

Synchronization in Learning in Periodic Zero-Sum Games Triggers Divergence from Nash equilibrium

Yuma Fujimoto^{1,2,3} †, Kaito Ariu¹, Kenshi Abe^{1,4}

¹CyberAgent

²University of Tokyo

³Soken University

⁴University of Electro-Communications

[†]fujimoto.yuma1991@gmail.com

Abstract

Learning in zero-sum games studies a situation where multiple agents competitively learn their strategy. In such multi-agent learning, we often see that the strategies cycle around their optimum, i.e., Nash equilibrium. When a game periodically varies (called a “periodic” game), however, the Nash equilibrium moves generically. How learning dynamics behave in such periodic games is of interest but still unclear. Interestingly, we discover that the behavior is highly dependent on the relationship between the two speeds at which the game changes and at which players learn. We observe that when these two speeds synchronize, the learning dynamics diverge, and their time-average does not converge. Otherwise, the learning dynamics draw complicated cycles, but their time-average converges. Under some assumptions introduced for the dynamical systems analysis, we prove that this behavior occurs. Furthermore, our experiments observe this behavior even if these assumptions are removed. This study discovers a novel phenomenon, i.e., synchronization, and gains insight widely applicable to learning in periodic games.

Code — https://github.com/CyberAgentAILab/periodic_games_synchronization

Extended version — <https://arxiv.org/abs/2408.10595>

1 Introduction

Learning in games discusses how multiple agents optimize their strategies in the repetition of games (Fudenberg and Levine 1998). Their optimal strategies are usually characterized by Nash equilibrium (Nash Jr 1950), where every player cannot increase its payoff by other strategies. When their payoffs compete with each other, i.e., in zero-sum games, it is difficult to achieve equilibrium by naive learning algorithms, such as gradient descent-ascent (GDA). This is because one’s optimal strategy crucially depends on the other’s strategy. Their strategies typically draw a complex trajectory, such as a cycle around the equilibrium, without convergence. Thus, dynamical systems analysis is often introduced in learning in games to understand such complex behavior (Sato, Akiyama, and Farmer 2002; Piliouras et al.

2014; Bloembergen et al. 2015; Mertikopoulos and Sandholm 2016; Mertikopoulos, Papadimitriou, and Piliouras 2018; Bailey and Piliouras 2019; Fujimoto, Ariu, and Abe 2024a,b).

Learning in games typically assumes that the same game is repeatedly played, and the dynamical systems in learning (say, learning dynamics) may be even more complex when the game changes with time (called “time-varying” games). These time-varying games currently attract much attention (Fiez et al. 2021; Zhang et al. 2022; Duvocelle et al. 2023; Yan, Zhao, and Zhou 2023; Anagnostides et al. 2023; Feng et al. 2023, 2024). One of the major factors causing such time-varyingness is the periodic effect in the real-world environment, such as daily and seasonal cycles. The games that change periodically are called “periodic” games. Although such periodic games are significant, they have been studied only in a few papers (Fiez et al. 2021; Feng et al. 2023, 2024). Moreover, these studies focus on a special class of periodic games whose equilibrium is invariant over time. Since the Nash equilibrium is the target of learning, the movement of the equilibrium is expected to impact the learning dynamics significantly¹. To summarize, the learning dynamics in general periodic games with their Nash equilibrium time-varying are crucial but still unexplored.

Besides the trajectory of the learning dynamics, the behavior of the “time-average”, defined as the average of the trajectory, is also of interest. In time-invariant zero-sum games, this time-average is known to converge to the Nash equilibrium over time. This convergence property is intuitive because, in GDA, the trajectory cycles around the Nash equilibrium, and the deviations from the equilibrium cancel each other in the long run. However, since the learning dynamics are expected to be complex in periodic games equipped with time-varying equilibrium, this time-average convergence becomes a non-trivial problem. The dynamical systems analysis is also helpful to examine the time-average convergence, which has been usually proved by the no-regret property (Banerjee and Peng 2005; Zinkevich et al. 2007; Daskalakis, Deckelbaum, and Kim 2011).

We tackle a dynamical systems analysis in general peri-

¹Indeed, it was said that “An interesting question arises — in periodic games where there is no common equilibrium.” in (Feng et al. 2024)

odic games and discuss the time-average convergence. To the best of our knowledge, this is the first study to focus on the phenomenon triggered by the synchronization of learning with periodic games. Interestingly, when the speeds of learning and time-varying games synchronize, the players' strategies diverge from the Nash equilibrium, and their time-average does not converge. Otherwise, they cycle with complicated trajectories, and their time-average converges. We practically show that this phenomenon is intrinsic by giving a simple example, i.e., periodic matching pennies. We prove that this phenomenon holds in 2×2 matrix games with arbitrary wave shapes of periodic games. Furthermore, our experiments support that this phenomenon is valid in more general settings: 1) independent of the action numbers, 2) with the boundary constraint of strategy spaces. This study finds a novel phenomenon in learning in periodic games and extracts insight to understand the learning dynamics theoretically.

2 Setting

2.1 Normal-Form Games

Let us introduce normal-form games between two players, X and Y . Every round, they independently choose their actions from given sets, $\mathcal{A} = \{a_i\}_{1 \leq i \leq m_X}$ and $\mathcal{B} = \{b_j\}_{1 \leq j \leq m_Y}$, respectively. When they choose a_i and b_j , they receive the payoffs of u_{ij} and v_{ij} , respectively. We define their payoff matrices as $\mathbf{U} := (u_{ij})_{ij}$ and $\mathbf{V} := (v_{ij})_{ij}$. We assume that the summation of their payoffs is zero, i.e., $\mathbf{V} = -\mathbf{U}$.

In normal-form games, players choose their actions following their strategies, denoted as $\mathbf{x} = (x_i)_i \in \mathcal{X} := \Delta^{m_X-1}$ and $\mathbf{y} = (y_j)_j \in \mathcal{Y} := \Delta^{m_Y-1}$. Here, x_i means the probability that X chooses action a_i . X 's expected payoff is given by $u(\mathbf{x}, \mathbf{y}) := \mathbf{x}^T \mathbf{U} \mathbf{y} = \sum_i \sum_j x_i y_j u_{ij}$. By the definition of zero-sum games, Y 's expected payoff is computed as $v(\mathbf{x}, \mathbf{y}) = -u(\mathbf{x}, \mathbf{y})$.

2.2 Periodic Games

This study considers a situation where the game can fluctuate periodically. Thus, the payoff matrix depends on time $U = U(t)$ under the following constraint.

Definition 1 (Periodic game). *A periodic game satisfies $U(t) = U(t + 2\pi/\omega)$ with some $\omega \in \mathbb{R}$ for all t .*

Here, the period is $2\pi/\omega$, where ω represents the frequency of the periodic game.

2.3 Gradient Descent-Ascent

Learning in games discusses a process where each player sequentially optimizes their strategies with time t . One of the representative learning algorithms is GDA, described as

$$\mathbf{x} = \arg \max_{\mathbf{x} \in \mathcal{X}} \left[\mathbf{x} \cdot \mathbf{x}^\dagger - \frac{\|\mathbf{x}\|_2^2}{2} \right], \quad \mathbf{x}^\dagger = + \frac{\partial u(\mathbf{x}, \mathbf{y})}{\partial \mathbf{x}}, \quad (1)$$

$$\mathbf{y} = \arg \max_{\mathbf{y} \in \mathcal{Y}} \left[\mathbf{y} \cdot \mathbf{y}^\dagger - \frac{\|\mathbf{y}\|_2^2}{2} \right], \quad \mathbf{y}^\dagger = - \frac{\partial u(\mathbf{x}, \mathbf{y})}{\partial \mathbf{y}}. \quad (2)$$

This algorithm is also known as the Follow the Regularized Leader (FTRL) with the Euclidean regularizer.

In the interior of the strategy space, Eqs. (1) and (2) are rewritten as

$$\frac{d}{dt} \begin{pmatrix} \mathbf{x} \\ \mathbf{y} \end{pmatrix} = \begin{pmatrix} \mathbf{O}_X & +\mathbf{P}_X \mathbf{U} \\ -\mathbf{P}_Y \mathbf{U}^T & \mathbf{O}_Y \end{pmatrix} \begin{pmatrix} \mathbf{x} \\ \mathbf{y} \end{pmatrix}, \quad (3)$$

$$\mathbf{P}_X := \mathbf{I}_X - \frac{\mathbf{1}_X \mathbf{1}_X^T}{m_X}, \quad \mathbf{P}_Y := \mathbf{I}_Y - \frac{\mathbf{1}_Y \mathbf{1}_Y^T}{m_Y}, \quad (4)$$

(see Appendix B.1 for derivation). Here, the subscripts of X and Y mean that the dimension of the vector or matrix is m_X and m_Y , respectively. Furthermore, \mathbf{O}_X and \mathbf{O}_Y are zero matrices, \mathbf{I}_X and \mathbf{I}_Y are identity matrices, and $\mathbf{1}_X$ and $\mathbf{1}_Y$ are all-ones vectors. Here, \mathbf{P}_X and \mathbf{P}_Y mean the projection to the strategy space by the Euclidean regularizer.

2.4 Eigenvalue of Learning

Let us analyze Eq. (3). For simplicity, we assume that the payoff matrix \mathbf{U} is time-invariant. Since Eq. (3) is a linear dynamical system (Hirsch, Smale, and Devaney 2013), its solution is characterized by the eigenvalues of

$$\mathbf{A} := \begin{pmatrix} \mathbf{O}_X & +\mathbf{P}_X \mathbf{U} \\ -\mathbf{P}_Y \mathbf{U}^T & \mathbf{O}_Y \end{pmatrix}. \quad (5)$$

Here, note that if the projection operator (\mathbf{P}_X and \mathbf{P}_Y) is removed, \mathbf{A} is a skew-Hermitian matrix, all of which eigenvalues are known to zero or purely imaginary. Interestingly, this projection operator does not affect the property of eigenvalues, and we can prove that all the eigenvalues of \mathbf{A} are zero or purely imaginary (see Appendix A for detailed calculation).

For the convenience of notation, consider $m_X = m_Y =: m$, then the eigenvalues of \mathbf{A} are denoted as $\{\pm i\alpha^{(j)} \mid j = 1, 2, \dots, m\}$ with $\alpha^{(j)} \in \mathbb{R}$ (purely imaginary eigenvalues). We can rearrange them so that they satisfy $\alpha^{(1)} \geq \alpha^{(2)} \geq \dots \geq \alpha^{(m)} = 0$. Here, at least one pair of the eigenvalues is zero ($\alpha^{(m)} = 0$) because both $\mathbf{P}_X \mathbf{U}$ and $\mathbf{P}_Y \mathbf{U}^T$ degenerate in \mathbf{A} . Now, these purely imaginary eigenvalues mean that the dynamics of Eq. (3) are given by the summation of circular functions. In particular, $\alpha^{(j)}$ corresponds to the cycling speed of each circular function. We simply call $\alpha^{(j)}$ as "eigenvalue" throughout this paper. We remark that the cycling behavior characterized by purely imaginary eigenvalues is consistent with the time-invariance of distance from the Nash equilibrium (Piliouras et al. 2014) and the Hamiltonian structure (Bailey and Piliouras 2019).

3 Example: Eigenvalue Invariant Game

In order to intuitively understand the behavior of learning in periodic games, we now consider the simplest periodic game, named eigenvalue invariant game, which provides a phenomenon beyond an existing class of games equipped with the time-invariant Nash equilibrium (Fiez et al. 2021; Feng et al. 2023, 2024). This eigenvalue invariant game is given by two 2×2 matrices of $\bar{\mathbf{U}} \in \mathbb{R}^{2 \times 2}$ and $\Delta \mathbf{U} \in \mathbb{R}^{2 \times 2}$ and formulated as follows.

Example 1 (Eigenvalue invariant game). *For given two payoff matrices $\bar{\mathbf{U}} = (\bar{u}_{ij})_{ij} \in \mathbb{R}^{2 \times 2}$ and $\Delta \mathbf{U} = (\Delta u_{ij})_{ij} \in$*

$\mathbb{R}^{2 \times 2}$ with the constraint of $\Delta u_{11} - \Delta u_{12} - \Delta u_{21} + \Delta u_{22} = 0$, the eigenvalue invariant game $\mathbf{U}(t)$ is formulated as

$$\mathbf{U}(t) = \bar{\mathbf{U}} + \Delta \mathbf{U} \cos \omega t. \quad (6)$$

Here, $\bar{\mathbf{U}}$ is the time average of periodic game $\mathbf{U}(t)$, while $\Delta \mathbf{U}$ is the amplitude of the oscillation. Thus, this periodic game oscillates between $\bar{\mathbf{U}} + \Delta \mathbf{U}$ and $\bar{\mathbf{U}} - \Delta \mathbf{U}$. The constraint for $\Delta \mathbf{U}$ means that the eigenvalue is time-invariant, as shown later.

3.1 Analysis of Eigenvalue Invariant Game

We now analyze learning in these eigenvalue invariant games. First, since $\mathbf{x} = (x_1, x_2) \in \Delta^1$ should be satisfied in 2×2 matrix games, \mathbf{x} can be described by a single value $x \in [0, 1]$ as $\mathbf{x} = (x, 1 - x)$. In a similar manner, we only consider $y \in [0, 1]$ for $\mathbf{y} = (y, 1 - y)$. Thus, the learning dynamics is described only by the two variables, $x(t)$ and $y(t)$.

By directly calculating the eigenvalues of Eq. (5), the eigenvalues of learning are obtained as

$$\alpha^{(1)} = \frac{1}{2}(\bar{u}_{11} - \bar{u}_{12} - \bar{u}_{21} + \bar{u}_{22}) =: \alpha, \quad (7)$$

$$\alpha^{(2)} = 0, \quad (8)$$

(see Appendix B.2 for derivation). Furthermore, the Nash equilibrium of the time-average of the periodic game, denoted as (\bar{x}^*, \bar{y}^*) , is given by

$$\bar{x}^* := \frac{-\bar{u}_{11} + \bar{u}_{22}}{2\alpha}, \quad \bar{y}^* := \frac{-\bar{u}_{12} + \bar{u}_{22}}{2\alpha}. \quad (9)$$

Next, the deviation of the Nash equilibrium caused by the oscillation of the game, denoted as $(\Delta x^*, \Delta y^*)$, is also given by

$$\Delta x^* := \frac{-\Delta u_{21} + \Delta u_{22}}{2\alpha}, \quad \Delta y^* := \frac{-\Delta u_{12} + \Delta u_{22}}{2\alpha}. \quad (10)$$

Thus, the Nash equilibrium oscillates between $(\bar{x}^* + \Delta x^*, \bar{y}^* + \Delta y^*)$ and $(\bar{x}^* - \Delta x^*, \bar{y}^* - \Delta y^*)$, which is beyond the scope of the previous studies (Fiez et al. 2021; Feng et al. 2023, 2024).

Finally, the learning dynamics, i.e., Eq. (3), are described as

$$\dot{x}(t) = +\alpha(y(t) - \bar{y}^* - \Delta y^* \cos \omega t), \quad (11)$$

$$\dot{y}(t) = -\alpha(x(t) - \bar{x}^* - \Delta x^* \cos \omega t). \quad (12)$$

Interestingly, these equations correspond to the dynamics of forced pendulum without dissipation (Mawhin 2004). In other words, the competitive learning between the players can be interpreted as an oscillation inherent in a pendulum, whereas the vibration of the Nash equilibrium can be as the external force into the pendulum.

3.2 Solution of Learning Dynamics

The learning dynamics, Eqs. (11) and (12), are solvable. In the case of $\alpha/\omega = 1$, the solution is

$$x(t) = \bar{x}^* + \frac{1}{2}\Delta x^*(\omega t \sin \omega t + \cos \omega t) - \frac{1}{2}\Delta y^*\omega t \cos \omega t + c_1 \cos \alpha t + c_2 \sin \alpha t, \quad (13)$$

$$y(t) = \bar{y}^* + \frac{1}{2}\Delta y^*(\omega t \sin \omega t + \cos \omega t) + \frac{1}{2}\Delta x^*\omega t \cos \omega t + c_2 \cos \alpha t - c_1 \sin \alpha t, \quad (14)$$

where c_1 and c_2 are uniquely determined by the initial state condition (see Appendix B.3 for the detailed expression of c_1 and c_2). This solution means that $x(t)$ and $y(t)$ diverges from the average Nash equilibrium (\bar{x}^*, \bar{y}^*) at a speed of $\sqrt{|\Delta x^*|^2 + |\Delta y^*|^2}\omega/2$.

On the other hand, in the case of $\alpha/\omega \neq 1$, the solution is

$$x(t) = \bar{x}^* + \frac{\alpha^2}{\alpha^2 - \omega^2}\Delta x^* \cos \omega t + \frac{\alpha\omega}{\alpha^2 - \omega^2}\Delta y^* \sin \omega t + c_1 \cos \alpha t + c_2 \sin \alpha t, \quad (15)$$

$$y(t) = \bar{y}^* + \frac{\alpha^2}{\alpha^2 - \omega^2}\Delta y^* \cos \omega t - \frac{\alpha\omega}{\alpha^2 - \omega^2}\Delta x^* \sin \omega t + c_2 \cos \alpha t - c_1 \sin \alpha t. \quad (16)$$

Here, $x(t)$ and $y(t)$ consist of two periodic solutions with the frequencies of α and ω . Thus, the solution qualitatively changes depending on whether $\alpha/\omega = 1$ or $\neq 1$.

3.3 Time-Average Analysis

Under fixed zero-sum payoff matrices, the dynamics of GDA are known to converge to the Nash equilibrium (called time-average convergence). We now discuss time-average convergence under periodic payoff matrices. The time-average of \mathbf{X} 's strategy until time $T > 0$ is defined as

$$\bar{x}(T) := \frac{1}{T} \int_0^T x(t) dt. \quad (17)$$

Interestingly, depending on whether $\alpha/\omega = 1$ or $\neq 1$, the time-average has different properties, as shown in the following. See Appendix C for the full proofs of all the theorems.

Theorem 1 (Time-average cycling in $\alpha/\omega = 1$). *In the eigenvalue invariant games, when $\alpha/\omega = 1$ holds, $\bar{x}(T)$ and $\bar{y}(T)$ cycle around the Nash equilibrium.*

Proof sketch. This theorem is proved by directly calculating the integral of Eqs. (13) and (14). Intuitively, since the solution diverges as it oscillates, its time-average is influenced by $x(t)$ of large time $t \simeq T$ and never converges. \square

Theorem 2 (Time-average convergence in $\alpha/\omega \neq 1$). *In the eigenvalue invariant games, when $\alpha/\omega \neq 1$ holds, $\bar{x}(T)$ and $\bar{y}(T)$ converge to the Nash equilibrium.*

Proof sketch. This theorem is proved by directly calculating the integral of Eqs. (15) and (16), where each cycle is canceled out. \square

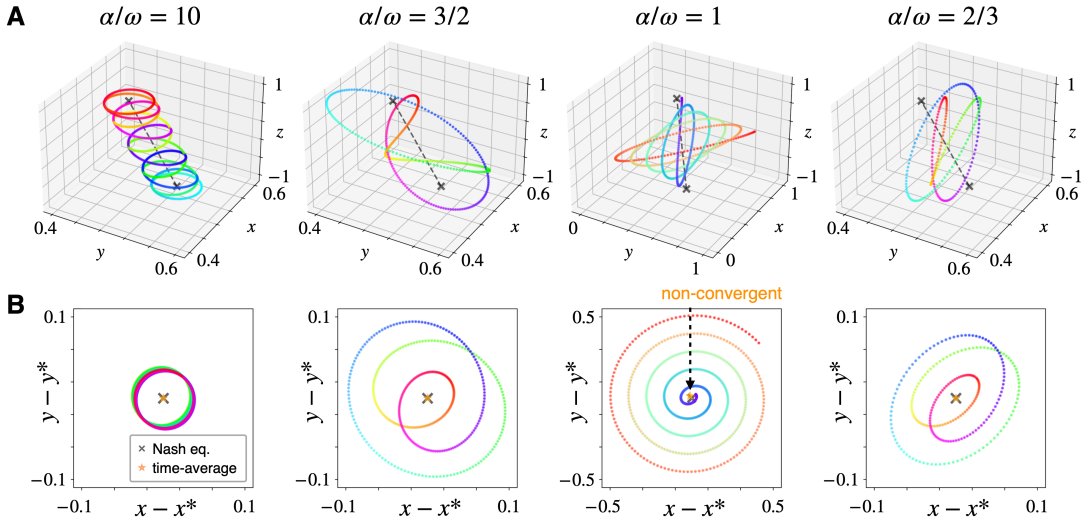


Figure 1: Learning dynamics in the periodic game of matching pennies with the time-varying Nash equilibrium. (A) The trajectories of the learning dynamics. The panels show the cases of $\alpha/\omega = 10, 2/3, 1$, and $2/3$, from left to right. In each panel, the x -, y -, z -axes indicate $x(t)$, $y(t)$, and $z(t)$, respectively. The black broken line shows the trajectory of the Nash equilibrium, i.e., $(x^*(t), y^*(t))$, and the cross markers show the edges of the oscillation of this equilibrium. In the cases of $\alpha/\omega \neq 1$, the rainbow color shows the time of a single cycle. In $\alpha/\omega = 1$, the rainbow color shows the passing of time from blue to red. (B) The projection of the trajectories of Panels A to the plane of $x(t) - x^*(t)$ and $y(t) - y^*(t)$. In each panel, the black cross marker shows the projection of the Nash equilibrium. The color corresponds to that of Panels A. The orange star markers show the time-average value of the plotted trajectory. The value does not correspond to the equilibrium in $\alpha/\omega = 1$. Otherwise, it corresponds. To simulate the learning dynamics, we use the Runge-Kutta fourth-order method with a step size of $1/40$. The initial strategies are set to the Nash equilibrium of the time-average game.

3.4 Visualization and Interpretation

Let us visualize the learning dynamics (see Fig. 1) and interpret their behavior. This section focuses on a periodic game formulated by $\bar{U} = ((1, -1), (-1, 1))$ and $\Delta U = ((1/10, 0), (0, -1/10))$ in Exm. 1. Here, \bar{U} shows Matching Pennies with the only equilibrium of $(\bar{x}^*, \bar{y}^*) = (1/2, 1/2)$. This periodic game has a time-invariant eigenvalue of $\alpha = 2$. However, its Nash equilibrium oscillates with the amplitude of $(\Delta x^*, \Delta y^*) = (-1/40, -1/40)$.

First, let us see the case of $\alpha/\omega = 10$. This case roughly means that the players learn ten times faster than the game changes. In other words, it approximately corresponds to a classical situation where a game is time-invariant. Thus, the learning dynamics are also approximated as the classical dynamics, where the strategies of both players almost cycle around the Nash equilibrium.

Next, we see the case of $\alpha/\omega = 3/2$, where the speed of game change is comparable but not equal to that of learning. In this case, the learning interacts with the game change and provides complex dynamics (see Panel A). Although the trajectory cycles around the Nash equilibrium many times, the distance from the Nash equilibrium gets closer and farther as the equilibrium moves. After a while, the trajectory returns to its initial state and thus is periodic as described by Eqs. (15) and (16). In addition, the average of the trajectory corresponds to the equilibrium (see Panel B). The above applies also to the case when the game changes faster than the players learn (e.g., $\alpha/\omega = 2/3$ in the panels).

Finally, let us focus on the case of $\alpha/\omega = 1$, where the speed of game change is equal to that of learning. In other words, the cycling behavior by learning synchronizes the oscillation of the game. Thus, each time the game oscillates once, the amplitude of the cycle is amplified a little (see Panel A). By this amplification accumulated, the trajectory diverges from the Nash equilibrium as expressed in Eqs. (13) and (14). Furthermore, the average of the trajectory does not correspond to the Nash equilibrium.

4 Theory on Eigenvalue Varying Games

We now extend the insight obtained from the eigenvalue invariant game (i.e., Exm. 1) to more general settings. The eigenvalue invariant game is generalized in the following two points: 1) its eigenvalue time-varying and 2) the wave shape. In the definition, we assume the smoothness of the wave shape to take the time integral of the payoff matrix.

Definition 2 (2×2 smooth periodic games). 2×2 smooth periodic games are defined as any smooth $\{U(t)\}_{0 \leq t < 2\pi/\omega} \in \mathbb{R}^{2 \times 2} \times [0, 2\pi/\omega)$, satisfying $U(t) = U(t + 2\pi/\omega)$.

4.1 Solution of Learning Dynamics

Below, we use the notation of

$$2f(t) := u_{11}(t) - u_{12}(t) - u_{21}(t) + u_{22}(t), \quad (18)$$

$$2g(t) := -u_{12}(t) + u_{22}(t), \quad (19)$$

$$2h(t) := -u_{21}(t) + u_{22}(t). \quad (20)$$

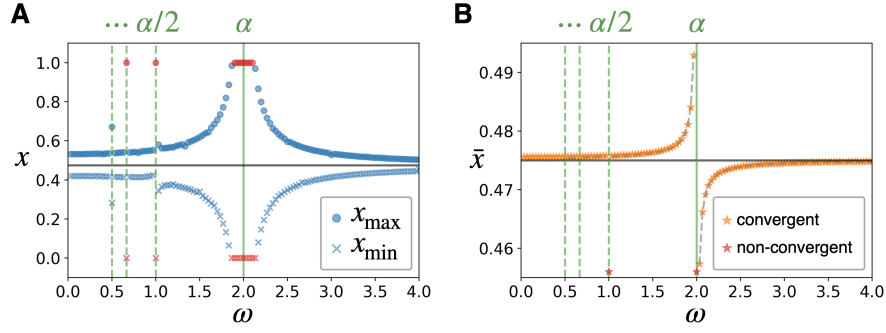


Figure 2: Experiments for learning dynamics in the eigenvalue invariant game for $\bar{U} = ((1.1, -1), (-1, 0.9))$, $\Delta U = ((0.2, 0), (0, 0))$ without the constraint of their eigenvalues fixed. (A) The maximum (circle) and minimum (cross) values of $x(t)$ within $0 \leq t \leq T = 3 \times 10^4$. Here, the blue markers indicate that the value exists in the interior of $[0, 1]$, while the red exterior. (B) The time-average \bar{x} in the last time T . The orange marker is plotted when the time-average sufficiently converges, i.e., the time-average moves less than 10^{-3} in the last 10^2 time. The red marker means that the time-average does not converge and take the value outside of the y-axis. The gray broken line shows the analytical solution. The numerical method and parameter is the same as Fig. 1.

All of these functions are periodic with frequency ω . Here, we see that $f(t)$ represents the eigenvalue of learning and varies over time. By using these equations, we describe the learning dynamics in Def. 2 as

$$\dot{x}(t) = +f(t)y(t) - g(t), \quad (21)$$

$$\dot{y}(t) = -f(t)x(t) + h(t). \quad (22)$$

These simultaneous differential equations are solved as

$$\begin{aligned} x(t) = & \int_0^t h(\tau) \sin(F(t) - F(\tau)) d\tau \\ & - \int_0^t g(\tau) \cos(F(t) - F(\tau)) d\tau \\ & + c_1 \cos F(t) + c_2 \sin F(t), \end{aligned} \quad (23)$$

$$\begin{aligned} y(t) = & \int_0^t g(\tau) \sin(F(t) - F(\tau)) d\tau \\ & + \int_0^t h(\tau) \cos(F(t) - F(\tau)) d\tau \\ & + c_2 \cos F(t) - c_1 \sin F(t). \end{aligned} \quad (24)$$

Here, c_1 and c_2 are uniquely determined by the initial state condition (see Appendix B.4 for detailed expression of c_1 and c_2). In addition, we defined $F(t) := \int_0^t f(\tau) d\tau$. This $F(t)$ is not periodic, but other than the linear term $\Delta F(t) := F(t) - \bar{f}t$ is periodic with frequency ω . This linear term corresponds to the average eigenvalue, i.e., $\alpha = \bar{f}$.

4.2 Time-Average Analysis

Let us analyze the time-average of the learning dynamics, i.e., Eqs. (23) and (24). Since we assume general smooth periodic functions for f , g , and h , we cannot analytically calculate the integral of $x(t)$, as different from the case of Exm. 1. Thus, we have to devise a way to evaluate $x(T)$ (see the following proof sketch for details). Now, when $\alpha/\omega \in \mathbb{N}$ holds, we prove that the time-average diverges generically (see Appendix D for some exceptions).

Theorem 3 (Time-average divergence in $\alpha/\omega \in \mathbb{N}$). *In 2×2 smooth periodic games with $\alpha/\omega \in \mathbb{N}$, $\bar{x}(T)$ and $\bar{y}(T)$ diverge over time T .*

Proof sketch. As different from the proofs of Thms. 1 and 2, the integral of the trajectory (Eqs. (23) and (24)) cannot be calculated directly. In the case of $\alpha/\omega \in \mathbb{N}$, however, we can easily evaluate the time-average by paying attention to the periodicity of $\sin F(\cdot)$ and $\cos F(\cdot)$ with frequency ω . We divide the ranges of integrals, i.e., $t \in [0, T]$ and $\tau \in [0, t]$, into the intervals of $2\pi/\omega$, and the time-average is also divided into the following four terms: 1) the term of diverges which is $O(T)$ in the time-average, 2) the term oscillates which is $O(1)$, 3) the term converges which is $O(1)$, and 4) the term is negligible which is $O(1/T)$. In general functions, f , g , and h , the divergence term takes a finite value, and thus we prove that the time-average is $O(T)$ and thus diverges with time. \square

Otherwise, the time-average converges as follows.

Theorem 4 (Time-average convergence in $\alpha/\omega \notin \mathbb{N}$). *In 2×2 smooth periodic games with $\alpha/\omega \notin \mathbb{N}$, $\bar{x}(T)$ and $\bar{y}(T)$ converge over time T .*

Proof sketch. We consider the same division of the ranges of the integrals as proof of Thm. 3 and evaluate the four terms. In the case of $\alpha/\omega \notin \mathbb{N}$, however, this evaluation is much more difficult because we cannot generally use the periodicity of $\sin F(\cdot)$ and $\cos F(\cdot)$. By utilizing a property of the non-periodicity, the divergence term falls into $O(1)$, while the oscillation term into $O(1/T)$. Thus, the time-average converges with neither divergence nor oscillation. \square

4.3 Experimental Verification for Theory

Let us see that Thms. 3 and 4 hold in experiments. It is enough to simply extend Exm. 1 not to satisfy the constraint for ΔU , and we use the example of $\bar{U} = ((1.1, -1), (-1, 0.9))$ and $\Delta U = ((0.2, 0), (0, 0))$. This example is calculated as $f(t) = 2 + 0.1 \cos \omega t$ and $g(t) =$

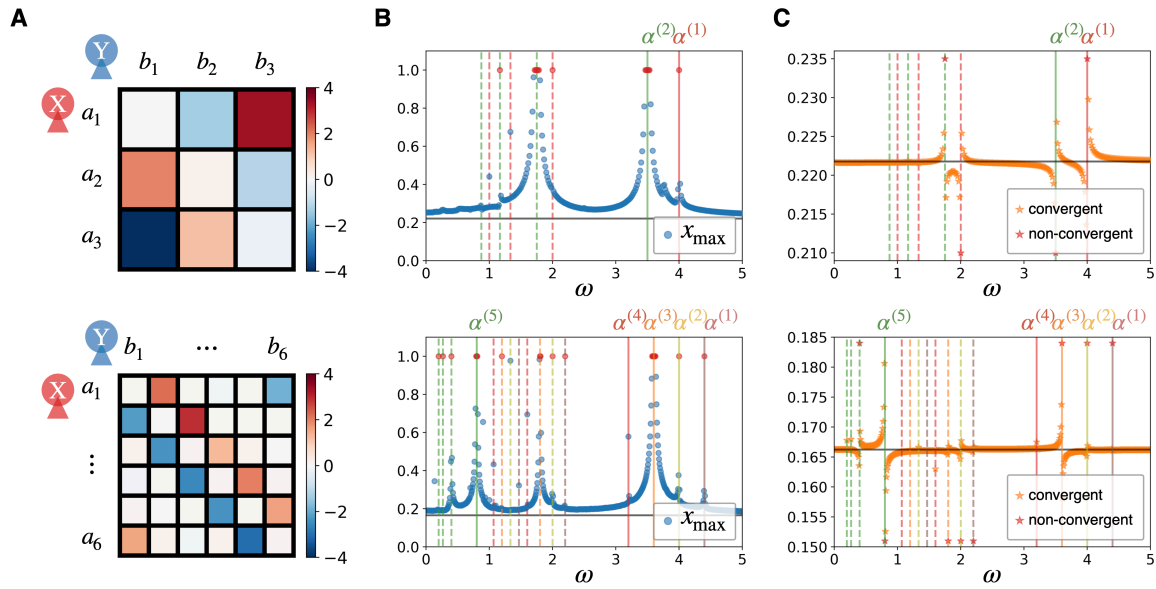


Figure 3: Learning dynamics in games with more than two actions. The upper panels show an example of 3×3 matrix games, while the lower 6×6 . (A) \bar{U} , i.e., the time-average payoff matrix is visualized. The row (a_i) and column (b_i) indicate X's and Y's actions, respectively. The darker red shows that X receives more payoff, while the darker blue shows that Y receives more. Here, note that a non-dominant structure is given in each payoff matrix, where each action wins one of the opponent's actions but loses another. We also remark that all the elements of the payoff matrix are non-zero (colored by light red or blue). We further consider the perturbation in the payoff matrix. Each element are independently perturbed by the four waves of $\cos \omega t$, $\sin \omega t$, $\cos 2\omega t$, and $\sin 2\omega t$ with the amplitude of a random number $\sim \mathcal{N}(0, 0.04^2)$ in the upper panels and $\sim \mathcal{N}(0, 0.02^2)$ in the lower. (B) The maximum value of $x_1(t)$ in overall time $0 \leq t \leq T = 3 \times 10^4$. The meanings of the markers and axes are the same as Fig. 2-A. (C) The average value in the last time T . We regard the time-average as convergent when it moves less than 10^{-3} in the last 10^3 time. The meanings of the markers and axes are the same as Fig. 2-B. The numerical method and parameter is the same as Fig. 1.

$h(t) = 0.95$, showing that not only the Nash equilibrium but the eigenvalue oscillates. Fig. 2 shows comprehensive experiments for various ω .

First, Panel A shows the maximum and minimum values of $x(t)$ for the whole t . We see that $x(t)$ always continues to oscillate. Especially in $\alpha/\omega \in \mathbb{N}$, the amplitude of this oscillation is large. Elsewhere, the amplitude is small.

Next, Panel B shows the time-average $\bar{x}(T)$ for sufficiently large T . We observe that the time-average diverges in $\alpha/\omega \in \mathbb{N}$, reflecting Thm. 3. Here, however, note that when α/ω takes a sufficiently large integer ($\alpha/\omega \geq 3$ in the panel), the divergence is weak and judged to be convergent. The time-average converges in $\alpha/\omega \notin \mathbb{N}$, reflecting Thm. 4.

5 Experimental Results

So far, for theoretical analyses, we have assumed 1) two-action games and 2) no boundary in the strategy spaces. In this section, we experimentally demonstrate that independent of these assumptions, the following ‘‘synchronization phenomenon’’ occurs.

- When the frequency (ω) synchronizes with the eigenvalues $\alpha^{(j)}$, the dynamics diverge from the Nash equilibrium, and their time-average never converges.
- Otherwise, the dynamics do not diverge, and their time-average converges.

As follows, we show experiments 1) for various action numbers and 2) with the boundary constraint of the strategy space. See also Appendix E.1 for non-smooth periodic games.

5.1 Various Action Numbers

3×3 matrix games: We consider the case when each player can take three actions (see the upper panels in Fig. 3). Here, note that the periodic small perturbation of the payoff matrix is considered, which causes the slight vibration of the interior Nash equilibrium. As different from 2×2 games, 3×3 ones have multiple non-zero eigenvalues, $\alpha^{(1)} \geq \alpha^{(2)} > 0$. From Panel B, we see that when $\alpha^{(j)}/\omega \in \mathbb{N}$ holds for either $j = 1$ or 2 , the learning dynamics diverge from the Nash equilibrium, and otherwise, they cycle around the equilibrium. Panel C shows that when $\alpha^{(j)}/\omega \in \mathbb{N}$ holds, the time-average of the dynamics deviates from the Nash equilibrium, while otherwise, it converges to near the equilibrium.

6×6 matrix games: Even if the game has a larger number of actions, similar results can be obtained. The lower panels of Fig. 3 show an example of a 6×6 matrix game with an interior equilibrium that fluctuates with time (see Panel A). This matrix game is equipped with five non-zero

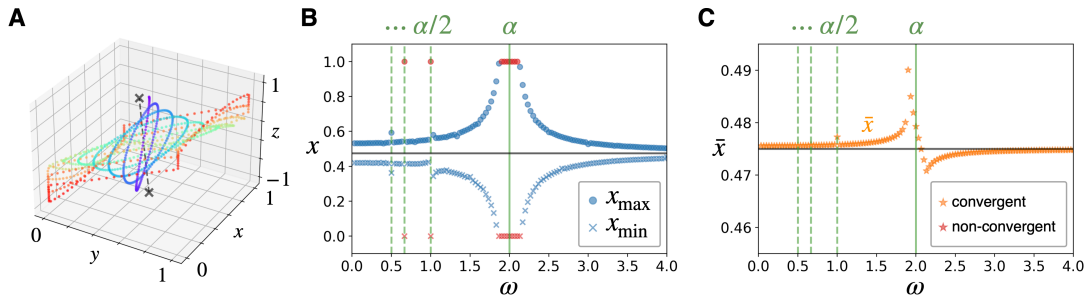


Figure 4: Learning dynamics under the boundary constraint of Eqs. (1) and (2). All the parameters of the game and data are the same as Fig. 2. (A) The trajectory of learning dynamics in $\omega = \alpha$. The meanings of plots and axes are the same as Fig. 1-A. (B) The maximum value of $x(t)$ in overall time $0 \leq t \leq T = 10^4$. The meanings of the markers and axes are the same as Fig. 2-A. (C) The average value in the last time T . We regard the time-average as convergent when it moves less than 10^{-3} in the last 10^2 time. The meanings of the markers and axes are the same as Fig. 2-B. The numerical method is the same as Fig. 1 but the step size is $1/(4 \times 10^3)$.

eigenvalues. Although the number of the eigenvalues is different, Panels B and C similarly demonstrate that the synchronization occurs only when $\alpha^{(j)}/\omega \in \mathbb{N}$ holds for either $j = 1, 2, 3, 4$, or 5.

Intuition from experiments: Intuitively, the above synchronization can occur in general two-player zero-sum games. As already explained, such games only have purely imaginary eigenvalues in general. Since the learning dynamics in general two-player zero-sum games are fixed in the Nash equilibrium, they are solved as the summation of multiple cycles around the equilibrium with frequencies $\alpha^{(j)}$. Thus, the oscillation of the equilibrium in a periodic game can resonate with these cycles. Furthermore, this synchronization is expected to occur in polymatrix games equipped with more than two players. These polymatrix games are known to have cycles in the learning dynamics. Indeed, we have observed the synchronization in three-player Matching Pennies, an example of polymatrix games (see Appendix E.2).

5.2 Boundary Constraint

Next, we consider the boundary constraint given by the Euclidean regularizer in Eqs. (1) and (2). The learning dynamics are different from Eq. (3) only on the boundary of the strategy spaces, i.e., $x, y \in \{0, 1\}$ (see Panel A). Despite the existence of the boundary constraint, in $\alpha/\omega \in \mathbb{N}$ and its neighbor, one's strategy oscillates with a large amplitude (see Panel B). Elsewhere, the amplitude is small. However, since the amplitude is bounded by the boundary constraint, the time-average of one's strategy always converges, even in $\alpha/\omega \in \mathbb{N}$ (see Panel C).

6 Conclusion

This study examined learning in periodic games where the Nash equilibrium is time-varying. We discussed the dynamics of learning and their time-average convergence/non-convergence. Notably, this study focused on the synchronization of learning with the periodic change of games. We

identified a phenomenon where the learning dynamics qualitatively change depending on whether or not the synchronization occurs. When the synchronization occurs, learning dynamics diverge from the Nash equilibrium, and their time average does not converge. Otherwise, the dynamics enter a complex cycle, but their time-average converges. We proved this phenomenon in a wide range of games, but with a limitation of 2×2 matrices, the smooth periodicity of the game, and the absence of the boundary constraint of the strategy space. Our experiments demonstrated that this phenomenon universally occurs, regardless of these limitations.

One of the remaining problems is that the time-average converges elsewhere than the Nash equilibrium of the game's time-average. This problem is contrary to classical time-invariant games but consistent with (Fiez et al. 2021). As a future work, it would be interesting to consider how to track the time-varying Nash equilibrium or converge to the Nash equilibrium of the time-average of periodic games. Another problem is to see and analyze the learning dynamics of other algorithms, such as optimistic (Rakhlín and Sridharan 2013; Syrgkanis et al. 2015; Daskalakis et al. 2018; Daskalakis and Panageas 2019; Cai, Oikonomou, and Zheng 2022), extra-gradient (Korpelevich 1976; Mertikopoulos et al. 2019; Lee and Kim 2021; Cai, Oikonomou, and Zheng 2022), and negative momentum (Polyak 1964; Gidel et al. 2019; Kovachki and Stuart 2021; Zhang and Wang 2021; Hemmat et al. 2023), where the strategies themselves converge to the Nash equilibrium, called last-iterate convergence. The behavior of such algorithms is already known under periodic games when the Nash equilibrium does not move (Feng et al. 2023, 2024), but not when it moves. In addition, it would be interesting to analyze the synchronization in polymatrix games (Bergman and Fokin 1998; Cai and Daskalakis 2011; Bailey and Piliouras 2019) more deeply. It is non-trivial how synchronization occurs when a polymatrix game can be divided into multiple games equipped with different frequencies and eigenvalues. This study finds a new phenomenon in learning in periodic games and gives insight into the phenomenon, providing a theoretical basis for these future problems.

Acknowledgments

We thank Sosuke Ito for the fruitful discussions on linear dynamical systems. K. Ariu is supported by JSPS KAKENHI Grant No. 23K19986.

References

- Anagnostides, I.; Panageas, I.; Farina, G.; and Sandholm, T. 2023. On the convergence of no-regret learning dynamics in time-varying games. In *NeurIPS*, 16367–16405.
- Bailey, J. P.; and Piliouras, G. 2019. Multi-Agent Learning in Network Zero-Sum Games is a Hamiltonian System. In *AAMAS*, 233–241.
- Banerjee, B.; and Peng, J. 2005. Efficient no-regret multiagent learning. In *AAAI*, 41–46.
- Bergman, L.; and Fokin, I. 1998. On separable non-cooperative zero-sum games. *Optimization*, 44(1): 69–84.
- Bloembergen, D.; Tuyls, K.; Hennes, D.; and Kaisers, M. 2015. Evolutionary dynamics of multi-agent learning: A survey. *Journal of Artificial Intelligence Research*, 53: 659–697.
- Cai, Y.; and Daskalakis, C. 2011. On minmax theorems for multiplayer games. In *SODA*, 217–234.
- Cai, Y.; Oikonomou, A.; and Zheng, W. 2022. Finite-time last-iterate convergence for learning in multi-player games. In *NeurIPS*, 33904–33919.
- Daskalakis, C.; Deckelbaum, A.; and Kim, A. 2011. Near-optimal no-regret algorithms for zero-sum games. In *SODA*, 235–254.
- Daskalakis, C.; Ilyas, A.; Syrgkanis, V.; and Zeng, H. 2018. Training GANs with Optimism. In *ICLR*.
- Daskalakis, C.; and Panageas, I. 2019. Last-Iterate Convergence: Zero-Sum Games and Constrained Min-Max Optimization. In *ITCS*.
- Duvocelle, B.; Mertikopoulos, P.; Staudigl, M.; and Vermeulen, D. 2023. Multiagent online learning in time-varying games. *Mathematics of Operations Research*, 48(2): 914–941.
- Feng, Y.; Fu, H.; Hu, Q.; Li, P.; Panageas, I.; Wang, X.; et al. 2023. On the last-iterate convergence in time-varying zero-sum games: Extra gradient succeeds where optimism fails. In *NeurIPS*, 21933–21944.
- Feng, Y.; Li, P.; Panageas, I.; and Wang, X. 2024. Last-iterate Convergence Separation between Extra-gradient and Optimism in Constrained Periodic Games. In *UAI*.
- Fiez, T.; Sim, R.; Skoulakis, S.; Piliouras, G.; and Ratliff, L. 2021. Online learning in periodic zero-sum games. In *NeurIPS*, volume 34, 10313–10325.
- Fudenberg, D.; and Levine, D. K. 1998. *The theory of learning in games*, volume 2. MIT press.
- Fujimoto, Y.; Ariu, K.; and Abe, K. 2024a. Global Behavior of Learning Dynamics in Zero-Sum Games with Memory Asymmetry. *arXiv preprint arXiv:2405.14546*.
- Fujimoto, Y.; Ariu, K.; and Abe, K. 2024b. Nash Equilibrium and Learning Dynamics in Three-Player Matching m -Action Games. *arXiv preprint arXiv:2402.10825*.
- Gidel, G.; Hemmat, R. A.; Pezeshki, M.; Le Priol, R.; Huang, G.; Lacoste-Julien, S.; and Mitliagkas, I. 2019. Negative momentum for improved game dynamics. In *AISTATS*, 1802–1811.
- Hemmat, R. A.; Mitra, A.; Lajoie, G.; and Mitliagkas, I. 2023. LEAD: Min-Max Optimization from a Physical Perspective. *Transactions on Machine Learning Research*.
- Hirsch, M. W.; Smale, S.; and Devaney, R. L. 2013. *Differential equations, dynamical systems, and an introduction to chaos*. Academic press.
- Korpelevich, G. M. 1976. The extragradient method for finding saddle points and other problems. *Matecon*, 12: 747–756.
- Kovachki, N. B.; and Stuart, A. M. 2021. Continuous time analysis of momentum methods. *Journal of Machine Learning Research*, 22(17): 1–40.
- Lee, S.; and Kim, D. 2021. Fast extra gradient methods for smooth structured nonconvex-nonconcave minimax problems. In *NeurIPS*, volume 34, 22588–22600.
- Mawhin, J. 2004. Global results for the forced pendulum equation. In *Handbook of differential equations: Ordinary differential equations*, volume 1, 533–589. Elsevier.
- Mertikopoulos, P.; Lecouat, B.; Zenati, H.; Foo, C.-S.; Chandrasekhar, V.; and Piliouras, G. 2019. Optimistic mirror descent in saddle-point problems: Going the extra (gradient) mile. In *ICLR*.
- Mertikopoulos, P.; Papadimitriou, C.; and Piliouras, G. 2018. Cycles in adversarial regularized learning. In *SODA*, 2703–2717.
- Mertikopoulos, P.; and Sandholm, W. H. 2016. Learning in games via reinforcement and regularization. *Mathematics of Operations Research*, 41(4): 1297–1324.
- Nash Jr, J. F. 1950. Equilibrium points in n -person games. *Proceedings of the National Academy of Sciences*, 36(1): 48–49.
- Piliouras, G.; Nieto-Granda, C.; Christensen, H. I.; and Shamma, J. S. 2014. Persistent patterns: Multi-agent learning beyond equilibrium and utility. In *AAMAS*, 181–188.
- Polyak, B. T. 1964. Some methods of speeding up the convergence of iteration methods. *Ussr computational mathematics and mathematical physics*, 4(5): 1–17.
- Rakhlin, S.; and Sridharan, K. 2013. Optimization, learning, and games with predictable sequences. In *NeurIPS*.
- Sato, Y.; Akiyama, E.; and Farmer, J. D. 2002. Chaos in learning a simple two-person game. *Proceedings of the National Academy of Sciences*, 99(7): 4748–4751.
- Syrgkanis, V.; Agarwal, A.; Luo, H.; and Schapire, R. E. 2015. Fast convergence of regularized learning in games. In *NeurIPS*.
- Yan, Y.-H.; Zhao, P.; and Zhou, Z.-H. 2023. Fast rates in time-varying strongly monotone games. In *ICML*, 39138–39164.
- Zhang, G.; and Wang, Y. 2021. On the suboptimality of negative momentum for minimax optimization. In *AISTATS*, 2098–2106.

Zhang, M.; Zhao, P.; Luo, H.; and Zhou, Z.-H. 2022. No-regret learning in time-varying zero-sum games. In *ICML*, 26772–26808.

Zinkevich, M.; Johanson, M.; Bowling, M.; and Piccione, C. 2007. Regret minimization in games with incomplete information. In *NeurIPS*, 1729–1736.