Learning Nash Equilibria in Large Populations With Constrained Strategy Switching

Semih Kara, *Student Member, IEEE*, and Nuno C. Martins, *Senior Member, IEEE*

***Abstract*—We consider a large population of learning agents that interact noncooperatively by selecting strategies from a common set. Each strategy has a payoff, assigned by a strictly concave potential game. The agents repeatedly revise their strategies according to a learning rule that models how they seek alternatives with higher payoffs. Our objective is to determine when the population learns the Nash equilibrium of the game, meaning its strategy profile asymptotically converges to this equilibrium. Unlike previous work assuming unrestricted strategy switching, here we tackle the case where only certain strategies are accessible from certain others, characterized by a strategy graph that is connected but possibly incomplete. Through Lyapunov's method, we prove that modifications based on KL-divergence to either the payoffs or the learning rules ensure the strategy profile's near-global convergence to the Nash equilibrium. We highlight the practical significance of our findings and provide a numerical validation.**

***Index Terms*—Game theory, graphs, nonlinear systems.**

I. Introduction

We consider a large population of agents that interact by selecting strategies from a common set $V := \{1, \dots, n\}$. The agents are *nondescript* and are characterized by a continuum of unit mass, implying that their distribution x on V represents the population's strategy profile. The strategy profile set is $\Delta := \{x \in [0,1]^n \mid \sum_{i=1}^n x_i = 1\}$ and $x_i(t)$ is the proportion of the population selecting strategy i at time t. A memoryless map $\mathcal{F} : \mathbb{R}^n \to \mathbb{R}^n$ specifies a payoff mechanism, which we call the *game*, that acts to generate a payoff vector as $\mathcal{F} : x(t) \mapsto p(t)$. The payoff for strategy i at time t is $p_i(t)$. Each agent follows one strategy at a time, which the agent can revise and subsequently change—causing $x(t)$ to vary over time. The revision mechanism is modelled by a *learning rule* $\rho : \Delta \times \mathbb{R}^n \to \mathbb{R}^{n\times n}_{\geq 0}$ that quantifies the population's preferences in terms of the *rates of switching among strategies*, leading to the following dynamics for $t \geq 0$ and $i \in V$:

$$\dot{x}_i(t) = \underbrace{\sum_{j\in\mathcal{I}_i} x_j(t)\rho_{ji}(x(t),p(t))}_{\text{Inflow to } i} - \underbrace{\sum_{j\in\mathcal{O}_i} x_i(t)\rho_{ij}(x(t),p(t))}_{\text{Outflow from } i}. \quad (1)$$

Here, $\mathcal{O}_i \subseteq V$ is the set of strategies available to the agents following i, and $\mathcal{I}_i := \{j \in V \mid i \in \mathcal{O}_j\}$ is the set of strategies from which the agents can switch to i. Using the edge set $E := \{(i,j) \in V \times V \mid j \in \mathcal{O}_i\}$ we define the *strategy graph* $G := (V, E)$.

A. From Finite Number of Agents to the Mean Field Limit

We adopt the population game and evolutionary dynamic approach [1], which derives (1) as the deterministic mean field approximation of a realistic stochastic model having a large but finite number of agents. This finite-agent model accounts for asynchronous strategy updates, and the learning rules (including those we will propose) are often easy to implement and do not require knowledge of $\mathcal{F}$. The analysis in [2], [3] establishes comparisons with error bounds between the solutions of (1) and realizations of the finite-agent stochastic model. These comparisons justify the practical relevance of our framework, facilitating its applications in fields such as traffic management [4], electricity demand regulation [5], [6], task allocation [7], [8], distributed extremum seeking [9], [10], and communication networks [11].

In [3], the authors prove that the globally asymptotically stable equilibria of (1) are accurate predictors of the long-term strategic profiles for these *large but finite populations*. Thus, a common theme in population games research is to assume a structure on $\mathcal{F}$ and identify learning rules that make the asymptotically stable equilibria of (1) coincide with the Nash equilibria of $\mathcal{F}$. We share the same goal in this letter.

B. Incomplete Strategy Graphs, Related Work and Gaps

The novelty of our work lies in considering an incomplete strategy graph G. Unlike most studies, which assume that the agents can switch between any two strategies (implying that G is complete), we allow only certain strategies to be accessible from certain others, effectively eliminating the corresponding edges in G. Allowing for an incomplete G makes it possible to impose constraints on the strategic behavior of agents, as required in many applications. Section IV illustrates an application in which the strategies are regions that the agents

Manuscript received 8 March 2024; revised 9 May 2024; accepted 23 May 2024. Date of publication 31 May 2024; date of current version 21 June 2024. This work was supported in part by the Air Force Office of Scientific Research (AFOSR) under Grant FA95502310467, and in part by NSF under Grant ECCS 2139713 and Grant CNS 2135561. Recommended by Senior Editor M. Guay. *(Corresponding author: Semih Kara.)*

The authors are with the ECE Department and the ISR, University of Maryland at College Park, College Park, MD 20742 USA (e-mail: skara@umd.edu; nmartins@umd.edu).

Digital Object Identifier 10.1109/LCSYS.2024.3408102

can occupy, and an incomplete G represents the available paths between the regions that the agents can take. Additional examples include games with spatial/informational constraints [10], or belief formation [12] in which agents can only switch between similar beliefs.

There is no existing work ascertaining when Nash equilibrium learning as discussed in Section I-A is achieved for an incomplete G. The state of the art[1] has considered certain graph structures[2] [8], studied local stability of Nash equilibria [10], [15], or established global stability of sets that may not include Nash equilibria [16], [17]. Specifically, these articles conclude that common learning rules and payoff mechanisms, well known for guaranteeing global asymptotic stability of Nash equilibria for complete G, fail when G is incomplete. Importantly, [16] illustrates this with an example in which x converges to a non-Nash point.

C. Contributions: Guaranteed Learning of Nash Equilibria

Our main contribution is a new class of learning rules that achieve *near-global asymptotic stability of the Nash equilibrium of a strictly concave potential game under any connected and possibly incomplete G*. These rules ensure that x converges to the Nash equilibrium of $\mathcal{F}$ from *any interior initial state*. We also allow for G to be directed.

Our approach merges methods from [17] and [18]. The study by [17] addresses incomplete G, demonstrating that a modification of the well-known Smith learning rule [4] results in the asymptotic stability of a Gibbs measure in any potential game. To stabilize Nash equilibria instead of a Gibbs measure, we employ a different modification, inspired by the Kullback-Leibler (KL) regularization idea in [18]. We apply this regularization to a broad class of learning rules [19] that includes the Smith rule (used in [17]) as a particular case, models a wide range of strategic behaviors, and allows for fully decentralized implementation. We also employ proof techniques from [18]. Nevertheless, noting that [18] assumes complete G and uses a different learning rule (called perturbed best response), we alter these techniques accordingly.

II. The Framework and Problem Description

A. The Model

Our aim is to investigate the equilibrium stability of (1). In this system, we assume that $p = \mathcal{F}(x)$ and $\mathcal{F}$ is a potential game with a potential function f, meaning that $\mathcal{F} = \nabla f$.

For a given strategy graph $G = (V, E)$, the presence of an edge (i, j) in E means that the agents following strategy i can switch to j if they so choose. Throughout this letter, we assume that G is connected and, in general, incomplete.

An important concept in game theory is Nash equilibria, which is defined [1] for $\mathcal{F}$ as

$$\mathbb{NE}(\mathcal{F}) := \{\xi \in \Delta \mid (\xi - \zeta)^T \mathcal{F}(\xi) \geq 0, \ \forall \zeta \in \Delta\}.$$

So, at a Nash equilibrium, a positive share of the population plays a strategy only if it offers the highest payoff. Distinctively, potential games have an additional connection to Nash equilibria, relevant to distributed optimization [20]:

Remark 1: When $\mathcal{F}$ is a potential game, $\mathbb{NE}(\mathcal{F})$ coincides with the points that satisfy the Karush-Kuhn-Tucker conditions for the problem of maximizing f over Δ. Hence, if f is *concave*, then

$$\mathbb{NE}(\mathcal{F}) = \arg\max_{\xi \in \Delta} f(\xi).$$

Moreover, if f is strictly concave, then $\arg\max_{\xi \in \Delta} f(\xi)$ is unique. We will denote this unique maximizer as $x^{\mathbb{NE}}$.

Another crucial element of the framework is the learning rule[3] (or rule for short) ρ. A well-researched class of rules that is particularly relevant to this letter is the pairwise comparison (PC) class [19], [21]. We say that ρ is a PC learning rule if it is sign-preserving, meaning that for all $i, j \in V$, $\pi \in \mathbb{R}^n$ it satisfies

$$\pi_j - \pi_i > 0 \Rightarrow \rho_{ij}(\xi, \pi) > 0,$$
$$\pi_j - \pi_i \leq 0 \Rightarrow \rho_{ij}(\xi, \pi) = 0.$$

Therefore, agents following these learning rules do not switch to strategies that offer lower payoffs. We will denote an arbitrary PC rule by $\phi : \Delta \times \mathbb{R}^n \to \mathbb{R}^{n \times n}_{\geq 0}$. A celebrated example is the Smith rule $\phi_{ij}(\xi, \pi) = \max\{\pi_j - \pi_i, 0\}$, which was introduced by [4] to analyze traffic flow on roadway networks. In cases where G is complete, PC rules are known to guarantee the global asymptotic stability of $\mathbb{NE}(\mathcal{F})$ (for any potential game) [19].

Remark 2: We can embed the effects of G into ρ by defining $\rho_{ij} \equiv 0$ for $(i, j) \notin E$. Under this interpretation, ρ can be a PC rule only when G is complete. Although earlier results on PC rules [19] are rather general, they followed this interpretation, implicitly requiring completeness.

B. Problem Description

In the classical setting, there are learning rules that secure the global asymptotic stability of $\mathbb{NE}(\mathcal{F})$ [1]. Nevertheless, similar guarantees are lacking when G is incomplete (this absence holds true for all $\mathcal{F}$, not just potential games). For instance, even when ρ is the Smith rule, f is strictly concave, and G is connected, [16] demonstrates that x can converge to a point that does not belong to $\mathbb{NE}(\mathcal{F})$. So, our goal is to find learning rules that ensure the near-global asymptotic stability of $\mathbb{NE}(\mathcal{F})$ under any G.

The following observation is essential to our approach: If ρ is a PC rule and x remains in $int(\Delta) := \{\xi \in (0, 1)^n \mid \sum_{i=1}^n \xi_i = 1\}$, then x converges to $\mathbb{NE}(\mathcal{F})$. Thus, our learning rules blend the PC class with an effect that makes the boundary of Δ "repellent," where this effect fades near $\mathbb{NE}(\mathcal{F})$.

In [17], the authors impose a similar effect on the Smith rule by subtracting the gradient of the entropy of x from the payoffs. This modification makes the boundary of Δ repellent, however it does not disappear near $\mathbb{NE}(\mathcal{F})$, leading [17] to

[1] We note that there are also articles studying multi-population interactions over networks [13], [14]. These articles assume complete strategy graphs within each population, distinguishing their focus from ours.

[2] [8] extends a standard assumption on the agents' revision times and shows that the resulting x can be analyzed via an augmented strategy set with a specific graph structure. Their results only cover this graph structure.

[3] Often also called strategy revision protocol.

conclude the asymptotic stability of a Gibbs measure.[4] In a different context (i.e., assuming that G is complete), [18] introduces a KL-divergence regularization model. Specifically, the idea in [18] is to subtract $\nabla D(x(t)\|y(t))$ from the payoffs and find an appropriate *update rule for the regularization parameter* y, where $D(\xi\|\theta) := \sum_{i=1}^{n} \xi_i \ln(\xi_i/\theta_i)$ denotes the KL-divergence for any $\xi, \theta \in int(\Delta)$.

We find that this modification achieves the effect that we need. Hence, fusing the ideas in [17] and [18], we propose KL-regularized PC learning rules.

Definition 1: We say that ρ is a *KL-regularized PC rule* if for all $i, j \in V$, $\xi \in int(\Delta)$ and $\pi \in \mathbb{R}^n$ it satisfies

$$\rho_{ij}(\xi, \pi) = \phi_{ij}(\xi, \pi - \eta \nabla D(\xi\|\theta))$$

in which ϕ is a PC rule, $\theta \in int(\Delta)$, and $\eta > 0$. For notational convenience, we define $\tilde{\pi}^{\eta,\theta} := \pi - \eta \nabla D(\xi\|\theta)$.

The resulting dynamics for x is as follows:

$$\dot{x} = \mathcal{V}^{G,PC}\Big(x, \tilde{\mathcal{F}}^{\eta,y}(x)\Big) \tag{2}$$
$$\tilde{\mathcal{F}}^{\eta,y}(x) := \mathcal{F}(x) - \eta \nabla D(x\|y)$$

where $\mathcal{V}^{G,PC}: \Delta \times \mathbb{R}^n \to \mathbb{R}^n$ is given for all $i \in V$, $\xi \in \Delta$ and $\pi \in \mathbb{R}^n$ by

$$\mathcal{V}_i^{G,PC}(\xi, \pi) := \sum_{j \in \mathcal{I}_i} \xi_j \phi_{ji}(\xi, \pi) - \sum_{j \in \mathcal{O}_i} \xi_i \phi_{ij}(\xi, \pi).$$

Observe from (2) that we can interpret KL-regularization as replacing $\mathcal{F}$ with the "perturbed" (or KL-regularized) payoff mechanism $\tilde{\mathcal{F}}^{\eta,y}$. This mechanism has the characteristics below for any fixed value of y as $\theta \in int(\Delta)$.

Remark 3: Each point in $\mathbb{NE}(\tilde{\mathcal{F}}^{\eta,\theta})$ is an equilibrium of $\dot{x} = \mathcal{V}^{G,PC}(x, \tilde{\mathcal{F}}^{\eta,\theta}(x))$ and belongs to $int(\Delta)$. Moreover, if f is concave, then $\tilde{f}^{\eta,\theta} := f - \eta D(\cdot\|\theta)$ is strictly concave and $\mathbb{NE}(\tilde{\mathcal{F}}^{\eta,\theta}) = \arg\max_{\xi \in \Delta} \tilde{f}^{\eta,\theta}(\xi)$ is unique. In this case, we will denote $\mathbb{NE}(\tilde{\mathcal{F}}^{\eta,\theta})$ as $x^{\mathbb{PNE}_{\eta,\theta}}$. Notice that, $x^{\mathbb{PNE}_{\eta,\theta}} = x^{\mathbb{NE}}$ if and only if $\theta = x^{\mathbb{NE}}$.

A pivotal step in the approach of [18] is to determine an update rule for y that ensures its convergence to the same Nash equilibrium as x. To find such update rule, we again resort to [18], which suggests the algorithm below. Essentially, this algorithm iteratively updates $y(t)$ as $x(t)$ at t that satisfy

$$\max_{\zeta \in \Delta} \Big\{(\zeta - x(t))^T \big(\mathcal{F}(x(t)) - \eta \nabla D(x(t)\|x(\underline{t}))\big)\Big\} \leq \frac{\eta}{2} D\big(x(t)\|x(\underline{t})\big), \tag{3}$$

in which $\underline{t}$ is the previous update time. We discuss the underlying intuition, existence of the update times, and convergence properties of y in Section III.

Overall, the payoff mechanism, KL-regularized PC rule and Algorithm 1 operate within the feedback configuration in Fig. 1. Hereafter, we prove that if f is strictly concave and $x^{\mathbb{NE}} \in int(\Delta)$, then $x^{\mathbb{NE}}$ is an asymptotically stable equilibrium of the resulting dynamics, and the corresponding region of attraction is the entire set $int(\Delta)$.

Remark 4: It is also possible to view KL-regularization from a mechanism design perspective. Consider that there is a coordinator that has access to x and p. As depicted in Fig. 2, this coordinator can compute and transmit $\tilde{\mathcal{F}}^{\eta,y}$, incorporating the KL-regularization step into the payoff assignment. Crucially, this enables the agents to act in a fully decentralized manner, as they can implement a PC rule using only $\tilde{\mathcal{F}}^{\eta,y}$, without needing to know x. Additionally, the agents can implement certain PC rules (e.g., the Smith rule) using only the "local" regularized payoffs, given for agents following strategy i by $\{\tilde{\mathcal{F}}_j^{\eta,y} \mid j \in \mathcal{O}_i\}$. We demonstrate this concept in Section IV.

[4] In [17], the Gibbs measure ξ has the form $\xi_i = e^{(\mathcal{F}_i(\xi)/\eta)} / \sum_{j=1}^{n} e^{(\mathcal{F}_j(\xi)/\eta)}$ for all $i \in V$.

Algorithm 1: Update Rule for y

Input: $\eta > 0$, $x(t)$, t
Output: $y(t)$
Internal State: θ_l
if $t = 0$ **then**
  $l \leftarrow 0$
  $\theta_l \leftarrow x(0)$ ▷ Initialize θ_l as $x(0)$
end
if $\max_{\zeta \in \Delta} \{(\zeta - x(t))^T (\mathcal{F}(x(t)) - \eta \nabla D(x(t)\|\theta_l))\}$ $\leq \frac{\eta}{2} D(x(t)\|\theta_l)$ **then**
  $l \leftarrow l + 1$
  $\theta_l \leftarrow x(t)$ ▷ Remember $x(t)$
end
$y(t) \leftarrow \theta_l$ ▷ Reset y to θ_l

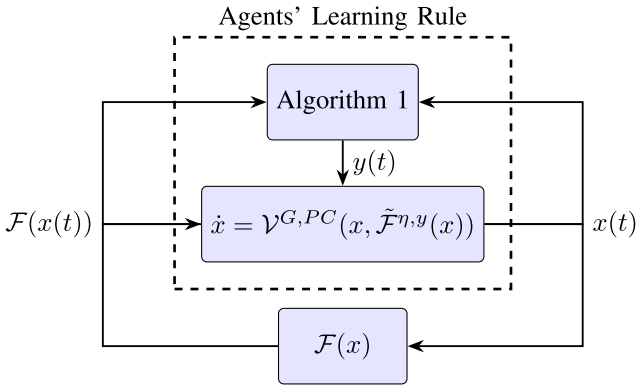


Fig. 1. The agents' learning rule and $\mathcal{F}$ as a feedback system.

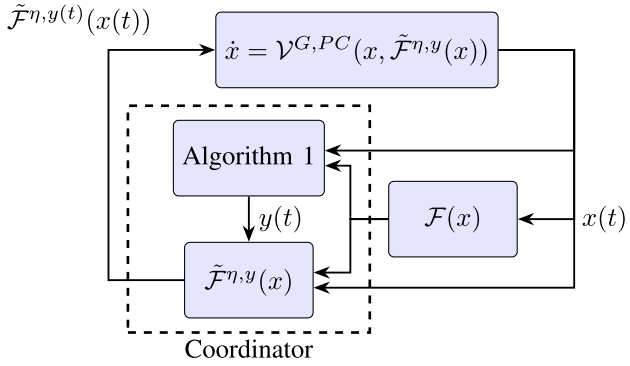


Fig. 2. Coordinator implementation of KL-regularization.

III. Main Results

We analyze (2) in two steps: We first obtain results under a fixed value of y, then we include the effects of Algorithm 1.

A. Analysis Under a Fixed Regularization Parameter

Throughout this section, we assume that y has a fixed value $\theta \in int(\Delta)$. As $\nabla D(\cdot\|\theta)$ is unbounded on Δ, the first question is whether x is well-defined for all $t \geq 0$. A trivial modification of [22, Lemma 6] shows that the answer is affirmative whenever $x(0) \in int(\Delta)$.

Lemma 1: For each θ and x^0 that belong to $int(\Delta)$, the initial value problem $\dot{x} = \mathcal{V}^{G,PC}(x, \tilde{\mathcal{F}}^{\eta,\theta}(x))$; $x(0) = x^0$ has a unique solution. Moreover, there exists a compact set $K(x^0, \theta) \subset int(\Delta)$ such that $x(t) \in K(x^0, \theta)$ for all $t \geq 0$.

Proof: It suffices to repeat the proof of [22, Lemma 6] with the $\epsilon_1, \ldots, \epsilon_n$ values below:

$$\epsilon_1 := \frac{1}{2}\min\left\{\frac{1}{1+\gamma}, \min_{i\in V}\left\{\min\{\theta_i, x_i^0\}\right\}\right\}, \quad \epsilon_l := \frac{\epsilon_{l-1}}{1+\gamma}.$$

where $v^* := \max_{\xi\in\Delta, i,j\in V}\{\mathcal{F}_i(\xi) - \mathcal{F}_j(\xi)\}$, $\gamma := \theta^* e^{v^*/\eta}$, and $\theta^* := \max_{i,j\in V}\{\theta_i/\theta_j\}$. ∎

Having proven existence, we now address stability.

Lemma 2: For any $\theta \in int(\Delta)$, if $\mathcal{F}$ has a strictly concave potential, then $x^{\mathbb{PNE}_{\eta,\theta}}$ is an asymptotically stable equilibrium of $\dot{x} = \mathcal{V}^{G,PC}(x, \tilde{\mathcal{F}}^{\eta,\theta}(x))$, with region of attraction $int(\Delta)$.

Proof: We will prove Lemma 2 by showing that $\mathcal{L}(\cdot) := -\tilde{f}^{\eta,\theta}(\cdot) + \tilde{f}^{\eta,\theta}(x^{\mathbb{PNE}_{\eta,\theta}})$ is a Lyapunov function for $\dot{x} = \mathcal{V}^{G,PC}(x, \tilde{\mathcal{F}}^{\eta,\theta}(x))$ on $int(\Delta)$.

From Remark 3, it trivially follows that $\mathcal{L}$ is positive semi-definite and $\mathcal{L}(x^{\mathbb{PNE}_{\eta,\theta}}) = 0$. As for the derivative condition, it holds for all $\xi \in int(\Delta)$ that

$$\begin{aligned}
&(\nabla\mathcal{L}(\xi))^T\mathcal{V}^{G,PC}\left(\xi, \tilde{\mathcal{F}}^{\eta,\theta}(\xi)\right)\\
&= -\sum_{i=1}^{n}\tilde{\mathcal{F}}_i^{\eta,\theta}(\xi)\sum_{j\in\mathcal{I}_i}\xi_j\phi_{ji}\left(\xi, \tilde{\mathcal{F}}^{\eta,\theta}(\xi)\right)\\
&\quad + \sum_{i=1}^{n}\tilde{\mathcal{F}}_i^{\eta,\theta}(\xi)\sum_{j\in\mathcal{O}_i}\xi_i\phi_{ij}\left(\xi, \tilde{\mathcal{F}}^{\eta,\theta}(\xi)\right)\\
&= -\sum_{(j,i)\in E}\tilde{\mathcal{F}}_i^{\eta,\theta}(\xi)\xi_j\phi_{ji}\left(\xi, \tilde{\mathcal{F}}^{\eta,\theta}(\xi)\right)\\
&\quad + \sum_{(j,i)\in E}\tilde{\mathcal{F}}_j^{\eta,\theta}(\xi)\xi_j\phi_{ji}\left(\xi, \tilde{\mathcal{F}}^{\eta,\theta}(\xi)\right)\\
&= -\sum_{(j,i)\in E}\left(\tilde{\mathcal{F}}_i^{\eta,\theta}(\xi) - \tilde{\mathcal{F}}_j^{\eta,\theta}(\xi)\right)\xi_j\phi_{ji}\left(\xi, \tilde{\mathcal{F}}^{\eta,\theta}(\xi)\right). \quad (4)
\end{aligned}$$

Recall that ϕ is non-negative and that $\phi_{ji}(\xi, \tilde{\mathcal{F}}^{\eta,\theta}(\xi)) > 0$ if and only if $\tilde{\mathcal{F}}_i^{\eta,\theta}(\xi) > \tilde{\mathcal{F}}_j^{\eta,\theta}(\xi)$. This, together with $\xi \in int(\Delta)$, imply that (4) is non-positive. Hence, the final step is to show that (4) is 0 only when $\xi = x^{\mathbb{PNE}_{\eta,\theta}}$.

Let $[i^k]$ denote the set of strategies that offer the k-th highest payoff. If $\xi \neq x^{\mathbb{PNE}_{\eta,\theta}}$, then $[i^2] \neq \emptyset$. Therefore, when $\xi \neq x^{\mathbb{PNE}_{\eta,\theta}}$, the connectivity of G guarantees the existence of nonempty $[i^l]$, $[i^m]$ with $l > m$ such that $(j^*, i^*) \in E$ for some $j^* \in [i^l]$ and $i^* \in [i^m]$. Also, note that $\xi_{j^*} > 0$ because $\xi \in int(\Delta)$. Consequently, we have

$$\begin{aligned}
&-\sum_{(j,i)\in E}\left(\tilde{\mathcal{F}}_i^{\eta,\theta}(\xi) - \tilde{\mathcal{F}}_j^{\eta,\theta}(\xi)\right)\xi_j\phi_{ji}\left(\xi, \tilde{\mathcal{F}}^{\eta,\theta}(\xi)\right)\\
&\leq -\left(\tilde{\mathcal{F}}_{i^*}^{\eta,\theta}(\xi) - \tilde{\mathcal{F}}_{j^*}^{\eta,\theta}(\xi)\right)\xi_{j^*}\phi_{j^*i^*}\left(\xi, \tilde{\mathcal{F}}^{\eta,\theta}(\xi)\right) < 0.
\end{aligned}$$

∎

With Lemma 2, we establish the near-global (convergence from any $x(0) \in int(\Delta)$) asymptotic stability of the "perturbed equilibria" $x^{\mathbb{PNE}_{\eta,\theta}}$. Thus, we would have our target stability guarantee if $x^{\mathbb{PNE}_{\eta,y}} \equiv x^{\mathbb{NE}}$. However, according to Remark 3, this equivalence holds only if $y \equiv x^{\mathbb{NE}}$. This is where Algorithm 1 comes into play: To ensure that the regularization parameter y converges to $x^{\mathbb{NE}}$.

B. Updating the Regularization Parameter

In this section, we focus on Algorithm 1. This updating scheme was originally proposed in [18, Sec. IV.B], based on the following intuition.

From the definition of Nash equilibria in Section II, observe that $\max_{\zeta\in\Delta}\{(\zeta - x(t))^T(\mathcal{F}(x(t)) - \eta\nabla D(x(t)\|\theta))\}$ is small if and only if $x(t)$ is close to $\mathbb{NE}(\tilde{\mathcal{F}}^{\eta,\theta})$. So, Algorithm 1 updates y when x is sufficiently close to the perturbed Nash equilibria. In essence, this causes x to approach to a new perturbed equilibria in each epoch. Pivotally, under some types of payoff mechanisms, the perturbation diminishes at each step, resulting in the convergence of x to a Nash equilibrium. In [18], the authors prove this fading effect for so-called contractive games (further details are present in the proof of the upcoming Theorem 1).

Note that, under the assumptions of Lemma 2, there is an increasing sequence of time instances at which y is updated. More precisely:

Remark 5: When f is strictly concave and $x(0) \in int(\Delta)$, there is an increasing sequence $t_0 = 0$, $(t_l)_{l=1}^{\infty}$ satisfying

$$\max_{\zeta\in\Delta}\left\{(\zeta - x(t_{l+1}))^T\tilde{\mathcal{F}}^{\eta,x(t_l)}(x(t_{l+1}))\right\} \leq \frac{\eta}{2}D(x(t_{l+1})\|x(t_l)), \quad (5)$$

where $x : [t_l, t_{l+1}) \to \Delta$ is the solution of

$$\dot{x} = \mathcal{V}^{G,PC}\left(x, \tilde{\mathcal{F}}^{\eta,x(t_l)}\right)$$

with initial condition $x(t_l)$.

This is a straightforward observation because y gets updated when x is close enough to the perturbed equilibrium, and Lemma 2 ensures that x approaches this equilibrium. To elaborate, Lemma 2 guarantees that the left hand side of (3) tends to 0 as t increases. Meanwhile, the right hand side of (3) converges to a positive value, provided $x(0) \neq x^{\mathbb{NE}}$. Therefore, (3) will hold at some point, triggering an update. Furthermore, Lemma 1 implies that y remains in $int(\Delta)$. Repeating these steps for each updated value of y confirms the validity of Remark 5. We refer to [18] for further details.

Remark 6: Periodic execution of Algorithm 1 at intervals of any $T > 0$ maintains the validity of Remark 5. Hence, it is not necessary for Algorithm 1 to operate at every t.

C. Asymptotic Stability of $x^{\mathbb{NE}}$

Up to this point, we analyzed (2) under fixed y. Now, we consider the feedback interconnection of (2) and Algorithm 1. The following is our main result.

Theorem 1: Assume that $\mathcal{F}$ has a strictly concave potential, $x^{\mathbb{NE}} \in int(\Delta)$ and y is determined by Algorithm 1. Then, $x^{\mathbb{NE}}$ is an asymptotically stable equilibrium of (2), with the region

of attraction $int(\Delta)$. Additionally, it holds for any $x(0) \in int(\Delta)$ that

$$\lim_{t\to\infty} \|y(t) - x^{\mathbb{NE}}\| = 0.$$

In other words, if f is strictly concave, $x^{\mathbb{NE}} \in int(\Delta)$ and the agents follow a KL-regularized PC rule, then $x^{\mathbb{NE}}$ is a near-globally asymptotically stable equilibrium of (2).

Proof: In what follows, we assume that $x(0) \in int(\Delta)$ and verify three conjectures: (i) y converges to $x^{\mathbb{NE}}$, (ii) x converges to $x^{\mathbb{NE}}$, and (iii) $x^{\mathbb{NE}}$ is stable. Our proofs for (i) and (ii) draw heavily from the proof of [18, Lemma 4], whereas our treatment of (iii) (i.e., the stability component) extends beyond the scope of analysis in [18].

We begin by establishing the convergence of y. In Part I of the proof of [18, Lemma 4], the authors show the following: If $\mathcal{F}$ is strictly contractive and $(t_l)_{l=1}^{\infty}$ satisfies Remark 5, then $\lim_{t\to\infty} \|y(t) - x^{\mathbb{NE}}\| = 0$ for any initial state $x(0) \in int(\Delta)$, where $\mathcal{F}$ is said to be strictly contractive if $(\xi - \zeta)^T(\mathcal{F}(\xi) - \mathcal{F}(\zeta)) < 0$ for all $\xi, \zeta \in \Delta$ with $\xi \neq \zeta$. Now, observe that the strict concavity of f yields

$$f(\xi) - f(\zeta) < \mathcal{F}(\zeta)^T(\xi - \zeta),$$
$$f(\zeta) - f(\xi) < \mathcal{F}(\xi)^T(\zeta - \xi),$$

implying that

$$0 > \mathcal{F}(\zeta)^T(\zeta - \xi) - \mathcal{F}(\xi)^T(\zeta - \xi) = (\zeta - \xi)^T(\mathcal{F}(\zeta) - \mathcal{F}(\xi)).$$

Consequently, every $\mathcal{F}$ with a strictly concave potential is strictly contractive. So, we can leverage Part I of the proof of [18, Lemma 4] to conclude that $\lim_{t\to\infty} \|y(t) - x^{\mathbb{NE}}\| = 0$.

Next, we show the convergence of x to $x^{\mathbb{NE}}$ from any $x(0) \in \text{int}(\Delta)$. As in Algorithm 1, let θ_l denote the value of y over $[t_l, t_{l+1})$. We know from the proof of Lemma 2 that $-f(x(t)) + \eta D(x(t)\|\theta_l)$ is decreasing over $t \in [t_l, t_{l+1})$ for each $l \in \mathbb{N}$. Note also that $-f(\theta_l) + \eta D(\theta_l\|\theta_l) = -f(\theta_l)$ and $\lim_{l\to\infty} \theta_l = x^{\mathbb{NE}}$. Combining these findings, we can assert that for any $\epsilon > 0$, there exists $L \in \mathbb{N}$ such that the following inequalities hold for each $l \geq L$ and $t \in [t_l, t_{l+1})$:

$$f\left(x^{\mathbb{NE}}\right) - f(x(t)) + \eta D(x(t)\|\theta_l) \leq f\left(x^{\mathbb{NE}}\right) - f(\theta_l) \leq \epsilon. \tag{6}$$

Now, let us take L as above and assume, for contradiction, that x does not converge to $x^{\mathbb{NE}}$. Then, for all $\delta > 0$, there is an increasing sequence $(s_m)_{m=1}^{\infty}$ such that $\|x(s_m) - x^{\mathbb{NE}}\| > \delta$ for all $m \in \mathbb{N}$. Since $(s_m)_{m=1}^{\infty}$ is increasing, there exist $m' \in \mathbb{N}$ and $l \geq L$ for which $s_{m'} \in [t_l, t_{l+1})$. Recall that f is strictly concave and $x^{\mathbb{NE}} = \arg\max_{\xi\in\Delta} f(\xi)$. Therefore,

$$\left|f\left(x^{\mathbb{NE}}\right) - f(x(s_{m'})) + \eta D(x(s_{m'})\|\theta_l)\right| \geq \left|f\left(x^{\mathbb{NE}}\right) - f(x(s_{m'}))\right|,$$

and we can choose the δ above so that $f(x^{\mathbb{NE}}) - f(x(s_{m'})) > \epsilon$. However, $l \geq L$ and $s_{m'} \in [t_l, t_{l+1})$ imply that (6) should hold with $t = s_{m'}$, yielding a contradiction.

Finally, we prove that $x^{\mathbb{NE}}$ is stable. So, given any $\delta > 0$, we want to find $\epsilon > 0$ such that $\|x(t) - x^{\mathbb{NE}}\| \leq \delta$ holds for all $t \geq 0$ whenever $\|x(0) - x^{\mathbb{NE}}\| \leq \epsilon$. Let us define

$$B_\delta := \{\xi \in \Delta \mid \|\xi - x^{\mathbb{NE}}\| \leq \delta\}$$

and take $c^\delta > 0$ such that

$$\Gamma := \left\{\xi \in int(\Delta) \;\middle|\; \left|f\left(x^{\mathbb{NE}}\right) - f(\xi)\right| \leq c^\delta\right\} \subset B_\delta.$$

Observe that such c^δ exists because f is strictly concave and $x^{\mathbb{NE}} = \arg\max_{\xi\in\Delta} f(\xi)$. Without loss of generality, we assume δ to be small enough that $\Gamma \subset int(\Delta)$, which is possible because $x^{\mathbb{NE}} \in int(\Delta)$. Now, let us set $x(0) = x^0$ for any $x^0 \in \Gamma$. As we noted earlier, we know from Lemma 2 that $f(x(t)) - \eta D(x(t)\|x^0)$ is increasing over $[t_0, t_1)$, where t_0 is 0 and t_1 is the first update time for y. Therefore,

$$f(x(t)) - \eta D\left(x(t)\|x^0\right) \geq f\left(x^0\right) - \eta D\left(x^0\|x^0\right) = f\left(x^0\right)$$

for all $t \in [t_0, t_1]$. This, together with $f(x^0) \geq f(x^{\mathbb{NE}}) - c^\delta$, implies for all $t \in [t_0, t_1]$ that

$$f(x(t)) \geq f(x(t)) - \eta D\left(x(t)\|x^0\right) \geq f\left(x^0\right) \geq f\left(x^{\mathbb{NE}}\right) - c^\delta.$$

Hence, $x(t) \in \Gamma$ for all $t \in [t_0, t_1]$. To see that x remains in Γ through $[t_1, \infty)$, observe that the initial state of the system on the epoch $[t_1, t_2]$ is $x(t_1)$, which again belongs to Γ. Thus, the preceding discussion applies with $[t_0, t_1]$ replaced by $[t_l, t_{l+1}]$ for any $l \in \mathbb{N}$. As a result, $x(0) \in \Gamma$ guarantees that $x(t) \in \Gamma \subset B_\delta$ for all $t \geq 0$. ■

Remark 7: Theorem 1 not only ensures that x *converges to* $x^{\mathbb{NE}}$ *from any initial state in* $int(\Delta)$, but also implies that $x^{\mathbb{NE}}$ is a *stable equilibrium*. As we outlined in Section I-A, this enables us to apply the analysis from [3] to obtain convergence guarantees to $x^{\mathbb{NE}}$ when the population size is large but finite.

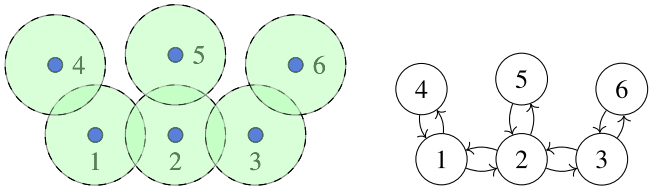
Remark 8: We remind that, throughout this letter, we assume G to be connected. Without this assumption, there always exist $x(0)$ from which x fails to approach $x^{\mathbb{NE}}$, regardless of ρ and the payoff mechanism.

IV. Numerical Example

In this section, we illustrate an application of Theorem 1 via a decentralized resource allocation problem. Consider that there are 6 regions, each associated with a transmitter's range. These transmitters assign rewards to their regions based on how the agents are distributed across them. Suppose that the agents can only migrate between intersecting regions, i.e., they must always remain within the range of at least one transmitter. We assume that the regions have the configuration in Fig. 3. Our goal is to find a payoff mechanism and a learning rule that ultimately lead to an equal allocation of agents across all regions.

Theorem 1 tells that we can achieve our goal with a KL-regularized PC rule and a potential game with a strictly concave f that satisfies $\arg\max_{\xi\in\Delta} f(\xi) = \underline{1}/6$ ($\underline{1}$ denotes the 6-dimensional vector of ones).

Thus, we take $f(\xi) = -\sum_{i=1}^{6}(\xi_i - 1/6)^2/2$, yielding the potential game $\mathcal{F}(\xi) = \nabla f(\xi) = \underline{1}/6 - \xi$. As for the learning rule, we use the KL-regularized Smith rule with $\eta = 0.05$, given by $\rho_{ij}(\xi, \pi) = \max\{\tilde{\pi}_j^{0.05,\theta} - \tilde{\pi}_i^{0.05,\theta}, 0\}$. We plot the



(a) Transmitter coverage (b) Strategy graph

Fig. 3. Transmitter configurations and the resulting G.

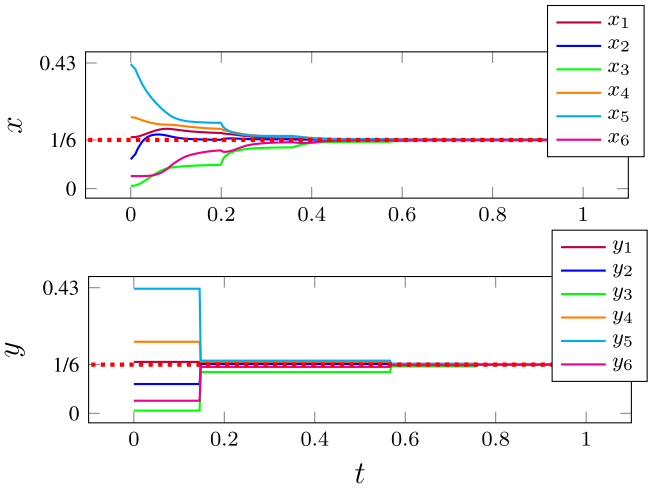


Fig. 4. Time domain plots of x and y.

resulting x and y (from a random $x(0)$) in Fig. 4, which verifies that both converge to $\underline{1}/6$.

Remark 9: To implement the KL-regularized Smith rule, the agents need to access the payoffs from their own and neighboring regions, along with the corresponding y values. Therefore, if the transmitters collectively calculate y and each transmitter i broadcasts $\tilde{\mathcal{F}}_i^{\eta,y}$, then the agents can rely solely on local information.

Remark 10: Observe in Fig. 4 that y changes instantly at discrete time points. This demonstrates the discussion in Section III-B: Once (5) holds and y is reset to x, it takes time for (5) to be satisfied again under the new y value, resulting in a constant y during this time frame.

Remark 11: Observe how the spatial constraint associated with the transmitters' ranges naturally corresponds to an incomplete G. This highlights the pertinence of constrained strategy switching in games with spatial and informational constraints. Another related example can involve multi-agent path planning, as the discretization of feasible paths can be modeled by an incomplete G.

V. Conclusion

We studied a setting of large-scale multi-agent decision-making, wherein only certain strategies are reachable from certain others, as characterized by a strategy graph G. By combining pairwise comparison rules with a recently introduced KL-regularization concept, we proposed a new class of learning rules that ensure near-global asymptotic stability of Nash equilibria for any connected G. We also validated our results using a decentralized resource allocation problem. Future avenues for research could involve extensions to time-varying G and broader payoff structures.

References

[1] W. H. Sandholm, *Population Games and Evolutionary Dynamics*. Cambridge, MA, USA: MIT Press, 2010.
[2] S. Kara and N. C. Martins, "Differential equation approximations for population games using elementary probability," 2023, *arXiv:2312.07598*.
[3] M. Benaïm and J. Weibull, "Deterministic approximation of stochastic evolution in games," *Econometrica*, vol. 71, no. 3, pp. 873–903, 2003.
[4] M. J. Smith, "The stability of a dynamic model of traffic assignment—An application of a method of Lyapunov," *Transp. Sci.*, vol. 18, no. 3, pp. 245–252, Aug. 1984.
[5] P. Srikantha and D. Kundur, "Resilient distributed real-time demand response via population games," *IEEE Trans. Smart Grid*, vol. 8, no. 6, pp. 2532–2543, Nov. 2017.
[6] N. Quijano, C. Ocampo-Martinez, J. Barreiro-Gomez, G. Obando, A. Pantoja, and E. Mojica-Nava, "The role of population games and evolutionary dynamics in distributed control systems," *IEEE Control Syst. Mag.*, vol. 37, no. 1, pp. 70–97, Feb. 2017.
[7] S. Park, Y. D. Zhong, and N. E. Leonard, "Multi-robot task allocation games in dynamically changing environments," in *Proc. IEEE Int. Conf. Robot. Autom. (ICRA)*, 2021, pp. 8678–8684.
[8] S. Kara, N. C. Martins, and M. Arcak, "Population games with Erlang clocks: Convergence to Nash equilibria for pairwise comparison dynamics," in *Proc. IEEE 61st Conf. Decis. Control (CDC)*, 2022, pp. 7688–7695.
[9] J. I. Poveda and N. Quijano, "Shahshahani gradient-like extremum seeking," *Automatica*, vol. 58, pp. 51–59, Aug. 2015.
[10] J. Barreiro-Gomez, G. Obando, and N. Quijano, "Distributed population dynamics: Optimization and control applications," *IEEE Trans. Syst., Man, Cybern., Syst.*, vol. 47, no. 2, pp. 304–314, Feb. 2017.
[11] E. Altman, Y. Hayel, and H. Kameda, "Evolutionary dynamics and potential games in non-cooperative routing," in *Proc. 5th Int. Symp. Model. Optim. Mobile, Ad Hoc Wireless Netw. Workshops*, 2007, pp. 1–5.
[12] A. Bizyaeva, A. Franci, and N. E. Leonard, "Nonlinear opinion dynamics with tunable sensitivity," *IEEE Trans. Autom. Control*, vol. 68, no. 3, pp. 1415–1430, Mar. 2023.
[13] G. Como, F. Fagnani, and L. Zino, "Imitation dynamics in population games on community networks," *IEEE Trans. Control Netw. Syst.*, vol. 8, no. 1, pp. 65–76, Mar. 2021.
[14] J. Martinez-Piazuelo, C. Ocampo-Martinez, and N. Quijano, "Distributed Nash equilibrium seeking in strongly contractive aggregative population games," *IEEE Trans. Autom. Control*, early access, Oct. 2, 2023, doi: 10.1109/TAC.2023.3321208.
[15] J. Barreiro-Gomez, G. Obando, A. Pantoja, and H. Tembine, "Heterogeneous multi-population evolutionary dynamics with migration constraints," *IFAC-PapersOnLine*, vol. 53, no. 2, pp. 16852–16857, 2020.
[16] S. Tan, Y. Wang, and A. V. Vasilakos, "Distributed population dynamics for searching generalized Nash equilibria of population games with graphical strategy interactions," *IEEE Trans. Syst., Man, Cybern., Syst.*, vol. 52, no. 5, pp. 3263–3272, May 2022.
[17] S.-N. Chow, W. Li, J. Lu, and H. Zhou, "Population games and discrete optimal transport," *J. Nonlin. Sci.*, vol. 29, no. 3, pp. 871–896, Oct. 2018.
[18] S. Park and N. E. Leonard, "Learning with delayed payoffs in population games using Kullback-Leibler divergence regularization," 2023, *arXiv:2306.07535*.
[19] W. H. Sandholm, "Pairwise comparison dynamics and evolutionary foundations for Nash equilibrium," *Games*, vol. 1, no. 1, pp. 3–17, 2010.
[20] J. Barreiro-Gomez, N. Quijano, and C. Ocampo-Martinez, "Constrained distributed optimization: A population dynamics approach," *Automatica*, vol. 69, pp. 101–116, Jul. 2016.
[21] S. Kara and N. C. Martins, "Pairwise comparison evolutionary dynamics with strategy-dependent revision rates: Stability and δ-passivity," *IEEE Trans. Control Netw. Syst.*, vol. 10, no. 4, pp. 1656–1668, Dec. 2023.
[22] S.-N. Chow, W. Li, and H. Zhou, "Entropy dissipation of Fokker-Planck equations on graphs," 2017, *arXiv:1701.04841*.