Nash Equilibrium and Learning Dynamics in Three-Player Matching m-Action Games

Extended Abstract*

Yuma Fujimoto
CyberAgent
Tokyo, Japan
fujimoto.yuma1991@gmail.com

Kaito Ariu
CyberAgent
Tokyo, Japan
kaito_ariu@cyberagent.co.jp

Kenshi Abe
CyberAgent
Tokyo, Japan
abekenshi1224@gmail.com

ABSTRACT

Learning in games discusses the processes where multiple players learn their optimal strategies through the repetition of game plays. The dynamics of learning between two players in zero-sum games, such as Matching Pennies, where their benefits are competitive, have already been well analyzed. However, it is still unexplored and challenging to analyze the dynamics of learning among three players. In this study, we formulate a minimalistic game where three players compete to match their actions with one another. Although interaction among three players diversifies and complicates the Nash equilibria, we fully analyze the equilibria. We also discuss the dynamics of learning based on some famous algorithms categorized into Follow the Regularized Leader. From both theoretical and experimental aspects, we characterize the dynamics by categorizing three-player interactions into three forces to synchronize their actions, switch their actions rotationally, and seek competition.

KEYWORDS

Non-Cooperative Games, Multi-Agent Learning, Evolutionary Game

ACM Reference Format:
Yuma Fujimoto, Kaito Ariu, and Kenshi Abe. 2025. Nash Equilibrium and Learning Dynamics in Three-Player Matching m-Action Games: Extended Abstract. In *Proc. of the 24th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2025), Detroit, Michigan, USA, May 19 – 23, 2025*, IFAAMAS, 3 pages.

1 INTRODUCTION

Dynamical systems approach is often taken for analyzing learning in games [1, 2, 27, 28]. This is because gradient-based algorithms fail to converge to the Nash equilibrium in zero-sum games, where the utility functions of two agents conflict. Indeed, a representative class of learning algorithms, Follow the Regularized Leader (FTRL) [17, 18, 22], which is tied to replicator dynamics [3, 7, 12, 12, 21, 26] and gradient ascent [6, 11, 24, 29], cannot stop a cycling behavior even in a simple game like Matching Pennies [2, 3]. This cycling behavior is understood based on the Bregman divergence [1, 17, 20], which corresponds to the distance from the equilibrium. Dynamical systems are also pivotal to discuss convergent algorithms to the equilibrium [4, 8]. This convergence is understood based on some Lyapunov functions, which globally decrease with time.

Despite such a thorough understanding of two-player games, three-player games are difficult to understand in general. Classically, this difficulty is seen in Jordan's game [13], which is a three-player version of Matching Pennies. In this game, the divergence from the equilibrium is observed [9, 15, 16, 23]; the distance from the equilibrium is no longer conserved. This divergence is not seen in Matching Pennies. Thus, the motivation to study complex dynamics in learning in three-player games is established [14, 19]. In addition, the Nash equilibria of three-player games are hard to fully analyze [5], even when the games are zero-sum.

Our contribution: This study proposes Three-Player Matching m-Action (m-3MA) game as an extension of Matching Pennies. Despite the Nash equilibrium becomes complex, we fully analyze it. We further introduce the continuous-time FTRL, characterize it by a Lyapunov function V and the Bregman divergence G, and interpret it based on three parameters, α, β, and γ.

2 PRELIMINARY

We now formulate m-3MA games (see Fig. 1). Let X, Y, and Z denote three players. Every round, they independently determine their actions from the same m-action set, $\mathcal{A} = \{a_1, \cdots, a_m\}$. Players who choose the same action interact with each other. This interaction follows a three-way deadlock relationship among them: X wins Y, Y wins Z, but Z wins X. They receive their scores. When only two of them interact, the winner and loser are determined following the three-way deadlock relationship, and the winner's and loser's scores are a and b, respectively. Players who chose a different action from the others receive the default payoff of c. If all three players

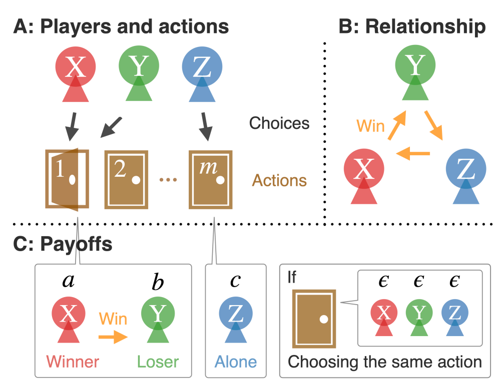

Figure 1: Illustration of m-3MA games.

*The full version is available at https://arxiv.org/abs/2402.10825

Proc. of the 24th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2025), Y. Vorobeychik, S. Das, A. Nowé (eds.), May 19 – 23, 2025, Detroit, Michigan, USA. © 2025 International Foundation for Autonomous Agents and Multiagent Systems (www.ifaamas.org).

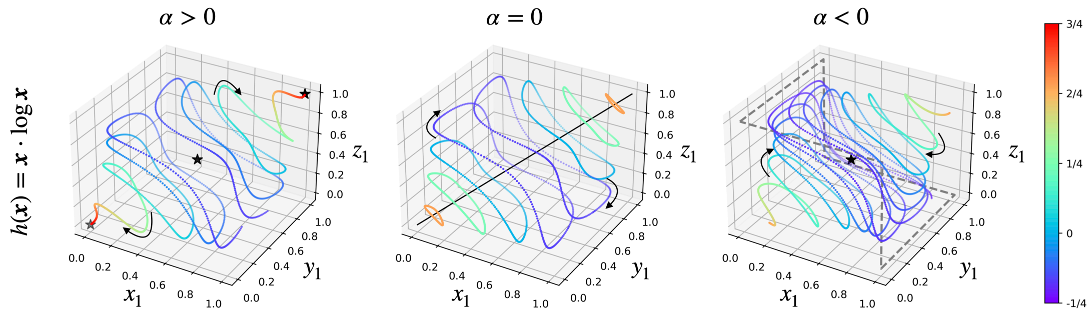

Figure 2: Learning dynamics in m-3MA game for $m = 2$. The color indicates the value of V.

take the same action, they commonly receive the scores of ϵ. Here, we assume that the winner's and loser's scores are highest and lowest, respectively, i.e., $b < c < a$ and $b < \epsilon < a$.

Let $\boldsymbol{x} := (x_1, \cdots, x_m) \in \Delta^{m-1}$ (the $m-1$ dimensional simplex) denote X's strategy, where x_i is the probability that X chooses action a_i. Similarly, Y's and Z's strategies are denoted by $\boldsymbol{y} \in \Delta^{m-1}$ and $\boldsymbol{z} \in \Delta^{m-1}$, respectively. When players follow such strategies, X's expected payoff is given by

$$u(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}) = \epsilon \sum_i x_i y_i z_i + a \sum_i x_i y_i \bar{z}_i + b \sum_i x_i \bar{y}_i z_i + c \sum_i x_i \bar{y}_i \bar{z}_i,$$

where we defined $\bar{\mathcal{X}} := 1 - \mathcal{X}$ for arbitrary variable $\mathcal{X}$. Y's and Z's expected payoffs are also described as $u(\boldsymbol{y}, \boldsymbol{z}, \boldsymbol{x})$ and $u(\boldsymbol{z}, \boldsymbol{x}, \boldsymbol{y})$, respectively. These m-3MA games are characterized by three parameters, $\alpha := \epsilon - c$, $\beta := a - b > 0$, and $\gamma := a + b - 2c$.

3 NASH EQUILIBRIUM

The Nash equilibrium of m-3MA is defined as the set of strategies $(\boldsymbol{x}^*, \boldsymbol{y}^*, \boldsymbol{z}^*)$ which maximize their expected payoffs, respectively.

It is difficult to derive equilibrium in three-player games [5], and indeed, there are few successful studies [10, 25]. Nevertheless, we can fully analyze all the Nash equilibria and interpret them as follows.

Theorem 1 (Main properties of the Nash equilibria). *First, the following property always holds.*

- **(Player symmetry)** *For any Nash equilibrium, all three players take the same strategy, i.e., $\boldsymbol{x}^* = \boldsymbol{y}^* = \boldsymbol{z}^*$.*

Furthermore, the region of the Nash equilibria has the following properties.

- **(Neutral equilibria)** *When $\alpha = \gamma = 0$, all the strategies in the simplex Δ^{m-1} can be the Nash equilibria.*
- **(Pure-strategy equilibria)** *$\mathcal{N}_{\mathrm{P}}(m) = \{\boldsymbol{e}_1, \cdots, \boldsymbol{e}_m\}$ are the Nash equilibrium strategies, if and only if $\alpha \geq 0$.*
- **(Uniform-choice equilibrium)** *$\mathcal{N}_{\mathrm{U}}(m) = \{\boldsymbol{1}/m\}$ is always the Nash equilibrium strategy.*

4 LEARNING DYNAMICS

We introduce the continuous-time FTRL;

$$\boldsymbol{x} = \boldsymbol{q}(\boldsymbol{x}^\dagger), \quad \dot{\boldsymbol{x}}^\dagger = \frac{\partial u}{\partial \boldsymbol{x}}, \quad \boldsymbol{q}(\boldsymbol{x}^\dagger) = \arg\max_{\boldsymbol{x}} \left\{ \boldsymbol{x}^\dagger \cdot \boldsymbol{x} - h(\boldsymbol{x}) \right\}.$$

Here, $h(\boldsymbol{x})$ is "regularizer", a penalty term in projecting the updated strategy back to its strategy space. Several representative examples are the entropic regularizer $h(\boldsymbol{x}) = \boldsymbol{x} \cdot \log \boldsymbol{x}$ and the Euclidean regularizer $h(\boldsymbol{x}) = \|\boldsymbol{x}\|_2^2/2$.

To investigate learning dynamics given by the continuous-time FTRL, we introduce G and V as

$$G(\boldsymbol{x}^\dagger, \boldsymbol{y}^\dagger, \boldsymbol{z}^\dagger) := \Sigma_{\mathrm{cyc}} \max_{\boldsymbol{x}} \{\boldsymbol{x}^\dagger \cdot \boldsymbol{x} - h(\boldsymbol{x})\} - \boldsymbol{x}^\dagger \cdot \boldsymbol{1}/m,$$

$$V(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}) := \Sigma_i \, x_i y_i z_i - 1/m^2.$$

Here, Σ_{cyc} indicates the cyclic sum for three players. In other words, $\Sigma_{\mathrm{cyc}} \, \mathcal{F}(\boldsymbol{x}) = \mathcal{F}(\boldsymbol{x}) + \mathcal{F}(\boldsymbol{y}) + \mathcal{F}(\boldsymbol{z})$ holds for arbitrary function $\mathcal{F}$. We now explain the meanings of G and V. First, G is known to be conserved under zero-sum games [17] and corresponds to the Bregman divergence from the uniform-choice equilibrium. Next, V means the probability that all three players choose the same action, in other words, the degree of synchronization of their action choices.

We now consider m-3MA with the case of $m = 2$. Since $m = 2$ holds, $x_2 = 1 - x_1$, $y_2 = 1 - y_1$, and $z_2 = 1 - z_1$ hold so that we can describe the learning dynamics by only the three variables of (x_1, y_1, z_1). The continuous-time FTRL are independent of γ and interpreted as follows.

Theorem 2 (Global behavior of dynamics). *In m-3MA with $m = 2$, the continuous-time FTRL with the entropic and Euclidean regularizers gives the following properties in general.*

- *When $\alpha = 0$, both G and V are conserved in the trajectory.*
- *When $\alpha > 0$, the trajectory asymptotically converges to the states of maximum V, i.e., either of the fixed points.*
- *When $\alpha < 0$, the trajectory asymptotically converges to the states of minimum V, i.e., the heteroclinic cycle.*

For general m, we capture the intuitions of α, β, and γ as follows.

- α contributes to synchronization. When $\alpha > 0$ (resp. $\alpha < 0$), the players learn to synchronize (desynchronize) their actions.
- β contributes to rotation. The larger β is, the faster the cycling behavior of the learning is.
- γ matters only for $m > 2$ and contributes to the frequency of competition. When $\gamma > 0$, the players prune their action choices to two. When $\gamma < 0$, they decentralize their action choices.

ACKNOWLEDGMENTS

K. Ariu is supported by JSPS KAKENHI Grant No. 23K19986.

REFERENCES

[1] James P Bailey and Georgios Piliouras. 2019. Multi-Agent Learning in Network Zero-Sum Games is a Hamiltonian System. In *AAMAS*. 233–241.

[2] Daan Bloembergen, Karl Tuyls, Daniel Hennes, and Michael Kaisers. 2015. Evolutionary dynamics of multi-agent learning: A survey. *Journal of Artificial Intelligence Research* 53 (2015), 659–697.

[3] Tilman Börgers and Rajiv Sarin. 1997. Learning through reinforcement and replicator dynamics. *Journal of Economic Theory* 77, 1 (1997), 1–14.

[4] Yun Kuen Cheung and Georgios Piliouras. 2020. Chaos, extremism and optimism: Volume analysis of learning in games. In *NeurIPS*.

[5] Constantinos Daskalakis and Christos H Papadimitriou. 2005. Three-player games are hard. In *Electronic colloquium on computational complexity: ECCC*.

[6] Ulf Dieckmann and Richard Law. 1996. The dynamical theory of coevolution: a derivation from stochastic ecological processes. *Journal of mathematical biology* 34 (1996), 579–612.

[7] Daniel Friedman. 1991. Evolutionary games in economics. *Econometrica: Journal of the Econometric Society* (1991), 637–666.

[8] Yuma Fujimoto, Kaito Ariu, and Kenshi Abe. 2024. Memory Asymmetry Creates Heteroclinic Orbits to Nash Equilibrium in Learning in Zero-Sum Games. In *AAAI*.

[9] Andrea Gaunersdorfer and Josef Hofbauer. 1995. Fictitious play, Shapley polygons, and the replicator equation. *Games and Economic Behavior* 11, 2 (1995), 279–303.

[10] William C Grant. 2023. Correlated Equilibrium and Evolutionary Stability in 3-Player Rock-Paper-Scissors. *Games* 14, 3 (2023), 45.

[11] Josef Hofbauer and Karl Sigmund. 1990. Adaptive dynamics and evolutionary stability. *Applied Mathematics Letters* 3, 4 (1990), 75–79.

[12] Josef Hofbauer, Karl Sigmund, et al. 1998. *Evolutionary games and population dynamics*. Cambridge university press.

[13] James S Jordan. 1993. Three problems in learning mixed-strategy Nash equilibria. *Games and Economic Behavior* 5, 3 (1993), 368–386.

[14] Robert D Kleinberg, Katrina Ligett, Georgios Piliouras, and Éva Tardos. 2011. Beyond the Nash Equilibrium Barrier. In *ICS*.

[15] Kevin A McCabe, Arijit Mukherji, and David E Runkle. 2000. An experimental study of information and mixed-strategy play in the three-person matching-pennies game. *Economic Theory* 15 (2000), 421–462.

[16] Richard Mealing and Jonathan L Shapiro. 2015. Convergence of Strategies in Simple Co-Adapting Games. In *FOGA*.

[17] Panayotis Mertikopoulos, Christos Papadimitriou, and Georgios Piliouras. 2018. Cycles in adversarial regularized learning. In *SODA*.

[18] Panayotis Mertikopoulos and William H Sandholm. 2016. Learning in games via reinforcement and regularization. *Mathematics of Operations Research* 41, 4 (2016), 1297–1324.

[19] Sai Ganesh Nagarajan, Sameh Mohamed, and Georgios Piliouras. 2018. Three body problems in evolutionary game dynamics: Convergence, periodicity and limit cycles. In *AAMAS*.

[20] Georgios Piliouras, Carlos Nieto-Granda, Henrik I Christensen, and Jeff S Shamma. 2014. Persistent patterns: Multi-agent learning beyond equilibrium and utility. In *AAMAS*.

[21] Yuzuru Sato, Eizo Akiyama, and J Doyne Farmer. 2002. Chaos in learning a simple two-person game. *Proceedings of the National Academy of Sciences* 99, 7 (2002), 4748–4751.

[22] Shai Shalev-Shwartz and Yoram Singer. 2006. Convex repeated games and Fenchel duality. In *NeurIPS*.

[23] Jeff S Shamma and Gürdal Arslan. 2005. Dynamic fictitious play, dynamic gradient play, and distributed convergence to Nash equilibria. *IEEE Trans. Automat. Control* 50, 3 (2005), 312–327.

[24] Satinder Singh, Michael J Kearns, and Yishay Mansour. 2000. Nash Convergence of Gradient Dynamics in General-Sum Games.. In *UAI*.

[25] Duane Szafron, Richard G Gibson, and Nathan R Sturtevant. 2013. A parameterized family of equilibrium profiles for three-player kuhn poker.. In *AAMAS*.

[26] Peter D Taylor and Leo B Jonker. 1978. Evolutionary stable strategies and game dynamics. *Mathematical biosciences* 40, 1-2 (1978), 145–156.

[27] Karl Tuyls, Pieter Jan'T Hoen, and Bram Vanschoenwinkel. 2006. An evolutionary dynamical analysis of multi-agent learning in iterated games. In *AAMAS*.

[28] Karl Tuyls and Ann Nowé. 2005. Evolutionary game theory and multi-agent reinforcement learning. *The Knowledge Engineering Review* 20, 1 (2005), 63–90.

[29] Martin Zinkevich. 2003. Online convex programming and generalized infinitesimal gradient ascent. In *ICML*.