Last-iterate Convergence of Heterogeneous Learning in Time-Varying Zero-Sum Games

Extended Abstract

Junyue Zhang
Academy of Mathematics and Systems Science, Chinese Academy of Sciences
Beijing, China
zhangjunyue@amss.ac.cn

Xinxiang Guo
Academy of Mathematics and Systems Science, Chinese Academy of Sciences
Beijing, China
guoxinxiang@amss.ac.cn

Yi Feng
Aarhus University
Aarhus, Denmark
yi.feng@cs.au.dk

Xiao Wang
MoE Key Laboratory of Interdisciplinary Research of Computation and Economics, Shanghai University of Finance and Economics
Shanghai, China
wangxiao@sufe.edu.cn

Yifen Mu
State Key Laboratory of Mathematical Sciences (SKLMS), Academy of Mathematics and Systems Science, Chinese Academy of Sciences
Beijing, China
mu@amss.ac.cn

ABSTRACT

In this paper, we study heterogeneous learning in time-varying zero-sum games, where one player employs Mirror Descent (MD), a family covering numerous well-known algorithms such as Hedge and Gradient Descent, while the opponent best-responds. We extend the commonly studied time-varying games to a broader class, which captures essential features that influence the evolution of heterogeneous dynamics, and prove last-iterate convergence of MD. Notably, MD converges whereas Optimistic Gradient Descent Ascent fails in periodic games; MD also converges in convergent perturbed games but under weaker assumptions than those in prior work [10]. Numerical simulations validate our theoretical results and demonstrate that heterogeneous learning improves stability and accelerates convergence compared to standard homogeneous approaches. This provides the first theoretical and empirical support for heterogeneous learning dynamics in time-varying games.

KEYWORDS

Last-iterate convergence; Time-varying games; Heterogeneous learning; Zero-sum games

ACM Reference Format:
Junyue Zhang, Xinxiang Guo, Yi Feng, Xiao Wang, and Yifen Mu. 2026. Last-iterate Convergence of Heterogeneous Learning in Time-Varying Zero-Sum Games: Extended Abstract. In *Proc. of the 25th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2026), Paphos, Cyprus, May 25 – 29, 2026*, IFAAMAS, 3 pages. https://doi.org/10.65109/VFNO5777

Proc. of the 25th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2026), C. Amato, L. Dennis, V. Mascardi, J. Thangarajah (eds.), May 25 – 29, 2026, Paphos, Cyprus. © 2026 International Foundation for Autonomous Agents and Multiagent Systems (www.ifaamas.org). https://doi.org/10.65109/VFNO5777

1 INTRODUCTION

Beyond time-averaged convergence, recent research has increasingly focused on the day-to-day dynamics of learning algorithms in games. Unfortunately, the dynamics of most classical algorithms under self-play, such as Mirror Descent (MD) and Follow-The-Regularized-Leader, have been shown to be non-convergent and even chaotic in time-invariant zero-sum games [2, 22]. Optimistic variants of these algorithms incorporate a predictive term into their update rules, which leverage the intrinsic homogeneous structure of self-play, and thereby attain last-iterate convergence in time-invariant zero-sum games [3, 7, 8, 21, 26].

However, the assumption of time-invariant payoffs means a stationary learning environment, which, as demonstrated in [4, 9, 19], is often unrealistic in practice. Many real-world scenarios exhibit non-stationary environments where agents' utilities evolve dynamically [16, 20]. From a technical perspective, time-varying games present more challenging settings for achieving last-iterate convergence, since the continually changing payoffs can diffuse the updates of learning algorithms. Indeed, recent studies have shown that even optimistic variants, such as Optimistic Gradient Descent Ascent (OGDA), fail to achieve last-iterate convergence in such environments [10, 11, 13].

Notably, all of the aforementioned studies focus on homogeneous learning, where both players employ the same algorithm to update their strategies simultaneously. The failure of homogeneous learning in both time-invariant and time-varying games suggests that these frameworks may have inherent limitations with respect to last-iterate convergence, motivating a shift of attention toward heterogeneous learning.

In this paper, heterogeneous learning is referred to the paradigm where players in a game update their strategies using different ways. It has garnered increasing attention due to its broad practical applications [1, 14, 15, 17, 18, 23–25]. However, the performance of heterogeneous learning in time-varying games remains largely unexplored. In particular, whether heterogeneous learning admits last-iterate convergence under dynamically changing payoffs is still unknown. This gap motivates us to consider the following question:

Will heterogeneous learning attain last-iterate convergence in time-varying zero-sum games?

2 DECOMPOSABLE TIME-VARYING ZERO-SUM GAMES

Let $\{C_t\}_{t\geq 0}$ be a sequence of payoff matrices defining a time-varying zero-sum game. At each round t, the row player (Player X) chooses $x_t \in \Delta(\mathcal{I})$ and the column player (Player Y) chooses $y_t \in \Delta(\mathcal{J})$, and the column player's payoff is $x_t^\top C_t y_t$.

We introduce a novel class of time-varying zero-sum games called *decomposable time-varying zero-sum games*, where C_t admits the decomposition:

$$C_t = A_t + \mathcal{E}_t, \tag{1}$$

with the following properties:

(1) The Nash policy sets $\{\mathcal{X}^*_{A_t}\}$ for player X of *non-vanishing term* A_t have limiting points:
$$\mathcal{X}^* := \left\{x^* \mid \exists \{x_t^*\}_{t\geqslant 0} \text{ with } x_t^* \in \mathcal{X}^*_{A_t}, \text{ s.t. } x_t^* \to x^*\right\} \neq \emptyset.$$
And there exists $x^* \in \mathcal{X}^*$ such that:
$$\inf_{x_t^* \in \mathcal{X}^*_{A_t}} \operatorname{dist}(x_t^*, x^*) = O(1/t^q), \quad q > 0.$$

(2) The *vanishing term* $\mathcal{E}_t$ converges to $\mathbf{O}$ with rate ε and the *non-vanishing term* A_t is uniformly bounded by $M > 0$:
$$\|\mathcal{E}_t\|_1 = O(1/t^\varepsilon), \varepsilon > 0, \quad \|A_t\|_1 \leqslant M.$$

Remarkably, this new class of time-varying games covers several important cases in the literature:

Example 1 (Periodic games). *The periodic games studied in [10, 12, 13] can be represented as $C_t = A_t + \mathbf{O}$ where $A_{t+T} = A_t$ for some period $T > 0$ and $\mathcal{X}^* = \mathcal{X}^*_{A_t}$ for all $t > 0$.*

Example 2 (Convergent perturbed games). *The convergent perturbed games considered in [6, 9, 10] correspond to the case where $C_t = A + \mathcal{E}_t$ with $\|\mathcal{E}_t\|_1 = O(1/t^\varepsilon)$.*

3 MAIN RESULTS

In this section, we propose a general scheme of heterogeneous learning method as shown by Algorithm 1, where one player employs MD and the other best responds to the MD player, and establish its last-iterate convergence in time-varying zero-sum games.

Assumption 1. *For decomposable time-varying zero-sum games with decomposition* (1)*, the Player X in Algorithm 1 chooses the step size $1/\lambda_t$ that fulfill standard stochastic approximation requirements [5] and can bound the accumulated variants:*

$$\sum_{t=0}^{\infty} \frac{1}{\lambda_t} = \infty, \quad \sum_{t=0}^{\infty} \frac{1}{\lambda_t^2} < \infty, \quad \sum_{t=0}^{\infty} \frac{t^{-\min\{\varepsilon,q\}}}{\lambda_t} < \infty.$$

We now present our main convergence results.

Theorem 1. *Let $\{x_t\}$ be the strategy sequence generated by Algorithm 1. Then under Assumption 1, the following result holds:*

$$V_t(x_t) - v_t^* = o\left(\frac{\lambda_t}{t}\right),$$

where $V_t(x) := \max_{y\in\Delta(\mathcal{J})} x^\top C_t y$, and $v_t^* := \min_{x\in\Delta(\mathcal{I})} V_t(x)$.

Algorithm 1 Mirror Descent learning against Best Response

Require: Initial strategy x_0, distance-generating function ψ, time-varying games $\{C_t\}$.

for $t = 0, 1, 2, \ldots$ **do**

  Take the Best Response y_t to x_t with respect to C_t,

  Update x_{t+1} based on Mirror Descent:

$$x_{t+1} = \arg\min_{x\in\Delta(\mathcal{I})} \{\langle C_t y_t, x - x_t\rangle + \lambda_t D_\psi(x, x_t)\}.$$

The semi-duality gap $V_t(x_t) - v_t^*$ provides a natural performance measure, evaluating how well x_t performs against the most adversarial opponent in each round. Theorem 1 demonstrates that the MD player can learn robust strategies even under such adversarial conditions. The robustness stems from the gradient information provided by alternating Best Response adversaries. The MD method, with increasing λ_t under Assumption 1, effectively mitigates estimation errors induced by time-varying payoffs, enabling $\{x_t\}$ to track the evolving Nash policy sets $\mathcal{X}^*_{C_t}$.

In this following, we empirically verify the convergence of heterogeneous learning dynamics in a randomly generated convergent perturbed games, and show how the decay rate of the step size sequence influences the convergence behavior. We generate a random 100×200 payoff matrix C with entries from $(-5, 5)$ as the limiting game. Figure 1 presents the semi-duality gap sequences for Hedge and PGD under different decay rates of the step size. With a constant step size (i.e., no decay), the gap sequences under both Hedge and PGD fail to converge. In contrast, appropriately decaying step sizes ensure convergence, and a larger step size leads to slower convergence, consistent with Theorem 1.

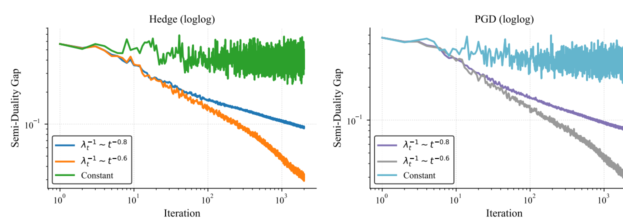

Figure 1: Convergence of heterogeneous learning under different step-size rates in convergent perturbed games.

ACKNOWLEDGMENTS

The research of Junyue Zhang, Xinxiang Guo and Yifen Mu is supported by the Natural Science Foundation of China under Grant T2293770, the National Key Research and Development Program of China under grant No.2022YFA1004600, and the Key Laboratory of Interdisciplinary Research of Computation and Economics, Shanghai University of Finance and Economics, Ministry of Education. The research of Xiao Wang is supported by National Key R&D Program of China under (2023YFA1009500), the Fundamental Research Funds for the Central Universities, Shanghai Sci-tech Co-research Program (25HB2703800).

REFERENCES

[1] Yuksel Arslantas, Ege Yuceel, Yigit Yalin, and Muhammed O. Sayin. 2025. Convergence of Heterogeneous Learning Dynamics in Zero-Sum Stochastic Games. *IEEE Trans. Automat. Control* 70, 11 (2025), 7523–7537. https://doi.org/10.1109/TAC.2025.3576039

[2] James P. Bailey and Georgios Piliouras. 2018. Multiplicative Weights Update in Zero-Sum Games. In *Proceedings of the 2018 ACM Conference on Economics and Computation* (Ithaca, NY, USA) *(EC '18)*. Association for Computing Machinery, New York, NY, USA, 321–338. https://doi.org/10.1145/3219166.3219235

[3] Yang Cai, Argyris Oikonomou, and Weiqiang Zheng. 2022. Tight Last-Iterate Convergence of the Extragradient and the Optimistic Gradient Descent-Ascent Algorithm for Constrained Monotone Variational Inequalities. arXiv:2204.09228 [math.OC] https://arxiv.org/abs/2204.09228

[4] Adrian Rivera Cardoso, Jacob Abernethy, He Wang, and Huan Xu. 2019. Competing Against Nash Equilibria in Adversarially Changing Zero-Sum Games. In *Proceedings of the 36th International Conference on Machine Learning (Proceedings of Machine Learning Research, Vol. 97)*, Kamalika Chaudhuri and Ruslan Salakhutdinov (Eds.). PMLR, San Diego, CA, USA, 921–930. https://proceedings.mlr.press/v97/cardoso19a.html

[5] Han-Fu Chen. 2002. *Stochastic Approximation and Its Applications*. Springer New York, NY.

[6] Yanzheng Chen and Jun Yu. 2025. Classic but Everlasting: Traditional Gradient-Based Algorithms Converge Fast Even in Time-Varying Multi-Player Games. In *International Conference on Learning Representations*, Y. Yue, A. Garg, N. Peng, F. Sha, and R. Yu (Eds.), Vol. 2025. 41119–41155. https://proceedings.iclr.cc/paper_files/paper/2025/file/65b721a1df04c1098567f70d483d6468-Paper-Conference.pdf

[7] Constantinos Daskalakis, Andrwe Ilyas, Vasilis Syrgkanis, and Haoyang Zeng. 2018. Training GANs with Optimism. In *ICLR 2018 Conference* (iclr 2018 conference ed.). https://www.microsoft.com/en-us/research/publication/training-gans-optimism/

[8] Constantinos Daskalakis and Ioannis Panageas. 2019. Last-Iterate Convergence: Zero-Sum Games and Constrained Min-Max Optimization. In *10th Innovations in Theoretical Computer Science Conference (ITCS 2019) (Leibniz International Proceedings in Informatics (LIPIcs), Vol. 124)*, Avrim Blum (Ed.). Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl, Germany, 27:1–27:18. https://doi.org/10.4230/LIPIcs.ITCS.2019.27

[9] Benoit Duvocelle, Panayotis Mertikopoulos, Mathias Staudigl, and Dries Vermeulen. 2023. Multiagent Online Learning in Time-Varying Games. *Mathematics of Operations Research* 48, 2 (2023), 914–941. https://doi.org/10.1287/moor.2022.1283 arXiv:https://doi.org/10.1287/moor.2022.1283

[10] Yi Feng, Hu Fu, Qun Hu, Ping Li, Ioannis Panageas, Bo Peng, and Xiao Wang. 2023. On the last-iterate convergence in time-varying zero-sum games: extra gradient succeeds where optimism fails. In *Proceedings of the 37th International Conference on Neural Information Processing Systems* (New Orleans, LA, USA) *(NIPS '23)*. Curran Associates Inc., Red Hook, NY, USA, Article 961, 12 pages.

[11] Yi Feng, Ping Li, Ioannis Panageas, and Xiao Wang. 2024. Last-iterate convergence separation between extra-gradient and optimisim in constrained periodic games. In *Proceedings of the Fortieth Conference on Uncertainty in Artificial Intelligence* (Barcelona, Spain) *(UAI '24)*. JMLR.org, Barcelona, Spain, Article 63, 32 pages.

[12] Tanner Fiez, Ryann Sim, Stratis Skoulakis, Georgios Piliouras, and Lillian Ratliff. 2021. Online learning in periodic zero-sum games. *Advances in Neural Information Processing Systems* 34 (2021), 10313–10325.

[13] Yuma Fujimoto, Kaito Ariu, and Kenshi Abe. 2025. Synchronization in learning in periodic zero-sum games triggers divergence from nash equilibrium. In *Proceedings of the Thirty-Ninth AAAI Conference on Artificial Intelligence and Thirty-Seventh Conference on Innovative Applications of Artificial Intelligence and Fifteenth Symposium on Educational Advances in Artificial Intelligence (AAAI'25/IAAI'25/EAAI'25)*. AAAI Press, Article 2587, 9 pages. https://doi.org/10.1609/aaai.v39i22.34485

[14] Xinxiang Guo and Yifen Mu. 2025. Globally Optimal Strategy against the Hedge Algorithm in Repeated Games. *Journal of Global Optimization* 93, 3 (Nov. 2025), 861–895. https://doi.org/10.1007/s10898-025-01554-w

[15] Xinxiang Guo, Yifen Mu, and Xiaoguang Yang. 2025. Periodicity in Hedge-Myopic System and an Asymmetric NE-Solving Paradigm for Two-Player Zero-Sum Games. *Dynamic Games and Applications* (May 2025). https://doi.org/10.1007/s13235-025-00649-z

[16] F P Kelly, A K Maulloo, and D K H Tan. 1998. Rate Control for Communication Networks: Shadow Prices, Proportional Fairness and Stability. *Journal of the Operational Research Society* 49, 3 (March 1998), 237–252. https://doi.org/10.1057/palgrave.jors.2600523

[17] Edward Lockhart, Marc Lanctot, Julien Pérolat, Jean-Baptiste Lespiau, Dustin Morrill, Finbarr Timbers, and Karl Tuyls. 2019. Computing approximate equilibria in sequential adversarial games by exploitability descent. In *Proceedings of the 28th International Joint Conference on Artificial Intelligence* (Macao, China) *(IJCAI'19)*. AAAI Press, Macao, China, 464–470.

[18] Aleksander Madry, Aleksandar Makelov, Ludwig Schmidt, Dimitris Tsipras, and Adrian Vladu. 2019. Towards Deep Learning Models Resistant to Adversarial Attacks. arXiv:1706.06083 [stat.ML] https://arxiv.org/abs/1706.06083

[19] Tung Mai, Milena Mihail, Ioannis Panageas, Will Ratcliff, Vijay Vazirani, and Peter Yunker. 2018. Cycles in Zero-Sum Differential Games and Biological Diversity. In *Proceedings of the 2018 ACM Conference on Economics and Computation* (Ithaca, NY, USA) *(EC '18)*. Association for Computing Machinery, New York, NY, USA, 339–350. https://doi.org/10.1145/3219166.3219227

[20] Panayotis Mertikopoulos, E. Veronica Belmega, Romain Negrel, and Luca Sanguinetti. 2017. Distributed Stochastic Optimization via Matrix Exponential Learning. *IEEE Transactions on Signal Processing* 65, 9 (2017), 2277–2290. https://doi.org/10.1109/TSP.2017.2656847

[21] Panayotis Mertikopoulos, Bruno Lecouat, Houssam Zenati, Chuan-Sheng Foo, Vijay Chandrasekhar, and Georgios Piliouras. 2018. Optimistic mirror descent in saddle-point problems: Going the extra (gradient) mile. arXiv:1807.02629 [cs.LG] https://arxiv.org/abs/1807.02629

[22] Panayotis Mertikopoulos, Christos Papadimitriou, and Georgios Piliouras. 2018. Cycles in adversarial regularized learning. In *Proceedings of the Twenty-Ninth Annual ACM-SIAM Symposium on Discrete Algorithms* (New Orleans, Louisiana) *(SODA '18)*. Society for Industrial and Applied Mathematics, USA, 2703–2717.

[23] Hadi Salman, Andrew Ilyas, Logan Engstrom, Ashish Kapoor, and Aleksander Madry. 2020. Do adversarially robust imagenet models transfer better? *Advances in Neural Information Processing Systems* 33 (2020), 3533–3545.

[24] Dimitris Tsipras, Shibani Santurkar, Logan Engstrom, Alexander Turner, and Aleksander Madry. 2019. Robustness May Be at Odds with Accuracy. arXiv:1805.12152 [stat.ML] https://arxiv.org/abs/1805.12152

[25] Oriol Vinyals, Igor Babuschkin, Wojciech M Czarnecki, Michaël Mathieu, Andrew Dudzik, Junyoung Chung, David H Choi, Richard Powell, Timo Ewalds, Petko Georgiev, et al. 2019. Grandmaster level in StarCraft II using multi-agent reinforcement learning. *nature* 575, 7782 (2019), 350–354.

[26] Chen-Yu Wei, Chung-Wei Lee, Mengxiao Zhang, and Haipeng Luo. 2021. Linear Last-iterate Convergence in Constrained Saddle-point Optimization. arXiv:2006.09517 [cs.LG] https://arxiv.org/abs/2006.09517