

# A system-level approach to generalized feedback Nash equilibrium seeking in partially observed games

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

**Abstract**—This work proposes an algorithm for seeking generalized feedback Nash equilibria (GFNE) in noncooperative dynamic games. The focus is on cyber-physical systems with dynamics which are linear, stochastic, potentially unstable, and partially observed. We employ System Level Synthesis (SLS) to reformulate the problem as the search for an equilibrium profile of closed-loop responses to noise, which can then be used to reconstruct a stabilizing output-feedback policy. Under this setup, we leverage monotone operator theory to design a GFNE-seeking algorithm capable to enforce closed-loop stability, operational constraints, and communication constraints onto the control policies. This algorithm is amenable to numerical implementation and we provide conditions for its convergence. We demonstrate our approach in a simulated experiment on the noncooperative stabilization of a decentralized power grid.

**Index Terms**—Noncooperative games, Feedback Nash equilibrium, Monotone operators, System level synthesis

## I. INTRODUCTION

CYBER-PHYSICAL systems are often comprised of many interacting subsystems each operated by a self-interested decision-making agent. Each agent should control its subsystem according to a feedback policy which is optimal, given its local objective, while still satisfying global requirements, given the policies applied to the other subsystems. However, due to the noncooperative and multi-objective nature of the problem, a traditional approach to policy synthesis becomes impractical. Dynamic game theory offers an alternative framework that explicitly accounts for the behavior of these rational agents when choosing their policies [1]. Under this framework, the task translates to the search for a profile of feedback policies (the *strategies*) which are feasible and agreeable to all agents (the *players*); that is, a strategic equilibrium. Game-theoretical methods for online decision-making have found success in many applications, such as power systems [2], [3], [4], communication networks [5], and multi-agent robotics [6], [7], [8].

Due to their generality, a pertinent class of problems concern noncooperative dynamic games in which the underlying system is stochastic, potentially unstable, and partially observed. In such problems, players choose output-feedback policies to jointly stabilize the whole system while using (noisy) partial measurements of its internal state. Their choice can be further restricted by coupled constraints on their control actions, on

the state responses, and on the policies themselves (e.g., encoding communication delays). A relevant solution concept is the *generalized feedback Nash equilibrium* (GFNE, [1], [9]): A set of policies which are, simultaneously, the best strategy for each player given the others' choices. Despite prevalent, Nash equilibria are notoriously hard to compute [10] and there are still no systematic solutions to the aforementioned class of dynamic games. The current practice for solving dynamic games includes dynamic programming [11], [12], [13], [14], [15], complementarity methods [16], [17], and *ad-hoc* heuristics [18], [19], [20]. These methods are mostly limited to state-feedback problems and cannot address either closed-loop stability, operational constraints, or the design of the policy structure. Moreover, they focus on solving the game for the sake of analyzing how the agents behave when operating under equilibrium policies; often not discussing how these agents would themselves learn such policies. *Equilibrium-seeking algorithms*, when players actively search for an equilibrium solution, mimic the behavior of real rational agents and thus provide practical routines for decentralized policy learning.

Recently, monotone operator theory has gained attention as a unifying framework for designing equilibrium-seeking algorithms for many applications [21], [22], [23], [24], [25]. Under a refined solution concept, the *variational generalized Nash equilibrium* (vGNE), noncooperative games can be reformulated as monotone inclusion problems and then approached by fixed-point methods which often translate to equilibrium-seeking routines. This operator-theoretical approach not only allows the solution of noncooperative games with non-trivial information structures, a major challenge in the field [26], but also simplifies the convergence analysis of such routines. Regrettably, its application is mostly limited to static games due to technical challenges arising in GFNE-seeking problems: The space of output-feedback policies (i.e., mappings from measurement to action signals) is generally not endowed with an inner product, necessary for defining *variational* equilibria. In addition, the numerics common to most vGNE-seeking routines (e.g., computing pseudo-gradient and projections) are cumbersome when the strategy spaces are infinite-dimensional; thus limiting their scalability. However, if such issues can be overcome, monotone operator theory stands out as a promising platform for designing GFNE-seeking algorithms for dynamic games. This is the purpose of this study.

In this work, we propose a GFNE-seeking algorithm for partially observed stochastic dynamic games. Our algorithm is centred on the equivalent representation of linear output-

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

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

feedback policies as their corresponding *closed-loop responses* (or *system level responses*) to disturbances. Under this setup, players design policies indirectly by instead choosing desired system level responses that serve as their parametrization. Using the System Level Synthesis (SLS, [27]) methodology, we translate the original dynamic game into a static game whose strategy space consists of closed-loop responses that explicitly enforce stability (for the whole networked system), operational constraints (on the state and control signals), and communication constraints (for the output-feedback policies). This strategy space is convex and finite-dimensional, making the associated static game amenable to numerical solutions. We formulate a monotone inclusion problem to obtain a vGNE for this static game, which is then used to directly reconstruct a GFNE solution to the original problem; effectively enabling the design of a GFNE-seeking algorithm. In this direction, we design a routine in which players improve their policies, in parallel, by updating their system level parametrization, while a coordinator ensures that constraints are satisfied. Using results from operator theory, we derive convergence certificates for this algorithm. In summary, our contributions are:

- (i) A principled approach to solve noncooperative dynamic games using monotone operator theory. We propose an equilibrium-seeking algorithm based on the equivalent representation of GFNE as their corresponding closed-loop responses to external disturbances;
- (ii) A configuration of our method to address stability, operational, and informational constraints in partially observed systems. We focus on the (not restrictive) case of policies subjected to actuation and communication delays;
- (iii) A formal analysis of the convergence properties of our algorithm and a specific implementation of its numerics to improve its efficiency and scalability.

We demonstrate our GFNE-seeking algorithm in a simulated problem: The noncooperative control of a decentralized power grid. The routine is executed alongside the operation of the system and we show that players approach a GFNE while still complying with operational and communication constraints.

This article builds on our previous work [28] which addressed GFNE-seeking in state-feedback problems (i.e., fully-observed games) using the *best-response dynamics* algorithm. Our results here provide a significant extension not only by considering the more general class of output-feedback problems, but also by applying monotone operator theory to design a scalable *operator splitting* algorithm with well-understood convergence properties. Moreover, while our previous study focused on systems affected by bounded disturbances, here we consider the case of unbounded (e.g., Gaussian) noise.

The paper is structured as follows: Section II reviews noncooperative games and operator splitting algorithms. Section III presents the SLS-based reformulation of GFNE problems and our proposed equilibrium-seeking algorithm for their solution. Section IV demonstrates our proposal in an exemplary problem and, finally, Section V concludes the article.

#### A. Notation

We use Latin and Greek letters to denote vectors (lowercase) and mappings (uppercase). Sets are in calligraphic font;

exceptions are the usual  $\mathbb{R}$  and  $\mathbb{N}$ . Given  $\mathcal{I} \subseteq \mathbb{N}$ , sequences are denoted  $x = (x_t)_{t \in \mathcal{I}}$  or  $x = (x_t)_{t=0}^T$  if  $\mathcal{I} = \{0, \dots, T\}$ . For  $p \in (0, \infty)$ , we define the set of  $N_x$ -dimensional sequences  $\ell_p^{N_x}(\mathcal{I}) = \{x \mid \|x\|_{\ell_p} = (\sum_{t \in \mathcal{I}} \|x_t\|^p)^{\frac{1}{p}} < \infty\}$ , with  $\ell_\infty^{N_x}$  the space of all bounded sequences.  $\mathcal{L}(\mathcal{X}, \mathcal{Y})$  is the set of bounded linear operators  $A : \mathcal{X} \rightarrow \mathcal{Y}$ . We use the standard definitions of Hardy spaces  $\mathcal{H}_\infty$  and  $\mathcal{RH}_\infty$ , with  $\frac{1}{z}\mathcal{RH}_\infty$  denoting the set of real-rational strictly proper transfer functions. Some standard signals and operators used in this paper include: The impulse signal  $\delta = (\delta_t)_{t \in \mathbb{N}}$ , the identity operator  $I$  and matrix  $I_N$  (with  $e_i$  denoting its  $i$ -th column), and the normal cone  $N_S$  and projection  $\text{proj}_S$  operators for a closed convex set  $S \subseteq \mathbb{R}^N$ .

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 often write  $s = (s^p, s^{-p})$  to highlight the  $p$ -th element; this should not be interpreted as a reordering. Similarly, for any set  $\mathcal{S} = \prod_{p \in \mathcal{P}} \mathcal{S}^p$ , we define the product  $\mathcal{S}^{-p} = \prod_{\bar{p} \in \mathcal{P} \setminus \{p\}} \mathcal{S}^{\bar{p}}$ . A mapping  $F : \mathbb{R}^N \rightrightarrows \mathbb{R}^M$  is  $M_F$ -strongly-monotone and  $L_F$ -Lipschitz continuous with  $0 < M_F \leq L_F < \infty$  if

$$\langle u - v, x - y \rangle \geq M_F \|x - y\|_2^2 \text{ and } \|u - v\|_2 \leq L_F \|x - y\|_2,$$

for any  $x, y \in \mathbb{R}^N$ ,  $u \in F(x)$  and  $v \in F(y)$ .  $F$  is a contraction if  $L_F < 1$ , nonexpansive if  $L_F = 1$ , monotone if the first inequality only holds for  $M_F = 0$ , and maximally monotone if no other monotone operator properly contains it. Finally,  $\text{fix}(F) = \{x \mid x \in F(x)\}$  and  $\text{zer}(F) = \{x \mid 0 \in F(x)\}$  are, respectively, the set of fixed-points and zeroes of  $F$ .

## II. PRELIMINARIES: NONCOOPERATIVE GAMES AND EQUILIBRIUM SEEKING VIA OPERATOR SPLITTING

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

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

defines the problem in which *players*  $p \in \mathcal{P} = \{1, \dots, N_P\}$  each decides on a strategy  $s^p \in S^p(s^{-p}) \subseteq S^p$  to minimize an objective function  $L^p : \mathcal{S}^1 \times \dots \times \mathcal{S}^{N_P} \rightarrow \mathbb{R}$ . The strategy sets  $S^p \subseteq \mathbb{R}^{N_s^p}$  ( $\forall p \in \mathcal{P}$ ) define the actions available to the players, with  $S^p : \mathcal{S}^{-p} \rightrightarrows S^p$  restricting this choice based on their rivals' strategies. As such, the problem is coupled through both objective functions and feasible action sets. Finally, the players are assumed to be non-cooperative and acting simultaneously.

A strategy profile  $s = (s^1, \dots, s^{N_P}) \in \mathcal{S} = \mathcal{S}^1 \times \dots \times \mathcal{S}^{N_P}$ , characterizes a solution to  $\mathcal{G}$  when it is agreeable to all players. If acting rationally, a reasonable assumption is that players might prefer to (myopically) choose a best-response to their rivals's strategies. Formally, the mapping  $BR^p : \mathcal{S}^{-p} \rightrightarrows S^p$ ,

$$BR^p(s^{-p}) := \arg \min_{s^p} \{L^p(s^p, s^{-p}) \mid s^p \in S^p(s^{-p})\} \quad (2)$$

defines all such strategies for each player  $p \in \mathcal{P}$ . A profile is a best-response for all players, simultaneously, when it is a fixed-point of the joint-best-response mapping  $BR : \mathcal{S} \rightrightarrows \mathcal{S}$  defined as  $BR(s) := BR^1(s^{-1}) \times \dots \times BR^{N_P}(s^{-N_P})$ . This motivates a formal solution concept for non-cooperative games known as the *generalized Nash equilibrium* (GNE, [9]).

**Definition 1.** A profile  $s^* = (s^{1*}, \dots, s^{N_P*}) \in \mathcal{S}$  satisfying  $s^* \in BR(s^*)$  is a *generalized Nash equilibrium* (GNE) of  $\mathcal{G}$ .

Under this definition, the game  $\mathcal{G}$  is solved when no player can improve its outcome by unilaterally deviating from its current strategy. The set of fixed-points  $\text{fix}(BR)$  comprise all the solutions to  $\mathcal{G}$ . As such, the existence and uniqueness of a GNE depend explicitly on the properties of the objectives  $\{L^p\}_{p \in \mathcal{P}}$  and constraints  $\{S^p\}_{p \in \mathcal{P}}$  functions. Hereafter, we ensure that the problems being discussed are well-posed by making the following assumptions on their primitives:

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

- (a) the objective  $L^p$  is continuously differentiable on  $\mathcal{S}$ , and  $L^p(\cdot, s^{-p})$  is strictly convex on  $S^p$  for all  $s^{-p} \in S^{-p}$ .
- (b) the strategy set  $S^p$  is nonempty, compact, and convex. The mapping  $S^p : S^{-p} \rightrightarrows S^p$  takes the form

$$S^p(s^{-p}) := \{s^p \in S^p \mid (s^p, s^{-p}) \in \mathcal{S}^{\text{global}}\}, \quad (3)$$

given a compact convex constraint set  $\mathcal{S}^{\text{global}} \subseteq \mathbb{R}^{N_s}$ . Moreover, they satisfy  $\mathcal{S}_{\mathcal{G}} = (S^1 \times \dots \times S^{N_P}) \cap \mathcal{S}^{\text{global}} \neq \emptyset$ .

Under Assumption 1,  $BR : \mathcal{S}_{\mathcal{G}} \rightarrow \mathcal{S}_{\mathcal{G}}$  is a continuous function from a nonempty compact convex set onto itself: The existence of GNE (i.e.,  $\text{fix}(BR) \neq \emptyset$ ) is ensured by the Brouwer's fixed-point theorem [29]. In practice, these conditions imply that players' best-responses are always unique and their actions are coupled only through a common constraint. Specifically, Eq. (3) can be interpreted as the competition for a limited shared resource (e.g., bandwidth in a network or goods in a supply-chain). Although restrictive, these assumptions still cover a broad class of problems of practical relevance.

We investigate GNE-seeking algorithms for solving  $\mathcal{G}$ . Namely, we are interested in fixed-point methods for players to learn a strategy profile  $s^* \in \text{fix}(BR)$ . Assuming that  $\mathcal{G}$  can be repeated, with  $s_k = (s_k^1, \dots, s_k^{N_P})$  being the strategies played at the  $k$ -th episode, we seek an *update rule*  $T : \mathcal{S}_{\mathcal{G}} \rightarrow \mathcal{S}_{\mathcal{G}}$  such that  $s_{k+1} = T(s_k)$  converges to an equilibrium  $s^* \in \text{fix}(BR)$ . Conforming to realistic scenarios, such a mapping must be

- *semi-decentralized*, in the sense that players compute their updates independently, with minimal coordination;
- *based on private information*, in the sense that players do not query their rivals' objectives and strategy sets.

A natural choice consists on the operator  $T = (1-\eta)I + \eta BR$  for some  $\eta \in (0, 1)$ , since  $s_{k+1} = (1-\eta)s_k + \eta BR(s_k)$  converges to a fixed-point  $s^* \in \text{fix}(BR)$  (a GNE, by definition) when  $BR$  is either a nonexpansive or contractive operator [30]. This method, *Best-Response Dynamics* (BRD), satisfies the above requirements. However, the Lipschitz properties of  $BR$  are difficult to establish for most generalized Nash equilibrium problems and convergence might fail even in simple cases [28].

A class of equilibrium-seeking algorithms which are computationally tractable and have straightforward convergence properties can be obtained by adopting a refined solution concept: The *variational generalized Nash equilibrium* (vGNE, [9]).

**Definition 2.** A profile  $s^* = (s^{1*}, \dots, s^{N_P*}) \in \mathcal{S}$  satisfying

$$\langle F(s^*), s - s^* \rangle \geq 0, \quad \forall s \in \mathcal{S}_{\mathcal{G}}, \quad (4)$$

given the pseudo-gradient  $F(s) = (\nabla_{s^p} L^p(s^p, s^{-p}))_{p \in \mathcal{P}}$ , is a *variational generalized Nash equilibrium* (vGNE) of game  $\mathcal{G}$ . Moreover, we let  $\text{VI}(BR)$  denote the set of all vGNE of  $\mathcal{G}$ .

The term vGNE is justified since  $\text{VI}(BR) \subseteq \text{fix}(BR)$  [9]. Despite the implication that some GNE can potentially be lost, the variational approach always leads to equilibrium solutions which are fair, in the sense that the “shared resource” described by  $\mathcal{S}^{\text{global}}$  is uniformly distributed among players; it is indeed a refinement [31]. The following standing assumption is taken:

**Assumption 2.** The pseudo-gradient  $F$  is maximal monotone.

Under this assumption, obtaining a vGNE  $s^* \in \text{VI}(BR)$  is equivalent to the monotone inclusion problem

$$\text{Find } s^* \in \mathbb{R}^{N_s} \text{ such that } 0 \in F(s^*) + N_{\mathcal{S}_{\mathcal{G}}}(s^*), \quad (5)$$

which can be solved via operator splitting methods [30], [32]. In this work, we tackle Problem (5) using *forward-backward splitting* (FBS), a well-established method for solving monotone inclusion problems. This method and its application to vGNE-seeking are overviewed in the following.

#### A. vGNE-seeking via FBS

Because  $F$  is maximal monotone and single-valued, some algebra allows us to obtain the equivalence

$$0 \in F(s^*) + N_{\mathcal{S}_{\mathcal{G}}}(s^*) \iff s^* = (I + \eta N_{\mathcal{S}_{\mathcal{G}}})^{-1}(I - \eta F)(s^*),$$

for any constant  $\eta > 0$ . In turn, this implies

$$\text{VI}(BR) = \text{zer}(F + N_{\mathcal{S}_{\mathcal{G}}}) = \text{fix}(T) \quad (6)$$

for the operator  $T = (I + \eta N_{\mathcal{S}_{\mathcal{G}}})^{-1} \circ (I - \eta F)$ . This operator is a candidate update rule since the iterations  $s_{k+1} = T(s_k)$  converge to a (unique) solution  $s^* \in \text{VI}(BR)$  whenever  $T$  is a contraction. Due to its composition form, this iteration can be split into a forward and backward step, respectively

$$\underbrace{\begin{bmatrix} s_{k+1/2}^1 \\ \vdots \\ s_{k+1/2}^{N_P} \end{bmatrix}}_{s_{k+1/2}} = \underbrace{\begin{bmatrix} s_k^1 - \eta \nabla_{s^1} L^1(s_k^1, s_k^{-1}) \\ \vdots \\ s_k^{N_P} - \eta \nabla_{s^{N_P}} L^{N_P}(s_k^{N_P}, s_k^{-N_P}) \end{bmatrix}}_{(I - \eta F)(s_k)}; \quad (7a)$$

$$s_{k+1} = \text{proj}_{\mathcal{S}_{\mathcal{G}}}(s_{k+1/2}), \quad (7b)$$

since  $(I + \eta N_{\mathcal{C}})^{-1} = \text{proj}_{\mathcal{C}}$  for any closed convex set  $\mathcal{C}$  [30]. Finally, we obtain the semi-decentralized routine for vGNE-Seeking via FBS outlined in Algorithm 1. At each  $k > 0$ , this routine consists of two main instructions:

- (1) Simultaneously, players propose new strategies to improve their objectives while disregarding the constraints;
- (2) A coordinator collects these proposals and then computes (and broadcasts) the closest permissible strategies.

---

#### Algorithm 1: vGNE-Seeking via FBS

---

- 1 Initialize  $s_0 = (s_0^1, \dots, s_0^{N_P})$ ;
  - 2 **for**  $k = 0, 1, 2, \dots$  **do**
  - 3     **for each player**  $p \in \mathcal{P}$  **do**
  - 4          $s_{k+1/2}^p := s_k^p - \eta \nabla_{s^p} L^p(s_k^p, s_k^{-p})$ ;
  - 5     **for coordinator do**
  - 6          $s_{k+1} := \text{proj}_{\mathcal{S}_{\mathcal{G}}}(s_{k+1/2})$ ;
-

In practice, this routine is interpreted as players moving towards best-response strategies while a coordinator ensures that no constraint is violated. In addition to this convenient interpretation, Algorithm 1 benefits a realistic vGNE-seeking by requiring only the profiles  $\{s_k\}_{k \in \mathbb{N}}$  to be public: The players do not access the objectives  $\{L^p\}_{p \in \mathcal{P}}$  and constraint  $\{S^p\}_{p \in \mathcal{P}}$  functions, except for their own. This information needs only to be available to the coordinator. We remark that the routine can still be specialized to have the players estimate their rivals' strategies by exchanging information [21], [23], [22], [24].

As previously mentioned,  $T(s_k) \rightarrow s^* \in \mathbf{VI}(BR)$  when  $T$  is a contraction. Because the operator  $\text{proj}_{S_G} = (I + \eta N_{S_G})^{-1}$  is nonexpansive (as  $S_G$  is nonempty, closed and convex), this translates into requiring that the forward operator  $(I - \eta F)$  be contractive. The following result can thus be established:

**Lemma 1.** *Let the pseudo-gradient  $F$  be  $L_F$ -Lipschitz and  $M_F$ -strongly-monotone. If  $\eta \in (0, 2M_F/L_F^2)$ , then Algorithm 1 converges linearly to the unique  $s^* \in \mathbf{VI}(BR)$  with rate*

$$\lim_{k \rightarrow \infty} \frac{\|s_{k+1} - s^*\|_2}{\|s_k - s^*\|_2} = \sqrt{1 - \eta(2M_F - \eta L_F^2)} \quad (8)$$

from any initial strategy profile  $s_0 \in \mathbb{R}^{N_s}$ .

*Proof.* Since  $\mathbf{VI}(BR) = \text{zer}(F + N_{S_G})$ , the proof is as in [30, Proposition 26.16] with  $A = N_{S_G}$  and  $B = F$ .  $\square$

A corollary to this result is that Problem (5) has a unique solution when  $\kappa_F = L_F/M_F \in (0, \infty)$ . This constant is the *condition number* of  $F$ , as it describes the sensitivity of this operator to changes in its arguments. The linear convergence of Algorithm 1, implying that each strategy update decreases the distance to the vGNE solution by (at least) a constant factor, can still be slow if  $\kappa_F \gg 0$ . In such cases, the game (or, more directly, the pseudo-gradient  $F$ ) is said to be ill-conditioned. We also note that choosing a learning rate  $\eta \in (0, 2M_F/L_F^2)$  requires full knowledge of the pseudo-gradient. In this work, it is assumed that the coordinator is responsible for computing and broadcasting this parameter to the players.

### B. Noncooperative dynamic games

We are interested in  $N_P$ -player dynamic stochastic games,

$$\mathcal{G}_{\infty}^{\text{LQ}} := (\mathcal{P}, \{U^p\}_{p \in \mathcal{P}}, \{J^p\}_{p \in \mathcal{P}}), \quad (9)$$

defined by linear dynamics and measurement model

$$\Sigma_{\mathcal{G}} : \begin{cases} x_{t+1} = Ax_t + B_w w_t + \sum_{p \in \mathcal{P}} B_u^p u_t^p; \\ y_t = Cx_t + D_w w_t. \end{cases} \quad (10a)$$

$$(10b)$$

with initial state  $x_0 = 0$ . The dynamics (Eq. 10a) describe how the state of the game,  $x = (x_t)_{t \in \mathbb{N}}$ , evolves in response to the disturbances  $w = (w_t)_{t \in \mathbb{N}}$  and player's actions  $u^p = (u_t^p)_{t \in \mathbb{N}}$ ,  $p \in \mathcal{P}$ . The measurement process (Eq. 10b) describes how a public observation signal,  $y = (y_t)_{t \in \mathbb{N}}$ , is formed by (noisy) partial emissions of the state. We consider bounded signals, that is,  $x \in \ell_{\infty}^{N_x}(\mathbb{N})$ ,  $u^p \in \ell_{\infty}^{N_u^p}(\mathbb{N})$ ,  $w \in \ell_{\infty}^{N_w}(\mathbb{N})$ , and  $y \in \ell_{\infty}^{N_y}(\mathbb{N})$ . Finally, we assume  $w$  to be a white noise process satisfying  $Ew_t = 0$  and  $E(w_{t+\tau} w_t^T) = \delta_{\tau} I_{N_w}$  for all  $t, \tau \in \mathbb{N}$ .

In this class of dynamic games, each player decides a plan of action  $u^p \in U^p(u^{-p})$  to minimize an objective functional

$$J^p(u^p, u^{-p}) := E \left[ \sum_{t=0}^{\infty} \left\| \begin{bmatrix} W_x^p & W_u^p \end{bmatrix} \begin{bmatrix} x_t \\ u_t^p \end{bmatrix} \right\|_2^2 \right], \quad (11)$$

given weighting matrices  $W_x^p \in \mathbb{R}^{N_z \times N_x}$  and  $W_u^p \in \mathbb{R}^{N_z \times N_u}$  with dimension  $N_z \geq N_x + N_u$ . This objective corresponds to a standard *linear-quadratic regulator* control objective and can be interpreted as the expected energy  $E\|z\|_{\ell_2}^2$  of a performance signal  $z = (W_x^p x_t + W_u^p u_t^p)_{t \in \mathbb{N}}$  [33]. Note that  $J^p(\cdot)$  depends on other players' actions  $u^{-p}$  via the state signal  $x$  (Eq. 10a). The mapping  $U^p : \mathcal{U}^{-p} \rightrightarrows \mathcal{U}^p$  restricts the  $p$ -th player's strategies based on its rivals' choices, with  $\mathcal{U}^p$  being the set of all permissible strategies. As in the static case, players might prefer to act according to best-responses,

$$BR^p(u^{-p}) := \arg \min_{u^p} \{J^p(u^p, u^{-p}) \mid u^p \in U^p(u^{-p})\},$$

so that a (open-loop) GNE solution to  $\mathcal{G}_{\infty}^{\text{LQ}}$  is understood as an action profile  $u^* = (u^{1*}, \dots, u^{N_P*}) \in \mathcal{U} = \mathcal{U}^1 \times \dots \times \mathcal{U}^{N_P}$ , for which no player unilaterally benefits by deviating. Formally, this corresponds to a fixed-point  $u^* \in \text{fix}(BR)$ . The following assumptions are thus taken for this class of games:

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

- (a) *the weighting matrix  $W_u^p$  is full-column-rank and it satisfies  $(W_u^p)^T(W_x^p) = 0$  and  $(W_x^p)^T(W_u^p) = 0$ ;*
- (b) *the noise-filtering matrices  $B_w$  and  $D_w$  are full-row-rank and they satisfy  $(B_w)(D_w)^T = 0$  and  $(D_w)(B_w)^T = 0$ .*
- (c) *the mapping  $U^p : \mathcal{U}^{-p} \rightrightarrows \mathcal{U}^p$  takes the form*

$$U^p(u^{-p}) := \{u^p \in \mathcal{U}^p \mid (u^p, u^{-p}) \in \mathcal{U}^{\text{global}}\},$$

given the local  $\mathcal{U}^p$  and global  $\mathcal{U}^{\text{global}}$  constraint sets

$$\mathcal{U}^p = \{u^p \in \ell_{\infty}^{N_u^p}(\mathbb{N}) \mid G_{u^p} u_t^p \preceq 1, t \in \mathbb{N}\};$$

$$\mathcal{U}^{\text{global}} = \{u \in \ell_{\infty}^{N_u}(\mathbb{N}) \mid G_x x_t + G_u u_t \preceq 1, t \in \mathbb{N}\},$$

where we implicitly use the fact that  $x = F_u u + F_w w$  for causal linear operators  $(F_u, F_w)$  induced by Eq. (10a). Moreover,  $\mathcal{U}_{\mathcal{G}} = (\mathcal{U}^1 \times \dots \times \mathcal{U}^{N_P}) \cap \mathcal{U}^{\text{global}} \neq \emptyset$ .

These conditions are analogous to Assumption 1: They are to ensure that  $\text{fix}(BR) \neq \emptyset$ . In this case, however, the joint policy  $u = (u^1, \dots, u^{N_P})$  must also be internally stabilizing to avoid  $x_t \rightarrow \infty$  (and thus,  $J^p(u^p, u^{-p}) \not\prec \infty$  for some  $p \in \mathcal{P}$ ). Hereafter, we characterize an equilibrium as admissible only if it stabilizes the game. The following assumption ensures that (not necessarily  $\mathcal{U}_{\mathcal{G}}$ -feasible) stabilizing action profiles exist.

**Assumption 4.** *The system  $(A, B_u, C)$ , given the input matrix  $B_u = [B_u^1 \dots B_u^{N_P}]$ , is both controllable and observable.*

In this work, we are not interested in open-loop equilibria  $u^* \in \text{fix}(BR)$ . A plan of action with such an information structure is undesirable, as the game becomes sensible to noise disturbances and decision errors [1]. Conversely, feedback policies  $u^p = K^p(y)$ , for some  $K^p : \ell_{\infty}^{N_y} \rightarrow \ell_{\infty}^{N_u^p}$ , can detect such errors and adapt the plan of action accordingly. Thus, we let each  $p$ -th player's actions be represented by the policies

$$u^p := K^p y, \quad K^p : y \mapsto \Phi_K^p * y, \quad (12)$$

given linear causal operators  $K^p \in S^p(K^{-p}) \subseteq \mathcal{L}(\ell_\infty^{N_y}, \ell_\infty^{N_u})$  with convolution kernels  $\Phi_K^p = (\Phi_{K,n}^p)_{n \in \mathbb{N}}$ . The mapping  $S^p$  restricts the strategies to policies that are stabilizing and satisfy some  $\mathcal{G}_\infty^{\text{LQ}}$ -related restrictions (e.g., actuation and communication delays). In this setup, the  $p$ -th player's best-responses are the solutions to the synthesis problem

$$\underset{u^p := K^p y}{\text{minimize}} \quad \mathbb{E} \left[ \sum_{t=0}^{\infty} \left\| \begin{bmatrix} W_x^p & W_u^p \end{bmatrix} \begin{bmatrix} x_t \\ u_t^p \end{bmatrix} \right\|_2^2 \right] \quad (13a)$$

$$\text{subject to} \quad \Sigma_{\mathcal{G}} \text{ (Eq. 10) with } u^{\bar{p}} = K^{\bar{p}} y \ (\forall \bar{p} \in \mathcal{P}), \quad (13b)$$

$$K^p \in S^p(K^{-p}), \quad (13c)$$

$$\begin{bmatrix} G_x & G_u \\ e_p^\top \otimes G_{u^p} \end{bmatrix} \begin{bmatrix} x_t \\ u_t \end{bmatrix} \preceq 1. \quad (13d)$$

We denote the solutions to Problem (13) as  $BR_K^p(K^{-p})$ , given the joint policy  $K^{-p} : y \mapsto (K^{\bar{p}} y)_{\bar{p} \in \mathcal{P} \setminus \{p\}}$ . As before, the map

$$BR_K(K) := BR_K^1(K^{-1}) \times \cdots \times BR_K^{N_P}(K^{-N_P})$$

denotes the set of jointly-best-response policies. This induces a version of the game  $\mathcal{G}_\infty^{\text{LQ}}$  in which a solution is now understood as a *policy profile*  $K := (K^1, \dots, K^{N_P}) \in \mathcal{L}(\ell_\infty^{N_y}, \ell_\infty^{N_u})$  which is agreeable to all players. The solution concept that naturally arises is the generalized *feedback* Nash equilibrium (GFNE).

**Definition 3.** A profile  $K^* = (K^{1*}, \dots, K^{N_P*})$  satisfying  $K^* \in BR_K(K^*)$  is a *generalized feedback Nash equilibrium* (GFNE) of the dynamic game  $\mathcal{G}_\infty^{\text{LQ}}$ .

Although GFNEs and GNEs are similar concepts, designing GFNE-seeking routines is considerably more challenging: Since Problem (13) is infinite-dimensional and stochastic, it cannot be approached by the FBS routine from Section II-A. Indeed, even defining a vGFNE is not straightforward, as the strategy space  $\mathcal{L}(\ell_\infty^{N_y}, \ell_\infty^{N_u})$  is a Banach space, and therefore it is not equipped with an inner product [29]. In the following, we propose to overcome these issues by considering a specific, but equivalent, representation of the policies  $K^p$ , for all  $p \in \mathcal{P}$ . We show how this representation can be leveraged to recast the best-response mappings  $BR_K^p$  ( $\forall p \in \mathcal{P}$ ) as finite-dimensional optimization problems, thus enabling the design of vGFNE-seeking algorithms for solving dynamic games.

### III. vGFNE-SEEKING VIA SYSTEM LEVEL SYNTHESIS

We propose an equivalent representation of control policies,  $\{K^p\}_{p \in \mathcal{P}}$ , as system-level responses to noise,  $\{\Phi^p\}_{p \in \mathcal{P}}$ . Using System Level Synthesis (SLS), optimal stabilizing policies can be synthesized through tractable convex optimization problems whose domains are Hilbert spaces [27]. A class of *variational generalized feedback Nash equilibrium* (vGFNE) problems can then be formalized using this equivalent representation.

In this section, we introduce our SLS-based approach to vGFNE-Seeking in partially observed games. We first recast the original best-response mappings  $\{BR_K^p\}_{p \in \mathcal{P}}$  as equivalent *system-level best-response mappings*  $\{BR_\Phi^p\}_{p \in \mathcal{P}}$  defined by finite-horizon deterministic convex programs (Section III-A). The system-level best-response mappings are then used to define a general class of vGFNE-Seeking problems that can be solved through operator splitting, ultimately enabling the solution of the class of dynamic games  $\mathcal{G}_\infty^{\text{LQ}}$  (Section III-B).

#### A. System-level best-response mappings

We start by assuming a stabilizing profile  $(K^1, \dots, K^{N_P})$ , ensured by Assumption 4. Each policy is associated with a transfer matrix  $\hat{K}^p \in \mathcal{RH}_\infty$ ,  $\hat{K}^p = \sum_{n=0}^{\infty} \frac{1}{z^n} \Phi_{K,n}^p$ , defining the feedback policy  $\hat{u}^p = \hat{K}^p \hat{y}$  in the frequency domain. From the  $\mathbb{Z}$ -transform of the system  $\Sigma_{\mathcal{G}}$  (Eq. 10), we have

$$z\hat{x} = A\hat{x} + B_w\hat{w} + \sum_{p \in \mathcal{P}} B_u^p \hat{u}^p; \quad (14a)$$

$$\hat{u}^p = \hat{K}^p (C\hat{x} + D_w\hat{w}), \quad \forall p \in \mathcal{P}. \quad (14b)$$

The signals  $(\hat{x}, \hat{u}^1, \dots, \hat{u}^{N_P})$  can be posed in terms of  $\hat{w}$  as

$$\begin{bmatrix} \hat{x} \\ \hat{u}^1 \\ \vdots \\ \hat{u}^{N_P} \end{bmatrix} = \begin{bmatrix} \hat{\Phi}_{xx} & \hat{\Phi}_{xy} \\ \hat{\Phi}_{ux}^1 & \hat{\Phi}_{uy}^1 \\ \vdots & \vdots \\ \hat{\Phi}_{ux}^{N_P} & \hat{\Phi}_{uy}^{N_P} \end{bmatrix} \begin{bmatrix} \hat{\delta}_x \\ \hat{\delta}_y \end{bmatrix}, \quad (15)$$

where  $\hat{\delta}_x = B_w\hat{w}$  and  $\hat{\delta}_y = D_w\hat{w}$ , and

$$\hat{\Phi}_{xx} = ((zI - A) - \sum_{p \in \mathcal{P}} B_u^p \hat{K}^p C)^{-1};$$

$$\hat{\Phi}_{xy} = \sum_{p \in \mathcal{P}} \hat{\Phi}_{xx} B_u^p \hat{K}^p;$$

$$\hat{\Phi}_{ux}^p = \hat{K}^p C \hat{\Phi}_{xx};$$

$$\hat{\Phi}_{uy}^p = \hat{K}^p + \sum_{\bar{p} \in \mathcal{P}} \hat{K}^{\bar{p}} C \hat{\Phi}_{xx} B_u^{\bar{p}} \hat{K}^{\bar{p}}.$$

The transfer matrices  $(\hat{\Phi}_{xx}, \hat{\Phi}_{xy}, \hat{\Phi}_{ux}^p, \hat{\Phi}_{uy}^p)$  are *system level responses* or *closed-loop maps*: They describe how the game's state and the players' actions respond to the noise. For ease of notation, we define  $\hat{\Phi}_{ux} = (\hat{\Phi}_{ux}^p)_{p \in \mathcal{P}}$  and  $\hat{\Phi}_{uy} = (\hat{\Phi}_{uy}^p)_{p \in \mathcal{P}}$ . Under this representation, the condition that the policy profile  $\hat{K} = (\hat{K}^1, \dots, \hat{K}^{N_P})$  is stabilizing can be expressed directly in terms of the system level response  $(\hat{\Phi}_{xx}, \hat{\Phi}_{xy}, \hat{\Phi}_{ux}, \hat{\Phi}_{uy})$ . Moreover, each policy  $\hat{u}^p = \hat{K}^p \hat{y}$  ( $\forall p \in \mathcal{P}$ ) can be represented directly in terms of these closed-loop maps. Formally:

**Theorem 2** (System level parametrization). *Consider the system  $\Sigma_{\mathcal{G}}$  (Eq. 14) under output-feedback policies  $\hat{u}^p = \hat{K}^p \hat{y}$  for all players  $p \in \mathcal{P}$ . The following statements are true:*

(a) *The policy profile  $\hat{K} = (\hat{K}^1, \dots, \hat{K}^{N_P})$  is stabilizing if its system level responses  $(\hat{\Phi}_{xx}, \hat{\Phi}_{xy}, \hat{\Phi}_{ux}, \hat{\Phi}_{uy})$  satisfy*

$$[zI - A \quad -B_u^1 \cdots -B_u^{N_P}] \begin{bmatrix} \hat{\Phi}_{xx} & \hat{\Phi}_{xy} \\ \hat{\Phi}_{ux} & \hat{\Phi}_{uy} \end{bmatrix} = [I \quad 0]; \quad (16a)$$

$$\begin{bmatrix} \hat{\Phi}_{xx} & \hat{\Phi}_{xy} \\ \hat{\Phi}_{ux} & \hat{\Phi}_{uy} \end{bmatrix} \begin{bmatrix} zI - A \\ -C \end{bmatrix} = \begin{bmatrix} I \\ 0 \end{bmatrix}; \quad (16b)$$

$$\hat{\Phi}_{xx}, \hat{\Phi}_{xy}, \hat{\Phi}_{ux} \in \frac{1}{z} \mathcal{RH}_\infty, \quad \hat{\Phi}_{uy} \in \mathcal{RH}_\infty. \quad (16c)$$

(b) *The responses  $(\hat{\Phi}_{xx}, \hat{\Phi}_{xy}, \hat{\Phi}_{ux}, \hat{\Phi}_{uy})$  satisfying Eq. (16) are achieved by policies  $\hat{K}^p = \hat{\Phi}_{uy}^p - \hat{\Phi}_{ux}^p \hat{\Phi}_{xx}^{-1} \hat{\Phi}_{xy}$  ( $p \in \mathcal{P}$ ) which can be implemented as in Figure 1, that is,*

$$z\hat{\xi} = \tilde{\Phi}_{xx}\hat{\xi} + \tilde{\Phi}_{xy}\hat{y}; \quad (17a)$$

$$\hat{u}^p = \tilde{\Phi}_{ux}^p \hat{\xi} + \tilde{\Phi}_{uy}^p \hat{y}, \quad (17b)$$

with  $\tilde{\Phi}_{xx} = z(I - z\hat{\Phi}_{xx})$ ,  $\tilde{\Phi}_{ux}^p = z\hat{\Phi}_{ux}^p$ , and  $\tilde{\Phi}_{xy} = -z\hat{\Phi}_{xy}$ . Moreover,  $\hat{K} = (\hat{K}^1, \dots, \hat{K}^{N_P})$  is internally stabilizing.

*Proof.* The proof is identical to that of [27, Theorem 5.1] by letting  $B_2 = [B_u^1 \cdots B_u^{N_P}]$  and  $C_2 = C$ .  $\square$

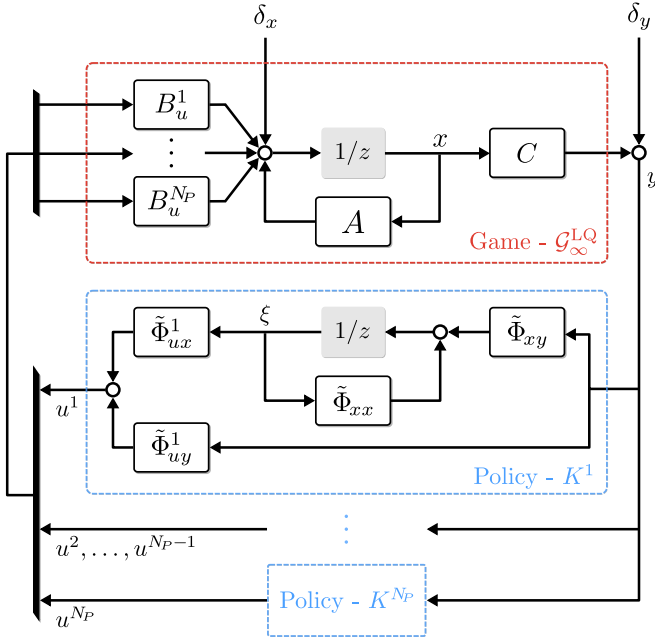


Fig. 1. The feedback structure for the policies  $\hat{K}^p = \hat{\Phi}_{uy}^p - \hat{\Phi}_{ux}^p \hat{\Phi}_{xx}^{-1} \hat{\Phi}_{xy} = \tilde{\Phi}_{uy}^p + \tilde{\Phi}_{ux}^p (zI - \tilde{\Phi}_{xx})^{-1} \tilde{\Phi}_{xy}$ , for all  $p \in \mathcal{P}$ , according to Eq. (17).

Favouring of a time-domain exposition, we will refer to the system level responses mostly through their spectral factors. In this case, Theorem 2(a) implies that a stabilizing policy must have its system level responses satisfying the system

$$\Sigma_\Phi : \begin{cases} \Phi_{xx,n+1} = A\Phi_{xx,n} + \sum_{p \in \mathcal{P}} B_u^p \Phi_{ux,n}^p; \\ \Phi_{xy,n+1} = A\Phi_{xy,n} + \sum_{p \in \mathcal{P}} B_u^p \Phi_{uy,n}^p, \end{cases} \quad (18a)$$

$$(18b)$$

and the “dual” system

$$\Sigma_\Phi^* : \begin{cases} \Phi_{xx,n+1} = \Phi_{xx,n} A + \Phi_{xy,n} C; \\ \Phi_{ux,n+1}^p = \Phi_{ux,n}^p A + \Phi_{uy,n}^p C, \quad (\forall p \in \mathcal{P}), \end{cases} \quad (19a)$$

$$(19b)$$

with  $\Phi_{xx,1} = I_{N_x}$ , and  $\Phi_{xx,0} = \Phi_{xy,0} = \Phi_{ux,0}^p = 0$  ( $\forall p \in \mathcal{P}$ ) due to strict causality. A time-domain characterization of each policy  $\hat{K}^p$  can also be obtained in terms of these components by applying an inverse  $\mathbb{Z}$ -transform to Eq. (17):

**Corollary 2.1.** *The policy  $\hat{K}^p$  from Eq. (17) is defined by the kernel  $\Phi_K^p = \Phi_{uy}^p - \Phi_{ux}^p * \Phi_{xx}^{-1} * \Phi_{xy}$  and is implemented as*

$$\xi_{t+1} = - \sum_{n=0}^t \Phi_{xx,n+2} \xi_{t-n} - \sum_{n=0}^t \Phi_{xy,n+1} y_{t-n}; \quad (20a)$$

$$u_t^p = \sum_{n=0}^t \Phi_{ux,n+1}^p \xi_{t-n} + \sum_{n=0}^t \Phi_{uy,n}^p y_{t-n}. \quad (20b)$$

using an auxiliary internal state  $\xi = (\xi_n)_{n \in \mathbb{N}}$  with  $\xi_0 = 0$ .

The system level parametrization enables a methodology for policy synthesis consisting of searching the space of stabilizing policies directly through  $(\Phi_{xx}, \Phi_{xy}, \Phi_{ux}^p, \Phi_{uy}^p)$ , for all  $p \in \mathcal{P}$ . In this direction, consider that the players plan their actions by choosing a desired system level response  $\Phi_u^p = (\Phi_{ux}^p, \Phi_{uy}^p)$  to noise, then implementing their policy as in Corollary 2.1. The original best-response mappings  $\{BR_K^p\}_{p \in \mathcal{P}}$  are then

equivalent to system-level best-response mappings  $\{BR_\Phi^p\}_{p \in \mathcal{P}}$  corresponding to the solution set of problems of the form

$$\text{minimize}_{\Phi_{ux}^p, \Phi_{uy}^p} \sum_{n=0}^{\infty} \left\| \begin{bmatrix} W_x^p & W_u^p \end{bmatrix} \begin{bmatrix} \Phi_{xx,n} & \Phi_{xy,n} \\ \Phi_{ux,n}^p & \Phi_{uy,n}^p \end{bmatrix} \begin{bmatrix} B_w \\ D_w \end{bmatrix} \right\|_F^2 \quad (21a)$$

$$\text{subject to } \Sigma_\Phi \text{ (Eq. 18)}, \quad \Sigma_\Phi^* \text{ (Eq. 19)}, \quad (21b)$$

$$\begin{bmatrix} \Phi_{xx,n} & \Phi_{xy,n} \\ \Phi_{ux,n}^p & \Phi_{uy,n}^p \end{bmatrix} \in \mathcal{S}_\Phi^p, \quad (21c)$$

$$\begin{bmatrix} G_x & G_u \\ \tilde{G}_{u^p} \end{bmatrix} \begin{bmatrix} \Phi_{xx} & \Phi_{xy} \\ \Phi_{ux} & \Phi_{uy} \end{bmatrix} \begin{bmatrix} B_w \\ D_w \end{bmatrix} * w_n \preceq 1, \quad (21d)$$

where we consider  $\tilde{G}_{u^p} = e_p^\top \otimes G_{u^p}$  to simplify its statement. The objective (Eq. 21a) and operational constraints (Eq. 21d) are obtained from the definition of the system level responses. The original structural constraints (Eq. 13c) are translated into (i) the system level parametrization (Eq. 21b) enforcing internal stability, and (ii) the system level constraints (Eq. 21c) enforcing  $\mathcal{G}_\infty^{\text{LQ}}$ -related restrictions via the policy’s system-level parameters. Note that player’s policies are linked through the  $(\Phi_{xx}, \Phi_{xy})$  responses. We refer to [27] for more details on how Problem (21) can be derived from Problem (13).

The mappings  $\{BR_\Phi^p\}_{p \in \mathcal{P}}$  are still intractable, as Problem (21) is also infinite-dimensional and stochastic. In this case, however, it can be made tractable by a specific choice of the structural constraints (Eq. 21c) and a robust formulation of the operational constraints (Eq. 21d); as shown in the following.

**Structural constraints:** The problems in  $\{BR_\Phi^p\}_{p \in \mathcal{P}}$  can be made finite-dimensional by restricting the choice of  $K^p$  to policies with finite-impulse response (FIR) kernels. Using the sets  $\mathcal{S}_\Phi^p = (\mathcal{S}_{\Phi,n}^p)_{n \in \mathbb{N}}$ , this can be done by enforcing

$$\begin{bmatrix} \Phi_{xx,n} & \Phi_{xy,n} \\ \Phi_{ux,n}^p & \Phi_{uy,n}^p \end{bmatrix} = 0, \quad \text{for } n \geq N. \quad (22)$$

In practice, this corresponds to enforcing the terminal conditions  $\Phi_{xx,N} = \Phi_{xy,N} = \Phi_{ux,N}^p = \Phi_{uy,N}^p = 0$ , then restricting Problem (21) to the finite sequences of spectral factors  $(\Phi_{xx,n}, \Phi_{xy,n}, \Phi_{ux,n}^p)_{n=1}^{N-1}$  and  $(\Phi_{uy,n}^p)_{n=0}^{N-1}$ . Additionally, FIR kernels simplify the implementation of  $K^p$  by fixing the total number of summand terms in Corollary 2.1.

**Remark 1.** *The terminal constraints induced by Eq. (22) cannot be satisfied when the system is not controllable and observable. While they can be relaxed, the stability and optimality properties of the resulting controller are well-understood only for the state-feedback case [27], [34].*

The remaining constraint sets  $(\mathcal{S}_{\Phi,n}^p)_{n=0}^N$  are designed to impose structure onto  $K^p$  through their system-level parametrization (Corollary 2.1). We consider actuation and communication delays encoded by the sparsity constraints

$$\text{Sp} \begin{bmatrix} \Phi_{xx,n} & \Phi_{xy,n} \\ \Phi_{ux,n}^p & \Phi_{uy,n}^p \end{bmatrix} = \text{Sp} \begin{bmatrix} A^{\alpha_n} & A^{\alpha_n} C^\top \\ B_u^{\alpha_n} A^{\alpha_n} & B_u^{\alpha_n} A^{\alpha_n+1} C^\top \end{bmatrix}, \quad (23)$$

for  $n = 0, \dots, N$ , with  $\alpha_n = \max(0, \lfloor (n - d_a)/d_c \rfloor)$  and  $\text{Sp}[\cdot]$  denoting the support of a matrix. Under this setup, the policies  $K^p$  ( $p \in \mathcal{P}$ ) satisfy that the  $i$ -th state component

$[x]_i = ([x_t]_i)_{t \in \mathbb{N}}$  is only affected by the  $i$ -th action component  $[B_u u^p]_i = ([B_u u_t^p]_i)_{t \in \mathbb{N}}$  after an *actuation delay* of  $d_a \geq 0$  steps. Additionally, it specifies that measurements propagate in the communication network (which has the same topology as the physical system) with a *communication delay* of  $d_c > 0$ . These parameters are often determined by the hardware of the underlying control system and communication infrastructure. We refer to policies satisfying this information pattern as  $(d_a, d_c)$ -*delayed feedback policies*. This choice is not restrictive and the results in this paper generalize to a broader class of informational constraints (such as those discussed in [35]). Finally, we note that the ability to impose sparsity to feedback policies, a central feature of the SLS framework, enables the solution of GFNE problems with asymmetric information patterns; a major challenge in the field [36], [26].

*Operational constraints:* The mapping  $BR_{\Phi}^p$  is stochastic due to the original operational constraints (Eq. 13d) becoming random inequalities in the system-level problem (Eq. 21d). The problem can be made deterministic by instead enforcing

$$\text{prob} \left( \begin{bmatrix} G_x & G_u \\ & \tilde{G}_{u^p} \end{bmatrix}_{i,:} \begin{bmatrix} \Phi_{xx} & \Phi_{xy} \\ \Phi_{ux} & \Phi_{uy} \end{bmatrix} \begin{bmatrix} B_w \\ D_w \end{bmatrix} * w_n \leq 1 \right) \geq \rho, \quad (24)$$

for every  $i$ -th row of the matrix  $G^p = \text{col}([G_x \ G_u], [0 \ \tilde{G}_{u^p}])$ , given a probability  $\rho \in (0.5, 1)$ . Since  $w$  is a white-noise process and  $\Phi_w = (\Phi_{xx} B_w + \Phi_{xy} D_w, \Phi_{ux} B_w + \Phi_{uy} D_w)$  is FIR, Eq. (24) can be expressed as  $\text{prob}(G_{\Phi,i}^p \tilde{w} \leq 1) \geq \rho$  given the random vector  $\tilde{w}$  with  $E\tilde{w} = 0$  and  $E(\tilde{w}\tilde{w}^\top) = I_{N_{N^p}}$ , and the block-matrix  $G_{\Phi,i}^p = [[G^p]_{i,:} \Phi_{w,0} \ \dots \ [G^p]_{i,:} \Phi_{w,N}]$ . The constraints Eq. (21d) can thus be realized using the cumulative distribution function of each  $G_{\Phi,i}^p \tilde{w}$  ( $i = 1, \dots, N_{G^p}$ ).

A common assumption is to consider the random vector  $\tilde{w}$  to be Gaussian, i.e.,  $\tilde{w} \sim \text{Normal}(0, I)$ . Using standard results from optimization theory [37], and expanding  $G_{\Phi,i}^p$ , Eq. (24) is then equivalent to the second-order conic (SOC) constraint

$$\left\| \left( \begin{bmatrix} G_x & G_u \\ & \tilde{G}_{u^p} \end{bmatrix}_{i,:} \begin{bmatrix} \Phi_{xx} & \Phi_{xy} \\ \Phi_{ux} & \Phi_{uy} \end{bmatrix} \begin{bmatrix} B_w \\ D_w \end{bmatrix} \right)^\top \right\|_{\ell_2} \leq \frac{1}{Q(\rho)} \quad (25)$$

with  $Q : [0, 1] \rightarrow \mathbb{R}$  being the quantile function of the standard Normal distribution (which is only positive for  $\rho > 0.5$ ). We remark that these constraints cannot ensure that the synthesized policy satisfies the original  $U^p$  for all realizations of the noise  $w$ ; even without the Gaussian assumption. However, at the risk of conservativeness, the best-response mapping  $BR_{\Phi}^p$  can still be designed to ensure with arbitrarily high probability  $\rho > 0.5$  that such constraints will be satisfied during operation.

### B. vGFNE-seeking via SLS and FBS

In this section, to simplify notation, we momentarily define  $\Phi_x = [\Phi_{xx} \ \Phi_{xy}]$ ,  $\Phi_u^p = [\Phi_{ux}^p \ \Phi_{uy}^p]$ , and  $W_w = (B_w, D_w)$ . Due to the equivalence between  $BR_K$  and  $BR_{\Phi}$ , it is clear that a stabilizing policy profile is a GFNE (i.e.,  $K^* \in BR_K(K^*)$ ) if its associated system level response is a fixed-point of  $BR_{\Phi}$  (i.e.,  $\Phi_u^* \in BR_{\Phi}(\Phi_u^*)$ ). The concept of a *variational generalized feedback Nash equilibrium* can be defined via the system level parametrization of the output-feedback policies:

### Algorithm 2: vGFNE-Seeking via FBS

---

```

1 Initialize  $\Phi_{u|0} = (\Phi_{u|0}^1, \dots, \Phi_{u|0}^{N_P})$ ;
2 for  $k = 0, 1, 2, \dots$  do
3   for each player  $p \in \mathcal{P}$  do
4      $\hat{K}_k^p := \hat{\Phi}_{uy|k}^p - \hat{\Phi}_{ux|k}^p \hat{\Phi}_{xx|k}^{-1} \hat{\Phi}_{xy|k}^p$ ;
5      $\Phi_{u|k+1/2}^p := \Phi_{u|k}^p - \eta \nabla_{\Phi_u^p} J_{\Phi}^p(\Phi_{u|k}^p, \Phi_{u|k}^{-p})$ ;
6   for coordinator do
7      $\Phi_{u|k+1} := \text{proj}_{\mathcal{U}_{\Phi, \mathcal{G}}}(\Phi_{u|k+1/2})$ ;

```

---

A policy profile  $K^* = (\hat{\Phi}_{uy}^* - \hat{\Phi}_{ux}^* (\hat{\Phi}_{xx}^*)^{-1} \hat{\Phi}_{xy}^*)_{p \in \mathcal{P}}$  is a vGFNE when the response profile  $\Phi_u^* = (\Phi_u^{1*}, \dots, \Phi_u^{N_P*})$  is a vGNE of the (static) game defined by the objectives

$$J_{\Phi}^p(\Phi_u^p, \Phi_u^{-p}) = \sum_{n=0}^N \left\| \begin{bmatrix} W_x^p & W_u^p \end{bmatrix} \begin{bmatrix} \Phi_{x,n} \\ \Phi_{u,n}^p \end{bmatrix} W_w \right\|_F^2$$

and the finite-dimensional and convex global feasible set

$$\begin{aligned} \mathcal{U}_{\Phi, \mathcal{G}} = \{ \Phi_u \in \ell_2[0, N] \mid \\ \Sigma_{\Phi} \text{ (Eq. 18)}, \ \Sigma_{\Phi}^* \text{ (Eq. 19)}, \ \Phi_x = \Phi_u = 0, \\ \text{Sp} \begin{bmatrix} \Phi_{x,n} \\ \Phi_{u,n} \end{bmatrix} = \text{Sp} \begin{bmatrix} A^{\alpha_n} & A^{\alpha_n} C^\top \\ B_u^\top A^{\alpha_n} & B_u^\top A^{\alpha_{n+1}} C^\top \end{bmatrix} \ (\forall n), \\ \left\| \begin{bmatrix} G_x & G_u \\ & \tilde{G}_{u^p} \end{bmatrix}_{i,:} \begin{bmatrix} \Phi_x \\ \Phi_u \end{bmatrix} W_w \right\|_{\ell_2} \leq \frac{1}{Q(\rho)} \ (\forall i) \}, \end{aligned}$$

with  $\tilde{G}_u = \text{blkdiag}(G_{u^1}, \dots, G_{u^{N_P}})$ . Since  $\Phi_x = (\Phi_{x,n})_{n=1}^N$  and  $\Phi_u^p = (\Phi_{u,n}^p)_{n=0}^N$  are FIR, these signals can be represented as (concatenated) matrices. The vGFNE-seeking problem can thus be formulated as the monotone inclusion problem

$$\begin{aligned} \text{Find } \Phi_u^* \in \mathbb{R}^{(N+1)N_u \times (N_x + N_y)} \\ \text{such that } 0 \in F_{\Phi}(\Phi_u^*) + N_{\mathcal{U}_{\Phi, \mathcal{G}}}(\Phi_u^*), \end{aligned} \quad (26)$$

given the pseudo-gradient  $F_{\Phi}(\Phi_u) = (\nabla_{\Phi_u^p} J_{\Phi}^p(\Phi_u^p, \Phi_u^{-p}))_{p \in \mathcal{P}}$ . Note that  $F_{\Phi}$  is maximal monotone because its components,  $\nabla_{\Phi_u^p} J_{\Phi}^p$  ( $\forall p \in \mathcal{P}$ ), are the gradients of closed convex proper functions [30]. The Problem (26) is a special instance of the monotone inclusion problem presented in Section II, and thus it can be solved via FBS. The resulting routine (Algorithm 2) will be referred to as the *vGFNE-Seeking via FBS* method<sup>1</sup>.

The distinct feature in this vGNE-seeking routine is the policy update (line 4): In practice, it corresponds to plugging  $(\Phi_{xx|k}, \Phi_{xy|k}, \Phi_{ux|k}^p, \Phi_{uy|k}^p)$  on the policy's implementation from Corollary 2.1. Importantly, Algorithm 2 does not require knowledge of the state  $x$  or input  $u^p$  signals. A direct consequence is that the players can perform this policy learning routine while simultaneously operating the underlying networked system. Specifically, the players can operate according to the policies  $K_k = (K_k^1, \dots, K_k^{N_P})$  until the coordinator processes their proposals  $\Phi_{u|k+1/2} = (\Phi_{u|k+1/2}^p)_{p \in \mathcal{P}}$  and then returns the responses  $\Phi_{u|k+1} = (\Phi_{u|k+1}^p)_{p \in \mathcal{P}}$  which they can use to update their policies. This enables vGFNE-seeking in games that cannot be interrupted for players to redesign their policies.

<sup>1</sup>The update-index  $k \in \mathbb{N}$  should not be mistaken with time-indices  $n \in \mathbb{N}$ . In particular,  $\Phi_{n|k}$  is the  $n$ -th factor of a kernel  $\Phi \in \ell_2(\mathbb{N})$  after  $k$  updates.

We finish the section with discussions on the convergence and computational properties of Algorithm 2.

*Convergence properties:* From Lemma 1, the algorithm converges to a unique solution if the learning rate satisfies  $\eta \in (0, 2M_{F_\Phi}/L_{F_\Phi}^2)$ , with  $L_{F_\Phi}$  and  $M_{F_\Phi}$  being the Lipschitz and strong-monotonicity constants of  $F_\Phi$ . Given the quadratic objectives  $J_\Phi^p$  ( $p \in \mathcal{P}$ ) and linear dynamics  $\Sigma_\Phi$  (Eq. 18), each player's gradient can be shown to have the affine form

$$\nabla_{\Phi_u^p} J_\Phi^p(\Phi_u) = 2 \sum_{\tilde{p} \in \mathcal{P}} (H_\Phi^{p\tilde{p}} H_\Phi^{p\tilde{p}}) \Phi_u^{\tilde{p}} (W_w W_w^\top) + h^p \quad (27)$$

given the block-triangular matrices  $H_\Phi^{p\tilde{p}} = [H_{\Phi,n,n'}^{p\tilde{p}}]_{0 \leq n, n' \leq N}$ ,

$$H_{\Phi,n,n'}^{p\tilde{p}} = \begin{cases} W_u^p & \text{if } n = n' \text{ and } p = \tilde{p} \\ W_x^p A^{(n-1)-n'} B_u^{\tilde{p}} & \text{if } n > n' \\ 0 & \text{otherwise} \end{cases} \quad (28)$$

and the appropriate constant matrix  $h^p \in \mathbb{R}^{N N_u \times (N_x + N_y)}$ . As such, also the pseudo-gradient  $F_\Phi(\Phi_u)$  must be an affine operator. We can then establish the following result:

**Theorem 3.** *The system-level pseudo-gradient  $F_\Phi$  is  $M_{F_\Phi}$ -strongly-monotone and  $L_{F_\Phi}$ -Lipschitz with constants*

$$M_{F_\Phi} = \lambda_{\min}(D_\Phi^\top H_\Phi + H_\Phi^\top D_\Phi) \sigma_{\min}^2(W_w), \quad (29a)$$

$$L_{F_\Phi} = 2\sigma_{\max}(D_\Phi^\top H_\Phi) \sigma_{\max}^2(W_w). \quad (29b)$$

given  $D_\Phi = \text{blkdiag}(H_\Phi^{pp})_{p \in \mathcal{P}}$  and  $H_\Phi = [H_\Phi^{p\tilde{p}}]_{p, \tilde{p} \in \mathcal{P}}$ .

*Proof.* See the Appendix.  $\square$

Interestingly, this implies that the pseudo-gradient  $F_\Phi$  is ill-conditioned if so is the noise-filtering matrix  $W_w = (B_w, D_w)$  and that convergence rates are not affected by the noise if  $W_w = I$ . Moreover, it suggests that open-loop unstable games (when  $\rho(A) \geq 1$ ) might require a careful tuning of  $W_x^p$  ( $\forall p$ ) to avoid the terms  $W_x^p A^n B_u^{\tilde{p}}$  ( $n = 1, \dots, N-1$ ) in Eq. (28) from exploding, then affecting the condition number of  $F_\Phi$ .

Applying Lemma 1 and Theorem 3, the following holds:

**Corollary 3.1.** *If  $\eta \in (0, 2M_{F_\Phi}/L_{F_\Phi}^2)$ , with  $(M_{F_\Phi}, L_{F_\Phi})$  as in Eq. (29), then Algorithm 2 converges linearly to the unique vGFNE  $\hat{K}^* = (\hat{\Phi}_{uy}^{p*} - \hat{\Phi}_{ux}^{p*} \hat{\Phi}_{xx}^{*-1} \hat{\Phi}_{xy}^*)_{p \in \mathcal{P}}$ , parametrized by the vGNE  $\Phi_u^* \in \mathbf{VI}(BR_\Phi)$ , with rate*

$$\lim_{k \rightarrow \infty} \frac{\|\Phi_{u|k+1} - \Phi_u^*\|_F}{\|\Phi_{u|k} - \Phi_u^*\|_F} = \sqrt{1 - \eta(2M_{F_\Phi} - \eta L_{F_\Phi}^2)} \quad (30)$$

from any initial  $\Phi_{u|0} \in \mathbb{R}^{(N+1)N_u \times (N_x + N_y)}$ .

In summary, whenever the (system-level) pseudo-gradient  $F_\Phi$  has a condition number  $\kappa_{F_\Phi} = L_{F_\Phi}/M_{F_\Phi} \in (0, \infty)$ , then the vGFNE-Seeking via FBS method converges linearly to the unique solution to the monotone inclusion Problem (26).

*Computation:* Numerically, the pseudo-gradient operator  $F_\Phi$  can be evaluated through automatic differentiation and the projection operator  $\text{proj}_{\mathcal{U}_{\Phi, \mathcal{G}}}$  by directly solving the associated optimization problem. However, the structure of this vGFNE

problem allows for more efficient numerics. In the case of  $F_\Phi = (\nabla_{\Phi_u^p} J_\Phi^p)_{p \in \mathcal{P}}$ , each component can be computed as

$$\nabla_{\Phi_u^p} J_\Phi^p(\cdot) = \left( 2(W_u^{p\top} W_u^p) \Phi_{u,n|k}^p (W_w W_w^\top) + 2(B_u^p)^\top \Delta_{x,n|k}^p (W_w W_w^\top) \right)_{n=0}^N, \quad (31)$$

with the sensitivities  $\Delta_{x|k}^p = (\Delta_{x,n|k}^p)_{n=0}^N$  obtained by first forward-propagating the  $\Phi_{x|k}$  responses,

$$\Phi_{x,n+1|k} = A \Phi_{x,n|k} + \sum_{p \in \mathcal{P}} B_u^p \Phi_{u,n|k}^p,$$

from  $\Phi_{x,0|k} = [I_{N_x} \ 0]$ , then by backward-propagating

$$\Delta_{x,n-1|k}^p = A^\top \Delta_{x,n|k}^p + (W_x^{p\top} W_x^p) \Phi_{x,n|k}$$

from  $\Delta_{x,N|k} = 0$ . These operations are not demanding and thus players are not required large computational resources to participate in this vGFNE-seeking routine. Moreover, the gradient  $\nabla_{\Phi_u^p} J_\Phi^p(\cdot)$  depends on other players'  $\Phi_{u|k}^{\tilde{p}}$  only through  $\Phi_{x|k}$ ; when this response is already available (e.g., provided by the coordinator), then each  $p$ -th player is capable of computing its proposed update  $\Phi_{u|k+1/2}^p$  using only private information.

In general, the projection map

$$\text{proj}_{\mathcal{U}_{\Phi, \mathcal{G}}}(\cdot) = \arg \min_{\Phi_u} \{ \|\Phi_u - \Phi_{u|k+1/2}\|_{\ell_2} \mid \Phi_u \in \mathcal{U}_{\Phi, \mathcal{G}} \} \quad (32)$$

is a sparse convex optimization that can be solved efficiently. However, the problem has  $(N+1)N_u(N_x + N_y)$  decision variables and  $NN_x(N_x + N_y)$  equality constraints, and thus scales poorly with the state-space dimensions  $(N_x, N_u, N_y)$  and FIR horizon  $N$ . Its computational burden can be alleviated by exploiting the structure of Problem (32) and using a distributed approach for its solution. Consider the problem

$$\underset{\Psi, \Lambda}{\text{minimize}} \quad J_\Phi^{(r)}(\Psi) + J_\Phi^{(c)}(\Lambda) \quad \text{subject to} \quad \Psi = \Lambda \quad (33)$$

defined by the extended-real-value functions

$$J_\Phi^{(r)}(\Psi) = \begin{cases} \frac{1}{2} \|\Psi - \Phi_{u,k+1/2}\|_{\ell_2}^2 & \text{if Eq. (19, 22) holds} \\ \infty & \text{otherwise} \end{cases}$$

$$J_\Phi^{(c)}(\Lambda) = \begin{cases} \frac{1}{2} \|\Lambda - \Phi_{u,k+1/2}\|_{\ell_2}^2 & \text{if Eq. (18, 22, 25) holds} \\ \infty & \text{otherwise} \end{cases}$$

The Problem (33) can be solved by the Alternating Direction Method of Multipliers (ADMM, [32]), and either  $\Psi^*$  or  $\Lambda^*$  used as the projection  $\Phi_{u|k+1}$ . The subproblem associated with  $J_\Phi^{(r)}$  considers only the dual dynamic constraints  $\Sigma_\Phi^*$  (Eq. 19) and the structural constraints from Eq. (22). The subproblem associated with  $J_\Phi^{(c)}$  considers the primal dynamic constraints  $\Sigma_\Phi$  (Eq. 18), the structural constraints from Eq. (22), and the operational constraints from Eq. (25). For many cases (e.g., diagonal weighting matrices and bound constraints), these subproblems are, respectively, row-wise and column-wise separable and can be reduced into smaller problems to be solved in parallel. This can lead to a substantial performance improvement, making Algorithm 2 scalable to large-scale problems. We refer to [38] for more details on partially separable SLS problems and their solution via ADMM.

#### IV. ILLUSTRATIVE APPLICATION: STABILIZATION OF A PARTIALLY OBSERVED POWER GRID

In the following, we study the efficacy and performance of the proposed vGFNE-Seeking via FBS (Algorithm 2) in a simulated application inspired by the decentralized control of power networks. The example is adapted from [38].

We consider a  $3 \times 3$  grid network of interconnected subsystems (Figure 2). Each  $p$ -th node represents a power system operated by a self-interested agent aiming to stabilize its *phase angle deviation*  $\theta^p$  and *frequency deviation*  $\dot{\theta}^p$  against disturbances. For this purpose, each agent has actuators capable of applying a load directly to  $\dot{\theta}^p$ . However, for budgetary reasons, each subsystem is equipped only with a phase measurement unit producing noisy readings of  $\theta$ . The measurements are publicly available through some communication network (which is assumed to have the same topology as in Figure 2).

Each  $p$ -th subsystem has the continuous-time dynamics

$$m^p \ddot{\theta}^p + d^p \dot{\theta}^p = - \sum_{\tilde{p} \in \mathcal{P}} \kappa^{p,\tilde{p}} (\theta^p - \theta^{\tilde{p}}) + u^p + \delta_x^p, \quad (34)$$

where  $u^p$  and  $\delta_x^p$  represent the controllable load and the external disturbances, respectively. The fixed parameters are sampled from  $m^p \sim \text{Uniform}(0.5, 1)$ ,  $d^p \sim \text{Uniform}(1, 1.5)$ , and  $\kappa^{p,\tilde{p}} = \kappa^{\tilde{p},p} \sim \text{Uniform}(0.5, 1)$  with  $\kappa^{p,\tilde{p}} = 0$  if the link  $j \rightarrow i$  is not present in the topology shown in Figure 2. The phase measurements of each  $p$ -th subsystem are represented by the signal  $y^p = \theta^p + \delta_y^p$ , with  $\delta_y^p$  being the measurement noise. The disturbances  $\delta_x^p$  and the measurement noise  $\delta_y^p$  are modelled as additive white Gaussian noise (AWGN) with standard deviations of 1 and 0.1, respectively.

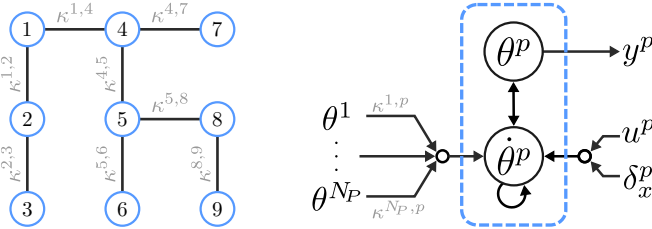


Fig. 2. Power grid: Schematic of the interconnected network (left) and the interactions within each  $p$ -th subsystem (right).

The agents seek to stabilize their associated subsystems while subject to operational and structural limitations: The bus between any two subsystems  $p, \tilde{p} \in \mathcal{P}$  is required to satisfy

$$-5 \leq \kappa^{p,\tilde{p}} (\theta^p(\tau) - \theta^{\tilde{p}}(\tau)) \leq 5, \quad \forall \tau \in \mathbb{R}, \quad (35)$$

to avoid overloading the corresponding channels. Moreover, the actuation and communication infrastructure is such that both the deployment of controls and the transmission of information occur with a time delay of 0.1 [units-of-time].

##### A. Game formulation and the vGFNE-seeking problem

This problem is formulated as a stochastic dynamic game

$$\mathcal{G}_{\infty}^{\text{LQ}} := (\mathcal{P}, \{U^p\}_{p \in \mathcal{P}}, \{J^p\}_{p \in \mathcal{P}}),$$

with players  $\mathcal{P} = \{1, \dots, 9\}$  and discrete-time dynamics

$$\Sigma_{\mathcal{G}} : \begin{cases} x_{t+1} = Ax_t + B_w w_t + \sum_{p \in \mathcal{P}} B_u^p u_t^p, & x_0 = 0; \\ y_t = Cx_t + D_w w_t. \end{cases}$$

defined by the block matrix  $A = [A_{ij}]_{1 \leq i, j \leq N_P}$ ,

$$A_{ij} = \begin{cases} \begin{bmatrix} 1 & 0.1 \\ -\frac{0.1 \sum_j k^{ij}}{m^i} & 1 - \frac{0.1 d^i}{m^i} \end{bmatrix} & \text{if } i = j \\ \begin{bmatrix} 0 & 0 \\ \frac{0.1 k^{ij}}{m^i} & 0 \end{bmatrix} & \text{otherwise} \end{cases}$$

and  $B_u^p = e_p \otimes [0 \quad \frac{0.1}{m^p}]^T$  ( $\forall p \in \mathcal{P}$ ),  $C = I_{N_P} \otimes [1 \quad 0]$ , and

$$(B_w, D_w) = \text{blkdiag}\left(I_{N_P} \otimes \begin{bmatrix} 0.01 & \\ & \frac{0.1}{m^p} \end{bmatrix}, 0.1 I_{N_y}\right).$$

This model is obtained by defining  $x = (\theta^p, \dot{\theta}^p)_{p \in \mathcal{P}}$ , then performing an Euler discretization to the state-space obtained from Eq. (34) using an interval of  $\Delta\tau = 0.1$  units-of-time. The global measurement signal is defined as  $y = (y^p)_{p \in \mathcal{P}}$ . To ensure Assumption 3(b), we also include a small artificial disturbance to the phase angle deviation  $\theta^p$ . In our experiment,  $\rho(A) = 1$  and thus the system is not stable. It is, however, both controllable and observable. The task of stabilizing each individual subsystem is encoded through the functional

$$J^p(u^p, u^{-p}) := \mathbb{E} \left[ \sum_{t=0}^{\infty} \left\| \begin{bmatrix} W_x^p & W_u^p \end{bmatrix} \begin{bmatrix} x_t \\ u_t^p \end{bmatrix} \right\|_2^2 \right],$$

with  $[W_x^p \ W_u^p] = \text{blkdiag}(e_p e_p^T \otimes 0.01 I_{N_x}, I_{N_u^p})$ . Finally, each  $p$ -th player's policy must be  $(d_a, d_c)$ -delayed with  $d_a = d_c = 1$  and it must satisfy the operational constraints

$$U^p(u^{-p}) = \{u^p \in \ell_{\infty}^{N_u}(\mathbb{N}) \mid G_x x_t \preceq 1, \ t \in \mathbb{N}\}, \quad (36)$$

with matrix  $G_x$  encoding the capacity limits from Eq. (35).

We simulate an execution of the game  $\mathcal{G}_{\infty}^{\text{LQ}}$  in which the players are seeking an equilibrium policy by adhering to the vGFNE-seeking routine in Algorithm 2. In this scenario, each player's policy is represented by a system level parametrization defined by FIR responses with  $N = 16$  spectral components. The responses are constrained to have the sparsity patterns defined in Eq. (23) and to satisfy a robust realization of the operational constraints with probability  $\rho = 0.975$  (Eq. 25). The policies are updated every  $\Delta k = 32$  time steps (that is,  $k = \lfloor t/\Delta k \rfloor$ ) with rate  $\eta = \frac{M_{F\Phi}}{L_{F\Phi}^2} = \frac{0.0002}{0.0004} = 0.5$ , computed and provided by the central coordinator to ensure convergence. We set  $\Phi_{u|0} = \text{proj}_{\mathcal{U}_{\Phi}, \mathcal{G}}(0)$  as the initial profile.

##### B. Simulation results and discussion

Due to numerical limitations, we interrupt the updates at the  $k_f$  first satisfying  $\|\Phi_{u|k_f+1} - \Phi_{u|k_f}\|_{\ell_2} / \|\Phi_{u|k_f}\|_{\ell_2} \leq 10^{-15}$ , when the policy updates become numerically negligible, and assume that  $\Phi_{u|k_f} \approx \Phi_u^* \in \mathbf{VI}(BR_{\Phi})$ . The convergence of the vGFNE-seeking algorithm to this fixed-point is shown in Figure 3. The predicted convergence based on the rate  $L_{(I-\eta F_{\Phi})} = \sqrt{1 - \eta(2M_F - \eta L_F^2)} \approx 0.99995$  is also displayed. The results show that, when adhering to Algorithm 2, players converge to a vGFNE solution to  $\mathcal{G}_{\infty}^{\text{LQ}}$  within  $k_f = 10^4$  policy updates (or  $t_f = 10^4 \times \Delta k$  time steps). Interestingly, despite slow, the actual convergence of our vGFNE-seeking routine is still faster than that predicted by Corollary 3.1 by a factor of almost 100. We remark that this does not disqualify

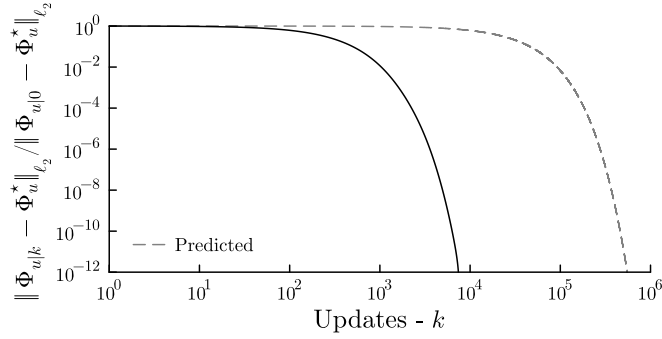


Fig. 3. Power grid: Convergence of the vGFNE-seeking routine.

the usefulness of this corollary, as its primary purpose is to certificate that  $\mathcal{G}_\infty^{\text{LQ}}$  has an unique solution to which Algorithm 2 converges. In principle, once convergence is ensured, a potential acceleration can be achieved by incorporating inertia into Algorithm 2 [39]. In this experiment, the seemingly slow convergence of the vGFNE-seeking routine is still not a practical issue: If the power grid is operated at a timescale of seconds, it would take  $(t_f \times \Delta\tau) = 32000$  seconds, or 8 hours, for players to achieve an equilibrium strategy. This is a relatively short period considering that such infrastructure is expected to be operational for years. Finally, we again emphasize that Algorithm 2 allows the policy updates to occur alongside (and independently of) the actual operation of the system; the players can continuously improve their policies without interrupting or destabilizing the grid.

In Figure 4, we show the first hour of operation of the power grid while the policies are being updated (that is, for the first  $t \leq 3600$  time steps or  $k \leq \lfloor 3600/\Delta k \rfloor = 112$  policy updates). We compare the closed-loop evolution with that obtained by an open-loop operation with  $u_t^p = 0$  for all players  $p \in \mathcal{P}$ . The results show that the policies  $\hat{K}_k^p := \hat{\Phi}_{uy|k}^p - \hat{\Phi}_{ux|k}^p \hat{\Phi}_{xx|k}^{-1} \hat{\Phi}_{xy|k}^p$  ( $\forall p \in \mathcal{P}$ ) stabilize the network against the external disturbances, as observed through the noisy measurements of  $\{\theta^p\}_{p \in \mathcal{P}}$ . Moreover, the closed-loop system is not destabilized by the implementation of a new control policy every  $\Delta k = 32$  time steps. The policies are also shown to enforce the robust realization of the operational constraints: The buses  $\kappa^{p,\bar{p}}(\theta_t^p - \theta_t^{\bar{p}})$ , between every  $p, \bar{p} \in \mathcal{P}$ , are simultaneously within the safety limits for approximately 97% of this simulation period. Under open-loop operation, the safety limits are violated for the majority of the simulation period. This performance is obtained at the expense of relatively large frequency loads  $\{u^p\}_{p \in \mathcal{P}}$ . Given the tuning  $\{W_x^p, W_u^p\}_{p \in \mathcal{P}}$ , this is by design the optimal actions that the players must deploy while seeking an equilibrium policy.

Finally, the output-feedback policies  $K_k^p$  ( $\forall p \in \mathcal{P}$ ) are also shown to implement a communication structure based on the topology in Figure 2. To demonstrate this property, we simulate the closed-loop responses from the vGFNE policy,  $K_{k_f} = (K_{k_f}^p)_{p \in \mathcal{P}}$ , when the system is initially at rest (i.e.,  $x_0 = 0$ ) and is then subjected to an impulse affecting the phase angle deviation of the central node (i.e.,  $B_w w_t = \delta_t e_9$ ). The results (Figure 5) show that the flow of information follows the network topology depicted in Figure 2: At  $t = 1$ , the

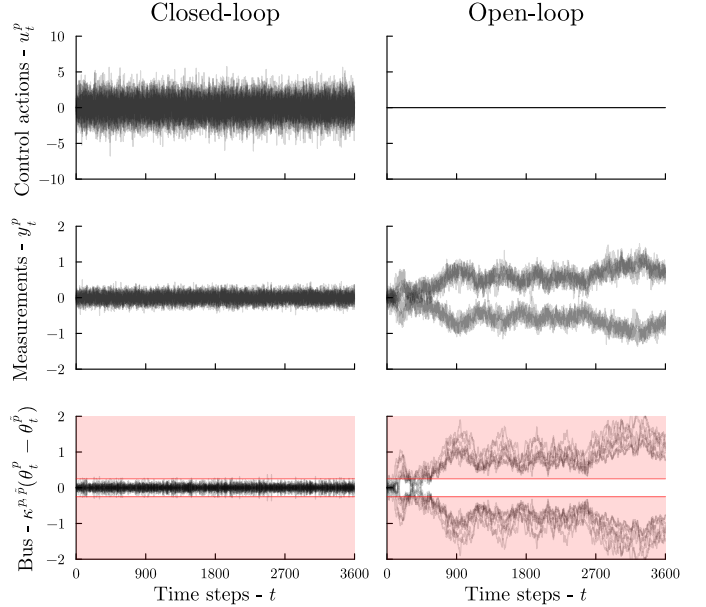


Fig. 4. Power grid: Closed-loop (left panels) and open-loop (right panels) responses in terms of frequency loads  $u^p$ , phase angle measurements  $y_t^p$ , and pairwise buses  $\kappa^{p,\bar{p}}(\theta_t^p - \theta_t^{\bar{p}})$ , for all  $p, \bar{p} \in \mathcal{P}$ . The signals from each subsystem are plotted superimposed. The shaded region indicates the unsafe operation region of the network channels.

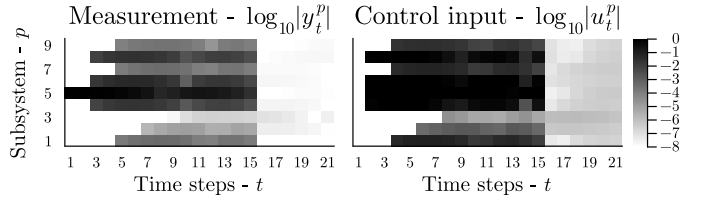


Fig. 5. Power grid: Closed-loop impulse response in terms of measurements  $y^p$  and control inputs  $u^p$ , for each individual subsystem  $p \in \mathcal{P}$ .

impulse is perceived at the 5-th subsystem and the observation  $y_1^5$  is propagated to its neighbors  $p \in \{4, 6, 8\}$ . The players  $p \in \{5, 4, 6, 8\}$  react to this disturbance after an action delay of  $d_a = 1$  time steps. At  $t = 3$ , its effect is perceived at the subsystems  $p \in \{4, 6, 8\}$  and measurements  $(y_1^p, y_2^p, y_3^p)_{p \in \{4, 6, 8\}}$  are transmitted to the two-hop neighbors  $p \in \{1, 7, 9\}$ . This information pattern proceeds until all the players  $p \in \mathcal{P}$  have been affected and are engaged in attenuating the disturbance. The magnitude of the applied control actions is shown to be higher the closer the subsystems are to the central node  $p = 5$ . Additionally, the policies are observed to implement *deadbeat* control: The state is stirred to zero instantaneously after  $N = 16$  control actions. This behavior (known as *time-localization* in the SLS literature [27]) explains the aforementioned large control actions, as deadbeat controllers are notoriously aggressive for small  $N$  [40]. Interestingly, this property also implies that disturbances can be attenuated before affecting subsystems that are more than  $N$ -hops distant from where they enter the network.

## V. CONCLUDING REMARKS

This work presented a GFNE-seeking algorithm for non-cooperative games with dynamics that are linear, stochastic,

potentially unstable, and partially observed. We first considered the equivalent representation of each player's policies as their corresponding closed-loop responses to the noise (given others' policies), then we designed a vGFNE-seeking routine based on well-established operator splitting algorithms. Under this system level parametrization, the GFNE-seeking algorithm is applicable to problems in which the policies are required to be stabilizing and to satisfy operational (on the state and input signals) and structural (on the policy itself) constraints. Using results from operator theory, we derived conditions for the existence and uniqueness of a vGFNE solution and established convergence certificates for our algorithm. After its main aspects are presented, the algorithm is demonstrated on the task of stabilizing a partially observed and decentralized power grid. The results demonstrate its efficacy: When adhering to this routine, players approach a GFNE solution while still complying with operational and communication constraints.

#### APPENDIX PROOF OF THEOREM 3

From Eq. (27), the pseudo-gradient has the compact form

$$F_\Phi(\Phi_u) = 2(D_\Phi^\top H_\Phi)\Phi_u(W_w W_w^\top) + h$$

given matrices  $D_\Phi = \text{blkdiag}(H_\Phi^{pp})_{p \in \mathcal{P}}$ ,  $H_\Phi = [H_\Phi^{pp}]_{p, \tilde{p} \in \mathcal{P}}$ , and  $h = (h^p)_{p \in \mathcal{P}}$ . In vectorized form, this operator becomes  $\text{vec}(F_\Phi(\Phi_u)) = 2((W_w W_w^\top) \otimes (D_\Phi^\top H_\Phi))\text{vec}(\Phi_u) + \text{vec}(g)$ . Since the strong-monotonicity constant of an affine operator  $F(x) = Ax + b$  is  $(1/2)\lambda_{\min}(A + A^\top)$  [32], we have that

$$\begin{aligned} M_{F_\Phi} &= \lambda_{\min}((W_w W_w^\top) \otimes D_\Phi^\top H_\Phi + ((W_w W_w^\top) \otimes D_\Phi^\top H_\Phi)^\top) \\ &= \lambda_{\min}((W_w W_w^\top) \otimes (D_\Phi^\top H_\Phi + H_\Phi^\top D_\Phi)) \\ &= \lambda_{\min}(D_\Phi^\top H_\Phi + H_\Phi^\top D_\Phi) \sigma_{\min}^2(W_w) \end{aligned}$$

where we used the distributivity of the Kronecker product, the fact that  $\lambda_{\min}(A \otimes B) = \lambda_{\min}(B)\lambda_{\min}(A)$ , and the property  $\lambda_{\min}(AA^\top) = \sigma_{\min}^2(A)$ , for any matrices  $A$  and  $B$  [41]. The tightest Lipschitz constant of  $F_\Phi$  is obtained from the spectral norm  $\|2(W_w W_w^\top) \otimes (D_\Phi^\top H_\Phi)\|_2$  [32], or, equivalently,

$$\begin{aligned} L_{F_\Phi} &= 2\sigma_{\max}((W_w W_w^\top) \otimes (D_\Phi^\top H_\Phi)) \\ &= 2\sigma_{\max}(D_\Phi^\top H_\Phi) \sigma_{\max}^2(W_w), \end{aligned}$$

using the fact that  $\lambda_{\max}(A \otimes B) = \lambda_{\max}(B)\lambda_{\max}(A)$  and that  $\lambda_{\max}(AA^\top) = \sigma_{\max}^2(A)$  for any matrices  $A$  and  $B$ .

#### REFERENCES

- [1] T. Başar and G. J. Olsder, *Dynamic noncooperative game theory*. SIAM, 1998.
- [2] L. Chen, N. Li, S. H. Low, and J. C. Doyle, "Two market models for demand response in power networks," in *2010 First IEEE International Conference on Smart Grid Communications*, pp. 397–402, 2010.
- [3] A.-H. Mohsenian-Rad, V. W. S. Wong, J. Jatskevich, R. Schober, and A. Leon-Garcia, "Autonomous demand-side management based on game-theoretic energy consumption scheduling for the future smart grid," *IEEE Transactions on Smart Grid*, vol. 1, no. 3, pp. 320–331, 2010.
- [4] 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.
- [5] L. Pavel, *Game Theory for Control of Optical Networks*. Birkhäuser, 2012.
- [6] J. F. Fisac, A. K. Akametalu, M. N. Zeilinger, S. Kaynama, J. Gillula, and C. J. Tomlin, "A general safety framework for learning-based control in uncertain robotic systems," *IEEE Transactions on Automatic Control*, vol. 64, no. 7, pp. 2737–2752, 2019.
- [7] 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.
- [8] M. Wang, Z. Wang, J. Talbot, J. C. Gerdes, and M. Schwager, "Game-theoretic planning for self-driving cars in multivehicle competitive scenarios," *IEEE Transactions on Robotics*, vol. 37, no. 4, pp. 1313–1325, 2021.
- [9] F. Facchinei and C. Kanzow, "Generalized Nash equilibrium problems," *Annals of Operations Res.*, vol. 175, no. 1, pp. 177–211, 2010.
- [10] 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.
- [11] 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.
- [12] 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.
- [13] 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.
- [14] A. Monti, B. Nortmann, T. Mylvaganam, and M. Sassano, "Feedback and open-loop Nash equilibria for LQ infinite-horizon discrete-time dynamic games," *SIAM Journal on Control and Optimization*, vol. 62, no. 3, pp. 1417–1436, 2024.
- [15] S. Xue, W. Zhang, B. Luo, and D. Liu, "Integral reinforcement learning-based dynamic event-triggered nonzero-sum games of USVs," *IEEE Transactions on Cybernetics*, vol. 55, no. 4, pp. 1706–1716, 2025.
- [16] 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.
- [17] 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.
- [18] 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*, pp. 1475–1481, IEEE, 2020.
- [19] 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*, pp. 2216–2222, IEEE, 2020.
- [20] S. Hall, G. Belgioioso, F. Dörfler, and D. Liao-McPherson, "Stability certificates for receding horizon games," *IEEE Transactions on Automatic Control*, pp. 1–8, 2025.
- [21] P. Yi and L. Pavel, "An operator splitting approach for distributed generalized Nash equilibria computation," *Automatica*, vol. 102, pp. 111–121, 2019.
- [22] K. Lu, G. Jing, and L. Wang, "Distributed algorithms for searching generalized Nash equilibrium of noncooperative games," *IEEE Transactions on Cybernetics*, vol. 49, no. 6, pp. 2362–2371, 2019.
- [23] G. Belgioioso, P. Yi, S. Grammatico, and L. Pavel, "Distributed generalized Nash equilibrium seeking: An operator-theoretic perspective," *IEEE Control Systems Magazine*, vol. 42, no. 4, pp. 87–102, 2022.
- [24] K. Lu and Q. Zhu, "Nonsmooth continuous-time distributed algorithms for seeking generalized Nash equilibria of noncooperative games via digraphs," *IEEE Transactions on Cybernetics*, vol. 52, no. 7, pp. 6196–6206, 2022.
- [25] G. Belgioioso, D. Liao-McPherson, M. H. de Badyn, S. Bolognani, R. S. Smith, J. Lygeros, and F. Dörfler, "Online feedback equilibrium seeking," *IEEE Transactions on Automatic Control*, vol. 70, no. 1, pp. 203–218, 2025.
- [26] T. Başar, "Stochastic differential games and intricacy of information structures," in *Dynamic Games in Economics*, pp. 23–49, Springer, 2014.
- [27] J. Anderson, J. C. Doyle, S. H. Low, and N. Matni, "System level synthesis," *Annual Reviews in Control*, vol. 47, pp. 364–393, 2019.
- [28] O. B. L. Neto, M. Mulas, and F. Corona, "SLS-BRD: A system-level approach to seeking generalized feedback Nash equilibria," *IEEE Transactions on Automatic Control*, vol. 70, no. 11, pp. 7173–7188, 2025.
- [29] J. B. Conway, *A course in functional analysis*. Springer, 2 ed., 2007.

- [30] H. H. Bauschke and P. L. Combettes, *Convex analysis and monotone operator theory in Hilbert spaces*. Springer, 2 ed., 2017.
- [31] A. A. Kulkarni and U. V. Shanbhag, “On the variational equilibrium as a refinement of the generalized Nash equilibrium,” *Automatica*, vol. 48, no. 1, pp. 45–55, 2012.
- [32] E. K. Ryu and W. Yin, *Large-scale convex optimization: algorithms & analyses via monotone operators*. Cambridge University Press, 2022.
- [33] K. Zhou and J. C. Doyle, *Essentials of robust control*. Prentice-Hall, 1998.
- [34] S. Dean, H. Mania, N. Matni, B. Recht, and S. Tu, “On the sample complexity of the linear quadratic regulator,” *Foundations of Computational Mathematics*, vol. 20, no. 4, pp. 633–679, 2020.
- [35] Y.-S. Wang, *A System Level Approach to Optimal Controller Design for Large-Scale Distributed Systems*. Dissertation (Ph.D.), California Institute of Technology, 2017.
- [36] A. Nayyar and T. Başar, “Dynamic stochastic games with asymmetric information,” in *51st IEEE Conference on Decision and Control*, pp. 7145–7150, IEEE, 2012.
- [37] S. P. Boyd and L. Vandenberghe, *Convex optimization*. Cambridge University Press, 2004.
- [38] Y.-S. Wang, N. Matni, and J. C. Doyle, “Separable and localized system-level synthesis for large-scale systems,” *IEEE Transactions on Automatic Control*, vol. 63, no. 12, pp. 4234–4249, 2018.
- [39] F. Iutzeler and J. M. Hendrickx, “A generic online acceleration scheme for optimization algorithms via relaxation and inertia,” *Optimization Methods and Software*, vol. 34, no. 2, pp. 383–405, 2017.
- [40] K. J. Åström and B. Wittenmark, *Computer-controlled systems: theory and design*. Prentice-Hall, 1997.
- [41] R. A. Horn and C. R. Johnson, *Topics in Matrix Analysis*. Cambridge University Press, 1991.