\\ \hat{\boldsymbol{K}}^{N_P}(zI - A - \sum_{p\in\mathcal{P}} B^p\hat{\boldsymbol{K}}^p)^{-1}\end{bmatrix}\hat{\boldsymbol{w}};\\ &= \begin{bmatrix}\hat{\boldsymbol{\Phi}}_{\boldsymbol{x}}\\ \hat{\boldsymbol{\Phi}}_{\boldsymbol{u}}^1\\ \vdots\\ \hat{\boldsymbol{\Phi}}_{\boldsymbol{u}}^{N_P}\end{bmatrix}\hat{\boldsymbol{w}},\end{aligned} \tag{23}$$
where the introduced transfer matrices $(\hat{\boldsymbol{\Phi}}_{\boldsymbol{x}}, \hat{\boldsymbol{\Phi}}_{\boldsymbol{u}}^1, \ldots, \hat{\boldsymbol{\Phi}}_{\boldsymbol{u}}^{N_P})$ are denoted as *system level responses* or *closed-loop maps*. Under this representation, the following result holds.

Theorem 4 (System level parametrisation, [36])**.** *Consider the dynamics Eq.* (22) *under state-feedback $\hat{\boldsymbol{u}}^p = \hat{\boldsymbol{K}}^p\hat{\boldsymbol{x}}$ ($\forall p \in \mathcal{P}$). The following statements are true:*

a) The affine space
$$\begin{bmatrix}zI - A & -B^1 & \cdots & -B^{N_P}\end{bmatrix}\begin{bmatrix}\hat{\boldsymbol{\Phi}}_{\boldsymbol{x}}\\ \hat{\boldsymbol{\Phi}}_{\boldsymbol{u}}^1\\ \vdots\\ \hat{\boldsymbol{\Phi}}_{\boldsymbol{u}}^{N_P}\end{bmatrix} = I, \tag{24}$$
with $\hat{\boldsymbol{\Phi}}_{\boldsymbol{x}}, \hat{\boldsymbol{\Phi}}_{\boldsymbol{u}}^1, \ldots, \hat{\boldsymbol{\Phi}}_{\boldsymbol{u}}^{N_P} \in \frac{1}{z}\mathcal{RH}_\infty$, parametrizes 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 $(\hat{\boldsymbol{K}}^1, \ldots, \hat{\boldsymbol{K}}^{N_P})$.

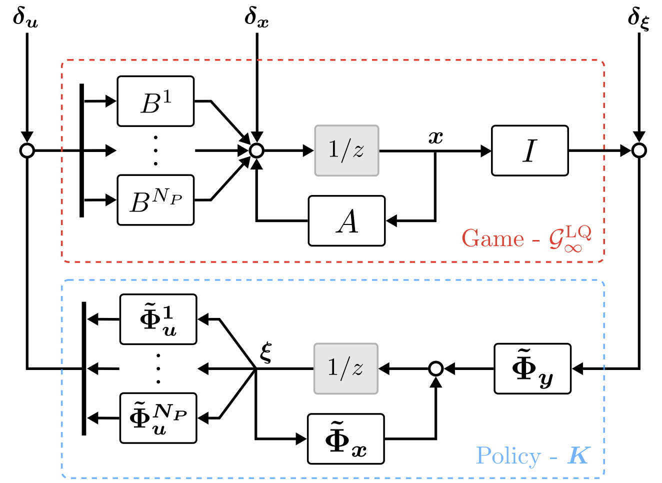

Fig. 2. Feedback structure for the policy $\hat{\boldsymbol{K}} = \tilde{\boldsymbol{\Phi}}_{\boldsymbol{u}}(zI - \tilde{\boldsymbol{\Phi}}_{\boldsymbol{x}})^{-1}\tilde{\boldsymbol{\Phi}}_{\boldsymbol{y}} = \hat{\boldsymbol{\Phi}}^p_{\boldsymbol{u}}\hat{\boldsymbol{\Phi}}_{\boldsymbol{x}}^{-1}$, equivalent to the internal representation in Eq. (25).

b) Any response $(\hat{\boldsymbol{\Phi}}_{\boldsymbol{x}}, \hat{\boldsymbol{\Phi}}^1_{\boldsymbol{u}}, \dots, \hat{\boldsymbol{\Phi}}^{N_P}_{\boldsymbol{u}})$ satisfying Eq. (24) is achieved by the policies $\hat{\boldsymbol{K}}^p = \hat{\boldsymbol{\Phi}}^p_{\boldsymbol{u}}\hat{\boldsymbol{\Phi}}_{\boldsymbol{x}}^{-1}$ ($\forall p \in \mathcal{P}$), which are internally stabilising and can be implemented as

$$z\hat{\boldsymbol{\xi}} = \tilde{\boldsymbol{\Phi}}_{\boldsymbol{x}}\hat{\boldsymbol{\xi}} + \hat{\boldsymbol{x}}; \tag{25a}$$

$$\hat{\boldsymbol{u}}^p = \tilde{\boldsymbol{\Phi}}^p_{\boldsymbol{u}}\hat{\boldsymbol{\xi}}, \tag{25b}$$

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

Proof. To simplify the notation, we momentarily define the matrices $B := [B^1 \ B^2 \ \cdots \ B^{N_P}]$, $\hat{\boldsymbol{\Phi}}_{\boldsymbol{u}} := \left[\hat{\boldsymbol{\Phi}}^{1^\intercal}_{\boldsymbol{u}} \ \hat{\boldsymbol{\Phi}}^{2^\intercal}_{\boldsymbol{u}} \cdots \hat{\boldsymbol{\Phi}}^{N_P^\intercal}_{\boldsymbol{u}}\right]^\intercal$, and $\hat{\boldsymbol{K}} := \left[\hat{\boldsymbol{K}}^{1^\intercal} \ \hat{\boldsymbol{K}}^{2^\intercal} \cdots \hat{\boldsymbol{K}}^{N_P^\intercal}\right]^\intercal$. For the first statement, consider internally stabilising policies $(\hat{\boldsymbol{K}}^1, \dots, \hat{\boldsymbol{K}}^{N_P})$ such that $\hat{\boldsymbol{u}}^p = \hat{\boldsymbol{K}}^p\hat{\boldsymbol{x}}$ ($\forall p \in \mathcal{P}$). Then, from Eq. (23), we have that

$$\begin{aligned}
&\begin{bmatrix} zI - A & -B \end{bmatrix} \begin{bmatrix} (zI - A - B\hat{\boldsymbol{K}})^{-1} \\ \hat{\boldsymbol{K}}(zI - A - B\hat{\boldsymbol{K}})^{-1} \end{bmatrix} \\
&= (zI - A)(zI - A - B\hat{\boldsymbol{K}})^{-1} - B\hat{\boldsymbol{K}}(zI - A - B\hat{\boldsymbol{K}})^{-1} \\
&= (zI - A - B\hat{\boldsymbol{K}})(zI - A - B\hat{\boldsymbol{K}})^{-1} \\
&= I.
\end{aligned}$$

For the second statement, we first show that $\hat{\boldsymbol{K}}$ achieves the desired response, and then that these policies are internally stabilising. Note that Eq. (24) implies that $\hat{\boldsymbol{\Phi}}_{\boldsymbol{x}}$ has the leading spectral component $\Phi_{x,1} = I_{N_x}$, which is invertible, and thus $\hat{\boldsymbol{\Phi}}_{\boldsymbol{x}}^{-1}$ exists. Then, $\hat{\boldsymbol{K}} = \hat{\boldsymbol{\Phi}}_{\boldsymbol{u}}\hat{\boldsymbol{\Phi}}_{\boldsymbol{x}}^{-1}$ is well-defined, and we have

$$\begin{aligned}
\hat{\boldsymbol{x}} &= (zI - A - B\hat{\boldsymbol{\Phi}}_{\boldsymbol{u}}\hat{\boldsymbol{\Phi}}_{\boldsymbol{x}}^{-1})^{-1}\hat{\boldsymbol{w}} \\
&= \hat{\boldsymbol{\Phi}}_{\boldsymbol{x}}\left((zI - A)\hat{\boldsymbol{\Phi}}_{\boldsymbol{x}} - B\hat{\boldsymbol{\Phi}}_{\boldsymbol{u}}\right)^{-1}\hat{\boldsymbol{w}} \\
&= \hat{\boldsymbol{\Phi}}_{\boldsymbol{x}}\hat{\boldsymbol{w}},
\end{aligned}$$

due to Eq. (24), and $\hat{\boldsymbol{u}} = \hat{\boldsymbol{K}}\hat{\boldsymbol{x}} = \hat{\boldsymbol{\Phi}}_{\boldsymbol{u}}\hat{\boldsymbol{\Phi}}_{\boldsymbol{x}}^{-1}\hat{\boldsymbol{\Phi}}_{\boldsymbol{x}}\hat{\boldsymbol{w}} = \hat{\boldsymbol{\Phi}}_{\boldsymbol{u}}\hat{\boldsymbol{w}}$. Thus, $\hat{\boldsymbol{K}}$ achieves the response $(\hat{\boldsymbol{\Phi}}_{\boldsymbol{x}}, \hat{\boldsymbol{\Phi}}_{\boldsymbol{u}})$, or, equivalently, $(\hat{\boldsymbol{K}}^1, \hat{\boldsymbol{K}}^2, \dots, \hat{\boldsymbol{K}}^{N_P})$ achieve $(\hat{\boldsymbol{\Phi}}_{\boldsymbol{x}}, \hat{\boldsymbol{\Phi}}^1_{\boldsymbol{u}}, \dots, \hat{\boldsymbol{\Phi}}^{N_P}_{\boldsymbol{u}})$. To show that this policy is internally stabilising, consider its equivalent representation $\hat{\boldsymbol{K}} = \tilde{\boldsymbol{\Phi}}_{\boldsymbol{u}}(zI - \tilde{\boldsymbol{\Phi}}_{\boldsymbol{x}})^{-1}\tilde{\boldsymbol{\Phi}}_{\boldsymbol{y}}$ with $\tilde{\boldsymbol{\Phi}}_{\boldsymbol{x}} = z(I - \hat{\boldsymbol{\Phi}}_{\boldsymbol{x}})$, $\tilde{\boldsymbol{\Phi}}_{\boldsymbol{u}} = z\hat{\boldsymbol{\Phi}}_{\boldsymbol{u}}$, and $\tilde{\boldsymbol{\Phi}}_{\boldsymbol{y}} = I$, as depicted in Figure 2. Introducing external perturbations $\{\delta_x, \delta_u, \delta_\xi\} \subseteq \ell_\infty$ into the game, it suffices to verify that the transfer matrices from $(\hat{\boldsymbol{\delta}}_{\boldsymbol{x}}, \hat{\boldsymbol{\delta}}_{\boldsymbol{u}}, \hat{\boldsymbol{\delta}}_{\boldsymbol{\xi}})$ to $(\hat{\boldsymbol{x}}, \hat{\boldsymbol{u}}, \hat{\boldsymbol{\xi}})$ are stable. In this case, we have

$$\begin{bmatrix} \hat{\boldsymbol{x}} \\ \hat{\boldsymbol{u}} \\ \hat{\boldsymbol{\xi}} \end{bmatrix} = \begin{bmatrix} \hat{\boldsymbol{\Phi}}_{\boldsymbol{x}} & \hat{\boldsymbol{\Phi}}_{\boldsymbol{x}}B & \hat{\boldsymbol{\Phi}}_{\boldsymbol{x}}(zI - A) \\ \hat{\boldsymbol{\Phi}}_{\boldsymbol{u}} & I + \hat{\boldsymbol{\Phi}}_{\boldsymbol{u}}B & \hat{\boldsymbol{\Phi}}_{\boldsymbol{u}}(zI - A) \\ \frac{1}{z}I & \frac{1}{z}B & \frac{1}{z}(zI - A) \end{bmatrix} \begin{bmatrix} \hat{\boldsymbol{\delta}}_{\boldsymbol{x}} \\ \hat{\boldsymbol{\delta}}_{\boldsymbol{u}} \\ \hat{\boldsymbol{\delta}}_{\boldsymbol{\xi}} \end{bmatrix}. \tag{26}$$

Since $\hat{\boldsymbol{\Phi}}_{\boldsymbol{x}}, \hat{\boldsymbol{\Phi}}_{\boldsymbol{u}} \in \frac{1}{z}\mathcal{RH}_\infty$, the transfer matrices in Eq. (26) are all stable. We thus conclude that the policy $\hat{\boldsymbol{K}}$ is internally stabilising, and so are the policies $(\hat{\boldsymbol{K}}^1, \hat{\boldsymbol{K}}^2, \dots, \hat{\boldsymbol{K}}^{N_P})$. $\square$

We refer to the system level responses through their kernels, $\boldsymbol{\Phi}_{\boldsymbol{x}} = (\Phi_{x,n})_{n\in\mathbb{N}} \in \ell_2(\mathbb{N})$ and $\boldsymbol{\Phi}^p_{\boldsymbol{u}} = (\Phi^p_{u,n})_{n\in\mathbb{N}} \in \ell_2(\mathbb{N})$, for all $p \in \mathcal{P}$. Due to strict causality, $\Phi_{x,0} = 0$ and $\Phi^p_{u,0} = 0$. From Theorem 4, the operators $\{\boldsymbol{K}^p \in \boldsymbol{\mathcal{C}}^p\}_{p\in\mathcal{P}}$ and the transfer matrices $\{\hat{\boldsymbol{K}}^p \in \mathcal{RH}_\infty\}_{p\in\mathcal{P}}$ are equivalent representations of the feedback policies. Hence, provided there is no confusion, we use exclusively the first notation. In particular, $\boldsymbol{K}^p = \boldsymbol{\Phi}^p_{\boldsymbol{u}}\boldsymbol{\Phi}_{\boldsymbol{x}}^{-1}$ ($p \in \mathcal{P}$) denotes the policy parametrised by $(\hat{\boldsymbol{\Phi}}_{\boldsymbol{x}}, \hat{\boldsymbol{\Phi}}^p_{\boldsymbol{u}})$ and $\boldsymbol{K} = (\boldsymbol{\Phi}^1_{\boldsymbol{u}}, \cdots, \boldsymbol{\Phi}^{N_P}_{\boldsymbol{u}})\boldsymbol{\Phi}_{\boldsymbol{x}}^{-1}$ denotes the corresponding profile. A time-domain characterisation of $\boldsymbol{K}$ is given in the following.

Corollary 4.1. *A policy $\boldsymbol{K}^p = \boldsymbol{\Phi}^p_{\boldsymbol{u}}\boldsymbol{\Phi}_{\boldsymbol{x}}^{-1}$ ($\forall p \in \mathcal{P}$) is defined by the kernel $\boldsymbol{\Phi}^p = \boldsymbol{\Phi}^p_{\boldsymbol{u}} * \boldsymbol{\Phi}_{\boldsymbol{x}}^{-1}$, and implemented as*

$$\xi_t = -\textstyle\sum_{\tau=1}^t \Phi_{x,\tau+1}\xi_{t-\tau} + x_t; \tag{27a}$$

$$u^p_t = \textstyle\sum_{\tau=0}^t \Phi^p_{u,\tau+1}\xi_{t-\tau}, \tag{27b}$$

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

The system level parametrisation enables a methodology for policy synthesis consisting of searching the space of stabilising policies (in)directly through $(\boldsymbol{\Phi}_{\boldsymbol{x}}, \boldsymbol{\Phi}^p_{\boldsymbol{u}})$, $p \in \mathcal{P}$. In particular, this parametrisation can be leveraged to reformulate the best-response dynamics maps in $\mathcal{G}^{\text{LQ}}_\infty$ as tractable numerical programs. In this direction, consider that players design stabilising policies $\boldsymbol{K} = (\boldsymbol{K}^1, \dots, \boldsymbol{K}^{N_P})$ by choosing their desired system level responses $\boldsymbol{\Phi}_{\boldsymbol{u}} = (\boldsymbol{\Phi}^1_{\boldsymbol{u}}, \dots, \boldsymbol{\Phi}^{N_P}_{\boldsymbol{u}})$ simultaneously. From the affine space Eq. (24), the signal $\boldsymbol{\Phi}_{\boldsymbol{x}}$, common to all players, satisfies the (deterministic) linear dynamics

$$\Phi_{x,n+1} = A\Phi_{x,n} + \sum_{p\in\mathcal{P}} B^p\Phi^p_{u,n}, \quad \Phi_{x,1} = I_{N_x}, \tag{28}$$

or $\boldsymbol{\Phi}_{\boldsymbol{x}} = \boldsymbol{F}_{\boldsymbol{\Phi}}\boldsymbol{\Phi}_{\boldsymbol{u}}$ given the causal affine operator

$$\boldsymbol{F}_{\boldsymbol{\Phi}} : \boldsymbol{\Phi}_{\boldsymbol{u}} \mapsto (I - \boldsymbol{S}_+\boldsymbol{A})^{-1}\Big(\sum_{p\in\mathcal{P}} \boldsymbol{S}_+\boldsymbol{B}^p\boldsymbol{\Phi}^p_{\boldsymbol{u}} + \boldsymbol{\delta}I_{N_x}\Big), \tag{29}$$

Using Parseval's Theorem [41], we substitute Eq. (23) into Eq. (21) to redefine the objective functionals of $\mathcal{G}^{\text{LQ}}_\infty$ as

$$\begin{aligned}
&J^p(\boldsymbol{\Phi}^p_{\boldsymbol{u}}, \boldsymbol{\Phi}^{-p}_{\boldsymbol{u}}) \\
&= \mathrm{E}\Big[\sum_{n=1}^\infty \Big(\|C^p\Phi_{x,n}w_n\|_F^2 + \|\textstyle\sum_{\tilde{p}\in\mathcal{P}} D^{p\tilde{p}}\Phi^{\tilde{p}}_{u,n}w_n\|_F^2\Big)\Big] \\
&= \sum_{n=1}^\infty \Big(\|C^p\Phi_{x,n}\Sigma_w^{\frac{1}{2}}\|_F^2 + \|\textstyle\sum_{\tilde{p}\in\mathcal{P}} D^{p\tilde{p}}\Phi^{\tilde{p}}_{u,n}\Sigma_w^{\frac{1}{2}}\|_F^2\Big).
\end{aligned}$$

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

$$\mathcal{G}^{\Phi}_\infty := (\mathcal{P}, \boldsymbol{\mathcal{C}}_x, \{\boldsymbol{\mathcal{C}}^p_u\}_{p\in\mathcal{P}}, \boldsymbol{\mathcal{W}}, \{J^p\}_{p\in\mathcal{P}}), \tag{30}$$

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

$$\begin{aligned} U_\Phi^p(\boldsymbol{\Phi}_{\boldsymbol{u}}^{-p}) \coloneqq \{\boldsymbol{\Phi}_{\boldsymbol{u}}^p \in \boldsymbol{\mathcal{C}}_u^p : (\boldsymbol{\Phi}_{\boldsymbol{u}}^p, \boldsymbol{\Phi}_{\boldsymbol{u}}^{-p}) \in \boldsymbol{F}_{\boldsymbol{\Phi}}^\dagger(\boldsymbol{\mathcal{C}}_x), \\ \boldsymbol{\Phi}_{\boldsymbol{u}}^p * \boldsymbol{w} \in U^p(\boldsymbol{\Phi}_{\boldsymbol{u}}^{-p} * \boldsymbol{w})\}, \end{aligned}$$

which incorporate the constraints U^p from the original $\mathcal{G}_\infty^{\text{LQ}}$. The sets $(\boldsymbol{\mathcal{C}}_x, \boldsymbol{\mathcal{C}}_u^p)$ are designed to enforce the policy constraints $\boldsymbol{K}^p \in \mathcal{C}^p$ directly through the kernels $(\boldsymbol{\Phi}_{\boldsymbol{x}}, \boldsymbol{\Phi}_{\boldsymbol{u}}^p)$: They are related as $\boldsymbol{\mathcal{C}}_u^p = \{\boldsymbol{K}^p \boldsymbol{\mathcal{C}}_x : \boldsymbol{K}^p \in \mathcal{C}^p\}$. We refer to a joint response $\boldsymbol{\Phi}_u = (\boldsymbol{\Phi}_{\boldsymbol{u}}^1, \ldots, \boldsymbol{\Phi}_{\boldsymbol{u}}^{N_P}) \in \boldsymbol{\mathcal{C}}_u$, $\boldsymbol{\mathcal{C}}_u = \boldsymbol{\mathcal{C}}_u^1 \times \cdots \times \boldsymbol{\mathcal{C}}_u^{N_P}$, 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

$$BR_\Phi^p(\boldsymbol{\Phi}_{\boldsymbol{u}}^{-p}) \coloneqq \underset{\boldsymbol{\Phi}_{\boldsymbol{u}}^p \in U_\Phi^p(\boldsymbol{\Phi}_{\boldsymbol{u}}^{-p})}{\arg\min} \; J^p(\boldsymbol{\Phi}_{\boldsymbol{u}}^p, \boldsymbol{\Phi}_{\boldsymbol{u}}^{-p}),$$

consisting of the set of closed-loop maps $\boldsymbol{\Phi}_{\boldsymbol{u}}^p$ which are best-responses to the maps of other players, $\boldsymbol{\Phi}_{\boldsymbol{u}}^{-p} = (\boldsymbol{\Phi}_{\boldsymbol{u}}^{\tilde{p}})_{\tilde{p} \in \mathcal{P} \setminus \{p\}}$. They are solutions to the system level synthesis problem

$$\begin{aligned}
\underset{\boldsymbol{\Phi}_{\boldsymbol{u}}^p}{\text{minimize}} \quad & \sum_{n=1}^{\infty} \Big(\|C^p \Phi_{x,n} \Sigma_w^{\frac{1}{2}}\|_F^2 + \Big\| \sum_{\tilde{p} \in \mathcal{P}} D^{p\tilde{p}} \Phi_{u,n}^{\tilde{p}} \Sigma_w^{\frac{1}{2}} \Big\|_F^2 \Big) & \text{(31a)} \\
\underset{\forall n \in \mathbb{N}^+}{\text{subject to}} \quad & \Phi_{x,n+1} = A\Phi_{x,n} + \textstyle\sum_{\tilde{p} \in \mathcal{P}} B^{\tilde{p}} \Phi_{u,n}^{\tilde{p}}, & \text{(31b)} \\
& (\Phi_x * w)_n \in \mathcal{X}, \quad (\Phi_u^p * w)_n \in \mathcal{U}^p, \\
& \qquad \big((\Phi_u^p * w)_n, (\Phi_u^{-p} * w)_n\big) \in \mathcal{U}_{\mathcal{G}}, & \text{(31c)} \\
& \boldsymbol{\Phi}_{\boldsymbol{x}} \in \boldsymbol{\mathcal{C}}_x, \quad \boldsymbol{\Phi}_{\boldsymbol{u}}^p \in \boldsymbol{\mathcal{C}}_u^p, & \text{(31d)} \\
& \Phi_{x,1} = I_{N_x}. & \text{(31e)}
\end{aligned}$$

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

Remark 1. *Problem* (31) *is independent of the matrix $\Sigma_w^{\frac{1}{2}}$ due to its linear-quadratic structure and the fact that $\Sigma_w \in \mathbb{S}^{N_x}$. Hence, we remove it from J^p ($\forall p$) to simplify notation.*

The best-responses $\{BR_\Phi^p\}_{p \in \mathcal{P}}$ are still intractable: *i)* They are defined by infinite-dimensional problems with no general solution and *ii)* that require full knowledge of the noise process $(w_n)_{n \in \mathbb{N}}$ to formulate the constraints U_Φ^p. In the following, we tackle both issues and provide a class of finite-dimensional robust optimisation programs that approximate Problem (31). We conclude the section by presenting a class of system level constraints which enforce a richer feedback information pattern.

Finite-horizon approximation: The programs in $\{BR_\Phi^p\}_{p \in \mathcal{P}}$ can be made finite-dimensional by restricting the closed-loop maps to the set of finite-impulse responses (FIR),

$$\begin{aligned}
\boldsymbol{\mathcal{C}}_x &= \{\boldsymbol{\Phi}_{\boldsymbol{x}} \in \ell_2[0, N] : \Phi_{x,n} \in \mathcal{C}_{x,n}, \; n \in [0, N), \; \Phi_{x,N} = 0\}; \\
\boldsymbol{\mathcal{C}}_u^p &= \{\boldsymbol{\Phi}_{\boldsymbol{u}}^p \in \ell_2[0, N) : \Phi_{u,n}^p \in \mathcal{C}_{u,n}^p, \; n \in [0, N)\},
\end{aligned}$$

given horizon $N \in [2, \infty)$. We enforce $\boldsymbol{\Phi}_{\boldsymbol{x}} \in \boldsymbol{\mathcal{C}}_x$ and $\boldsymbol{\Phi}_{\boldsymbol{u}}^p \in \boldsymbol{\mathcal{C}}_u^p$ in Problem (31) by adding the terminal constraint $\Phi_{x,N} = 0$, and restricting the spectral factors to satisfy $(\Phi_{x,n} \in \mathcal{C}_{x,n})_{n=1}^N$ and $(\Phi_{u,n}^p \in \mathcal{C}_{u,n}^p)_{n=1}^{N-1}$. The constraint sets $\mathcal{C}_{x,n} \subseteq \mathbb{R}^{N_x \times N_x}$ and $\mathcal{C}_{u,n}^p \subseteq \mathbb{R}^{N_u \times N_x}$ ($\forall p$) have no restriction other than being compact convex sets. The resulting problems are of the form

$$\begin{aligned}
\underset{\boldsymbol{\Phi}_{\boldsymbol{u}}^p = (\Phi_{u,n}^p)_{n=1}^{N-1}}{\text{minimize}} \quad & \sum_{n=1}^{N} \Big(\|C^p \Phi_{x,n}\|_F^2 + \| \textstyle\sum_{\tilde{p} \in \mathcal{P}} D^{p\tilde{p}} \Phi_{u,n}^{\tilde{p}} \|_F^2 \Big) & \text{(32a)} \\
\underset{\forall n \in [1, N)}{\text{subject to}} \quad & \Phi_{x,n+1} = A\Phi_{x,n} + \textstyle\sum_{\tilde{p} \in \mathcal{P}} B^{\tilde{p}} \Phi_{u,n}^{\tilde{p}}, & \text{(32b)} \\
& (\Phi_x * w)_n \in \mathcal{X}, \quad (\Phi_u^p * w)_n \in \mathcal{U}^p, \\
& \qquad \big((\Phi_u^p * w)_n, (\Phi_u^{-p} * w)_n\big) \in \mathcal{U}_{\mathcal{G}}, & \text{(32c)} \\
& \Phi_{x,n} \in \mathcal{C}_{x,n}, \quad \Phi_{u,n}^p \in \mathcal{C}_{u,n}^p, & \text{(32d)} \\
& \Phi_{x,1} = I_{N_x}, \quad \Phi_{x,N} = 0, & \text{(32e)}
\end{aligned}$$

which define finite-dimensional convex programs that can be solved numerically. Since $(\boldsymbol{\mathcal{C}}_x, \boldsymbol{\mathcal{C}}_u^p)$ are finite-dimensional, the operators $(\boldsymbol{F}_{\boldsymbol{\Phi}}, \boldsymbol{K})$ and the convolutions in Eq. (32c) can be represented by matrix multiplications: Consider any kernel $\boldsymbol{\Phi} \in \{\boldsymbol{\Phi}_{\boldsymbol{x}}\} \cup \{\boldsymbol{\Phi}^p, \boldsymbol{\Phi}_{\boldsymbol{u}}^p\}_{p \in \mathcal{P}}$ and some signal $\boldsymbol{z} \in \ell_\infty$. Then,

$$\boldsymbol{z}^{\text{out}} = \boldsymbol{\Phi} * \boldsymbol{z} \iff z^{\text{out}} = M_\Phi z,$$

given block-Toeplitz matrix $M_\Phi = [\Phi_{i-j}]_{i \in [1, 2N+1], j \in [1, N+1]}$ and (z, z^{out}) being a vector representation of these signals, that is, $z = \texttt{col}(z_0, \ldots, z_N)$ and $z^{\text{out}} = \texttt{col}(z_0^{\text{out}}, \ldots, z_{2N}^{\text{out}})$. Moreover, $z_n^{\text{out}} = (\Phi * z)_n = [M_\Phi]_n z$, where $[M_\Phi]_n$ is the n-th block-row of M_Φ. Thus, Problem (32) can be reformulated into a program requiring only matrix algebra. Finally, we remark that policies $(\boldsymbol{K}^p, p \in \mathcal{P})$ are implemented either by using Corollary 4.1 or by directly computing kernels $\boldsymbol{\Phi}^p = ([M_{\Phi_u^p}]_n \Phi_x^{-1})_{n \in \mathbb{N}}$.

Although realising Problem (31) into a tractable program, the Problem (32) is only feasible when the pair (A, B^p) is full-state controllable. This is a difficult requirement in multi-agent settings, as often $N_u^p \ll N_x$ for all $p \in \mathcal{P}$, leading to Problem (32) becoming overdetermined. 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

$$\begin{aligned}
\boldsymbol{\mathcal{C}}_x &= \{\boldsymbol{\Phi}_{\boldsymbol{x}} \in \ell_2[0, N] : \Phi_{x,n} \in \mathcal{C}_{x,n}, \; n \in [0, N), \; \|\Phi_{x,N}\|_F^2 \le \gamma\}; \\
\boldsymbol{\mathcal{C}}_u^p &= \{\boldsymbol{\Phi}_{\boldsymbol{u}}^p \in \ell_2[0, N) : \Phi_{u,n}^p \in \mathcal{C}_{u,n}^p, \; n \in [0, N)\},
\end{aligned}$$

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 i-th largest singular value of a matrix.

These are reffered to as the set of approximately FIR for $N > 0$. The p-th player best-response map thus takes the form

$$\underset{\boldsymbol{\Phi}_{\boldsymbol{u}}^p=(\Phi_{u,n}^p)_{n=1}^{N-1}}{\text{minimize}} \quad \sum_{n=1}^{N-1} \Big(\|C^p\Phi_{x,n}\|_F^2 + \| \textstyle\sum_{\tilde{p}\in\mathcal{P}} D^{p\tilde{p}}\Phi_{u,n}^{\tilde{p}}\|_F^2 \Big) + \|C^p\Phi_{x,N}\|_F^2 \tag{33a}$$

$$\underset{\forall n\in[1,N)}{\text{subject to}} \quad \Phi_{x,n+1} = A\Phi_{x,n} + \textstyle\sum_{\tilde{p}\in\mathcal{P}} B^{\tilde{p}}\Phi_{u,n}^{\tilde{p}}, \tag{33b}$$

$$(\Phi_x * w)_n \in \mathcal{X}, \quad (\Phi_u^p * w)_n \in \mathcal{U}^p, \quad \big((\Phi_u^p * w)_n, (\Phi_u^{-p} * w)_n\big) \in \mathcal{U}_{\mathcal{G}}, \tag{33c}$$

$$\Phi_{x,n} \in \mathcal{C}_{x,n}, \quad \Phi_{u,n}^p \in \mathcal{C}_{u,n}^p, \tag{33d}$$

$$\Phi_{x,1} = I_{N_x}, \quad \|\Phi_{x,N}\|_F^2 \leq \gamma. \tag{33e}$$

The solutions to Problem (33) approximate those of the infinite-horizon Problem (31): With respect to N, the performance of the former converges to that achieved by the latter [36]. In this case, feasibility only requires (A, B^p) stabilisable and a sufficiently large horizon N to ensure that $\|\Phi_{x,N}\|_F^2 \leq \gamma$ is achievable for some $\boldsymbol{\Phi}_{\boldsymbol{u}}^p \in U_\Phi^p(\boldsymbol{\Phi}_{\boldsymbol{u}}^{-p})$. Computationally, this is still a finite-dimensional convex problem which can be solved numerically. Here, we let $\widehat{BR}_\Phi^p : \boldsymbol{\mathcal{C}}_u^{-p} \rightrightarrows \boldsymbol{\mathcal{C}}_u^p$ be the solutions of Problem (33) parametrised by $\boldsymbol{\Phi}_{\boldsymbol{u}}^{-p}$. The map $\widehat{BR}_\Phi : \boldsymbol{\mathcal{C}}_u \rightrightarrows \boldsymbol{\mathcal{C}}_u$, $\widehat{BR}_\Phi(\boldsymbol{\Phi}_{\boldsymbol{u}}) = \widehat{BR}_\Phi^1(\boldsymbol{\Phi}_{\boldsymbol{u}}^{-1}) \times \cdots \times \widehat{BR}_\Phi^{N_P}(\boldsymbol{\Phi}_{\boldsymbol{u}}^{-N_P})$, is the joint *(approximately)best-response* to the system-level profile $\boldsymbol{\Phi}_{\boldsymbol{u}}$.

Conversely to BR_Φ, the fixed-points of $\widehat{BR}_\Phi$ do not coincide with the set of GNEs $\Omega_{\mathcal{G}_\infty^\Phi}$, but are rather contained in the set of ε-GNEs $\Omega_{\mathcal{G}_\infty^\Phi}^\varepsilon$ for some equilibrium gap $\epsilon > 0$. This is clear from the fact that the original Problem (31) and the approximation Problem (33) have different optimal values. Under certain conditions, this fact can be shown explicitly.

Theorem 5. *Consider a fixed-point $\boldsymbol{\Phi}_{\boldsymbol{u}}^\varepsilon \in \widehat{BR}_\Phi(\boldsymbol{\Phi}_{\boldsymbol{u}}^\varepsilon)$ and assume that $\|\Phi_{x,N}^\star\|_F^2 \leq \gamma$ for $\boldsymbol{\Phi}_{\boldsymbol{x}}^\star = \boldsymbol{F}_{\boldsymbol{\Phi}}\boldsymbol{\Phi}_{\boldsymbol{u}}^\star$ obtained from the original best-response $\boldsymbol{\Phi}_{\boldsymbol{u}}^\star \in BR_\Phi(\boldsymbol{\Phi}_{\boldsymbol{u}}^\varepsilon)$. Then, the profile $\boldsymbol{\Phi}_{\boldsymbol{u}}^\varepsilon = (\boldsymbol{\Phi}_{\boldsymbol{u}}^{1^\varepsilon}, \dots, \boldsymbol{\Phi}_{\boldsymbol{u}}^{N_P^\varepsilon})$ is an ε-GNE of $\mathcal{G}_\infty^\Phi$ satisfying*

$$J^p(\boldsymbol{\Phi}_{\boldsymbol{u}}^{p^\varepsilon}, \boldsymbol{\Phi}_{\boldsymbol{u}}^{-p^\varepsilon}) \leq \min_{\boldsymbol{\Phi}_{\boldsymbol{u}}^p \in U_\Phi^p(\boldsymbol{\Phi}_{\boldsymbol{u}}^{-p^\epsilon})} J^p(\boldsymbol{\Phi}_{\boldsymbol{u}}^p, \boldsymbol{\Phi}_{\boldsymbol{u}}^{-p^\varepsilon}) + \varepsilon \tag{34}$$

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

The equilibrium gap associated with $\boldsymbol{\Phi}_{\boldsymbol{u}}^\varepsilon \in \widehat{BR}_\Phi(\boldsymbol{\Phi}_{\boldsymbol{u}}^\varepsilon)$ is thus proportional to the parameter γ, assuming that the terminal constraint Eq. (33e) also holds for solutions to the original Problem (31). Hereafter, $U_\Phi^p : \boldsymbol{\mathcal{C}}_u^{-p} \rightrightarrows \boldsymbol{\mathcal{C}}_u^p$ $(\forall p)$ are assumed to incorporate the pair ($\boldsymbol{\mathcal{C}}_x$, $\boldsymbol{\mathcal{C}}_u^p$) defining the set of (soft-constrained) FIR approximations defined above for a $N > 0$. We write $U_{\Phi,N}^p$ whenever this needs to be made explicit (e.g., to distinguish this choice from a more general $\boldsymbol{\mathcal{C}}_x$ and $\boldsymbol{\mathcal{C}}_u^p$).

Robust operational constraints: From Assumption 3, the sets $\mathcal{X}$, $\mathcal{U}^p$, and $\mathcal{U}_{\mathcal{G}}$, can be expressed by linear inequalities,

$$\mathcal{X} = \{x_t \in \mathbb{R}^{N_x} : G_x x_t \leq g_x\};$$
$$\mathcal{U}^p = \{u_t^p \in \mathbb{R}^{N_u^p} : G_u^p u_t^p \leq g_u^p\};$$
$$\mathcal{U}_{\mathcal{G}} = \{u_t \in \textstyle\prod_{\tilde{p}\in\mathcal{P}} \mathbb{R}^{N_u^{\tilde{p}}} : \sum_{\tilde{p}\in\mathcal{P}} G_{\mathcal{G}}^{\tilde{p}} u_t^{\tilde{p}} \leq g_{\mathcal{G}}\},$$

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 $G_{\mathcal{G}}^{\tilde{p}} \in \mathbb{R}^{N_{\mathcal{U}_{\mathcal{G}}} \times N_u^{\tilde{p}}}$ $(\forall \tilde{p} \in \mathcal{P})$, and vectors $g_x \in \mathbb{R}_{\geq 0}^{N_{\mathcal{X}}}$, $g_u^p \in \mathbb{R}_{\geq 0}^{N_{\mathcal{U}}^p}$, and $g_{\mathcal{G}} \in \mathbb{R}_{\geq 0}^{N_{\mathcal{U}_{\mathcal{G}}}}$. 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

$$[G_x]_i(\Phi_x * w)_n \leq [g_x]_i, \quad i = 1, \dots, N_{\mathcal{X}}; \tag{35}$$
$$[G_u^p]_j(\Phi_u^p * w)_n \leq [g_u^p]_j, \quad j = 1, \dots, N_{\mathcal{U}^p}; \tag{36}$$
$$\textstyle\sum_{\tilde{p}\in\mathcal{P}} [G_{\mathcal{G}}^{\tilde{p}}]_l(\Phi_u^{\tilde{p}} * w)_n \leq [g_{\mathcal{G}}]_l, \quad l = 1, \dots, N_{\mathcal{U}_{\mathcal{G}}}, \tag{37}$$

with $([G_x]_i, [g_x]_i)$, $([G_u^p]_j, [g_u^p]_j)$, and $([G_{\mathcal{G}}^{\tilde{p}}]_l, [g_{\mathcal{G}}]_l)$ denoting, respectively, the i-th, j-th, and l-th rows of the corresponding matrices and vectors. In this work, players are assumed to synthesise policies that satisfy these constraints for any $\boldsymbol{w} \in \boldsymbol{\mathcal{W}}$. At the expense of conservativeness, we cast Problem (33) as a robust optimization problem by considering the worst-case realisation of the noise. Specifically, we reformulate Eq. (35),

$$\sum_{n'=0}^{N} \sup_{w\in\mathcal{W}} [G_x]_i \Phi_{x,n'} w \leq [g_x]_i, \quad i = 1, \dots, N_{\mathcal{X}} \tag{38}$$

and similarly for Eqs. (36)–(37), then exploit our knowledge of $\boldsymbol{\mathcal{W}}$ to obtain an analytical solution for the supremum.

A common instance of $\mathcal{G}_\infty^{\mathrm{LQ}}$ considers the problem in which $(w_n)_{n\in\mathbb{N}}$ is uniformly bounded. In these problems, the inequality Eq. (38) can be expressed in terms of a dual norm on the appropriate vector space. We highlight two important cases.

- $\mathcal{W}$ is an ellipsoid centred at zero. Formally, the set is

$$\mathcal{W} = \{w_t \in \mathbb{R}^{N_x} : \|Pw_t\|_2 \leq 1\},$$

given a matrix $P \in \mathbb{S}_{++}^{N_x}$. This corresponds to noise processes with uniformly bounded energy, as encoded by $\|\cdot\|_2$. For such cases, Eqs. (35)–(37) can be enforced by the second-order conic (SOC) constraints

$$\textstyle\sum_{n=0}^{N} \big\|([G_x]_i \Phi_{x,n} P^{-1})^\mathsf{T}\big\|_2 \leq [g_x]_i, \quad (\forall i);$$
$$\textstyle\sum_{n=0}^{N-1} \big\|([G_u^p]_j \Phi_{u,n}^p P^{-1})^\mathsf{T}\big\|_2 \leq [g_u^p]_j, \quad (\forall j);$$
$$\textstyle\sum_{n=0}^{N-1} \big\| \sum_{\tilde{p}\in\mathcal{P}} ([G_{\mathcal{G}}^{\tilde{p}}]_l \Phi_{u,n}^{\tilde{p}} P^{-1})^\mathsf{T}\big\|_2 \leq [g_{\mathcal{G}}]_l, \quad (\forall l).$$

- $\mathcal{W}$ is a polyhedron, symmetric around zero. Formally,

$$\mathcal{W} = \{w_t \in \mathbb{R}^{N_x} : \|Pw_t\|_\infty \leq 1\},$$

given a full-rank matrix $P \in \mathbb{R}^{N_{\mathcal{W}} \times N_x}$. This corresponds to noise processes with uniformly bounded intensity, as encoded by $\|\cdot\|_\infty$. For such cases, Eqs. (35)–(37) can be enforced by the first-order conic constraints

$$\textstyle\sum_{n=0}^{N} \big\|([G_x]_i \Phi_{x,n} P^\dagger)^\mathsf{T}\big\|_1 \leq [g_x]_i, \quad (\forall i);$$
$$\textstyle\sum_{n=0}^{N-1} \big\|([G_u^p]_j \Phi_{u,n}^p P^\dagger)^\mathsf{T}\big\|_1 \leq [g_u^p]_j, \quad (\forall j);$$
$$\textstyle\sum_{n=0}^{N-1} \big\| \sum_{\tilde{p}\in\mathcal{P}} ([G_{\mathcal{G}}^{\tilde{p}}]_l \Phi_{u,n}^{\tilde{p}} P^\dagger)^\mathsf{T}\big\|_1 \leq [g_{\mathcal{G}}]_l, \quad (\forall l).$$

Remark 2. *If $\mathcal{X} = \mathbb{R}^{N_x}$, $\mathcal{U}^p = \mathbb{R}^{N_u^p}$, and $\mathcal{U}_{\mathcal{G}} = \prod_{p\in\mathcal{P}} \mathbb{R}^{N_u^p}$, these constraints are trivially satisfied for any $w \in \boldsymbol{\mathcal{W}}$, and thus can be removed from Problem (31). 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 possible $\boldsymbol{w}$.*

Structural constraints: The sets $(\mathcal{C}_{x,n}, \mathcal{C}^p_{u,n})_{n\in\mathbb{N}^+}$ $(\forall p \in \mathcal{P})$ are designed to impose some structure directly on the policy $\boldsymbol{K}^p$ (Corollary 4.1), often in the form of sparsity constraints. An important class of such structural constraints encodes information patterns incurred by the presence of actuation and sensing delays: Let $\boldsymbol{K}^p \in \boldsymbol{\mathcal{C}}^p$ $(\forall p)$ satisfy the restrictions

$$\boldsymbol{\mathcal{C}}^p = \{\boldsymbol{K}^p \in \mathcal{L}(\boldsymbol{\mathcal{X}}, \boldsymbol{\mathcal{U}}^p) : \text{ Actions } [B^p(K^p x)_t]_i \text{ respectively affect and feedback the component } [x_t]_i \text{ with the delays } d_a, d_s > 0\}, \tag{39}$$

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

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

with $\mathtt{Sp}(\cdot)$ denoting the sparsity pattern of a matrix[1]. It can be shown that $\boldsymbol{K}^p = \boldsymbol{\Phi}^p_{\boldsymbol{u}} \boldsymbol{\Phi}_{\boldsymbol{x}}^{-1} \in \boldsymbol{\mathcal{C}}^p$ $(\forall p)$ if its parametrisation satisfy $(\Phi_{x,n} \in \mathcal{C}_{x,n})_{n\in\mathbb{N}^+}$ and $(\Phi^p_{u,n} \in \mathcal{C}^p_{u,n})_{n\in\mathbb{N}^+}$ with

$$\mathcal{C}_{x,n} = \{\Phi_{x,n} \in \mathbb{R}^{N_x \times N_x} : \Phi_{x,n} = S_{x,n} \odot \Phi_{x,n}\}; \tag{40a}$$
$$\mathcal{C}^p_{u,n} = \{\Phi^p_{u,n} \in \mathbb{R}^{N^p_u \times N_x} : \Phi^p_{u,n} = S^p_{u,n} \odot \Phi^p_{u,n}\}. \tag{40b}$$

The constraints Eq. (40) enforce that the closed-loop response to the noise obeys an information pattern induced by the dynamics of the game $\mathcal{G}^{\mathrm{LQ}}_\infty$. Specifically, $[S_{x,n}]_{i,j} = 0$ (respectively, $[S^p_{u,n}]_{i,j} = 0$) indicates that any disturbance to the j-th component of the state, $[x_t]_j$, does not affect $[x_{t+n}]_i$ (does not feedback into action $[u^p_{t+n}]_i$). Finally, note that $(\mathcal{C}_{x,n}, \mathcal{C}^p_{u,n})_{n\in\mathbb{N}^+}$ are compact convex sets, as they correspond to the kernels of the linear operators $(I - \boldsymbol{S}_x)$ and $(I - \boldsymbol{S}^p_u)$.

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 Nash equilibrium problems [40, 42]. For $(\mathcal{C}_{x,n}, \mathcal{C}^p_{u,n})_{n\in\mathbb{N}^+}$ from Eq. (40), conditions for the existence of GFNE (i.e., $\Omega^{\boldsymbol{K}}_{\mathcal{G}^{\mathrm{LQ}}_\infty} \neq \emptyset$), can be stated in terms of the delay parameters d_a and d_s. Hereafter, the best-responses $\{BR^p_\Phi\}_{p\in\mathcal{P}}$ (and $\{\widehat{BR}^p_\Phi\}_{p\in\mathcal{P}}$) are assumed to include these system-level constraints with $d_a = d_s = 1$.

B. System-level best-response dynamics

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

Theorem 6. *A policy profile $\boldsymbol{K}^* = (\boldsymbol{\Phi}^{1^*}_{\boldsymbol{u}}, \dots, \boldsymbol{\Phi}^{N^*_P}_{\boldsymbol{u}}) \boldsymbol{\Phi}^{*^{-1}}_{\boldsymbol{x}} \in \boldsymbol{\mathcal{C}}$, is a GFNE of $\mathcal{G}^{LQ}_\infty$ if $\boldsymbol{\Phi}^*_{\boldsymbol{u}} \in BR_\Phi(\boldsymbol{\Phi}^*_{\boldsymbol{u}})$, or, equivalently,*

$$\boldsymbol{\Phi}^{p^*}_{\boldsymbol{u}} \in BR^p_\Phi(\boldsymbol{\Phi}^{-p^*}_{\boldsymbol{u}}), \quad \forall p \in \mathcal{P}. \tag{41}$$

Proof. Consider an arbitrary fixed-point $\boldsymbol{\Phi}^*_{\boldsymbol{u}} \in BR^p_\Phi(\boldsymbol{\Phi}^*_{\boldsymbol{u}})$. From Theorem 4, we have $\boldsymbol{\Phi}^*_{\boldsymbol{x}} = \boldsymbol{F}_{\boldsymbol{\Phi}} \boldsymbol{\Phi}^*_{\boldsymbol{u}}$. Now, consider policies $\boldsymbol{K}^{p^*} = \boldsymbol{\Phi}^{p^*}_{\boldsymbol{u}} (\boldsymbol{\Phi}^*_{\boldsymbol{x}})^{-1}$, $p \in \mathcal{P}$. Clearly, $\boldsymbol{\Phi}^{p^*}_{\boldsymbol{u}} = \boldsymbol{K}^{p^*} \boldsymbol{\Phi}^*_{\boldsymbol{x}}$. As a consequence, for any $\boldsymbol{w} \in \boldsymbol{\mathcal{W}}$,

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

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

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

- The pair $(\boldsymbol{\Phi}^p_{\boldsymbol{x,k}}, \boldsymbol{\Phi}^p_{\boldsymbol{u,k}})$ defines the p-th player's parametrisation after $k \in \mathbb{N}$ updates, that is, $\boldsymbol{\Phi}^p_{\boldsymbol{x,k}} = (\Phi^p_{x,k,t})_{t\in\mathbb{N}}$ and $\boldsymbol{\Phi}^p_{\boldsymbol{u,k}} = (\Phi^p_{u,k,t})_{t\in\mathbb{N}}$. The update index $k \in \mathbb{N}$ should not be mistaken for the stage index $t \in \mathbb{N}$.
- Updating the policy $\boldsymbol{K}^p_{k+1}$ consists of either employing Corollary 4.1 with maps $(\boldsymbol{\Phi}^p_{\boldsymbol{x,k+1}}, \boldsymbol{\Phi}^p_{\boldsymbol{u,k+1}})$, or directly computing its kernel $\boldsymbol{\Phi}^p_{\boldsymbol{k+1}} = ([M_{\Phi^p_{u,k+1}}]_n \Phi^{-1}_{x,k+1})_{n\in\mathbb{N}}$.
- Responses $\{\boldsymbol{\Phi}^p_{\boldsymbol{x,k}}\}_{p\in\mathcal{P}}$ are most likely distinct at $k < \infty$, that is, $\boldsymbol{\Phi}^p_{\boldsymbol{x,k}} \neq \boldsymbol{\Phi}^{\tilde{p}}_{\boldsymbol{x,k}}$ for $p \neq \tilde{p}$. Consequently, the system level parametrisation (Eq. 24) might not hold for the profile $\boldsymbol{K}_k = \left(\boldsymbol{\Phi}^1_{\boldsymbol{u,k}} (\boldsymbol{\Phi}^1_{\boldsymbol{x,k}})^{-1}, \dots, \boldsymbol{\Phi}^{N_P}_{\boldsymbol{u,k}} (\boldsymbol{\Phi}^{N_P}_{\boldsymbol{x,k}})^{-1}\right)$ and it could lead to $\mathcal{G}^{\mathrm{LQ}}_\infty$ becoming unstable. However, this policy still satisfies a robust version of Theorem 4 when $\|\boldsymbol{\Phi}^p_{\boldsymbol{x,k}} - \boldsymbol{\Phi}^{\tilde{p}}_{\boldsymbol{x,k}}\|$ is sufficiently small for all $p, \tilde{p} \in \mathcal{P}$, [36].

The SLS-BRD induces an operator $T_\Phi = (1-\eta)I + \eta BR_\Phi$, which defines the global update $\boldsymbol{\Phi}_{\boldsymbol{u,k+1}}$ from $\boldsymbol{\Phi}_{\boldsymbol{u,k}}$. As before, T_Φ and BR_Φ share the same fixed-points: The GNEs $\Omega_{\mathcal{G}^\Phi_\infty}$. From Theorem 6, convergence to a response $\boldsymbol{\Phi}^*_{\boldsymbol{u}} \in T_\Phi(\boldsymbol{\Phi}^*_{\boldsymbol{u}})$ then implies convergence to a policy $\boldsymbol{K}^* \in \Omega^{\boldsymbol{K}}_{\mathcal{G}_\infty}$. Hence, this learning dynamics is a formal procedure for GFNE seeking.

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

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

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

[1]The operator $\mathtt{Sp}(A)$ produces a binary matrix such that $[\mathtt{Sp}(A)]_{i,j} = 1$ if $[A]_{i,j} \neq 0$ and $[\mathtt{Sp}(A)]_{i,j} = 0$ otherwise.

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

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

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

Convergence of the SLS-BRD: The convergence of Algorithm 4 depends on BR_Φ (or $\widehat{BR}_\Phi$, for its tractable version) being at least non-expansive (Theorems 2–3). Formally,

Theorem 7. *Let the map $\widehat{BR}_\Phi : \mathcal{C}_u \rightrightarrows \mathcal{C}_u$ be $L_{\widehat{BR}_\Phi}$-Lipschitz, with $L_{\widehat{BR}_\Phi} < 1$. 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 $\boldsymbol{\Phi}^*_{\boldsymbol{u}} \in \Omega^\varepsilon_{\mathcal{G}^\Phi_\infty}$ with rate*

$$\frac{\|\boldsymbol{\Phi}_{\boldsymbol{u},\boldsymbol{k}} - \boldsymbol{\Phi}^*_{\boldsymbol{u}}\|_{\ell_2}}{\|\boldsymbol{\Phi}_{\boldsymbol{u},\boldsymbol{0}} - \boldsymbol{\Phi}^*_{\boldsymbol{u}}\|_{\ell_2}} \leq \big((1-\eta) - \eta L_{\widehat{BR}_\Phi}\big)^k \tag{42}$$

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

$$\boldsymbol{H}^{p\tilde{p}} = (\boldsymbol{C}^p\boldsymbol{F}^p_{\boldsymbol{\Phi}} + \boldsymbol{D}^{pp})^\mathsf{T}(\boldsymbol{C}^p\boldsymbol{F}^{\tilde{p}}_{\boldsymbol{\Phi}} + \boldsymbol{D}^{p\tilde{p}}), \tag{43}$$

and $\boldsymbol{H}^{p,-p} = (\boldsymbol{H}^{p,\tilde{p}})_{\tilde{p}\in\mathcal{P}}$. Because $\{\boldsymbol{\Phi}^p_u\}_{p\in\mathcal{P}}$ are FIR, all the elements defined above have equivalent matrix representations. Using this notation, Problem (33) can be reformulated as

$$\begin{aligned}\underset{\boldsymbol{\Phi}^p_{\boldsymbol{u}}\in U^p_\Phi(\boldsymbol{\Phi}^{-p}_{\boldsymbol{u}})}{\text{minimize}}\ \ &\mathrm{Tr}\big[{\boldsymbol{\Phi}^p_{\boldsymbol{u}}}^\mathsf{T}\boldsymbol{H}^{pp}\boldsymbol{\Phi}^p_{\boldsymbol{u}}\\ &\quad + 2(\textstyle\sum_{\tilde{p}\in\mathcal{P}\setminus\{p\}}\boldsymbol{H}^{p\tilde{p}}\boldsymbol{\Phi}^{\tilde{p}}_{\boldsymbol{u}} + \boldsymbol{H}^{p0})^\mathsf{T}\boldsymbol{\Phi}^p_{\boldsymbol{u}}\big]\end{aligned} \tag{44}$$

The maps $BR^p(\boldsymbol{\Phi}^{-p}_{\boldsymbol{u}})$, $p \in \mathcal{P}$, thus correspond to the solution of quadratic programs with convex constraints $U^p_\Phi(\boldsymbol{\Phi}^{-p}_{\boldsymbol{u}})$, for $\boldsymbol{\Phi}^{-p}_{\boldsymbol{u}} \in \mathcal{C}^{-p}_u$. In the case of $\mathcal{W} = \{w_t \in \mathbb{R}^{N_x} : \|w_t\|_\infty \leq 1\}$, and without structural constraints, that is, $\boldsymbol{S}_x = I$ and $\boldsymbol{S}^p_u = I$,

$$\begin{aligned}U^p_\Phi(\boldsymbol{\Phi}^{-p}_{\boldsymbol{u}}) = \{&\boldsymbol{\Phi}^p_{\boldsymbol{u}} \in \ell_2[0,N) :\\ &\Phi_{x,n+1} = A\Phi_{x,n} + \textstyle\sum_{\tilde{p}\in\mathcal{P}} B^{\tilde{p}}\Phi^{\tilde{p}}_{u,n},\ \Phi_{x,1} = I_{N_x},\\ &\textstyle\sum_{n=0}^{N}\big\|([G_x]_i\Phi_{x,n})^\mathsf{T}\big\|_1 \leq [g_x]_i,\\ &\textstyle\sum_{n=0}^{N-1}\big\|([G^p_u]_j\Phi^p_{u,n})^\mathsf{T}\big\|_1 \leq [g^p_u]_j,\\ &\textstyle\sum_{n=0}^{N-1}\big\|\sum_{\tilde{p}\in\mathcal{P}}([G^{\tilde{p}}_{\mathcal{G}}]_l\Phi^{\tilde{p}}_{u,n})^\mathsf{T}\big\|_1 \leq [g_{\mathcal{G}}]_l\}.\end{aligned} \tag{45}$$

Using standard techniques from optimisation, the constraints Eq. (45) can be incorporated into Problem (44) as linear inequalities. The solutions for these problems, that is, the best-responses $\{\widehat{BR}^p_\Phi\}_{p\in\mathcal{P}}$, are thus piecewise affine in $\boldsymbol{\Phi}^{-p}_{\boldsymbol{u}} \in \mathcal{C}^{-p}_u$ [43]. Consequently, also $\widehat{BR}_\Phi$ must be piecewise affine. A local Lipschitz constant can then be derived for each region of $\mathcal{C}^{-p}_u$ that leads to a subset of the operational constraints being active. For non-generalised games (i.e., when $\mathcal{X} = \mathbb{R}^{N_x}$ and $\mathcal{U}_\mathcal{G} = \prod_{p\in\mathcal{P}}\mathbb{R}^{N^p_u}$), the structure of this affine mapping can be exploited to derive a global Lipschitz constant for $\widehat{BR}_\Phi$.

Theorem 8. *Consider $\mathcal{X} = \mathbb{R}^{N_x}$ and $\mathcal{U}_\mathcal{G} = \prod_{p\in\mathcal{P}}\mathbb{R}^{N^p_u}$. Then, The map $\widehat{BR}_\Phi$ is $L_{\widehat{BR}_\Phi}$-Lipschitz with*

$$L_{\widehat{BR}_\Phi} = \sqrt{\textstyle\sum_{p\in\mathcal{P}}(L^p_{\widehat{BR}_\Phi})^2}, \tag{46}$$

given the player-specific constants

$$L^p_{\widehat{BR}_\Phi} = \big(1 + \kappa(\boldsymbol{H}^{pp})\big)\|(\boldsymbol{H}^{pp})^\dagger\boldsymbol{H}^{p,-p}\|_{2\to2}, \tag{47}$$

with condition number $\kappa(\boldsymbol{H}^{pp}) = \|(\boldsymbol{H}^{pp})^\dagger\|_{2\to2}\|\boldsymbol{H}^{pp}\|_{2\to2}$.

The proof of Theorem 8 is extensive: The reader is referred to Appendix C for the full details. Importantly, this Lipschitz constant is not tight and thus $L_{\widehat{BR}_\Phi} < 1$ using Eq. (46) is only a sufficient (but not necessary) condition for Theorem 7 to hold in practice. Regardless, it highlights some intuitive, but non-trivial, facts about the convergence of the SLS-BRD:

- The condition $L_{\widehat{BR}_\Phi} < 1$ is implies that 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 the class of Jacobi iterative methods, where diagonal dominance of the linear system being solved is a strict requirement.
- The convergence rate of the SLS-BRD is inversely proportional to $\|D^{pp}\|^2_2$ $(\forall p)$ through $\boldsymbol{H}^{pp}$. As such, faster convergence is expected for games in which players apply strong penalties to their own actions.
- The convergence rate of the SLS-BRD depends on the number of players: Let $L^p_{\widehat{BR}_\Phi} = \sqrt{\alpha + \beta^p}$ for all $p \in \mathcal{P}$ given some $\alpha, \beta^p > 0$. Then, $L^2_{\widehat{BR}_\Phi} = (\alpha N_p)^2\sum_p \beta^p$. In large-scale games, players might need to become more conservative in order to ensure that $\widehat{BR}$ is a contraction.
- The convergence rate of the SLS-BRD is dominated by the slowest player: Whenever there exists a $p \in \mathcal{P}$ such that $L^p_{\widehat{BR}_\Phi} \gg L^{\tilde{p}}_{\widehat{BR}_\Phi}$ for all $\tilde{p} \in \mathcal{P}$, then $L_{\widehat{BR}_\Phi} \approx L^p_{\widehat{BR}_\Phi}$.
- In unstable dynamic games, the norms $\|\boldsymbol{H}^{pp}\|_{2\to2} = \|\boldsymbol{C}^p\boldsymbol{F}^p_{\boldsymbol{\Phi}} + \boldsymbol{D}^{pp}\|^2_{2\to2}$ increase with the FIR horizon N. In some cases, players might need to proportionally decrease the state-penalties C^p to ensure (or improve) convergence.

At the cost of interpretability, similar Lipschitz constants as in Eqs. (46)–(47) can be obtained also when $(\mathcal{X}, \mathcal{U}_\mathcal{G}, \mathcal{U}^p)$ and P are general, and when structural constraints are present. Moreover, we remark that Theorem 8 is obtained under the assumption that $\|\Phi_{x,N}\|^2_F < \gamma$ holds strictly. The best-response maps $\{\widehat{BR}^p_\Phi\}_{p\in\mathcal{P}}$ when this constraint is active, or when $\mathcal{W}$ is an ellipsoid, become quadratically-constrained quadratic programs and their Lipschitz properties are less intuitive.

IV. Example: Bi-directional chain game

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

$$A = \begin{bmatrix} 1 & 0.2 & & \\ -0.2 & \ddots & \ddots & \\ & \ddots & \ddots & 0.2 \\ & & -0.2 & 1 \end{bmatrix}; \left\{ B^p = \begin{bmatrix} \mathbf{0}_{6(p-1)\times 2} \\ I_2 \\ \mathbf{0}_{6(3-p)\times 2} \end{bmatrix} \right\}_{p\in\mathcal{P}}, \tag{48}$$

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

$$J^p(u^p, u^{-p}) = \text{E}\left[\sum_{t=0}^{\infty}\left(\beta^p\|x_t\|_2^2 + 10\|u_t^p\|_2^2\right)\right],$$

equivalent to Eq. (21) after setting $C^p = [\sqrt{\beta^p} I_{N_x}\ \mathbf{0}_{N_x\times N_u}]^\mathsf{T}$, $D^{pp} = [\mathbf{0}_{N_u\times N_x}\ \sqrt{10} I_{Nu}]^\mathsf{T}$, and $D^{p\tilde{p}} = 0$ for $\tilde{p} \in \mathcal{P}\backslash\{p\}$. We consider $\beta^1 = \beta^3 = 0.1$ and $\beta^2 = 0.01$. The players' actions are subjected to operational constraints, $\boldsymbol{u}^p \in U^p(\boldsymbol{u}^{-p})$, with

$$\begin{aligned} \mathcal{X} &= \mathbb{R}^{N_x}; \\ \mathcal{U}^p &= \{u_t^p \in \mathbb{R}^{N_u^p} : [\mathbf{1}_{N_u^p}\ -\mathbf{1}_{N_u^p}]^\mathsf{T} u_t^p \le [10\ 10]^\mathsf{T}\}; \\ \mathcal{U}_\mathcal{G} &= \textstyle\prod_{p\in\mathcal{P}} \mathbb{R}^{N_u^p}, \end{aligned}$$

enforcing $[u_t]_1 + [u_t]_2 \in [-10, 10]$ for every $t \in \mathbb{N}$ and $p \in \mathcal{P}$. We consider that players design their state-feedback policies, $\boldsymbol{K}^p = \boldsymbol{\Phi}_{\boldsymbol{u}}^p \boldsymbol{\Phi}_{\boldsymbol{x}}^{-1} \in \boldsymbol{\mathcal{C}}^p$, considering a FIR horizon of $N = 32$ and the constraints $(\Phi_{x,n} \in \mathcal{C}_{x,n})_{n=1}^N$ and $(\Phi_{u,n}^p \in \mathcal{C}_{u,n})_{n=1}^N$,

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

In this experiment, players learn ε-GFNE policies through the SLS-BRD (Algorithm 4) considering their (approximately)best-response maps, $\{\widehat{BR}_\Phi^p\}_{p\in\mathcal{P}}$, with factors $\gamma \in \{2^{-1}, 2^{-3}, 2^{-5}\}$. The policies are updated every $\Delta T = 1$ stages with learning rate $\eta = 1/2$. The initial profile, $\boldsymbol{K}_0 = (\boldsymbol{\Phi}_{\boldsymbol{u},\boldsymbol{0}}^1, \dots, \boldsymbol{\Phi}_{\boldsymbol{u},\boldsymbol{0}}^{N_P})\boldsymbol{\Phi}_{\boldsymbol{x},\boldsymbol{0}}^{-1}$, is obtained by projecting $\hat{\boldsymbol{\Phi}}_{\boldsymbol{u},\boldsymbol{0}} = \mathbf{0}$ into the feasible set,

$$\begin{aligned} \boldsymbol{\Phi}_{\boldsymbol{u},\boldsymbol{0}} = \underset{\boldsymbol{\Phi}_{\boldsymbol{u}}\in\boldsymbol{\mathcal{C}}_u}{\arg\min}\quad & \|\boldsymbol{\Phi}_{\boldsymbol{u}} - \hat{\boldsymbol{\Phi}}_{\boldsymbol{u},\boldsymbol{0}}\|_{\ell_2} \\ \underset{\forall p\in\mathcal{P}}{\text{subject to}}\quad & \boldsymbol{\Phi}_{\boldsymbol{x}} = \boldsymbol{F}_{\boldsymbol{\Phi}}\boldsymbol{\Phi}_{\boldsymbol{u}},\quad \|\Phi_{x,N}\|_F^2, \le \gamma \\ & \boldsymbol{\Phi}_{\boldsymbol{u}}^p * \boldsymbol{w} \in \boldsymbol{\mathcal{U}}^p. \end{aligned}$$

We simulate an execution of this game alongside the SLS-BRD for each $\gamma \in \{2^{-1}, 2^{-3}, 2^{-5}\}$. Due to numerical limitations, we interrupt the updates and consider that a fixed-point has been reached when $\|\boldsymbol{\Phi}_{u,k}^p - \boldsymbol{\Phi}_{u,k-1}^p\|_{\ell_2}^2 / \|\boldsymbol{\Phi}_{u,k}^p\|_{\ell_2}^2 \le 10^{-16}$ is first satisfied for all $p \in \mathcal{P}$. Under this learning dynamics, the convergence to a fixed-point $\boldsymbol{K}^* = (\boldsymbol{\Phi}_{\boldsymbol{u}}^{1^*}, \dots, \boldsymbol{\Phi}_{\boldsymbol{u}}^{N_P^*})\boldsymbol{\Phi}_{\boldsymbol{x}}^{-1}$ is shown in Figure 3. The results demonstrate that a GFNE is reached after 400, 900, and 1900, iterations for each respective value of γ. The convergence rate is observed to be geometric, indicating that the best-response maps in this scenario are contractive. In each case, the constraint $\|\Phi_{x,N}\|_F^2 \le \gamma$ remains active during most iterations: For this problem, the mapping $\widehat{BR}_\Phi$ is still shown to be non-expansive. However, convergence is observed to slow down as this restriction is made stricter. This is not surprising, as iterative improvement methods are known to converge to exact Nash equilibria (here, the case when $\gamma \to 0$) after exponentially many steps [44].

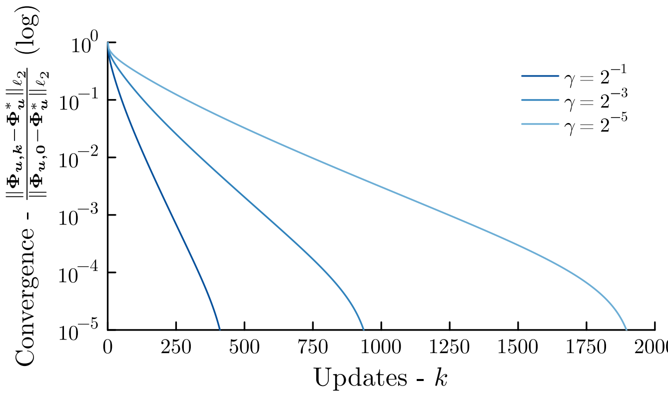

Fig. 3. N_P-Chain Game: Convergence of the SLS-BRD procedure.

The individual updates $\{\boldsymbol{\Phi}_{\boldsymbol{u},\boldsymbol{k}}^p\}_{p\in\mathcal{P}}$ from each player converge to the threshold accuracy as shown in Figure 4. In general, these local changes become numerically negligible at faster rate than the global convergence presented in Figure 3. Moreover, the convergence is observed to be similar for all players $p \in \mathcal{P}$, in each scenario, except for player $p = 2$ which shows a slightly slower convergence. This is expected, as players $p \in \{1, 3\}$ affect the dynamics of $\mathcal{G}_\infty^{\text{LQ}}$ in a symmetric fashion, whereas player $p = 2$ acts directly into the centre of the network. Finally, we observe the (normalised) pairwise distance between the responses $\{\boldsymbol{\Phi}_{\boldsymbol{x},\boldsymbol{k}}^p\}_{p\in\mathcal{P}}$, $\|\boldsymbol{\Phi}_{\boldsymbol{x},\boldsymbol{k}}^p - \boldsymbol{\Phi}_{\boldsymbol{x},\boldsymbol{k}}^{\tilde{p}}\|_{\ell_2} / \|\boldsymbol{\Phi}_{\boldsymbol{x},\boldsymbol{k}}^p\|_{\ell_2}$ for $k \in \mathbb{N}_+$, to decrease at a similar rate, Figure 4. These distances are shown to be relatively small since the initial stages of the game. The policies $\boldsymbol{K}_k = \left(\boldsymbol{\Phi}_{\boldsymbol{u},\boldsymbol{k}}^1(\boldsymbol{\Phi}_{\boldsymbol{x},\boldsymbol{k}}^1)^{-1}, \dots, \boldsymbol{\Phi}_{\boldsymbol{u},\boldsymbol{k}}^{N_P}(\boldsymbol{\Phi}_{\boldsymbol{x},\boldsymbol{k}}^{N_P})^{-1}\right)$ are thus expected to be stabilising during this learning dynamics.

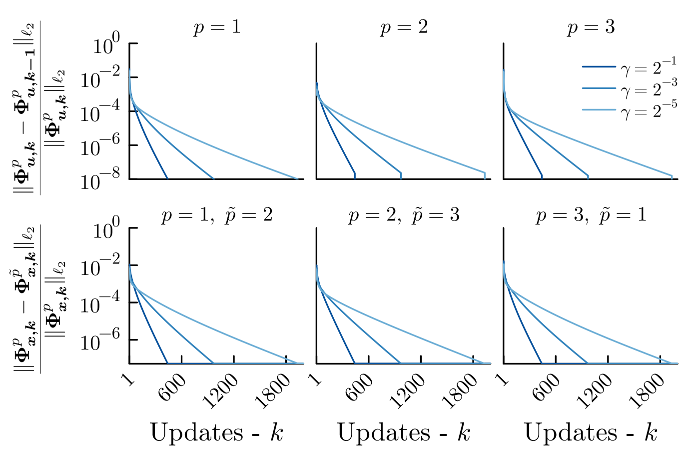

Fig. 4. N_P-Chain Game, log-scale: Normalised differences between updates $(\boldsymbol{\Phi}_{\boldsymbol{u},\boldsymbol{k}}^p, \boldsymbol{\Phi}_{\boldsymbol{u},\boldsymbol{k+1}}^p)$, top panels, and pairwise distances between responses $(\boldsymbol{\Phi}_{\boldsymbol{x},\boldsymbol{k}}^p, \boldsymbol{\Phi}_{\boldsymbol{x},\boldsymbol{k}}^{\tilde{p}})$, bottom panels, for all players $p, \tilde{p} \in \mathcal{P}$.

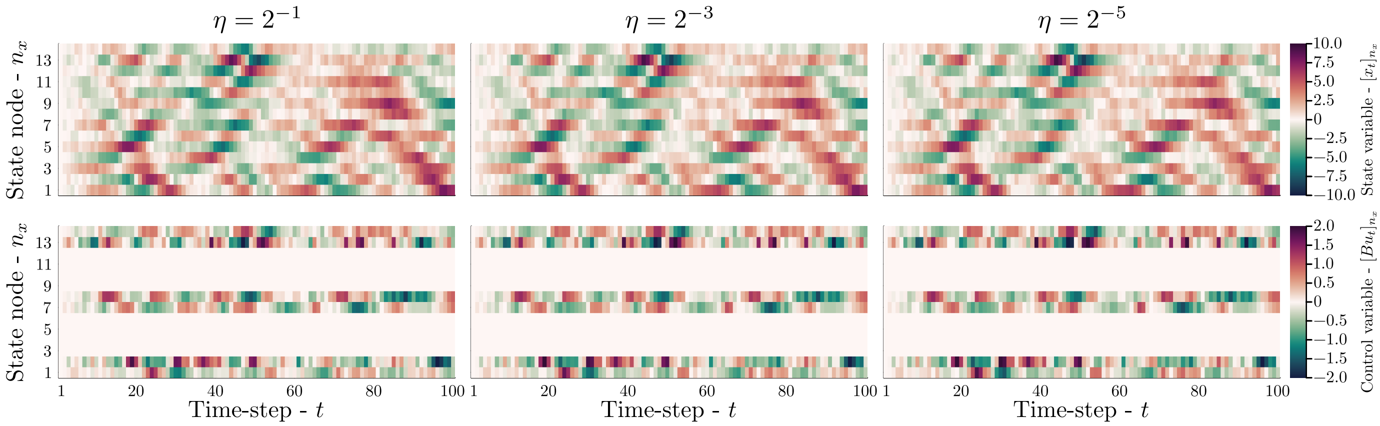

Fig. 5. N_P-Chain Game, $t \in (0, 100]$: Simulated state $\boldsymbol{x}$ (top panels) and control $\boldsymbol{Bu}$ (bottom panels) trajectories for each SLS-BRD execution with $\gamma \in \{2^{-1}, 2^{-3}, 2^{-5}\}$. The vertical axis represents each node in the chain-networked system.

The evolution of the game given the actions by each player is displayed in Figure 5 for its early stages. In all cases, the policies obtained through the SLS-BRD procedure are capable of jointly stabilising the state of the game, robustly against the random noise. Moreover, these policies are also shown to satisfy the operational constraints: Players' actions become (roughly) symmetrical to ensure that $[u_t]_1 + [u_t]_2 \approx 0$ for most stages. The enforcement of this strategy highlights the conservativeness resulting from the robust operational constraints. Noticeably, the simulated state- and control trajectories are very similar for each scenario. As such, a policy learning procedure considering $\gamma \to 1$ seems to be preferable in this experiment: It converges faster while achieving similar performance in practice.

V. Concluding remarks

This work presents the SLS-BRD, an algorithm for generalised feedback Nash equilibrium seeking in N_P-players non-cooperative games. The method is based on the class of best-response dynamics algorithms for decentralised learning and consists of players updating and announcing a parametrisation of their policies until converging to an equilibrium. Because not updating control actions explicitly, this learning dynamics can be performed simultaneously with the game's regular execution. Our approach leverages the System Level Synthesis framework to formulate each player's best-response map as the solution to robust finite-horizon optimal control problems. Using results from operator theory, we propose that convergence to a Nash equilibrium is geometric if certain conditions are met: Namely, that the best-response maps are contractive operators. The SLS-BRD benefits from this framework also by allowing richer information patterns to be enforced directly at the synthesis level. After the main theoretical and practical aspects are discussed, the algorithm is demonstrated on an exemplary problem from the decentralised control of multi-agent systems.

References

[1] T. Başar and G. J. Olsder, *Dynamic noncooperative game theory*. SIAM, 1998.

[2] F. Facchinei and C. Kanzow, "Generalized Nash equilibrium problems," *Annals of 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," *Communications of the ACM*, vol. 52, no. 2, pp. 89–97, 2009.

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

[5] J. S. Shamma and G. Arslan, "Dynamic fictitious play, dynamic gradient play, and distributed convergence to Nash equilibria," *IEEE Transactions on Automatic Control*, vol. 50, no. 3, pp. 312–327, 2005.

[6] J. Hofbauer and S. Sorin, "Best response dynamics for continuous zero-sum games," *Discrete and Continuous Dynamical Systems*, vol. 6, no. 1, pp. 215–224, 2006.

[7] J. R. Marden, G. Arslan, and J. S. Shamma, "Joint strategy fictitious play with inertia for potential games," *IEEE Transactions on Automatic Control*, vol. 54, no. 2, pp. 208–220, 2009.

[8] A. Cortés and S. Martínez, "Self-triggered best-response dynamics for continuous games," *IEEE Transactions on Automatic Control*, vol. 60, no. 4, pp. 1115–1120, 2014.

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

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

[11] A. Jafari, A. Greenwald, D. Gondek, and G. Ercal, "On no-regret learning, fictitious play, and nash equilibrium," in *ICML*, vol. 1, pp. 226–233, 2001.

[12] N. Cesa-Bianchi and G. Lugosi, *Prediction, learning, and games*. Cambridge university press, 2006.

[13] G. J. Gordon, A. Greenwald, and C. Marks, "No-regret learning in convex games," in *Proceedings of the 25th international conference on Machine learning*, pp. 360–367, 2008.

[14] J. Hartline, V. Syrgkanis, and E. Tardos, "No-regret learning in bayesian games," *Advances in Neural Information Processing Systems*, vol. 28, 2015.

[15] A. Celli, A. Marchesi, G. Farina, and N. Gatti, "No-regret learning dynamics for extensive-form correlated equilibrium," *Advances in Neural Information Processing Systems*, vol. 33, pp. 7722–7732, 2020.

[16] G. Belgioioso, D. Liao-McPherson, M. H. de Badyn, S. Bolognani, R. S. Smith, J. Lygeros, and F. Dörfler, "Online feedback equilibrium seeking," *arXiv preprint arXiv:2210.12088*, 2022.

[17] G. Scutari, S. Barbarossa, and D. Palomar, "Potential games: A framework for vector power control problems with coupled constraints," in *2006 IEEE International Conference on Acoustics Speech and Signal Processing Proceedings*, vol. 4, pp. IV–IV, 2006.

[18] 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 processing magazine*, vol. 29, no. 5, pp. 86–105, 2012.

[19] S. Maghsudi and S. Stańczak, "Channel selection for network-assisted d2d communication via no-regret bandit learning with calibrated forecasting," *IEEE Transactions on Wireless Communications*, vol. 14, no. 3, pp. 1309–1322, 2015.

[20] J. F. Fisac, E. Bronstein, E. Stefansson, D. Sadigh, S. S. Sastry, and A. D. Dragan, "Hierarchical game-theoretic planning for autonomous vehicles," in *2019 International conference on robotics and automation (ICRA)*, pp. 9590–9596, IEEE, 2019.

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

[22] Y. Chen and J. W. Vaughan, "A new understanding of prediction markets via no-regret learning," in *Proceedings of the 11th ACM conference on Electronic commerce*, pp. 189–198, 2010.

[23] M. Braverman, J. Mao, J. Schneider, and M. Weinberg, "Selling to a no-regret buyer," in *Proceedings of the 2018 ACM Conference on Economics and Computation*, pp. 523–538, 2018.

[24] D. Aussel, L. Brotcorne, S. Lepaul, and L. von Niederhäusern, "A trilevel model for best response in energy demand-side management," *European Journal of Operational Research*, vol. 281, no. 2, pp. 299–315, 2020.

[25] J. Engwerda, W. Van Den Broek, and J. M. Schumacher, "Feedback Nash equilibria in uncertain infinite time horizon differential games," in *Proceedings of the 14th International Symposium of Mathematical Theory of Networks and Systems, MTNS 2000*, pp. 1–6, Unknown Publisher, 2000.

[26] D. Vrabie and F. Lewis, "Integral reinforcement learning for online computation of feedback Nash strategies of nonzero-sum differential games," in *49th IEEE Conference on Decision and Control (CDC)*, pp. 3066–3071, IEEE, 2010.

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

[28] 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.

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

[30] 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 journal of Automatica Sinica*, vol. 1, no. 3, pp. 239–247, 2014.

[31] 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 *2020 IEEE international conference on robotics and automation (ICRA)*, pp. 1475–1481, IEEE, 2020.

[32] D. Fridovich-Keil, V. Rubies-Royo, and C. J. Tomlin, "An iterative quadratic method for general-sum differential games with feedback linearizable dynamics," in *2020 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 2216–2222, IEEE, 2020.

[33] A. Tanwani and Q. Zhu, "Feedback Nash equilibrium for randomly switching differential–algebraic games," *IEEE Transactions on Automatic Control*, vol. 65, no. 8, pp. 3286–3301, 2019.

[34] Z. Wang, R. Spica, and M. Schwager, "Game theoretic motion planning for multi-robot racing," in *Distributed Autonomous Robotic Systems: The 14th International Symposium*, pp. 225–238, Springer, 2019.

[35] F. Laine, D. Fridovich-Keil, C.-Y. Chiu, and C. Tomlin, "The computation of approximate generalized feedback nash equilibria," *SIAM Journal on Optimization*, vol. 33, no. 1, pp. 294–318, 2023.

[36] J. Anderson, J. C. Doyle, S. H. Low, and N. Matni, "System level synthesis," *Annual Reviews in Control*, vol. 47, pp. 364–393, 2019.

[37] I. L. Glicksberg, "A further generalization of the kakutani fixed theorem, with application to Nash equilibrium points," *Proceedings of the American Mathematical Society*, vol. 3, no. 1, pp. 170–174, 1952.

[38] E. K. Ryu and S. Boyd, "Primer on monotone operator methods," *Appl. comput. math*, vol. 15, no. 1, pp. 3–43, 2016.

[39] E. K. Ryu and W. Yin, *Large-scale convex optimization: algorithms & analyses via monotone operators*. Cambridge University Press, 2022.

[40] T. Başar, "Stochastic differential games and intricacy of information structures," in *Dynamic Games in Economics*, pp. 23–49, Springer, 2014.

[41] K. Zhou and J. C. Doyle, *Essentials of robust control*, vol. 104. Prentice hall Upper Saddle River, NJ, 1998.

[42] A. Nayyar and T. Başar, "Dynamic stochastic games with asymmetric information," in *2012 IEEE 51st IEEE conference on decision and control (CDC)*, pp. 7145–7150, IEEE, 2012.

[43] F. Borrelli, A. Bemporad, and M. Morari, *Predictive Control for Linear and Hybrid Systems*. Cambridge University Press, 2017.

[44] A. Skopalik and B. Vöcking, "Inapproximability of pure Nash equilibria," in *Proceedings of the fortieth annual ACM symposium on Theory of computing*, pp. 355–364, 2008.

[45] S. P. Boyd and L. Vandenberghe, *Convex optimization*. Cambridge University Press, 2004.

APPENDIX

A. Proof of Corollary 4.1

Corollary. *A policy $\boldsymbol{K}^p = \boldsymbol{\Phi}_{\boldsymbol{u}}^p \boldsymbol{\Phi}_{\boldsymbol{x}}^{-1}$ ($\forall p \in \mathcal{P}$) is defined by the kernel $\boldsymbol{\Phi}^p = \boldsymbol{\Phi}_{\boldsymbol{u}}^p * \boldsymbol{\Phi}_{\boldsymbol{x}}^{-1}$, and implemented as*

$$\xi_t = -\sum_{\tau=1}^{t} \Phi_{x,\tau+1}\xi_{t-\tau} + x_t; \tag{50a}$$
$$u_t^p = \sum_{\tau=0}^{t} \Phi_{u,\tau+1}^p \xi_{t-\tau}, \tag{50b}$$

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

Proof. The statement $\boldsymbol{\Phi}^p = \boldsymbol{\Phi}_{\boldsymbol{u}}^p * \boldsymbol{\Phi}_{\boldsymbol{x}}^{-1}$ follows directly from the inverse Z-Transform of $\hat{\boldsymbol{K}}^p = \hat{\boldsymbol{\Phi}}_{\boldsymbol{u}}^p \hat{\boldsymbol{\Phi}}_{\boldsymbol{x}}^{-1}$ and $\boldsymbol{\Phi}^p = (\Phi_n^p)_{n\in\mathbb{N}}$ being the convolution kernel of $\boldsymbol{K}^P$. The operations Eq. (50) are obtained as the inverse Z-Transform of Eq. (25) and the fact that $\mathcal{Z}^{-1}[z(I-\hat{\boldsymbol{\Phi}}_{\boldsymbol{x}})\hat{\boldsymbol{\xi}}] = \xi_{t+1} - \xi_t - \sum_{\tau=2}^{t+1}\Phi_{x,\tau}\xi_{t+1-\tau}$. □

B. Proof of Theorem 5

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

$$J^p(\boldsymbol{\Phi}_{\boldsymbol{u}}^{p^{\varepsilon}}, \boldsymbol{\Phi}_{\boldsymbol{u}}^{-p^{\varepsilon}}) \le \min_{\boldsymbol{\Phi}_{\boldsymbol{u}}^p \in U_{\Phi}^p(\boldsymbol{\Phi}_{\boldsymbol{u}}^{-p^{\epsilon}})} J^p(\boldsymbol{\Phi}_{\boldsymbol{u}}^{p}, \boldsymbol{\Phi}_{\boldsymbol{u}}^{-p^{\varepsilon}}) + \varepsilon \tag{51}$$

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

Proof. Let $\boldsymbol{\Phi}_{\boldsymbol{u}}^{\star} \in BR_{\Phi}(\boldsymbol{\Phi}_{\boldsymbol{u}}^{\varepsilon})$. We construct a candidate fixed-point by defining $\tilde{\boldsymbol{\Phi}}_{\boldsymbol{u}}^{\varepsilon} = (\Phi_{u,n}^{\star})_{n=1}^{N-1}$. Clearly, $\tilde{\boldsymbol{\Phi}}_{\boldsymbol{u}}^{p^{\varepsilon}}$ satisfies all the constraints in Problem (33) by construction and also $(\Phi_{x,n}^{\star})_{n=1}^{N} = \tilde{\boldsymbol{\Phi}}_{\boldsymbol{x}}^{\varepsilon} = \boldsymbol{F}_{\boldsymbol{\Phi}}\tilde{\boldsymbol{\Phi}}_{\boldsymbol{u}}^{\varepsilon}$ satisfies $\|\Phi_{x,N}^{\star}\|_F^2 \le \gamma$ by our assumption. Now, consider $\boldsymbol{\Phi}_{\boldsymbol{u}}^{\varepsilon} \in \widehat{BR}_{\Phi}(\boldsymbol{\Phi}_{\boldsymbol{u}}^{\varepsilon})$. From optimality,

$$J^p(\boldsymbol{\Phi}_{\boldsymbol{u}}^{p^{\varepsilon}}, \boldsymbol{\Phi}_{\boldsymbol{u}}^{-p^{\varepsilon}}) \le J^p(\tilde{\boldsymbol{\Phi}}_{\boldsymbol{u}}^{p^{\varepsilon}}, \boldsymbol{\Phi}_{\boldsymbol{u}}^{-p^{\varepsilon}}) \le \frac{1}{1-\gamma} J^p(\boldsymbol{\Phi}_{\boldsymbol{u}}^{\star}, \boldsymbol{\Phi}_{\boldsymbol{u}}^{-p^{\varepsilon}}), \tag{52}$$

where the second inequality derives from the fact that the quadratic objective functional is larger for the infinite-impulse response $\boldsymbol{\Phi}_{\boldsymbol{u}}^{\star}$ and that $\frac{1}{1-\gamma} > 1$. Finally, by definition,

$$\boldsymbol{\Phi}_{\boldsymbol{u}}^{p^{\star}} = \operatorname*{arg\,min}_{\boldsymbol{\Phi}_{\boldsymbol{u}}^p \in U_{\Phi}^p(\boldsymbol{\Phi}_{\boldsymbol{u}}^{-p^{\epsilon}})} J^p(\boldsymbol{\Phi}_{\boldsymbol{u}}^{p}, \boldsymbol{\Phi}_{\boldsymbol{u}}^{-p^{\varepsilon}})$$

and thus the inequality Eq. (51) follows directly from Eq. (52) by defining $\varepsilon = \max_{p\in\mathcal{P}} \gamma J^p(\boldsymbol{\Phi}_{\boldsymbol{u}}^{p^{\varepsilon}}, \boldsymbol{\Phi}_{\boldsymbol{u}}^{-p^{\varepsilon}})$. □

C. Proof of Theorem 8

In order to prove this theorem, we start by introducing some useful lemmas. Firstly, let $\mathcal{W} = \{w_t \in \mathbb{R}^{N_x} : \|w_t\|_{\infty} \le 1\}$ and assume $\|\Phi_{x,N}\|_F^2 \le \gamma$ is strictly satisfied, and define the operators $\{\boldsymbol{H}^{p\tilde{p}}\}_{p,\tilde{p}\in\mathcal{P}}$ as in Section III-B. In the following, we slightly abuse notation and apply matrix algebra directly to signals and operators: These should be understood as applied to their equivalent matrix representations.

Lemma 9. *Define the signal $\vec{\boldsymbol{\Phi}}_{\boldsymbol{u}}^p = (\texttt{vec}\ \Phi_{u,n}^p)_{n=1}^{N-1}$ for each $p \in \mathcal{P}$. The (approximately)best-response mappings satisfy $\boldsymbol{\Phi}_{\boldsymbol{u}}^{p^{\star}} \in \widehat{BR}_{\Phi}^{p}(\boldsymbol{\Phi}_{\boldsymbol{u}}^{-p})$, where $\boldsymbol{\Phi}_{\boldsymbol{u}}^{p^{\star}}$ is obtained from a solution of*

$$\begin{aligned}
\underset{\boldsymbol{t},\vec{\boldsymbol{\Phi}}_{\boldsymbol{u}}^p}{\text{minimize}} \quad & \vec{\boldsymbol{\Phi}}_{\boldsymbol{u}}^{p^{\mathsf{T}}} \widetilde{\boldsymbol{H}}^{pp} \vec{\boldsymbol{\Phi}}_{\boldsymbol{u}}^p + 2(\widetilde{\boldsymbol{H}}^{p,-p}\vec{\boldsymbol{\Phi}}_{\boldsymbol{u}}^{-p} + \widetilde{\boldsymbol{H}}^{p0})^{\mathsf{T}}\vec{\boldsymbol{\Phi}}_{\boldsymbol{u}}^p \\
\text{subject to} \quad & -\tilde{t} \le \widetilde{\boldsymbol{G}}_u^p \vec{\boldsymbol{\Phi}}_{\boldsymbol{u}}^p \le \tilde{t}, \\
& \boldsymbol{1}^{\mathsf{T}} t_j = [g_u^p]_j, \quad j = 1, \dots, N_{\mathcal{U}^p}
\end{aligned}$$

given the operators

$$\begin{aligned}
\widetilde{\boldsymbol{H}}^{p\tilde{p}} &= \texttt{blkdiag}(\boldsymbol{H}^{p\tilde{p}}, \dots, \boldsymbol{H}^{p\tilde{p}}) = I_{N_x} \otimes \boldsymbol{H}^{p\tilde{p}} \\
\widetilde{\boldsymbol{G}}_u^p &= \texttt{blkdiag}(G_u^p, \dots, G_u^p) = I_{(N-1)N_x} \otimes G_u^p,
\end{aligned}$$

for $\tilde{p} \in \mathcal{P}$. Here, t is an auxiliary block-vector of appropriate dimensions and $\tilde{t} = U^{\mathsf{T}} t$ for some unitary transformation U.

Proof. Consider Problem (33) with the aforementioned assumptions. Firstly, we convert the linear dynamics,

$$\begin{aligned}
&\texttt{vec}(\boldsymbol{\Phi}_{\boldsymbol{x}}) = \texttt{vec}(\boldsymbol{S}_+\boldsymbol{A}\boldsymbol{\Phi}_{\boldsymbol{x}} + \textstyle\sum_p \boldsymbol{S}_+\boldsymbol{B}^p\boldsymbol{\Phi}_{\boldsymbol{u}}^p + \boldsymbol{\delta} I_{N_x}), \\
&\Rightarrow \vec{\boldsymbol{\Phi}}_{\boldsymbol{x}} = (I_{N_x} \otimes \boldsymbol{S}_+\boldsymbol{A})\vec{\boldsymbol{\Phi}}_{\boldsymbol{x}} + \textstyle\sum_p (I_{N_x} \otimes \boldsymbol{S}_+\boldsymbol{B}^p)\vec{\boldsymbol{\Phi}}_{\boldsymbol{u}}^p + \tilde{\boldsymbol{\delta}}\vec{\Phi}_{x,1} \\
&\Rightarrow \vec{\boldsymbol{\Phi}}_{\boldsymbol{x}} = I_{N_x} \otimes (I - \boldsymbol{S}_+\boldsymbol{A})^{-1}\big(\textstyle\sum_p \boldsymbol{S}_+\boldsymbol{B}^p\vec{\boldsymbol{\Phi}}_{\boldsymbol{u}}^p + \tilde{\boldsymbol{\delta}}\vec{\Phi}_{x,1}\big) \\
&\Rightarrow \vec{\boldsymbol{\Phi}}_{\boldsymbol{x}} = I_{N_x} \otimes \big(\textstyle\sum_p \boldsymbol{F}_{\boldsymbol{\Phi}}^{\boldsymbol{p}}\vec{\boldsymbol{\Phi}}_{\boldsymbol{u}}^p + \boldsymbol{F}_{\boldsymbol{\Phi}}^0\big),
\end{aligned}$$

where $\vec{\Phi}_{x,1} = \texttt{vec}\ I_{N_x}$. After some standard manipulations, the objective functional becomes

$$\begin{aligned}
J^p(\cdot) = &\vec{\boldsymbol{\Phi}}_{\boldsymbol{u}}^{p^{\mathsf{T}}}(I_{N_x} \otimes \boldsymbol{H}^{pp})\vec{\boldsymbol{\Phi}}_{\boldsymbol{u}}^p \\
&+ 2\big((I_{N_x} \otimes \boldsymbol{H}^{p,-p})\vec{\boldsymbol{\Phi}}_{\boldsymbol{u}}^{-p} + (I_{N_x} \otimes \boldsymbol{H}^{p0})\big)^{\mathsf{T}}\vec{\boldsymbol{\Phi}}_{\boldsymbol{u}}^p \\
&+ (\text{affine terms}).
\end{aligned}$$

Consider now the vectorised constraints,

$$\begin{aligned}
&\textstyle\sum_{n=0}^{N-1} \|\texttt{vec}([G_u^p]_j \Phi_{u,n}^p)\|_1 \le [g_u^p]_j \\
&\quad\Rightarrow \textstyle\sum_{n=0}^{N-1} \|(I_{N_x} \otimes [G_u^p]_j)\vec{\Phi}_{u,n}^p\|_1 \le [g_u^p]_j, \\
&\quad\Rightarrow -t_{j,n} \le (I_{N_x} \otimes [G_u^p]_j)\vec{\Phi}_{u,n}^p \le t_{j,n}, \quad n = 1, \dots, N-1
\end{aligned}$$

where we introduced the auxiliary vector $t_{j,n} \in \mathbb{R}^{N_x}$ then reformulated the sum of 1-norms into its epigraph form by adding the constraint $\boldsymbol{1}^{\mathsf{T}} t_j = [g_u^p]_j$ with $t_j = \texttt{col}(t_{j,1}, \dots, t_{j,N-1})$. Equivalently, these inequalities can be written as

$$-t_j \le (I_{(N-1)N_x} \otimes [G_u^p]_j)\vec{\Phi}_u^p \le t_j, \quad \boldsymbol{1}^{\mathsf{T}} t_j = [g_u^p]_j.$$

Since $I_{(N-1)N_x} \otimes G_u^p = U\texttt{col}(I_{(N-1)N_x} \otimes [G_u^p]_j)_{j=1}^{N_{\mathcal{U}^p}}$ for some unitary matrix U, we concatenate all inequalities into

$$-U^{\mathsf{T}} t \le (I_{(N-1)N_x} \otimes G_u^p)\vec{\Phi}_u^p \le U^{\mathsf{T}} t,$$

with $t = \texttt{col}(t_1, \dots, t_{N_{\mathcal{U}^p}})$. Finally, we define $\{\widetilde{\boldsymbol{H}}^{p\tilde{p}}\}_{\tilde{p}\in\mathcal{P}}$ and $\widetilde{\boldsymbol{G}}_u^p$ as in the lemma's statement and reformulate the best-response mapping accordingly. □

The above result shows that $\{\widehat{BR}_{\Phi}^{p}\}_{p\in\mathcal{P}}$ are the solutions of parametric quadratic programs with linear constraints: They consist of piecewise affine operators. The following lemma relates the Lipschitz constants of a collection of affine operators with that of the operator defined by their concatenation.

Lemma 10. *Let $T = (T^1, \dots, T^{N_P}) : \mathcal{X} \to \mathcal{Y}$ be a mapping constructed from the affine operators $T^p x = A^p x + b^p$, $p \in \mathcal{P}$. Then, $L_T = \|(A^1, \dots, A^{N_P})\|_{2\to 2}$ is the (tightest) Lipschitz constant for T, which can be relaxed by the upper-bound*

$$L_T^2 \le \textstyle\sum_{p\in\mathcal{P}} \|A^p\|_{2\to 2}^2 \tag{54}$$

Proof. Firstly, $Tx = (A^1, \dots, A^{N_P})x + (b^1, \cdots, b^{N_P})$, that is, T is an affine operator with $A = (A^1, \dots, A^{N_P}) \in \mathcal{L}(\mathcal{X}, \mathcal{Y})$. From definition, the operator norm $\|(A^1, \dots, A^{N_P})\|_{2\to 2} = L_T$ is smallest Lipschitz constant for this mapping. Finally,

$$\begin{aligned} L_T^2 &= \|(A^1, \dots, A^{N_P})\|_{2\to 2}^2, \\ &= \|\textstyle\sum_{p\in\mathcal{P}} A^{p*} A^p\|_{2\to 2}, \\ &\le \textstyle\sum_{p\in\mathcal{P}} \|A^{p*} A^p\|_{2\to 2}, \\ &= \textstyle\sum_{p\in\mathcal{P}} \|A^p\|_{2\to 2}^2, \end{aligned}$$

since $(A^1, \dots, A^{N_P})^*(A^1, \dots, A^{N_P}) = \sum_{p\in\mathcal{P}} A^{p*} A^p$. ☐

Finally, we can proceed to prove the Theorem 7.

Theorem. *Consider $\mathcal{X} = \mathbb{R}^{N_x}$ and $\mathcal{U}_{\mathcal{G}} = \prod_{p\in\mathcal{P}} \mathbb{R}^{N_u^p}$. Then, The map $\widehat{BR}_\Phi$ is $L_{\widehat{BR}_\Phi}$-Lipschitz with*

$$L_{\widehat{BR}_\Phi} = \sqrt{\textstyle\sum_{p\in\mathcal{P}} (L^p_{\widehat{BR}_\Phi})^2}, \tag{55}$$

given the player-specific constants

$$L^p_{\widehat{BR}_\Phi} = \big(1 + \kappa(\boldsymbol{H}^{pp})\big) \|(\boldsymbol{H}^{pp})^\dagger \boldsymbol{H}^{p,-p}\|_{2\to 2}, \tag{56}$$

with condition number $\kappa(\boldsymbol{H}^{pp}) = \|(\boldsymbol{H}^{pp})^\dagger\|_{2\to 2} \|\boldsymbol{H}^{pp}\|_{2\to 2}$.

Proof. Consider the reformulation of the (approximately)best-response maps $\widehat{BR}_\Phi$ introduced in Lemma 9. Further, assume that the active constraints are known for each $\boldsymbol{\Phi}_{\boldsymbol{u}}^{-p} \in \boldsymbol{\mathcal{C}}_u^{-p}$: We let $\widetilde{\boldsymbol{G}}^p_{u,\mathcal{A}} = e_{\mathcal{A}} \widetilde{\boldsymbol{G}}^p_u$ denote the rows associated with the active inequalities indexed by $\mathcal{A}$, where the sign has been absorbed into the selector matrix $e_{\mathcal{A}}$. Using the KKT conditions [45], a solution of this problem is obtained from its Lagrangian as

$$\vec{\boldsymbol{\Phi}}^{p^\star}_{\boldsymbol{u}} = -(\widetilde{\boldsymbol{H}}^{pp})^\dagger (\widetilde{\boldsymbol{H}}^{p,-p} \vec{\boldsymbol{\Phi}}^{-p}_{\boldsymbol{u}} + \widetilde{\boldsymbol{H}}^{p0} + \tfrac{1}{2} \widetilde{\boldsymbol{G}}_u^{p^\intercal} \boldsymbol{\lambda}^{p^\star}), \tag{57}$$

where $\boldsymbol{\lambda}^{p^\star}$ is an optimal solution to the associated dual problem. From complementary slackness, $\lambda_i^{p^\star} = 0$ for all $i \notin \mathcal{A}$, and the non-zero lagrangian multiplier are

$$\begin{aligned} \boldsymbol{\lambda}^{p^\star}_{\mathcal{A}} &= \widetilde{\boldsymbol{G}}^p_{u,\mathcal{A}} \vec{\boldsymbol{\Phi}}^{p^\star}_{\boldsymbol{u}} - \tilde{t}_{\mathcal{A}} \\ &= -2\big[\widetilde{\boldsymbol{G}}^p_{u,\mathcal{A}} (\widetilde{\boldsymbol{H}}^{pp})^\dagger \widetilde{\boldsymbol{G}}^{p^\intercal}_{u,\mathcal{A}}\big]^\dagger \big[\widetilde{\boldsymbol{G}}^p_{u,\mathcal{A}} (\widetilde{\boldsymbol{H}}^{pp})^\dagger \widetilde{\boldsymbol{H}}^{p,-p}\big] \vec{\boldsymbol{\Phi}}^{-p}_{\boldsymbol{u}} \\ &\qquad + \text{(affine terms)}. \end{aligned} \tag{58}$$

Now, introduce the auxiliary operator $\widetilde{\boldsymbol{V}}^p = (\widetilde{\boldsymbol{G}}^p_{u,\mathcal{A}})^\dagger (\widetilde{\boldsymbol{G}}^p_{u,\mathcal{A}})$. After some algebra, we can combine Eqs. (57)–(58) into

$$\begin{aligned} \vec{\boldsymbol{\Phi}}^{p^\star}_{\boldsymbol{u}} &= -\Big\{ (\widetilde{\boldsymbol{H}}^{pp})^\dagger \widetilde{\boldsymbol{H}}^{p,-p} \\ &\qquad + (\widetilde{\boldsymbol{H}}^{pp})^\dagger \widetilde{\boldsymbol{V}}^{p^\intercal} \widetilde{\boldsymbol{H}}^{pp} \widetilde{\boldsymbol{V}}^p (\widetilde{\boldsymbol{H}}^{pp})^\dagger \widetilde{\boldsymbol{H}}^{p,-p} \Big\} \vec{\boldsymbol{\Phi}}^{-p}_{\boldsymbol{u}} \\ &\qquad + \text{(affine terms)}. \end{aligned} \tag{59}$$

Thus, the optimal solution takes the form of an affine operator $\vec{\boldsymbol{\Phi}}^{p^\star}_{\boldsymbol{u}} = (0, \boldsymbol{A}^{-p}) \vec{\boldsymbol{\Phi}}_{\boldsymbol{u}} + \boldsymbol{b}^p$ from $\boldsymbol{\mathcal{C}}_u$ to $\boldsymbol{\mathcal{C}}_u^p$, with $(\boldsymbol{A}^{-p}, \boldsymbol{b}^p)$ obtained from Eq. (59). The collective best-response is the concatenation of these solutions,

$$\vec{\boldsymbol{\Phi}}^\star_{\boldsymbol{u}} = \big((0, \boldsymbol{A}^{-1}), \dots, (0, \boldsymbol{A}^{-N_P})\big) \vec{\boldsymbol{\Phi}}_{\boldsymbol{u}} + (\boldsymbol{b}^1, \dots, \boldsymbol{b}^2).$$

Finally, note that converting this formula in terms of the original matrix-valued signal requires only a unitary transformation that preserves its Lipschitz properties. Thus, using Lemma 10,

$$L^2_{\widehat{BR}_\Phi} = \textstyle\sum_{p\in\mathcal{P}} \|(0, \boldsymbol{A}^{-p})\|^2_{2\to 2}.$$

Now, consider that

$$\begin{aligned} &\|(0, \boldsymbol{A}^{-p})\|_{2\to 2} \\ &\quad \le \big(1 + \kappa(\widetilde{\boldsymbol{H}}^{pp}) \|\widetilde{\boldsymbol{V}}^p\|^2_{2\to 2}\big) \|(\widetilde{\boldsymbol{H}}^{pp})^\dagger \widetilde{\boldsymbol{H}}^{p,-p}\|_{2\to 2} \\ &\quad = \big(1 + \kappa(\widetilde{\boldsymbol{H}}^{pp})\big) \|(\widetilde{\boldsymbol{H}}^{pp})^\dagger \widetilde{\boldsymbol{H}}^{p,-p}\|_{2\to 2} \\ &\quad = \big(1 + \kappa(\boldsymbol{H}^{pp})\big) \|(\boldsymbol{H}^{pp})^\dagger \boldsymbol{H}^{p,-p}\|_{2\to 2} \\ &\quad = L^p_{\widehat{BR}_\Phi} \end{aligned}$$

where for the first inequality we applied to Eq. (59) the triangle inequality and submultiplicative properties of operators norms, on the first equality we used the fact that $\|\widetilde{\boldsymbol{V}}^p\| = 1$, and for the last equality we used the fact that $\|I_m \otimes \boldsymbol{Z}\| = \|\boldsymbol{Z}\|$ for any operator norm, $m > 0$, and operator $\boldsymbol{Z}$. Finally, since

$$L^2_{\widehat{BR}_\Phi} \le \textstyle\sum_{p\in\mathcal{P}} (L^p_{\widehat{BR}_\Phi})^2,$$

we can relax this Lipschitz constant for $\widehat{BR}_\Phi$ by considering this expression to hold as an equality. ☐