2024 IEEE 63rd Conference on Decision and Control (CDC)
December 16-19, 2024. MiCo, Milan, Italy

Learning Equilibrium with Estimated Payoffs in Population Games

Shinkyu Park

Abstract— **We study a multi-agent decision problem in population games, where agents select from multiple available strategies and continually revise their selections based on the payoffs associated with these strategies. Unlike conventional population game formulations, we consider a scenario where agents must estimate the payoffs through local measurements and communication with their neighbors. By employing task allocation games – dynamic extensions of conventional population games – we examine how errors in payoff estimation by individual agents affect the convergence of the strategy revision process. Our main contribution is an analysis of how estimation errors impact the convergence of the agents' strategy profile to equilibrium. Based on the analytical results, we propose a design for a time-varying strategy revision rate to guarantee convergence. Simulation studies illustrate how the proposed method for updating the revision rate facilitates convergence to equilibrium.**

I. Introduction

Large-population game frameworks provide a systematic approach to studying the decision-making processes of multiple agents engaged in repeated strategic interactions. These frameworks find applications across a wide range of fields, including road congestion [1], [2], communication systems [3], [4], distributed control systems [5], and multi-robot task allocation [6]–[8], among others. In this work, we investigate a decision-making model within the population game framework [9], where agents utilize the model to select a strategy among a finite set of available options. This model consists of a *learning rule* (also referred to as a *revision protocol*) and *stochastic alarm clock*, defining how individual agents repeatedly revise their strategy selections, with the primary goal of learning the best strategies. Under this model, the agent strategy selection is influenced by information about payoffs of available strategies. Within the standard framework, it is generally assumed that all agents have perfect knowledge of these payoffs.

However, in many engineering applications, such as decentralized control systems [10], where sensing and decision-making are decentralized, this assumption requires well-established communication channels to ensure that all agents have complete knowledge of the payoffs. Otherwise, agents must estimate the payoffs based on local observations and communication with neighbors, and then make decisions on strategy selection based on their own payoff estimates. In this context, limited communication between agents leads to non-negligible estimation errors. This paradigm challenges one of the key assumptions of the standard population game formalism, which is critical for establishing convergence results [9].

In this paper, we adopt task allocation games [6], [11] and analyze the effect of errors in payoff estimation on the convergence of the agent strategy revision process. These games can be viewed as dynamic extensions of conventional population games, with potential applications in multi-robot systems research as demonstrated in [6]. In task allocation games, agents choose strategies to complete a set of predefined tasks, where the payoff for each strategy is determined by the amount of jobs remaining in its associated task. Unlike their conventional counterparts, the game has its own state, governed by a dynamical system model, to represent the remaining jobs for each task. In such dynamic game settings, if agents have limited capabilities for observing the game's state, they must estimate it and use this estimate to infer expected payoffs and select one of available strategies.

Relevant to our present work where agents are communicating with their neighbors defined by a graph for payoff estimation, the literature on games over graphs, as exemplified by [12], studies the long-term strategic interactions of multiple agents when social networks, represented by graphs, restrict the interactions between them. The study by [13] proposes a new population dynamics framework designed to model distributed information structures in population games. In this framework, the strategic decision-making of multiple agents is represented by a dynamical system model, concisely expressed as an ordinary differential equation (ODE) on a graph. Using this ODE model, the authors examine convergence of population dynamics.

As part of the series of works, the author of [14] proposes a framework to rigorously formulate the agent decision-making process in scenarios where each agent has limited access to information about the underlying game. This limitation results in noisy evaluations of payoffs for available strategies and limited knowledge of other agents' strategy selections. The work also evaluates equilibrium states of the process to predict the long-term behavior of the agents' decision-making. The work of [15] discusses analytical methods for assessing the long-term behavior of agents' noisy decision-making processes, described by stochastic dynamical system models. Notably, the main results illustrate how a *stochastically stable equilibrium* emerges as the noise level in the agent decision-making process approaches zero.

In our work, we analyze how payoff estimation errors affect the convergence of the agent decision-making process in task allocation games. Unlike previous studies that primarily focus on analyzing the impact of errors on the decision-

This work was supported by funding from King Abdullah University of Science and Technology (KAUST).

The author is with Electrical and Computer Engineering, King Abdullah University of Science and Technology (KAUST), Thuwal, 23955-6900, Kingdom of Saudi Arabia. shinkyu.park@kaust.edu.sa

making process, we propose a method to eliminate this influence in our problem setting. Specifically, we rigorously prove that with a decreasing strategy revision rate, agents' decision-making process effectively mitigates the effects of estimation errors, allowing them to asymptotically learn and attain the equilibrium corresponding to the optimal strategy selection. Based on this analysis, we propose a design for a stochastic alarm clock to ensure convergence. Our technical contributions are summarized as follows:

- Leveraging passivity-based analytical tools [16], [17] developed for population games, we analyze the impact of payoff estimation errors on the convergence of the agent decision-making process to its equilibrium state in task allocation games. In particular, we discuss how the influence of the estimation error can be mitigated by decreasing the rate of agents' strategy revision.
- Based on the analysis, we propose the design of a stochastic alarm clock using a non-homogeneous Poisson process to guarantee the convergence. Through simulations with a large number of agents, we illustrate our analytical results and evaluate the effectiveness of the proposed clock design.

The paper is organized as follows: In Section II, we formally introduce task allocation games and an agent decision-making model, specifying when and how each agent revises its strategy selection and how it estimates the payoffs for available strategies. In Section III, we present an analysis of how errors in payoff estimation affect convergence of the agent decision-making process in task allocation games, and propose a design for a non-homogeneous Poisson alarm clock that guarantees convergence. In Section IV, we present simulation results that illustrate both our analytical findings and the effectiveness of the proposed clock design.

II. Preliminaries and Problem Description

In task allocation games [6], there is a finite number n of tasks for a population of agents to carry out. Each agent can select one of available strategies to address the tasks. In this paper, for a concise presentation, we assume that the number of strategies and tasks is the same. We denote the amount of jobs to be completed associated with each task $i \in \{1, \cdots, n\}$ by a non-negative, time-dependent variable $q_i(t) \in [0, q_{\max}]$. We refer to $q(t) = (q_1(t), \cdots, q_n(t))$ as the *game state*. When $q_i(t)$ is below its maximum, $q_{\max}$, the variable increases at a constant rate as more jobs are assigned to the agents and decreases as the agents complete the jobs. This decrease is based on the agents' strategy selections, represented by the *population state* $x(t) = (x_1(t), \cdots, x_n(t))$, where each entry $x_i(t)$ denotes the fraction of the population selecting strategy i. Let $\mathbb{X}$ be the space of all viable population states, defined as $\mathbb{X} = \left\{x \in \mathbb{R}^n_+ \;\middle|\; \sum_{i=1}^n x_i = 1\right\}$, where $\mathbb{R}^n_+$ is the set of (element-wise) non-negative n-dimensional vectors.

A. Task Allocation Game Model

We adopt the following dynamical system representation from [6] to specify how $q_i(t)$ varies over time: If $q_i(t) < q_{\max}$, then

$$\dot{q}_i(t) = -\underbrace{\mathcal{F}_i(q_i(t), x_i(t))}_{\text{decrease rate}} + \underbrace{w_i}_{\text{increase rate}}, \; q_i(0) \geq 0, \tag{1}$$

and $\dot{q}_i(t) = \min\{0, -\mathcal{F}_i(q_i(t), x_i(t)) + w_i\}$ if $q_i(t) = q_{\max}$, where the latter condition ensures that $q_i(t)$ does not exceed its maximum $q_{\max}$. The function $\mathcal{F}_i : [0, q_{\max}] \times \mathbb{R}_+ \to \mathbb{R}_+$ is non-negative[1], and w_i is a positive constant. Also, to ensure that $q_i(t)$ is non-negative for all $t \geq 0$, the function $\mathcal{F}_i(q_i, x_i)$ is zero whenever $q_i = 0$.

According to (1), each agent can carry out one task at a time. The output of the game model is the payoff vector $p(t) = (p_1(t), \cdots, p_n(t))$, where we define each $p_i(t)$ as $p_i(t) = q_i(t)$ to associate $p_i(t)$ with the amount of remaining jobs in the i-th task. Consequently, under this definition of the payoff vector, agents can be incentivized to carry out the tasks with more remaining jobs. The following are assumptions we impose on (1).

Assumption 1: The function $\mathcal{F}_i : [0, q_{\max}] \times \mathbb{R}_+ \to \mathbb{R}_+$ is continuously differentiable and satisfies

$$\lim_{x_i \to \infty} \mathcal{F}_i(q_i, x_i) = \infty \text{ for any positive } q_i \tag{2a}$$

$$\frac{\partial \mathcal{F}_i}{\partial x_i}(q_i, x_i) > 0 \text{ for any positive } q_i, x_i \tag{2b}$$

$$\frac{\partial \mathcal{F}_i}{\partial q_i}(q_i, x_i) > 0 \text{ for any positive } q_i, x_i. \tag{2c}$$

In other words, the strategies are designed such that (2a) and (2b) imply that the more agents there are taking on the same task, the faster it can be completed. Additionally, (2c) suggests that the larger the amount of remaining jobs, the easier it is for the agents to locate and coordinate to complete them.

Note that, as we have proved in Lemma 1 in [18, Appendix A.1], under Assumption 1, the game model (1) is δ-*antipassive*, as defined in the appendix.

Assumption 2: The game model (1) has a unique equilibrium (q^*, x^*) in $[0, q_{\max})^n \times \mathbb{X}$ that satisfies $x_i^*(q_j^* - q_i^*) \leq 0, \; \forall i, j \in \{1, \cdots, n\}$.

As detailed in Section II-B, this implies the uniqueness of the equilibrium of (1) under the agent decision-making model considered in this work.

Example 1: As demonstrated in [6], the game model (1) can be applied to a multi-robot trash collection application. In this application, there are n spatially separated patches, with the variable $q_i(t)$ representing the volume of trash in the i-th patch. The function $\mathcal{F}_i$ is defined as

$$\mathcal{F}_i(q_i, x_i) = R_i \frac{e^{\alpha_i q_i} - 1}{e^{\alpha_i q_i} + 1} x_i^{\beta_i}, \tag{3}$$

where R_i, α_i, and β_i are positive parameters.

Remark 1: For brevity, we consider the scenario where the number of tasks and strategies is the same, and the payoff vector is defined as $p(t) = q(t)$. Also, as specified in (1),

[1]Note that even if the i-th portion $x_i(t)$ of the population state is within $[0, 1]$, we define the domain of $\mathcal{F}_i$ corresponding to $x_i(t)$ as the entire set $\mathbb{R}_+$ of non-negative real numbers.

every agent can undertake only one task at a time. However, this scenario can be extended to more general cases where the set of available strategies exceeds the number of tasks, and some strategies enable an agent to carry out multiple tasks simultaneously. In such scenarios, the payoff vector needs to be redefined as a non-trivial function of $q(t)$, e.g., $p(t) = Gq(t)$. Primary investigations of such extensions have been conducted in [11], and we consider adopting the analysis from this reference as a future research direction.

B. Learning Rule, Stochastic Alarm Clock, and Evolutionary Dynamics Model

Two central components of the agent decision-making model are the *learning rule* (also referred to as the *revision protocol*) and the *stochastic alarm clock*. The learning rule describes how each agent changes its strategy selection when given an opportunity and is typically defined as a function

$$\rho_{ji}(p,x) = \mathbb{P}\left(\text{agent switching strategy from } j \text{ to } i\right). \quad (4)$$

We consider the class of learning rules that depend only on p, i.e., $\rho_{ji}(p,x) = \rho_{ji}(p)$, and are Lipschitz continuous. That is, there exists a positive constant c for which the following inequality holds:

$$|\rho_{ji}(p) - \rho_{ji}(\bar{p})| \leq c \left\|p - \bar{p}\right\|_2, \ \forall p, \bar{p} \in \mathbb{R}^n. \quad (5)$$

Below is an example of an existing model that belongs to this class.

Example 2: Suppose ρ_{ji} is the Smith learning rule, i.e., $\rho_{ji}(p) = \varrho[p_i - p_j]_+ = \varrho \max(0, p_i - p_j)$ for $i \neq j$, originally investigated in transportation research [19]. Given that the range of each p_i is bounded, we select a constant ϱ to ensure that $\sum_{i=1}^n \varrho[p_i - p_j]_+ \leq 1$ holds for all j in $\{1, \cdots, n\}$. We can derive the following inequality and verify the Lipschitz continuity of the Smith learning rule:

$$|\rho_{ji}(p) - \rho_{ji}(\bar{p})| \leq \sqrt{2}\varrho \left\|p - \bar{p}\right\|_2. \quad (6)$$

While the learning rule defines how each agent revises its strategy, the stochastic alarm clock determines when the agent can make strategy revision using the learning rule. Typically, a homogeneous Poisson process $N(t)$ is utilized to define the clock [9, Chapter 10]. Specifically, at each ring of the clock, defined as any time t satisfying $N(t) - N(t-\delta) \geq 1, \ \forall \delta > 0$, the agent retains the opportunity to revise its strategy. As the Poisson processes assigned to the agents are independent, if they can assess the payoff vector p, the agents' strategy revision can be conducted in a decentralized manner.

Suppose these Poisson processes are identically distributed, and let λ define the rate of the processes. Consider the following ordinary differential equation:

$$\dot{x}_i(t) = \lambda \textstyle\sum_{j=1}^n x_j(t)\rho_{ji}(p(t)) - \lambda x_i(t) \sum_{j=1}^n \rho_{ij}(p(t)). \quad (7)$$

We refer to (7) as the *evolutionary dynamics model (EDM)*. It has been well-documented in [9, Chapters 5 and 10] that when the Poisson processes are independent and identically distributed and there is a sufficiently large number of agents, the solution of (7) serves as a good predictor of the population state with arbitrarily high accuracy. Throughout this work, we assume that (7) is δ-*passive*, as defined in [18, Appendix A.2]. We make the following assumptions regarding (7), which is widely known as *Nash stationarity* [17].

Assumption 3: With $p = q$, the following two conditions are equivalent.

1) $\sum_{j=1}^n x_j \rho_{ji}(p) - x_i \sum_{j=1}^n \rho_{ij}(p) = 0, \forall i \in \{1, \cdots, n\}$
2) $x_i(q_j - q_i) \leq 0, \ \forall i, j \in \{1, \cdots, n\}$

Note that when ρ_{ji} is defined as the Smith learning rule, (7) satisfies Assumption 3. Furthermore, in conjunction with Assumption 2, Assumption 3 implies that the feedback interconnection of (1) and (7), has a unique equilibrium.

In addition, according to Assumptions 2 and 3, the equilibrium state (q^*, x^*) satisfies $\mathcal{F}_i(q_i^*, x_i^*) = w_i, \ \forall i \in \{1, \cdots, n\}$ and $q_i^* = q_j^*, \ \forall i, j \in \{1, \cdots, n\}$. Consequently, we can infer that (q^*, x^*) is the optimal state at which the infinity norm $\|q(t)\|_\infty$ of the game state is minimized over the set $\mathbb{O}$ of stationary points of (1), defined by $\mathbb{O} = \{(q,x) \in [0, q_{\max}]^n \times \mathbb{X} \,|\, \mathcal{F}_i(q_i, x_i) = w_i, \ \forall i \in \{1, \cdots, n\}\}$. In other words, it holds that $\|q^*\|_\infty = \min_{(q,x)\in\mathbb{O}} \|q\|_\infty$.

C. Payoff Vector Estimation

In our problem formulation, instead of directly assessing $q(t)$, each agent k observes a function $y^{(k)}(t) = h^{(k)}(q(t))$ of the game state $q(t)$ and communicates with its neighbors to estimate $q(t)$. For instance, some agents may observe the full state $q(t)$, others might only take measurements of $q_i(t)$ associated with their strategy selection i, or yet others may not be able to collect any measurements of $q(t)$ at all. Since not every agent in the population can directly observe the full state $q(t)$, they adopt an estimation rule to infer the full state using their own observations $y^{(k)}(\tau), \tau \in [0, t]$ and information from their neighbors.

Motivated by the large literature on distributed state estimation [20], we consider that each agent k shares its own estimate $\hat{q}^{(k)}(t)$ of $q(t)$ whenever it can communicate with its neighbors. Let $\mathbb{N}_k(t)$ denote the set of neighbors of agent k at time t. Given its observation $y^{(k)}(t)$ of $q(t)$ and estimates $\{\hat{q}^{(l)}(t^-)\}_{l\in\mathbb{N}_k(t)}$ from its neighbors[2], the agent updates its estimate $\hat{q}^{(k)}(t)$ from which the estimate $\hat{p}^{(k)}(t)$ of the payoff vector $p(t)$ can be derived as $\hat{p}^{(k)}(t) = \hat{q}^{(k)}(t)$. We represent the estimation rule as

$$\hat{q}^{(k)}(t) = g\Big(\big\{\hat{q}^{(l)}(t^-)\big\}_{l\in\mathbb{N}_k(t)}, y^{(k)}(t)\Big). \quad (8)$$

We provide the following example to illustrate this.

Example 3: Given N agents in the population, suppose that only $N_{\text{leader}} (< N)$ leader agents can observe the game state $q(t)$, while the others cannot. Assuming that the agents are not aware of the game model (1), the estimation rule (8) can be implemented as

$$\hat{q}^{(k)}(t) = \begin{cases} q(t) & \text{if } k \text{ is a leader} \\ \frac{1}{|\mathbb{N}_k(t)|} \sum_{l\in\mathbb{N}_k(t)} \hat{q}^{(l)}(t^-) & \text{otherwise.} \end{cases} \quad (9)$$

[2]We use the notation $\hat{q}^{(l)}(t^-)$ to denote the estimate of agent l specifically before the agent updates its estimate at time t.

In other words, agent k sets its estimate $\hat{q}^{(k)}(t)$ to $q(t)$ if it is a leader and directly measures $q(t)$. Otherwise, the agent updates $\hat{q}^{(k)}(t)$ to the average of its neighbors' estimates.

The main results of this paper are applicable to any estimation rule g, provided it satisfies the following two assumptions: 1) Given that the game state $q(t)$ is bounded, the estimation error $\hat{q}^{(k)}(t) - q(t)$ also remains bounded for all $t \geq 0$, and 2) as the variation in $q(t)$ diminishes, i.e., $\|\dot{q}(t)\|_2 \to 0$ as $t \to \infty$, all agents in the population can asymptotically recover the full state of $q(t)$. We formally state these assumptions as follows:

Assumption 4: The estimation rule (8) satisfies the following conditions for every agent k:

$$\sup_{t\geq 0} \epsilon^{(k)}(t) < B_\epsilon \tag{10a}$$

$$\lim_{t\to\infty} \|\dot{q}(t)\|_2 = 0 \implies \lim_{t\to\infty} \epsilon^{(k)}(t) = 0, \tag{10b}$$

where $\epsilon^{(k)}(t) = \|\hat{q}^{(k)}(t) - q(t)\|_2$ and B_ϵ is a fixed positive constant.

Note that when the underlying communication graph is fixed and strongly connected, (9) satisfies Assumption 4.

III. Convergence Analysis and Revision Rate Update

We discuss how the estimation error $\epsilon^{(k)}(t)$ influences convergence of the population state and the game state. In a population of N agents, suppose each agent k adopts the learning rule $\rho_{ji}(\hat{p}^{(k)}(t))$ using its own payoff vector estimate $\hat{p}^{(k)}(t)$ for strategy revision. Let $\mathbb{M}_j^N(t)$ denote the set of agents adopting strategy j at time t. According to the definitions of the learning rule and the Poisson alarm clock, when the clock of an agent rings, the probability of that agent currently choosing strategy j and switching to strategy i can be specified as

$$x_j^N(t) \sum_{k\in\mathbb{M}_j^N(t)} \frac{\rho_{ji}(\hat{p}^{(k)}(t))}{|\mathbb{M}_j^N(t)|}, \tag{11}$$

where $x_j^N(t)$ is the fraction of the N-agent population adopting strategy j at time t, $\hat{p}^{(k)}(t) = \hat{q}^{(k)}(t)$ is agent k's estimate of the payoff vector $p(t)$, and $|\mathbb{M}_j^N(t)|$ denotes the cardinality of the set $\mathbb{M}_j^N(t)$. Note that if $\mathbb{M}_j^N(t)$ is empty, implying $x_j^N(t) = 0$, the expression in (11) is considered zero.

When the size of the population becomes arbitrarily large, that is, as N tends to infinity, we can derive

$$\lim_{N\to\infty} x_j^N(t) \sum_{k\in\mathbb{M}_j^N(t)} \frac{\rho_{ji}(\hat{p}^{(k)}(t))}{|\mathbb{M}_j^N(t)|} = x_j(t) \lim_{N\to\infty} \sum_{k\in\mathbb{M}_j^N(t)} \frac{\rho_{ji}(\hat{p}^{(k)}(t))}{|\mathbb{M}_j^N(t)|}, \tag{12}$$

where $x_j(t)$ represents the fraction of agents in the infinite population adopting strategy j. The limit $\lim_{N\to\infty} \sum_{k\in\mathbb{M}_j^N(t)} \frac{\rho_{ji}(\hat{p}^{(k)}(t))}{|\mathbb{M}_j^N(t)|}$ in the right-hand side of (12) represents the average strategy revision probability of all j-strategists. Assuming this limit exists and based on the definition of ρ_{ji} in (4), we can validate that

1) $0 \leq \lim_{N\to\infty} \sum_{k\in\mathbb{M}_j^N(t)} \frac{\rho_{ji}(\hat{p}^{(k)}(t))}{|\mathbb{M}_j^N(t)|} \leq 1$, and
2) $\sum_{i=1}^n \lim_{N\to\infty} \sum_{k\in\mathbb{M}_j^N(t)} \frac{\rho_{ji}(\hat{p}^{(k)}(t))}{|\mathbb{M}_j^N(t)|} = 1$.

Consequently, analogous to the derivation of (7), when agents revise their strategies based on estimates of the payoff vector and using Poisson alarm clocks with a rate of λ, the following EDM can be derived:

$$\begin{aligned}\dot{x}_i(t) &= \lambda \sum_{j=1}^n x_j(t) \lim_{N\to\infty} \sum_{k\in\mathbb{M}_j^N(t)} \frac{\rho_{ji}(\hat{p}^{(k)}(t))}{|\mathbb{M}_j^N(t)|} \\ &\quad - \lambda x_i(t) \sum_{j=1}^n \lim_{N\to\infty} \sum_{k\in\mathbb{M}_i^N(t)} \frac{\rho_{ij}(\hat{p}^{(k)}(t))}{|\mathbb{M}_i^N(t)|} \\ &= \lambda \sum_{j=1}^n x_j(t)\rho_{ji}(p(t)) - \lambda x_i(t) \sum_{j=1}^n \rho_{ij}(p(t)) + \lambda \xi_i(t),\end{aligned} \tag{13}$$

where $\xi_i(t)$ is defined as

$$\begin{aligned}\xi_i(t) &= \sum_{j=1}^n x_j(t) \left(\lim_{N\to\infty} \sum_{k\in\mathbb{M}_j^N(t)} \frac{\rho_{ji}(\hat{p}^{(k)}(t))}{|\mathbb{M}_j^N(t)|} - \rho_{ji}(p(t)) \right) \\ &\quad - x_i(t) \sum_{j=1}^n \left(\lim_{N\to\infty} \sum_{k\in\mathbb{M}_i^N(t)} \frac{\rho_{ij}(\hat{p}^{(k)}(t))}{|\mathbb{M}_i^N(t)|} - \rho_{ij}(p(t)) \right).\end{aligned}$$

According to its definition, $\xi_i(t)$ can be interpreted as the difference between the probability of agents' strategy revision based on the payoff vector $p(t)$ and the probability based on the estimates $\hat{p}^{(k)}(t)$.

By the Lipschitz continuity of ρ_{ji}, we compute the upper bound of $\xi_i(t)$ as

$$|\xi_i(t)| \leq c \sum_{j=1}^n x_j(t) \lim_{N\to\infty} \sum_{k\in\mathbb{M}_j^N(t)} \frac{\epsilon^{(k)}(t)}{|\mathbb{M}_j^N(t)|}. \tag{14}$$

Recall that $\epsilon^{(k)}(t)$ is defined as $\epsilon^{(k)}(t) = \|\hat{q}^{(k)}(t) - q(t)\|_2$. According to Assumption 4, since $\epsilon^{(k)}(t)$ is bounded, so does $\xi_i(t)$. Also, we can infer that $\xi_i(t)$ vanishes as does the estimation error $\epsilon^{(k)}(t)$ at every agent k.

A. Convergence to Equilibrium State

To visualize the effect of the estimation error on the convergence, we conduct simulations using the scenario

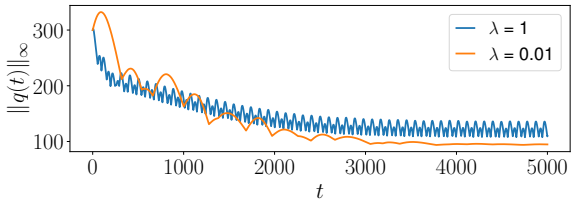

Fig. 1. Trajectories of the game state derived by the Smith learning rule in the task allocation game described in Example 1, where the agents are estimating the payoff vector using (9).

explained in Example 1, the Smith learning rule, explained in Example 2, and the estimation rule (9). As we use the same simulation setup as described in Section IV, except λ is fixed here, we refer to that section for details of the simulation setup.

The simulations are executed with two different revision rates, $\lambda = 0.01$ and 1.0, and the results are depicted in Fig. 1. When the revision rate is high, $\|q(t)\|_\infty$ experiences a steep decrease early on; however, in the long run, its value tends to oscillate and be higher compared to the scenario with a lower revision rate. Conversely, when the revision rate is low, the agents have fewer opportunities to revise their strategies, resulting in a larger overshoot in the early stages. Based on these observations, to ensure rapid convergence in the initial stages, the rate should be set sufficiently high; however, to achieve a lower value of $\|q(t)\|_\infty$ in the long term, the rate should be reduced.

Therefore, the optimal design of the strategy revision rate would necessitate a decrease over time. To provide a rigorous justification for our observation, we present the following theorem. The proof of the theorem is provided in [18, Appendix B].

Theorem 1: For a given revision rate λ, let $q_\lambda(t)$ and $x_\lambda(t)$ be the game and population states, respectively, determined by the feedback interconnection of the game model (1) and EDM (13). Under Assumptions 1-4, it holds that

$$\limsup_{t\to\infty} \left(\|q_\lambda(t) - q^*\|_2 + \|x_\lambda(t) - x^*\|_2\right) \to 0 \text{ as } \lambda \to 0,$$

where (q^*, x^*) is the unique equilibrium state of the closed loop system defined by (1) and (7) with $p(t) = q(t)$.

B. Revision Rate Update

As we stated in Theorem 1 and observed from the simulation results depicted in Fig. 1, starting with a high initial revision rate λ, followed by its proper regulation, ensures the convergence of $(q(t), x(t))$ to the equilibrium state, while particularly achieving steep convergence at the early stages of the game. The proof of the theorem utilizes the so-called *storage functions* $\mathcal{L}(q(t), x(t))$ and $\mathcal{S}(p(t), x(t))$ of the game model (1) and EDM (7), respectively, which are defined in [18, Appendix A], to construct a Lyapunov candidate function. Notably, by decreasing λ, we establish that both the storage functions diminish over time, thereby ensuring the convergence to the equilibrium state. This process requires knowledge of the game model (1) and its associated storage function $\mathcal{L}(q(t), x(t))$. However, if the game model (1) is unknown to the agents, this method of updating λ using the game model becomes infeasible.

Instead, we propose updating λ when the frequency of the agents revising to other strategies decreases, despite receiving revision opportunities. To implement this, we consider a time-varying revision rate $\lambda(t)$. Let $\{t_m\}_{m=1}^\infty$ and $\{\lambda_m\}_{m=1}^\infty$ be sequences of time instants and rates, respectively, where $\lambda(t)$ of each agent's Poisson alarm clock is defined as $\lambda(t) = \lambda_m$ for $t \in [t_m, t_{m+1})$. An agent with the most accurate estimates of $(q(t), x(t))$, such as the leader agents in Example 3, updates the revision rate to $\lambda_{m+1} = \gamma\lambda_m$ with $\gamma \in (0, 1)$ at time t_{m+1} if the following condition is met:

$$\nabla_x^T \mathcal{S}(p(t_{m+1}), x(t_{m+1}))\mathcal{V}(p(t_{m+1}), x(t_{m+1})) \geq -\epsilon, \quad (15)$$

where $\mathcal{V} = (\mathcal{V}_1, \cdots, \mathcal{V}_n)$ with $\mathcal{V}_i(p, x) = \sum_{j=1}^n x_j \rho_{ji}(p) - x_i \sum_{j=1}^n \rho_{ij}(p)$ and ϵ is a small positive constant. It then broadcasts the updated rate λ_{m+1} to other agents using the same communication graph employed for payoff vector estimation. According to [18, Eq. (23)], with small ϵ, (15) implies that $\mathcal{V}(p(t_{m+1}), x(t_{m+1}))$ becomes small. Consequently, the agents change their strategies less frequently.

Additionally, we require $t_{m+1} - t_m \geq \frac{\tau}{\lambda_m}$ to ensure that every agent receives a strategy revision opportunity with a certain probability dependent on a constant τ before the revision rate is updated. To understand this, by definition, $1 - e^{-\tau}$ represents the probability of a revision opportunity occurring within the time interval $[t_m, t_m + \frac{\tau}{\lambda_m})$. Therefore, the inequality requirement suggests that with a larger τ, it becomes increasingly likely that every agent will receive a revision opportunity before the revision rate updates at t_{m+1}.

IV. Simulations

To illustrate our analysis and also to validate the strategy revision rate update method, we design and carry out simulations based on Example 1 with $n = 3$. The parameters of (3) are set as $R_1 = R_2 = R_3 = 3.44$, $\alpha_1 = \alpha_2 = \alpha_3 = 0.036$, and $\beta_1 = \beta_2 = \beta_3 = 0.91$. The increase rate $w = (w_1, w_2, w_3)$ is specified as $w = (0.5, 1, 2)$. It can be verified that $\mathcal{F}_i$ satisfies Assumption 1. Additionally, by setting a sufficiently large value for $q_{\max}$, we can validate that Assumption 2 holds.

In simulation, there are 3000 agents in the population and they can share their estimates of the payoff vector via a strongly connected graph, generated by the Erdős-Rényi model with the edge formation probability $p = 0.1$. Among the entire population, we uniformly randomly select 10% of the total population as the leaders who can directly observe the game state $q(t)$, and the rest of the population cannot observe the payoff vector at all, as in the scenario explained in Example 3. Each agent k computes its own estimate $q^{(k)}(t)$ according to the estimation rule described in (9), and revises its strategy using the Smith learning rule $\rho_{ji}(p^{(k)}(t)) = \varrho[p_i^{(k)}(t) - p_j^{(k)}(t)]_+$ with $\varrho = 1/400$.[3]

The communication and observation of the game state by the agents occur at discrete time instants, specifically at $t = 1, 2, 3, \cdots$. For its strategy revision, each agent takes samples of time instants based on its Poisson alarm clock, and at each sampled time, it revises its strategy using the Smith learning rule and estimated payoff vector.

We iterate the simulation over a range of parameters for the strategy revision update method, fixing the initial conditions to $x(0) = (1/3, 1/3, 1/3)$ and $q(0) = (100, 200, 300)$, and setting the initial estimates of $q(0)$ by all agents to $\hat{q}^{(k)}(0) = (0, 0, 0)$. Fig. 2 illustrates the simulation results

[3]The choice $\varrho = 1/400$ is made to ensure that $\varrho \sum_{i=1}^n [p_i^{(k)}(t) - p_j^{(k)}(t)]_+ \leq 1$ throughout the simulations.

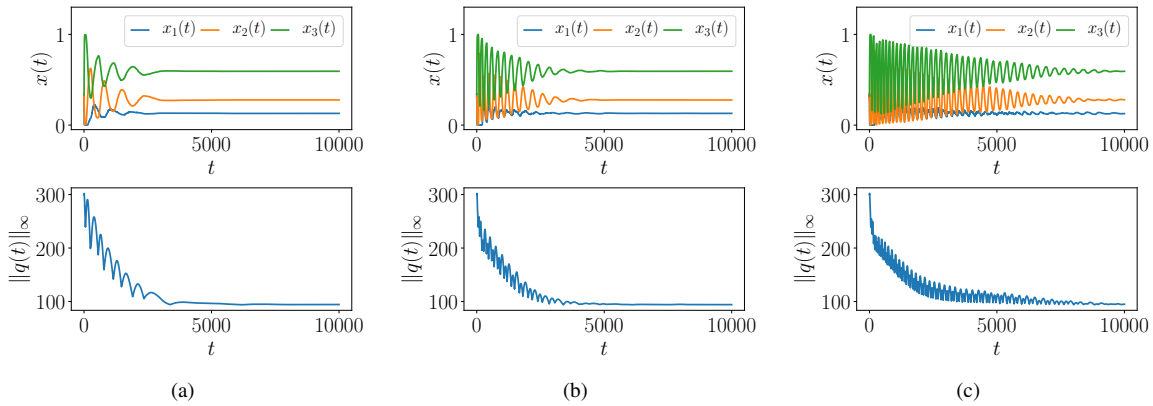

Fig. 2. Population state and game state trajectories when the rate of the Poisson alarm clock is updated according to the method described in Section III-B. The trajectories are examined using three different parameter choices: (a) $\gamma = 0.8, \tau = 0.2$, (b) $\gamma = 0.95, \tau = 1.0$, and (c) $\gamma = 0.99, \tau = 1.4$.

for three cases: (a) $\gamma = 0.8, \tau = 0.2$, (b) $\gamma = 0.95, \tau = 1.0$, and (c) $\gamma = 0.99, \tau = 1.4$, with the initial revision rate $\lambda(t_0)$ set to 1 in all three cases. We fix $\epsilon = 0.01$ as the left-hand side term in (15) approaches zero over time in the simulation. In all cases, the revision rates decrease over time, and we can observe the convergence of the population and game states to the equilibrium state, as we have discussed in Theorem 1. When the parameters γ and τ are small, the revision rate decreases fast, resulting in the substantial overshoot in the game state trajectory, as observed in Fig. 2(a). As those parameters increase, the game state tends to have a steeper decrease in the early stage, as observed in Fig. 2(b). However, if these parameters are too large, the reduction of the revision rate becomes slow and convergence to the equilibrium takes longer, as illustrated in Fig. 2(c).

V. Conclusion

We investigated a population game problem in which decision-making agents need to estimate the payoff vector for their strategy selections. By adopting task allocation games, we examined the effect of errors in the payoff vector estimation on the convergence to the equilibrium state. Our analysis showed how the adaptive rate of the agents' strategy revisions mitigates the impact of these errors on the convergence. Leveraging the analytical results, we proposed a design for time-varying strategy revision rates to ensure convergence. As a future direction of research, we plan to explore optimization of the time-varying revision rate design. Also, analyzing how the topology of the communication graph affects the convergence and investigating the optimal communication topology design are topics of our interest.

References

[1] N. Mehr and R. Horowitz, "How will the presence of autonomous vehicles affect the equilibrium state of traffic networks?" *IEEE Transactions on Control of Network Systems*, vol. 7, no. 1, pp. 96–105, 2020.

[2] S. H. Li, Y. Yu, N. I. Miguel, D. Calderone, L. J. Ratliff, and B. Açıkmeşe, "Adaptive constraint satisfaction for markov decision process congestion games: Application to transportation networks," *Automatica*, vol. 151, p. 110879, 2023.

[3] W. Saad, Z. Han, T. Basar, M. Debbah, and A. Hjorungnes, "Hedonic coalition formation for distributed task allocation among wireless agents," *IEEE Transactions on Mobile Computing*, vol. 10, no. 9, pp. 1327–1344, 2011.

[4] H. Tembine, E. Altman, R. ElAzouzi, and W. H. Sandholm, "Evolutionary game dynamics with migration for hybrid power control in wireless communications," in *47th IEEE Conference on Decision and Control (CDC)*, 2008, pp. 4479–4484.

[5] 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: The advantages of evolutionary game theory," *IEEE Control Systems Magazine*, vol. 37, no. 1, pp. 70–97, 2017.

[6] S. Park, Y. D. Zhong, and N. E. Leonard, "Multi-robot task allocation games in dynamically changing environments," in *2021 IEEE International Conference on Robotics and Automation (ICRA)*, 2021, pp. 8678–8684.

[7] I. Jang, H.-S. Shin, and A. Tsourdos, "Anonymous hedonic game for task allocation in a large-scale multiple agent system," *IEEE Transactions on Robotics*, vol. 34, no. 6, pp. 1534–1548, 2018.

[8] S. Amaya and A. Mateus, "Tasks allocation for rescue robotics: A replicator dynamics approach," in *Artificial Intelligence and Soft Computing*, L. Rutkowski, R. Scherer, M. Korytkowski, W. Pedrycz, R. Tadeusiewicz, and J. M. Zurada, Eds. Cham: Springer International Publishing, 2019, pp. 609–621.

[9] W. H. Sandholm, *Population Games and Evolutionary Dynamics*. MIT Press, 2011.

[10] A. Mahajan, N. C. Martins, M. C. Rotkowitz, and S. Yüksel, "Information structures in optimal decentralized control," in *2012 IEEE 51st IEEE Conference on Decision and Control (CDC)*, 2012, pp. 1291–1306.

[11] S. Park and J. Barreiro-Gomez, "Payoff mechanism design for coordination in multi-agent task allocation games," in *2023 62nd IEEE Conference on Decision and Control (CDC)*, 2023, pp. 8116–8121.

[12] G. Szabó and G. Fáth, "Evolutionary games on graphs," *Physics Reports*, vol. 446, no. 4, pp. 97–216, 2007.

[13] J. Barreiro-Gomez, G. Obando, and N. Quijano, "Distributed population dynamics: Optimization and control applications," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 47, no. 2, pp. 304–314, 2017.

[14] J. C. Harsanyi, "Games with incomplete information played by "bayesian" players, i-iii. part ii. bayesian equilibrium points," *Management Science*, vol. 14, no. 5, pp. 320–334, 1968.

[15] D. Foster and P. Young, "Stochastic evolutionary game dynamics*," *Theoretical Population Biology*, vol. 38, no. 2, pp. 219–232, 1990.

[16] M. J. Fox and J. S. Shamma, "Population games, stable games, and passivity," *Games*, vol. 4, pp. 561–583, Oct. 2013.

[17] S. Park, N. C. Martins, and J. S. Shamma, "From population games to payoff dynamics models: A passivity-based approach," in *2019 IEEE 58th Conference on Decision and Control (CDC)*, 2019, pp. 6584–6601.

[18] S. Park, "Learning equilibrium with estimated payoffs in population games," 2024. [Online]. Available: https://arxiv.org/abs/2407.06328

[19] M. J. Smith, "The stability of a dynamic model of traffic assignment—an application of a method of Lyapunov," *Transportation Science*, vol. 18, no. 3, pp. 245–252, 1984.

[20] F. F. Rego, A. M. Pascoal, A. P. Aguiar, and C. N. Jones, "Distributed state estimation for discrete-time linear time invariant systems: A survey," *Annual Reviews in Control*, vol. 48, pp. 36–56, 2019.