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

Periodicity in dynamical games driven by the Hedge algorithm and myopic best response

Xinxiang Guo[1], Yifen Mu[2] and Xiaoguang Yang[3]

***Abstract*— In this paper, we consider the $n \times n$ two-payer zero-sum repeated game in which one player (player X) employs the popular Hedge (also called multiplicative weights update) learning algorithm while the other player (player Y) adopts the myopic best response. The theoretical analysis on the dynamics of such game system is still rare, which is however of promising interests. We investigate the dynamics of such Hedge-myopic system by defining a metric $Q(\mathbf{x}_t)$, which measures the distance between the stage strategy $\mathbf{x}_t$ and Nash Equilibrium (NE) strategy of player X. We analyze the trend of $Q(\mathbf{x}_t)$ and prove that it is bounded and can only take finite values on the evolutionary path when payoffs are all rational numbers and the game has an interior NE. Based on this, we prove that the stage strategy sequence of both players are periodic after finite stages and the time-averaged strategy of player Y within one period is an exact NE strategy.**

I. INTRODUCTION

Game theory is widely used to model interactions and competitions among self-interested and rational agents in the real world [1], [2]. In such situations, the utility of each agent is determined by the actions of all the other agents, leading to the solution concept of equilibrium, in which Nash equilibrium (NE) is a central one. This makes the NE-solving one of the most significant problems in game theory, which however is very hard for general games.

The notion of NE [3] aims to describe a stable state where each participant makes the optimal choice considering the strategies of others and thus has no incentive to change the strategy unilaterally. Due to the mutual influence of participants in the game and the existence of possible multiple equilibrium points in different scenarios, NE-solving is a difficult problem, which has been proven to be PPAD-hard for general games [4], [5]. For the special two-player zero-sum games, linear programming [6] provides a powerful method to solve NE in polynomial time [7]. However, in practical scenarios, because of large scalability issues, imperfect information and the complexity of multiple-stage dynamics, two-player zero-sum game is still a subject of ongoing investigation and attracts attentions from researchers in different fields [8], [9]. Especially, participants are often not perfectly rational, leading to research which aimed at approximating NE from a learning perspective [10].

Concerning learning in games, there has been a long history and plentiful literature, where a lot of learning algorithms have been proposed according to different settings. For instance, under the Fictitious Play algorithm [10], the empirical distribution of actions taken by each player converges to NE if the stage game has generic payoffs and is 2×2 [11] or zero-sum game [12] or potential game [13]. When each player employs the no-regret algorithm [14] to determine their stage strategy in repeated games, their time-averaged strategy profile converges to the coarse correlated equilibrium in general-sum games [15] and to NE in two-player zero-sum games. In imperfect-information extensive-form games, Zinkevich et al. [16] proposed the counterfactual regret minimization (CFR) algorithm and proved its convergence to NE in the two-player zero-sum setting. Based on these methods, lots of variants were proposed and widely used in solving equilibrium in complicated games [17], [18], [19], [20]. Note that in all these works, every player in the game adopts the same learning algorithm and the convergence results are based on the term of time-averaged strategy.

However, further investigation on the learning dynamics shows that even in simple game models, basic learning algorithms can lead to highly complex behavior and may not converge [21], [22], [23], [24]. Palaiopanos et al. [25] discovered specific instances of 2×2 potential games where the behavior of multiplicative weights update (MWU) algorithm exhibits bifurcation at the critical value of its step size. Bailey and Piliouras [26] showed that in two-player zero-sum games, when both players adopt the MWU algorithm, the system dynamics deviate from equilibrium and converge towards the boundary of the strategy simplex. Mertikopoulos et al. [8] studied the regularized learning algorithms in two-player zero-sum games and proved the Poincaré recurrence of the system behavior, implying the impossibility of convergence to NE from any initial strategy profile. Generally speaking, there is no systematic framework for analyzing the limiting behavior of these repeated games [27], [28].

For the asymmetric case, as emphasized in [29], the limiting behavior of dynamical processes where players adhere to different update rules is an open question, even

*This work was supported by the Strategic Priority Research Program of Chinese Academy of Sciences under Grant XDA27030201, the National Key Research and Development Program of China under grant No.2022YFA1004600 and the Natural Science Foundation of China under Grant T2293770.

[1]Xinxiang Guo is with School of Mathematical Sciences, University of Chinese Academy of Sciences and The Key Laboratory of Systems and Control, Institute of Systems Science, Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing 100190, China. guoxinxiang@amss.ac.cn

[2]Yifen Mu is the corresponding author and with The Key Laboratory of Systems and Control, Institute of Systems Science, Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing 100190, China. mu@amss.ac.cn

[2]Xiaoguang Yang is with The Key Laboratory of Management, Decision and Information System, Institute of Systems Science, Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing 100190, China. xgyang@iss.ac.cn

for potential games. Therefore, related theoretical analysis of such system is extremely rare [30], [31]. Our previous work [32], [33] studied such dynamical system where one player employs the Hedge algorithm and the other player takes the globally or locally optimal strategy in finitely repeated two-player zero-sum games and proved its periodicity when the game is 2×2. Our investigation reveals that the detailed understanding about the dynamics can facilitate the design of novel algorithms with special properties, thereby suggesting a promising avenue for developing learning algorithms with better performance.

This paper will consider the general $n \times n$ zero-sum game and investigate the dynamics of repeated game under asymmetric updating rules. To be specific, we will study the dynamics of the repeated game where one player (player X) employs the Hedge algorithm to update his stage strategy and the other player (player Y) adopts the according myopic best response to the stage strategy of player X. The main contributions of this paper can be summarized as follows.

(1) This paper considers the Hedge-myopic system and investigates its dynamics by analyzing the trend of a quantity called $Q(\mathbf{x})$ based on the Kullback-Leibler divergence, which measures the distance between the stage strategy $\mathbf{x}$ and the NE strategy of player X.
(2) For the game with payoffs being rational numbers and with an interior NE, we prove that along the strategy sequence $\mathbf{x}_t$ the Q-sequence $Q(\mathbf{x}_t)$ is bounded and $\mathbf{x}_t$ can only take finite values on the evolutionary path. This implies that the strategy sequence $\mathbf{x}_t$ of player X will not converge to the NE strategy and justifies the finding in the literature.
(3) Using the dynamic property, this paper theoretically proves that the stage strategy sequences of both players are periodic after finite stages for the game with payoffs being rational numbers and with an interior NE. Additionally, the time-averaged strategy of player Y within one period is an exact NE strategy.

Paper Organization: Section II provides the preliminary knowledge and problem formulation; Section III presents the main results regarding the periodicity of the system behavior; Section IV concludes the paper.

II. Preliminary and Problem Formulation

A. Normal-form zero-sum Game

Consider a two-player zero-sum normal-form game Γ. The players are called player X and player Y. Suppose there are $n, n \geq 2$ feasible actions for each player. We denote the action set of player X by $\mathcal{I} = \{1, 2, \cdots, n\}$ and the action set of player Y by $\mathcal{J} = \{1, 2, \cdots, n\}$. For each action profile (i, j), the payoff obtained by player Y is a_{ij} and thus the payoff obtained by player X is $-a_{ij}$ since the game is zero-sum. Naturally, the payoff of the game is shown by a matrix $A = \{a_{ij}\}_{i \in \mathcal{I}, j \in \mathcal{J}}$. The matrix is called the payoff matrix for player Y and the loss matrix for player X. A mixed strategy of a player is a probability distribution over his/her action set. The mixed strategies of player X and player Y are denoted by $\mathbf{x}$ and $\mathbf{y}$, respectively. The bold font is used to emphasize that $\mathbf{x}$ and $\mathbf{y}$ are both vectors. Given the mixed strategy profile $(\mathbf{x}, \mathbf{y})$, the payoff of player Y is $\mathbf{x}^T A \mathbf{y}$ and the payoff of player X is $-\mathbf{x}^T A \mathbf{y}$. Denote their mixed strategy sets by $\Delta(\mathcal{I})$ and $\Delta(\mathcal{J})$, respectively.

A strategy profile $(\mathbf{x}, \mathbf{y})$ is a Nash equilibrium (NE) of the game if for any strategy $\mathbf{x}'$ and $\mathbf{y}'$,

$$\mathbf{x}^T A \mathbf{y} \geq \mathbf{x}^T A \mathbf{y}' \quad \text{and} \quad \mathbf{x}^T A \mathbf{y} \leq (\mathbf{x}')^T A \mathbf{y}.$$

Write the NE strategy profile as $(\mathbf{x}^*, \mathbf{y}^*)$. Then the value of the game is

$$v^* := (\mathbf{x}^*)^T A \mathbf{y}^*. \tag{1}$$

Denote the support of $\mathbf{x}^*$ by $\text{supp}(\mathbf{x}^*) = \{i \in \mathcal{I} : x_i^* > 0\}$ and the support of $\mathbf{y}^*$ by $\text{supp}(\mathbf{y}^*) = \{j \in \mathcal{J} : y_j^* > 0\}$. A NE is said to be interior (or fully-mixed) if $\text{supp}(\mathbf{x}^*) = \mathcal{I}$ and $\text{supp}(\mathbf{y}^*) = \mathcal{J}$.

A strategy profile $(\mathbf{x}, \mathbf{y})$ is called a $\varepsilon-$Nash equilibrium ($\varepsilon-$NE) if for all $\mathbf{x}' \in \Delta(\mathcal{I})$ and $\mathbf{y}' \in \Delta(\mathcal{J})$, we have

$$\mathbf{x}^T A \mathbf{y} \geq \mathbf{x}^T A \mathbf{y}' - \varepsilon \quad \text{and} \quad \mathbf{x}^T A \mathbf{y} \leq (\mathbf{x}')^T A \mathbf{y} + \varepsilon.$$

If the game admits an unique interior NE strategy $\mathbf{x}^*$ of player X, we denote the cross entropy between strategy $\mathbf{x}$ and $\mathbf{x}^*$ by function $Q(\mathbf{x})$, i.e.,

$$Q(\mathbf{x}) \triangleq -\sum_{i=1}^{n} x_i^* \ln x_i \tag{2}$$

where $0 < x_i < 1$ for all $i \in \mathcal{I}$, i.e., $\mathbf{x}$ belongs to the relative interior of region $\Delta(\mathcal{I})$. It is easy to see that $Q(\mathbf{x}) > 0$.

In the following, we will focus on infinitely repeated game. We denote $\mathbf{x}_t$, $\mathbf{y}_t$ the stage strategy of player X and player Y at time t respectively. Then, the instantaneous expected payoff of player Y is $\mathbf{x}_t^T A \mathbf{y}_t$. Below we will study the dynamic characteristics of the repeated games driven by the Hedge algorithm and the myopic best response.

B. Problem Formulation

First, let player X update his stage strategy according to the Hedge algorithm. Hedge algorithm is a popular no-regret learning algorithm proposed by Freund and Schapire[34]. Hedge algorithm is also known as weighted majority [35] or exponential weighted average prediction [14], or multiplicative weights update [36].

Specifically, the strategy of player X at time t is updated by

$$x_{i,t} = \frac{\exp(-\eta \sum_{\tau=1}^{t-1} e_i^T A \mathbf{y}_\tau)}{\sum_{j=1}^{n} \exp(-\eta \sum_{\tau=1}^{t-1} e_j^T A \mathbf{y}_\tau)}, \quad i = 1, 2, \cdots, n, \tag{3}$$

where η is called *the learning rate*, which is a constant parameter. In this paper, we assume that η is sufficiently small and determined by the payoff matrix A.

From formula (3), $\mathbf{x}_t$ is fully determined by $\mathbf{y}_1$, $\mathbf{y}_2$, ..., $\mathbf{y}_{t-1}$. Further, we can compute $\mathbf{x}_t$ from $\mathbf{x}_{t-1}$ and $\mathbf{y}_{t-1}$ and get

$$x_{i,t} = \frac{x_{i,t-1}\exp(-\eta e_i^T A\mathbf{y}_{t-1})}{\sum_{j=1}^n x_{j,t-1}\exp(-\eta e_j^T A\mathbf{y}_{t-1})}, \quad \forall\ i = 1, 2, \cdots, n. \tag{4}$$

Then, let player Y only consider maximizing her instantaneous expected payoff and take myopic best response to $\mathbf{x}_t$ at each time t. In most cases, the myopic best response is unique and is pure strategy. When the best response is not unique, we stipulate that player Y chooses the pure strategy with the smallest subscript in $\mathcal{J}$, i.e.,

$$\mathbf{y}_t = BR(\mathbf{x}_t) \triangleq \mathbf{y}^{\bar{j}}, \tag{5}$$

where $\bar{j} = \min\{j \in \mathcal{J} : \mathbf{x}_t^T A\mathbf{y}^j = \max_{\mathbf{y}\in\Delta(\mathcal{J})} \mathbf{x}_t^T A\mathbf{y}\}$ and $\mathbf{y}^j$ is a pure-strategy vector with only the j-th element is 1 and all the other elements are 0. Then, the strategy of player Y at each stage is well-defined and is pure strategy. Apparently $\mathbf{y}_t$ is totally determined by $\mathbf{x}_t$. Combined with (4), we know that $\mathbf{x}_{t+1}$ is fully determined by $\mathbf{x}_t$, with no randomness involved.

Given the action rule of player X and Y as above, the infinitely repeated game is intrinsically determined and the system is called *the Hedge-myopic system*. In this paper, we will study the dynamic characteristics of such a dynamical system and try to answer the questions like: Is the system periodic? Does the system converge?

III. Main Results

A. Rational Games with an Interior Equilibrium

In this section, we prove that for the game with payoffs being rational numbers and with an interior NE, the dynamics of the Hedge-myopic system is periodic after finite stages. Some proofs are omitted due to page limitation and see the full paper [37] for more details.

We state some assumptions below:

Assumption 1: The payoffs are all rational numbers and the game has an unique interior NE, denoted by $(\mathbf{x}^*, \mathbf{y}^*)$;

Assumption 2: The matrices $A_i \in \mathbb{R}^{n\times n}$ are all non-singular for $i = 1, 2, ..., n$ where A_i is defined as

$$\begin{pmatrix} a_{1,1} & a_{2,1} & \cdots & a_{n,1} \\ \vdots & & & \vdots \\ a_{1,i-1} & a_{2,i-1} & \cdots & a_{n,i-1} \\ a_{1,i+1} & a_{2,i+1} & \cdots & a_{n,i+1} \\ \vdots & & & \vdots \\ a_{1,n} & a_{2,n} & \cdots & a_{n,n} \\ 1 & 1 & \cdots & 1 \end{pmatrix}. \tag{6}$$

Recall formula (2) and (3), we denote

$$\begin{aligned} Q_t = Q(\mathbf{x}_t) &= -\sum_{i=1}^n x_i^* \ln x_{i,t} \\ &= \ln\left(\sum_{i=1}^n e^{\eta\sum_{\tau=1}^{t-1}(v^* - e_i^T A\mathbf{y}_\tau)}\right), \end{aligned} \tag{7}$$

where v^* is the value of the game. The quantity Q_t measures the level of resemblance between the strategy $\mathbf{x}_t$ and the NE strategy $\mathbf{x}^*$. We call $\{Q_t\}$ *the Q-sequence*.

By (7), we can calculate

$$Q_{t+1} - Q_t = \ln\left(\sum_{i=1}^n x_{i,t} e^{\eta(v^* - e_i^T A\mathbf{y}_t)}\right). \tag{8}$$

Then, we have the below theorem.

Theorem 3.1: If **Assumption 1** and **Assumption 2** hold, then there exists a positive number M_Q such that $0 < Q_t \le M_Q$ for all $t \ge 1$, i.e., the Q-sequence is bounded.

To prove Theorem 3.1, we need to study the sequence $\mathbf{x}_t$ and Q_t in details. To this end, define

$$D(\mathbf{x}) \triangleq \ln\left(\sum_{i=1}^n x_i e^{\eta(v^* - \mathbf{e}_i^T A\mathbf{y}_x)}\right), \quad \mathbf{x} \in \Delta(\mathcal{I}) \tag{9}$$

where $\mathbf{y}_x = BR(\mathbf{x}) \triangleq \mathbf{y}^{\bar{j}}$ with $\bar{j} = \min\{j \in \mathcal{J} : \mathbf{x}^T A\mathbf{y}^j = \max_{\mathbf{y}\in\Delta(\mathcal{J})} \mathbf{x}^T A\mathbf{y}\}$. Take $\mathbf{x} = \mathbf{x}_t$, we have $D(\mathbf{x}_t) = Q_{t+1} - Q_t$. Obviously, if $D(\mathbf{x}_t) < 0$, then $Q_{t+1} < Q_t$; if $D(\mathbf{x}_t) \ge 0$, then $Q_{t+1} \ge Q_t$.

Depending on whether $D(\mathbf{x})$ is positive, for the mixed strategy set $\Delta(\mathcal{I})$, define

$$Z_p \triangleq \{\mathbf{x} \in \Delta(\mathcal{I}) : D(\mathbf{x}) \ge 0\}$$

and

$$Z_n \triangleq \{\mathbf{x} \in \Delta(\mathcal{I}) : D(\mathbf{x}) < 0\}.$$

Easy to see that the NE strategy $\mathbf{x}^* \in Z_p$, hence $Z_p \neq \emptyset$.

To locate the regions Z_p and Z_n on $\Delta(\mathcal{I})$, we need the following inequality in probability theory.

Lemma 3.1 (Lemma A.1 of [14]): Let random variable $\mathbf{z}$ satisfying $a \le \mathbf{z} \le b$. Then, for any $s \in \mathbb{R}$,

$$\ln \mathbb{E}(e^{s\mathbf{z}}) \le s\mathbb{E}\mathbf{z} + \frac{s^2(b-a)^2}{8}.$$

Applying Lemma 3.1 , we can estimate $D(\mathbf{x})$ as below.

$$\begin{aligned} D(\mathbf{x}) &= \ln\left(\sum_{i=1}^n x_i e^{\eta(v^* - e_i^T A\mathbf{y}_x)}\right) \\ &\le \sum_{i=1}^n x_i \eta(v^* - e_i^T A\mathbf{y}_x) + \frac{\eta^2\delta^2}{8} \\ &= \eta(v^* - x^T A\mathbf{y}_x + \frac{\eta\delta^2}{8}) \end{aligned}$$

where $\delta = \max_{i\in\mathcal{I}, j\in\mathcal{J}} a_{i,j} - \min_{i\in\mathcal{I}, j\in\mathcal{J}} a_{i,j}$. Then, depending on whether $v^* - \mathbf{x}^T A\mathbf{y}_x + \frac{\eta\delta^2}{8}$ is positive, we further split the set $\Delta(\mathcal{I})$ into another two regions Z_u and Z_d where

$$Z_u \triangleq \{\mathbf{x} \in \Delta(\mathcal{I}) : v^* - \mathbf{x}^T A\mathbf{y}_x + \frac{\eta\delta^2}{8} \ge 0\}$$

and

$$Z_d \triangleq \{\mathbf{x} \in \Delta(\mathcal{I}) : v^* - \mathbf{x}^T A\mathbf{y}_x + \frac{\eta\delta^2}{8} < 0\}.$$

For the strategy $\mathbf{x} \in \Delta(\mathcal{I})$, if $\mathbf{x} \in Z_d$, immediately we have $D(\mathbf{x}) < 0$, i.e., $\mathbf{x} \in Z_u$. Hence, we have **Claim 1** below:

Claim 1: $Z_d \subset Z_n$ and $Z_p \subset Z_u$.

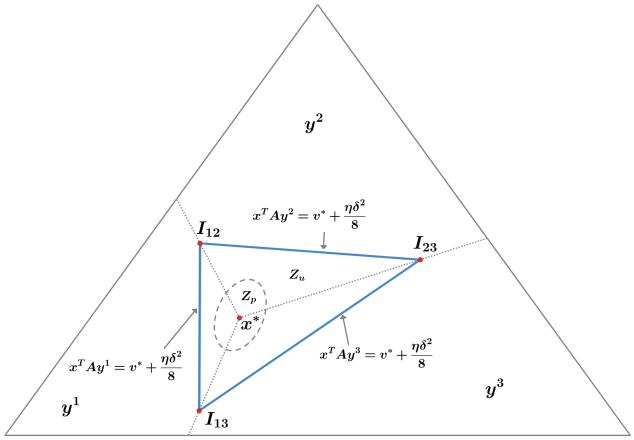

Fig. 1: Graphical illustration of Z_p, Z_u and NE for 3×3 games, whose payoff matrix is given in Example 3.1.

Figure 1 gives a graphical illustration about the region Z_p, Z_u, $\Delta(\mathcal{I})$ and interior NE strategy for a 3×3 game. In this figure, the entire triangular region represents the simplex $\Delta(\mathcal{I})$, and the red point in the center represents the NE strategy $\mathbf{x}^*$ of player X; the gray dashed lines divide the triangular region into three areas, labeled $\mathbf{y}^1, \mathbf{y}^2$, and $\mathbf{y}^3$ respectively, which represents the best response action of player Y to the strategies in that area; the triangular region enclosed by the blue solid line is region Z_u; within region Z_u, the elliptical region enclosed by the gray dashed line (the actual region Z_p may not necessarily be elliptical) is region Z_p, and the NE strategy $\mathbf{x}^*$ is located in region Z_p.

For the region $Z_u = \{\mathbf{x} \in \Delta(\mathcal{I}) : \mathbf{x}^T A \mathbf{y}_x \leq v^* + \frac{\eta\delta^2}{8}\}$, we claim that:

Claim 2: The region Z_u can be further rewritten as $Z_u = \{\mathbf{x} \in \Delta(\mathcal{I}) : \max_{\mathbf{y}^j} \mathbf{x}^T A \mathbf{y}^j \leq v^* + \frac{\eta\delta^2}{8}\} = \{\mathbf{x} \in \Delta(\mathcal{I}) : A^T \mathbf{x} \leq \mathbf{b}\}$, where $\mathbf{b} = [v^* + \frac{\eta\delta^2}{8}, v^* + \frac{\eta\delta^2}{8}, \cdots, v^* + \frac{\eta\delta^2}{8}]^T$. Thus, the region Z_u is a bounded polyhedron.

For the bounded polyhedron, we have the following result, called Representation of Bounded Polyhedra, which can be found in the Theorem 2.9 of [38].

Lemma 3.2: A bounded polyhedron is the set of all convex combinations of its vertices.

In Figure 1, the points I_{12}, I_{13} and I_{23} are the vertices of polyhedron Z_u. Thus by Lemma 3.2, every point in Z_u can be written as a convex combination of these vertex points. The purpose of assuming the learning rate η to be sufficiently small is to ensure that the vertex points of region Z_u lie in the strict interior of $\Delta(\mathcal{I})$. Then, we can prove that the polyhedron Z_u must lie in the strict interior of $\Delta(\mathcal{I})$.

Proposition 3.1: If $\mathbf{x} \in Z_u$, then the elements of $\mathbf{x}$ are uniformly lower bounded. That is, for all $\mathbf{x} \in Z_u$, there exists $\varepsilon_d > 0$ such that $x_i > \varepsilon_d$ for all $i \in \mathcal{I}$.

Based on Proposition 3.1, we can obtain the following corollary.

Corollary 3.1: In the Hedge-myopic system, if $\mathbf{x} \in Z_u$, then $Q(\mathbf{x})$ is upper bounded, i.e., there exists $M_p > 0$ such that $Q(\mathbf{x}) \leq M_p$ for all $\mathbf{x} \in Z_u$.

Proof: By Proposition 3.1, we know that for all $\mathbf{x} \in Z_u$,

$$Q(\mathbf{x}) = -\sum_{i=1}^{n} x_i^* \ln x_i \leq -\sum_{i=1}^{n} x_i^* \ln \varepsilon_d = -\ln \varepsilon_d,$$

i.e., the function $Q(\mathbf{x})$ is upper bounded in the region Z_u. ■

Since $Z_p \subset Z_u$, we immediately prove that $Q(\mathbf{x})$ is also upper bounded by M_p in the region Z_p.

Corollary 3.1 indicates the boundness of $Q(\mathbf{x})$ over Z_p and Z_u, which is a property in a spatial sense. The subsequent proof of Theorem 3.1 demonstrates that the boundness in the spatial sense actually implies boundness in the temporal sense.

Proof: (Proof of Theorem 3.1) By Corollary 3.1, if $\mathbf{x} \in Z_u$, then $Q(x) \leq M_p$.

Consider the sequence $\mathbf{x}_t$ in the Hedge-myopic system. Suppose that for some time t, $\mathbf{x}_t \in Z_p$. Then,

$$\begin{aligned} Q(\mathbf{x}_{t+1}) &= Q(\mathbf{x}_t) + \ln\left(\sum_{i=1}^{n} x_{i,t} e^{\eta(v^* - e_i^T A \mathbf{y}_t)}\right) \\ &\leq M_p + \ln\left(\sum_{i=1}^{n} x_{i,t} e^{\eta \max_{i\in\mathcal{I}}(v^* - e_i^T A \mathbf{y}_t)}\right) \\ &= M_p + \eta \max_{i\in\mathcal{I}}(v^* - e_i^T A \mathbf{y}_t) \\ &\leq M_p + \eta\delta_u, \end{aligned} \tag{10}$$

where $\delta_u = v^* - \min_{i\in\mathcal{I}, j\in\mathcal{J}} a_{ij}$.

Now, we consider different cases for the behavior of the strategy sequence $\{\mathbf{x}_t\}$.

Case 1: if the strategy sequence never goes into the region Z_p, which implies that $D(\mathbf{x}_t) < 0$ for all t, then we have $Q(\mathbf{x}_t) \leq Q(\mathbf{x}_1)$ for all t. By calculating, $Q(\mathbf{x}_1) = -\sum_{i=1}^{n} x_i^* \ln(1/n) = \ln n$. Hence, $Q(\mathbf{x}_t) \leq \ln n$ for all t.

Case 2: if the strategy sequence goes into the region Z_p at some time t', then for the strategy $\mathbf{x}_t$ before time t', we have $Q(\mathbf{x}_t) \leq \ln n$ for all $t < t'$. For the strategy $\mathbf{x}_t$ after time t':

(1). if $\mathbf{x}_t \in Z_p$, then we have $Q(\mathbf{x}_t) \leq M_p$;

(2). if $\mathbf{x}_t \notin Z_p$, then we can find an integer $k > 0$ such that $\mathbf{x}_{t-k} \in Z_p$ and $\mathbf{x}_{t-j} \notin Z_p$ for $0 < j < k$. Combining this with (10), we can obtain that $Q(\mathbf{x}_t) < Q(\mathbf{x}_{t-1}) < \cdots < Q_{t-k+1} \leq M_p + \eta\delta_u$ since $\mathbf{x}_{t-k} \in Z_p$.

Take $M_Q = \max\{\ln n, M_p + \eta\delta_u\}$, then the Q-sequence $\{Q_t\}_{t=1}^{\infty}$ is upper bounded by M_Q. ■

From Definition (2), for strategy $\mathbf{x}$, if some element x_i is near zero, the value of $Q(x)$ is near infinity. Hence, by Theorem 3.1, the elements of the stage strategy $\mathbf{x}_t$ cannot be too small since the Q-sequence is bounded. Based on this direct intuition, we can further prove the following theorem, which shrinks the range of possible values for $\mathbf{x}_t$ to be a finite set.

Theorem 3.2: If **Assumption 1** and **Assumption 2** hold, then stage strategy $\mathbf{x}_t$ for player X can only take finite values.

Remark 1: From the proof of Theorem 3.2, it can be observed that when there are irrational number elements in the payoff matrix, Theorem 3.2 no longer holds except for

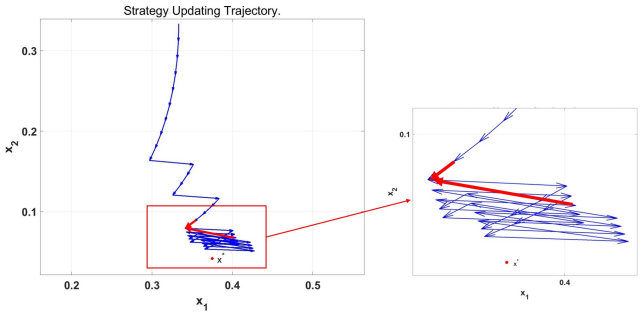

Fig. 2: The evolution of $\mathbf{x}_t$. The red point represents the NE strategy of player X.

the special case such as all elements are rational multiples of the same irrational number. This implies that the rationality of the payoff matrix is nearly an intrinsically necessary condition for periodicity.

By Theorem 3.2, in the Hedge-myopic system, player X can only adopt a finite number of mixed strategies. This directly leads to the periodicity of the dynamical system as stated below.

Theorem 3.3: In the Hedge-myopic system with **Assumption 1** and **Assumption 2** satisfied, after finite steps,

(1) the strategy sequence of player X and player Y enters a cycle (i.e., is periodic);
(2) the time-averaged strategy of player Y in one period is a NE strategy.

In other words, there exists T_s and T such that for all $t \geq T_s$, we have $\mathbf{x}_{t+T} = \mathbf{x}_t, \mathbf{y}_{t+T} = \mathbf{y}_t$, and $\sum_{k=1}^{T} \mathbf{y}_{t+k}/T = \mathbf{y}^*$, where $\mathbf{y}^*$ is a NE strategy of player Y.

Below we give an example to illustrate Theorem 3.3.

Example 3.1: Consider a 3×3 zero-sum game and the payoff matrix is taken as

$$A = \begin{pmatrix} -2 & 1 & 3 \\ 1 & 2 & -2 \\ 2 & 0 & -1 \end{pmatrix},$$

and $\eta = \frac{\ln 3}{625}$. In the Hedge-myopic system for this game, the evolution of $\mathbf{x}_t$ is shown in Figure 2. The X-axis represents the first element $x_{1,t}$, and the Y-axis represents the second element $x_{2,t}$.

From Figure 2, we can observe that basically $\mathbf{x}_t$ gradually approaches the NE strategy of player X. However, after reaching a certain range, $\mathbf{x}_t$ enters a cycle. In Figure 2, we mark a point such that two red arrows both point to it, meaning that a cycle is formed. Additionally, we can see that $\mathbf{x}_t$ does not converge to his NE strategy, no matter how long the game is repeated.

Before proving Theorem 3.3, we give Lemma 3.3 below by which we only need to prove the periodicity of $\mathbf{x}_t$ in order to prove the periodicity of the system.

Lemma 3.3: If $\mathbf{x}_t$ enters a cycle, then $\mathbf{y}_t$ also enters a cycle and her time-averaged strategy in a single cycle is a NE strategy.

Now, we can prove Theorem 3.3.

Proof: By Theorem 3.2, $\mathbf{x}_t$ can only take finite values. Then, using the pigeonhole principle, we can obtain that there must exist two stages, $t_1 < t_2$, such that $\mathbf{x}_{t_1} = \mathbf{x}_{t_2}$ because the game is repeated for infinite times. Since $\mathbf{x}_{t+1}$ is fully determined by $\mathbf{x}_t$, we have $\mathbf{x}_{t_1+k} = \mathbf{x}_{t_2+k}, \ \forall \ k = 1, 2, \cdots$, which implies that $\mathbf{x}_t$ is periodic from time t_1. Combining Lemma 3.3, we can prove Theorem 3.3. ■

This result has been proven for 2×2 games in our previous work [33], even though the Hedge-myopic system was not the main focus of that paper. For the 2×2 case, the stage strategy of player X (i.e., the player using Hedge algorithm) can be represented by a single number, which we call state. Thus, the updating process of the strategy sequence of player X can be depicted as alternatively adding two constant numbers to the state. By number theory, the stage sequence of player X is proven to be periodic after $O(1/\eta)$ stages.

However, for general $n \times n$ games, the degrees of freedom of the stage strategy of player X is $n-1$, which is larger than 1. The combination of numbers changes to the combination of vectors. Therefore, the method of using number theory to prove periodicity is no longer applicable.

Remark 2: The time required to enter a cycle depends on the parameter η. Due to the complexity of the problem, it is difficult to provide an explicit expression for it. We have studied it for 2×2 games in [33] where the time needed for the strategy sequences of both players to enter a cycle is $O(1/\eta)$.

Remark 3: In addition to periodicity, it is worth noting that the time-averaged strategy of player Y in a single cycle is a **precise** NE strategy! Compared with this, when both players adopt the no-regret learning algorithm in two-player zero-sum games, their time-averaged strategy profile converges to a NE when the time horizon goes to infinity, that is to say, only an approximate NE can be obtained. Moreover, in the Hedge-myopic system, not only can we obtain a precise NE, but we also only need to compute the time-averaged strategy in a single cycle whose length is far shorter than the whole time horizon.

B. Further Discussion about Non-periodicity

In this section, we discuss the effects of the assumptions on the periodicity of the system.

First, when there are irrational numbers in the NE of the game, the periodicity of the strategy sequence no longer holds. In the Hedge-myopic system, since player Y can only take pure strategies, the averaged strategy has only rational elements, so it is impossible for the averaged strategy to be a NE strategy. This explains why the periodicity no longer holds.

Next, when the game does not admit a unique interior equilibrium, whether the periodicity still holds is uncertain. Specifically, for some cases, the system is not periodic anymore and the strategy sequence of player X converge to the boundary of $\Delta(\mathcal{I})$ while for other cases, the repeated game does enter a special cycle.

These observations reveal that the dynamics of the Hedge-myopic system can vary significantly across different games. Because of the pages limitation, more examples and discussions can be seen in the complete paper [37].

IV. Conclusion and Future Work

In this paper, we study the repeated game between one player using the Hedge algorithm and the other using the myopic best response. We prove that within this framework, when payoffs are all rational numbers and the game has an interior NE, the dynamical game system enters a cycle after finite time. Moreover, within each period, the time-averaged strategy of the player using the myopic best response is an exact NE strategy.

In the future, we can explore the Hedge-myopic system for general games which have complex structures of NE. For $n \times m$ zero-sum games, we guess that the periodicity can be extended to the case where the Nash equilibrium strategy of player X is fully-mixed and its elements are all rational numbers. On the other hand, the periodicity of the Hedge-myopic system actually rules out the possibility of stage strategies converging to NE. Therefore, it is significant to consider how to modify the HBR paradigm so that the stage strategy can converge to NE strategy, i.e., the last-iterate property. We leave this also as future work.

References

[1] D. Fudenberg and J. Tirole, *Game theory*. MIT press, 1991.

[2] Y. Narahari, *Game theory and mechanism design*, vol. 4. World Scientific, 2014.

[3] J. F. Nash, "Equilibrium points in n-person games," *Proceedings of the national academy of sciences*, vol. 36, no. 1, pp. 48–49, 1950.

[4] C. H. Papadimitriou, "On the complexity of the parity argument and other inefficient proofs of existence," *Journal of Computer and system Sciences*, vol. 48, no. 3, pp. 498–532, 1994.

[5] 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.

[6] N. Karmarkar, "A new polynomial-time algorithm for linear programming," in *Proceedings of the sixteenth annual ACM symposium on Theory of computing*, pp. 302–311, 1984.

[7] J. van den Brand, "A deterministic linear program solver in current matrix multiplication time," in *Proceedings of the Fourteenth Annual ACM-SIAM Symposium on Discrete Algorithms*, pp. 259–278, SIAM, 2020.

[8] P. Mertikopoulos, C. Papadimitriou, and G. Piliouras, "Cycles in adversarial regularized learning," in *Proceedings of the twenty-ninth annual ACM-SIAM symposium on discrete algorithms*, pp. 2703–2717, SIAM, 2018.

[9] J. Perolat, R. Munos, J.-B. Lespiau, S. Omidshafiei, M. Rowland, P. Ortega, N. Burch, T. Anthony, D. Balduzzi, B. De Vylder, *et al.*, "From poincaré recurrence to convergence in imperfect information games: Finding equilibrium via regularization," in *International Conference on Machine Learning*, pp. 8525–8535, PMLR, 2021.

[10] D. Fudenberg and D. K. Levine, *The theory of learning in games*, vol. 2. MIT press, 1998.

[11] J. Robinson, "An iterative method of solving a game," *Annals of mathematics*, pp. 296–301, 1951.

[12] K. Miyasawa, *On the convergence of the learning process in a 2 x 2 non-zero-sum two-person game*. Princeton University Princeton, 1961.

[13] D. Monderer and L. S. Shapley, "Fictitious play property for games with identical interests," *J. Econ. Theory*, vol. 68, pp. 258–265, Jan. 1996.

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

[15] S. Hart and A. Mas-Colell, "A simple adaptive procedure leading to correlated equilibrium," *Econometrica*, vol. 68, no. 5, pp. 1127–1150, 2000.

[16] M. Zinkevich, M. Johanson, M. Bowling, and C. Piccione, "Regret minimization in games with incomplete information," *Advances in neural information processing systems*, vol. 20, 2007.

[17] N. Brown and T. Sandholm, "Superhuman ai for heads-up no-limit poker: Libratus beats top professionals," *Science*, vol. 359, no. 6374, pp. 418–424, 2018.

[18] N. Brown, A. Lerer, S. Gross, and T. Sandholm, "Deep counterfactual regret minimization," in *International conference on machine learning*, pp. 793–802, PMLR, 2019.

[19] M. Lanctot, K. Waugh, M. Zinkevich, and M. Bowling, "Monte carlo sampling for regret minimization in extensive games," *Advances in neural information processing systems*, vol. 22, 2009.

[20] M. Moravčík, M. Schmid, N. Burch, V. Lisỳ, D. Morrill, N. Bard, T. Davis, K. Waugh, M. Johanson, and M. Bowling, "Deepstack: Expert-level artificial intelligence in heads-up no-limit poker," *Science*, vol. 356, no. 6337, pp. 508–513, 2017.

[21] C. Papadimitriou and G. Piliouras, "From nash equilibria to chain recurrent sets: Solution concepts and topology," in *Proceedings of the 2016 ACM Conference on Innovations in Theoretical Computer Science*, pp. 227–235, 2016.

[22] G. Piliouras and J. S. Shamma, "Optimization despite chaos: Convex relaxations to complex limit sets via poincaré recurrence," in *Proceedings of the twenty-fifth annual ACM-SIAM symposium on Discrete algorithms*, pp. 861–873, SIAM, 2014.

[23] Y. Sato, E. Akiyama, and J. D. Farmer, "Chaos in learning a simple two-person game," *Proceedings of the National Academy of Sciences*, vol. 99, no. 7, pp. 4748–4751, 2002.

[24] G. Piliouras, C. Nieto-Granda, H. I. Christensen, and J. S. Shamma, "Persistent patterns: multi-agent learning beyond equilibrium and utility.," in *AAMAS*, pp. 181–188, 2014.

[25] G. Palaiopanos, I. Panageas, and G. Piliouras, "Multiplicative weights update with constant step-size in congestion games: Convergence, limit cycles and chaos," *Advances in Neural Information Processing Systems*, vol. 30, 2017.

[26] J. P. Bailey and G. Piliouras, "Multiplicative weights update in zero-sum games," in *Proceedings of the 2018 ACM Conference on Economics and Computation*, pp. 321–338, 2018.

[27] J. S. Jordan, "Three problems in learning mixed-strategy nash equilibria," *Games and Economic Behavior*, vol. 5, no. 3, pp. 368–386, 1993.

[28] L. Shapley, "Some topics in two-person games," *Advances in game theory*, vol. 52, pp. 1–29, 1964.

[29] O. Candogan, A. Ozdaglar, and P. A. Parrilo, "Dynamics in near-potential games," *Games and Economic Behavior*, vol. 82, pp. 66–90, 2013.

[30] Y. Arslantas, E. Yuceel, and M. O. Sayin, "Strategizing against q-learners: A control-theoretical approach," *arXiv preprint arXiv:2403.08906*, 2024.

[31] Y. Huang and Q. Zhu, "Deceptive reinforcement learning under adversarial manipulations on cost signals," in *Decision and Game Theory for Security: 10th International Conference, GameSec 2019, Stockholm, Sweden, October 30–November 1, 2019, Proceedings 10*, pp. 217–237, Springer, 2019.

[32] X. Guo and Y. Mu, "The optimal strategy against hedge algorithm in repeated games," *arXiv preprint arXiv:2312.09472*, 2023.

[33] X. Guo and Y. Mu, "Taking myopic best response against the hedge algorithm," in *2023 42nd Chinese Control Conference (CCC)*, pp. 8154–8158, IEEE, 2023.

[34] Y. Freund and R. E. Schapire, "A decision-theoretic generalization of on-line learning and an application to boosting," *Journal of computer and system sciences*, vol. 55, no. 1, pp. 119–139, 1997.

[35] N. Littlestone and M. K. Warmuth, "The weighted majority algorithm," *Information and computation*, vol. 108, no. 2, pp. 212–261, 1994.

[36] S. Arora, E. Hazan, and S. Kale, "The multiplicative weights update method: a meta-algorithm and applications," *Theory of computing*, vol. 8, no. 1, pp. 121–164, 2012.

[37] X. Guo, Y. Mu, and X. Yang, "Periodicity in hedge-myopic system and an asymmetric ne-solving paradigm for two-player zero-sum games," *arXiv preprint arXiv:2403.04336*, 2024.

[38] D. Bertsimas and J. N. Tsitsiklis, *Introduction to linear optimization*, vol. 6. Athena scientific Belmont, MA, 1997.