

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

Anonymous Author(s)

Submission Id: 599

## ABSTRACT

Last-iterate convergence has become a central topic in game theory, and seminal work has established that many homogeneous learning algorithms such as Extra-Gradient and Optimistic Gradient Descent Ascent (OGDA) attain it in time-invariant zero-sum games [9, 10, 18, 28]. However, recent studies reveal that these algorithms may fail to achieve last-iterate convergence in time-varying games, as they lack robustness to perturbed or periodically changing payoffs [8, 13, 14, 16]. Meanwhile, heterogeneous learning, widely used in practice (e.g., adversarial training [17] and AlphaStar [34]), has shown strongly effective empirical performance but still lacks theoretical justification. Motivated by this gap, 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 OGDA fails in periodic games; MD also converges in perturbed convergent games but under weaker assumptions than those in prior work [13]. Our results further extend to asymptotic zero-sum games, encompassing a broader class of time-varying non-zero-sum games. 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:

Anonymous Author(s). 2026. Last-iterate Convergence of Heterogeneous Learning in Time-Varying Zero-Sum Games. In *Proc. of the 25th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2026)*, Paphos, Cyprus, May 25 – 29, 2026, IFAAMAS, 9 pages.

## 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, 29]. 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 [4, 9, 10, 28, 37].

However, the assumption of time-invariant payoffs means a stationary learning environment, which, as demonstrated in [5, 12, 26], is often unrealistic in practice. Many real-world scenarios exhibit non-stationary environments where agents' utilities evolve dynamically, such as in auctions [22] and power control systems [27]. To address this limitation, a growing body of work has shifted focus from time-invariant to time-varying games, aiming to better capture the dynamics of real-world learning environments. 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 [13, 14, 16].

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. For example, in adversarial training, heterogeneous learning significantly enhances the robustness and performance of neural networks in classification tasks [25, 32, 33]. Similar techniques are also employed in the training of AlphaStar [34]. Beyond machine learning applications, heterogeneous learning has been leveraged to accelerate equilibrium computation and improve efficiency [1]. For instance, the exploitability descent (ED) algorithm [24] was proposed for extensive-form zero-sum games by combining gradient descent and best-response updates. Two-timescale gradient descent-ascent [23] can be viewed as a smoothed version of ED and has been applied to the training of GANs [20]. Besides, heterogeneous learning also provides novel insights into evolutionary behavior [3] and recent work has shown that playing best response against MWU can induce periodic dynamics, in sharp contrast to the chaotic dynamics observed in self-play [19]. Despite these advances, 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?

**Our Contributions.** We consider a specific heterogeneous learning setting and prove its last-iterate convergence. Specifically, one player adopts MD, a family of algorithms that encompasses two of the most popular learning algorithms, Hedge and Gradient Descent, while the other player takes Best Response. The detailed contributions listed below demonstrate the power of heterogeneous learning in handling perturbations and integrating common equilibrium information across time.

- In sharp contrast to the failure of state-of-the-art optimistic variants in time-varying games, heterogeneous learning is capable of handling a broader class of games beyond those commonly studied. To further explore the capability limits of heterogeneous learning, we introduce a new class of time-varying games, named decomposable time-varying games, from the perspective of game information decomposition. This class unifies both periodic games and convergent perturbed games as special cases, and reveals the essential informational structure that governs the evolution of dynamics.
- We establish the last-iterate convergence of heterogeneous learning with appropriately step sizes in decomposable time-varying games. The step size of MD plays a crucial role in enabling adaptation to the time-varying payoff information. Notably, the required condition on the perturbation term for convergence in heterogeneous learning is considerably milder than that of optimistic variants reported in the literature. To further examine the capability of heterogeneous learning, we extend our convergence results to asymptotic zero-sum games, a family of general-sum games, thereby substantially broadening the class of time-varying games that admit last-iterate convergence.
- Extensive simulations validate our theoretical findings and demonstrate that the proposed heterogeneous learning framework outperforms existing homogeneous methods in convergence rate and stability, including EG and OGDA. Further experiments on periodic games highlights the influence of payoff matrix on the behavior of the learning trajectory.

### 1.1 Comparison with Related Work

**Heterogeneous Learning.** The algorithmic setting considered in this paper shares similarities with those studied in prior work. Johanson et al. [21] proposed an asymmetric learning technique, CFR-BR, for imperfect-information extensive-form zero-sum games and proved its time-averaged convergence to Nash equilibrium. The exploitability descent algorithm [24], which combines policy gradient, a special case of MD, with best-response updates, guarantees best-iterate convergence. The two-timescale gradient descent-ascent algorithm [23] can be viewed as a smoothed version of exploitability descent and achieves time-averaged convergence in nonconvex-concave minimax problems. Guo et al. [19] further demonstrated the emergence of periodic patterns from the interplay between Hedge, another special case of MD, and best-response dynamics in zero-sum games. All of these studies focus on stationary learning environments and establish only time-averaged convergence. By contrast, this paper investigates heterogeneous learning

in time-varying games and establishes last-iterate convergence to Nash equilibrium strategies.

**Time-Varying Games.** Time-varying games are widely used to model games whose underlying payoff structures evolve with time. Previous works mainly consider the dynamics where the players use the same algorithms in the time-varying games. For the last-iterate convergence behaviors, Duong et al. [11] showed the convergence of MD dynamics on convergent strongly monotone games. Feng et al. [13, 14] studied the last-iterate convergence behaviors of several commonly used optimistic algorithms in the settings of perturbed convergent and periodic zero-sum games. Chen and Yu [8] generalized these results to multi-player general-sum games. Recent works of Fujimoto et al. [16] showed that the divergence behaviors of several algorithms will diverge from Nash equilibrium in general periodic games. However, the time-varying games examined in these prior studies assume relatively simple and slowly evolving dynamics. Moreover, a growing body of evidence has revealed negative results regarding the lack of last-iterate convergence in such dynamics. This work extends prior studies by generalizing the game settings and introducing heterogeneous learning dynamics beyond the homogeneous frameworks considered previously.

## 2 PRELIMINARIES

**Notations.** For any vector  $x \in \mathbb{R}^n$ , we denote the Euclidean norm by  $\|x\| := (\sum_i x_i^2)^{1/2}$  and the  $\ell_1$  norm by  $\|x\|_1 := \sum_i |x_i|$ . For any matrix  $A \in \mathbb{R}^{n \times m}$ , the  $\ell_1$  norm is  $\|A\|_1 := \max_j \sum_i |a_{ij}|$ , and the max norm is  $\|A\|_{\max} := \max_{i,j} |a_{ij}|$ .  $\mathbf{0}$  is the zero matrix, and the closure of a set  $X$  is denoted by  $\bar{X}$ .  $O(f(t))$  and  $o(f(t))$  denotes the order of magnitude function. Specifically,  $g(t) = O(f(t))$  implies that there exists a constant  $C > 0$  such that  $g(t) \leq Cf(t)$  for all  $t$  while  $g(t) = o(f(t))$  means  $g(t)/f(t) \rightarrow 0$ , as  $t \rightarrow \infty$ .

### 2.1 Time-Varying Zero-Sum Games

We consider a dynamic two-player zero-sum game, where Player X (row player) and Player Y (column player) repeatedly play matrix games. Let  $\mathcal{I} = \{1, 2, \dots, n\}$  and  $\mathcal{J} = \{1, 2, \dots, m\}$  denote their action sets, respectively. At each round  $t$ , the game is defined by a payoff matrix  $C_t \in \mathbb{R}^{n \times m}$ , where  $C_t$  represents the loss matrix of Player X (equivalently, the reward matrix for Player Y). The mixed strategies at round  $t$  are  $x_t \in \Delta(\mathcal{I})$  and  $y_t \in \Delta(\mathcal{J})$ , where  $\Delta(\mathcal{I})$  and  $\Delta(\mathcal{J})$  are the probability simplices over  $\mathcal{I}$  and  $\mathcal{J}$ , respectively.

If the sequence  $\{C_t\}$  is time-dependent, the game is called a *time-varying game*. Prior work has primarily studied two types of time-varying games: (i) *periodic games with consistent equilibria*, where there exists  $T \in \mathbb{N}^+$  such that  $C_{t+T} = C_t$  and all matrices within a period share the same set of Nash equilibria; (ii) *convergent perturbed games*, where there exists a limiting matrix  $C_\infty$  such that  $\lim_{t \rightarrow \infty} \|C_t - C_\infty\|_1 = 0$ . In the next section, we will introduce a broader class of time-varying games, which subsumes both the periodic and the convergent perturbed cases as special instances.

Given a matrix  $C$  and a mixed strategy  $x \in \Delta(\mathcal{I})$ , define the *Best Response* set-valued mapping  $\text{BR} : \Delta(\mathcal{I}) \times \mathbb{R}^{n \times m} \rightrightarrows \Delta(\mathcal{J})$  as

$$\text{BR}(x; C) := \left\{ y \in \Delta(\mathcal{J}) \mid x^\top C y = \max_{y' \in \Delta(\mathcal{J})} x^\top C y' \right\}.$$

We also define the value function  $V_C : \Delta(I) \rightarrow \mathbb{R}$  as

$$V_C(x) := x^\top C y, \quad y \in \text{BR}(x; C).$$

A strategy pair  $(x^*, y^*)$  is a Nash equilibrium of the game with matrix  $C$  if it satisfies

$$\begin{aligned} (x^*)^\top C y^* &\leq x^\top C y^*, \quad \forall x \in \Delta(I), \\ (x^*)^\top C y^* &\geq (x^*)^\top C y, \quad \forall y \in \Delta(J). \end{aligned} \quad (1)$$

And the *Nash policy set* of Player X is  $\mathcal{X}_C^* := \{x^* \in \Delta(I) \mid \exists y^* \in \Delta(J), \text{ s.t. } (x^*, y^*) \text{ forms an NE.}\}$  Based on the structure of the NE set in zero-sum games [10],  $\mathcal{X}_C^*$  is convex and closed.

The value of the game, denoted by  $v_C^*$ , is defined as  $v_C^* := (x^*)^\top C y^*$ . By the minimax theorem [36] and the definition of  $V_C$ , we have

$$v_C^* = \min_{x \in \Delta(I)} \max_{y \in \Delta(J)} x^\top C y = \min_{x \in \Delta(I)} V_C(x). \quad (2)$$

The quantity  $v_C^*$  is the minimum loss that Player X can guarantee in the worst case. Thus, for a given strategy  $x$ , the gap  $V_C(x) - v_C^*$  measures how far  $x$  is from the NE strategies. Due to its resemblance to the duality gap, we refer to it as the *semi-duality gap*.

## 2.2 Mirror Descent

Mirror Descent (MD), originally introduced by [30], is a powerful first-order optimization method that generalizes classical gradient descent to non-Euclidean geometries. It has become a fundamental method in convex optimization, online learning, and game-theoretic dynamics [31] due to its flexibility in handling constraint sets.

Let  $X \subseteq \mathbb{R}^n$  be a convex set. A central component of MD is the *distance-generating function*  $\psi : X \rightarrow \mathbb{R}$ , which is assumed to be strictly convex and differentiable on  $X$ . The *Bregman divergence* induced by  $\psi$  is defined as:

$$D_\psi(x, x') = \psi(x) - \psi(x') - \langle \nabla \psi(x'), x - x' \rangle,$$

which measures the discrepancy between  $x$  and  $x'$  in the geometry induced by  $\psi$ . Unlike Euclidean distance, Bregman divergence is typically asymmetric and better suited to structured domains  $X$  like simplex. Typical choices for  $\psi$  are the Euclidean norm  $\frac{1}{2}\|x\|^2$  and negative entropy function  $\sum_{i=1}^n x_i \log x_i$ , which will respectively lead to the projected gradient descent (PGD) algorithm and the Hedge algorithm.

At time  $t$ , given the current point  $x_t$  and a dual gradient-like vector  $g(x_t) \in \mathbb{R}^n$ , the next point generated by MD is

$$x_{t+1} = \arg \min_{x \in X} \{ \langle g(x_t), x - x_t \rangle + \lambda_t D_\psi(x, x_t) \}, \quad \forall t \geq 0, \quad (\text{MD})$$

where  $\lambda_t^{-1}$  is called the *step size*.

**EXAMPLE 1 (EUCLIDEAN SQUARE NORM).** When  $X$  is a general convex set, the Euclidean squared norm  $\psi(x) = \frac{1}{2}\|x\|^2$  is typically used. The induced MD is PGD and the update rule turns to

$$x_{t+1} = \text{Proj}_X(x_t - \lambda_t^{-1} g(x_t)),$$

where  $\text{Proj}$  is the projection operator.

**EXAMPLE 2 (NEGATIVE ENTROPY FUNCTION).** When  $X$  is the  $n$ -dimensional simplex  $\Delta_n$ , negative entropy function  $\psi(x) = \sum_{i=1}^n x_i \log x_i$  generates the KL-divergence  $D_\psi(x, y) = \sum_{i=1}^n x_i \log(x_i/y_i)$  and the

corresponding MD is the Hedge algorithm (also known as the multiplicative weights update). The update rule turns to

$$x_{i,t+1} = \frac{x_{i,t} e^{-\lambda_t^{-1} [g(x_t)]_i}}{\sum_{j=1}^n x_{j,t} e^{-\lambda_t^{-1} [g(x_t)]_j}},$$

where  $[g(x_t)]_i$  represents the  $i$ -th element of  $g(x_t)$ .

The following lemma provides a bound on the Bregman divergence before and after an update of MD. This bound is based on the “three-point identity” presented in [6, 28] and serves as a triangle inequality analog for Bregman divergence:

**LEMMA 1.** Let  $\psi$  be a  $\rho$ -strongly-convex and differentiable function on  $X$ . For any  $p, x \in X$  and  $g(x) \in \mathbb{R}^n$ , let  $x' = \arg \min_{z \in X} \{ \langle g(x), z - x \rangle + \lambda D_\psi(z, x) \}$ . Then, we have

$$D_\psi(p, x') \leq D_\psi(p, x) + \frac{1}{\lambda} \langle g(x), p - x \rangle + \frac{\|g(x)\|^2}{2\rho\lambda^2}.$$

## 3 SETUP AND ALGORITHM

In this work, we propose the following heterogeneous learning method (Algorithm 1), where one player employs Mirror Descent and the other best responds to the MD player.

---

### Algorithm 1 Mirror Descent learning against Best Response

---

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

- 1: **for**  $t = 0, 1, 2, \dots$  **do**
- 2:   Take the Best Response  $y_t$  to  $x_t$  with respect to  $C_t$ ,
- 3:   Update  $x_{t+1}$  based on Mirror Descent:

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


---

To illustrate that Algorithm 1 attains last-iterate convergence for broader range of time-varying zero-sum games. We introduce a novel class of time-varying zero-sum games called **decomposable time-varying zero-sum games** (later extended to non-zero-sum settings), where the payoff matrix  $C_t$  admits the decomposition:

$$C_t = A_t + \mathcal{E}_t, \quad (3)$$

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:

$$\mathcal{X}^* := \{x^* \mid \exists \{x_t^*\}_{t \geq 0} \text{ with } x_t^* \in \mathcal{X}_{A_t}^*, \text{ s.t. } x_t^* \rightarrow x^*\} \neq \emptyset.$$

And there exists  $x^* \in \mathcal{X}^*$  such that:

$$\inf_{x_t^* \in \mathcal{X}_{A_t}^*} \text{dist}(x_t^*, x^*) = O(1/t^q), \quad q > 0.$$

- (2) The *vanishing term*  $\mathcal{E}_t$  converges to  $\mathbf{0}$  with rate  $\varepsilon$  and the *non-vanishing term*  $A_t$  is uniformly bounded by  $M > 0$ :

$$\|\mathcal{E}_t\|_1 = O(1/t^\varepsilon), \varepsilon > 0, \quad \|A_t\|_1 \leq M.$$

Unlike previous studies of time-varying zero-sum games, our framework only requires the existence of limiting points of the Nash policy sets  $\mathcal{X}_{A_t}^*$  corresponding to the non-vanishing term. This property, which will be analyzed in detail later, guarantees

convergence of the heterogeneous learning process. Intuitively, these limiting points serve as directional guides for the strategy updates of the MD player, analogous to how attractors govern the evolution of dynamic systems.

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

**EXAMPLE 3 (PERIODIC GAMES).** *The periodic games studied in [13, 15, 16] 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 4 (CONVERGENT PERTURBED GAMES).** *The convergent perturbed games considered in [8, 12, 13] correspond to the case where  $C_t = A + \mathcal{E}_t$  with  $\|\mathcal{E}_t\|_1 = \mathcal{O}(1/t^\epsilon)$ .*

In the next section, we will establish convergence results under the following assumption, which means the MD player in Algorithm 1 can choose step sizes  $1/\lambda_t$  to control the impact of varying games during the learning process.

**ASSUMPTION 1.** *For decomposable time-varying zero-sum games with decomposition (3), the Player X in Algorithm 1 chooses the step size  $1/\lambda_t$  that fulfill standard stochastic approximation requirements [7] and can bound the accumulated variants:*

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

**REMARK 1.** *Selecting  $\lambda_t = t^p$  with  $\max\{1 - q, 1 - \epsilon, 1/2\} < p \leq 1$  suffices to satisfy the Assumption 1.*

*For the convergent time-varying games in Example 4, typical convergence results require the bounded accumulated perturbations (BAP) condition [13]:  $\sum_{t=0}^{\infty} \|\mathcal{E}_t\|_2 < \infty$ , which implies  $\|\mathcal{E}_t\|_1 = o(1/t)$ . However, Assumption 1 only requires  $\|\mathcal{E}_t\|_1 = \mathcal{O}(1/t^\epsilon)$  for some  $\epsilon > 0$ . This demonstrates that our assumption applies to a broader class of convergent time-varying games than previously considered.*

## 4 CONVERGENCE ANALYSIS

### 4.1 Convergence of Time-Varying Zero-Sum Games

We now present our main convergence results of the learning dynamics of Algorithm 1 in time-varying zero-sum games, which can be formulated as follows: Starting from an arbitrary initial strategy  $x_0 \in \Delta(\mathcal{I})$ , the policy profile updates for  $t \geq 0$  are given by:

$$\begin{aligned} y_t &\in \text{BR}(x_t; C_t), \\ x_{t+1} &= \arg \min_{x \in \Delta(\mathcal{I})} \{ \langle C_t y_t, x - x_t \rangle + \lambda_t D_\psi(x, x_t) \}. \end{aligned} \quad (\text{MD-BR})$$

**THEOREM 2.** *Let  $\{x_t\}$  be the strategy sequence generated by the dynamics (MD-BR). Then under Assumption 1, the following convergence result holds:*

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

where  $V_t(x) := V_{C_t}(x)$  is the value function of game  $C_t$  and  $v_t^*$  denotes the value of  $C_t$ .

**PROOF OF THEOREM 2.** To prove Theorem 2 (and same for following Theorems 6 and 7), the following lemma is vital to obtain the convergence of a bounded sequence:

**LEMMA 3.** *Let  $\{X_t\}$ ,  $\{Y_t\}$ , and  $\{Z_t\}$  be three nonnegative real-valued sequences satisfying the following conditions:*

- (1) *The inequality  $X_{t+1} \leq X_t + Y_t - Z_t$  holds for all  $t \geq 0$ ;*
- (2) *The sum of  $\{Y_t\}$  converges:  $\sum_{t=0}^{\infty} Y_t < \infty$ ,*

*then  $\sum_{t=0}^{\infty} Z_t < \infty$  and  $X_t$  converges.*

By Lemma 1, taking  $x' = x_{t+1}$ ,  $x = x_t$ ,  $p = x^* \in \mathcal{X}^*$  leads to

$$D_\psi(x^*, x_{t+1}) \leq D_\psi(x^*, x_t) + \frac{1}{\lambda_t} \langle C_t y_t, x^* - x_t \rangle + \frac{\|C_t y_t\|^2}{2\rho\lambda_t^2}. \quad (4)$$

Inspired by Lemma 3, we focus on the estimate of  $\langle C_t y_t, x^* - x_t \rangle$ . Let  $\bar{x}_t = \arg \min_{y \in \mathcal{X}_{A_t}^*} \text{dist}(y, x^*)$  and  $x_t^* \in \mathcal{X}_{C_t}^*$ , we have:

$$\begin{aligned} &\langle C_t y_t, x^* - x_t \rangle \\ &= \langle (A_t + \mathcal{E}_t) y_t, x^* \rangle - \langle C_t \text{BR}(x_t^*; C_t), x_t^* \rangle + v_t^* - \langle C_t y_t, x_t \rangle \\ &= \langle A_t y_t, x^* - \bar{x}_t \rangle + \langle A_t y_t, \bar{x}_t \rangle + \langle \mathcal{E}_t y_t, x^* \rangle \\ &\quad - \langle C_t \text{BR}(x_t^*; C_t), x_t^* \rangle + v_t^* - V_t(x_t) \\ &\leq M \text{dist}(x^*, x_t^*) + \|\mathcal{E}_t\|_1 + \langle A_t \text{BR}(\bar{x}_t; A_t), \bar{x}_t \rangle \\ &\quad - \langle (A_t + \mathcal{E}_t) \text{BR}(x_t^*; A_t), x_t^* \rangle + (v_t^* - V_t(x_t)) \\ &\leq M \text{dist}(x^*, x_t^*) + 2\|\mathcal{E}_t\|_1 + (v_t^* - V_t(x_t)). \end{aligned}$$

By taking this estimate into (4), we have:

$$\begin{aligned} D_\psi(x^*, x_{t+1}) &\leq D_\psi(x^*, x_t) + \frac{1}{\lambda_t} (M \text{dist}(x^*, x_t^*) + 2\|\mathcal{E}_t\|_1) \\ &\quad + \frac{v_t^* - V_t(x_t)}{\lambda_t} + \frac{\|C_t y_t\|^2}{2\rho\lambda_t^2}. \end{aligned}$$

Based on Assumption 1, directly applying Lemma 3 leads to:

$$\sum_{t=0}^{\infty} \frac{(V_t(x_t) - v_t^*)}{\lambda_t} < \infty,$$

which implies that  $V_t(x_t) - v_t^* = o(\lambda_t/t)$ .  $\square$

In contrast to [13], which focuses on the policy gradient norm  $\Delta_t = \|A^\top x_t\|_2 + \|A y_t\|_2$ , and [12], which considers the Euclidean distance  $\|x_t - \mathcal{X}_t^*\|_2$  between  $x_t$  and the Nash policy set, we employ the difference  $V_t(x_t) - v_t^*$  between the current strategy's value and the game value as the convergence metric. This choice is motivated by two key observations in our setting:

- (1) The equilibrium sets  $\mathcal{X}_t^*$  are time-varying, and their limiting set  $\mathcal{X}^*$  lacks desired properties to make Euclidean distance  $\text{dist}(x_t, \mathcal{X}^*)$  well-defined.
- (2) In matrix games with simplex-constrained strategies, the gradient norm  $\Delta_t = \|A^\top x_t\|_2$  of Player X may remain positive even at Nash policies. Thus, the zero-gradient characterization of Nash equilibria in [13] for unconstrained games does not apply to this setting.

The value difference  $V_t(x_t) - v_t^*$  provides a natural performance measure, evaluating how well  $x_t$  performs against the most adversarial opponent in each round. Theorem 2 demonstrates that the mirror descent player can learn robust strategies even under such adversarial conditions.

The robustness stems from the gradient information provided by alternating Best Response adversaries, as revealed in Theorem 2's estimation of  $\langle C_t y_t, x^* - x_t \rangle$ . The MD method, with increasing  $\lambda_t$  under Assumption 1, effectively mitigates estimation errors induced

by time-varying terms, enabling  $\{x_t\}$  to track the evolving Nash policy sets  $\mathcal{X}_{C_t}^*$ . This explains the difference convergence results between learning paradigms: while homogeneous simultaneous updates (with identical MD or Best Response dynamics) fail to converge in zero-sum games, our heterogeneous alternating approach guarantees provable convergence.

This mechanism differs fundamentally from extra-gradient methods (EG) and optimistic gradient descent ascent (OGDA), which rely on predicting opponent policies - a cooperative approach unnatural for zero-sum settings. Numerical results in the following section further demonstrate the superior convergence performance of our heterogeneous alternating dynamics (MD-BR) compared to these methods.

These theoretical insights have practical implications for individual strategy learning. After obtaining approximate Nash policies through self-play [35], introducing adversarial opponents can rapidly enhance strategy robustness through the proposed learning framework.

When examining time-varying games from previous literature, Theorem 2 remains applicable given appropriate decomposition of  $C_t$ :

**COROLLARY 4.** *For periodic time-varying games that maintain consistent equilibrium structures, as decomposed in Example 3, the sequence  $\{x_t\}$  generated by (MD-BR) with  $\lambda_t = t^p$ ,  $p \in (1/2, 1]$ , achieves  $V_t(x_t) - v_t^* = o(1/t^{1-p})$ .*

**COROLLARY 5.** *For convergent perturbed games with decomposition  $C_t = A + \mathcal{E}_t$  in Example 4, choosing  $\lambda_t = t^p$  with  $\max\{1/2, 1 - \varepsilon\} < p \leq 1$  yields  $V_t(x_t) - v_t^* = o(1/t^{1-p})$  for the generated sequence  $\{x_t\}$ . Furthermore, it can be shown that  $V_A(x_t) - v_A^* = o(1/t^{1-p})$ .*

Corollary 5 establishes that  $V_A(x_t) - v_A^* \rightarrow 0$ , demonstrating convergence of  $\{x_t\}$  to the convex and compact set  $\mathcal{X}_A^*$ . The following theorem further strengthens this result by proving pointwise convergence to specific elements of  $\mathcal{X}_A^*$ .

**THEOREM 6 (CONVERGENCE OF STRATEGY SEQUENCE).** *For the convergent perturbed games in Example 4, let  $\{x_t\}$  be the strategy sequence generated by the dynamics (MD-BR). If the Bregman divergence  $D_\psi$  in (MD-BR) satisfies the following Bregman reciprocity property:*

$$D_\psi(x, x_t) \rightarrow 0 \quad \text{when} \quad x_t \rightarrow x,$$

*then under Assumption 1, there exists a Nash policy  $x^* \in \mathcal{X}_A^*$  such that:*

$$\lim_{t \rightarrow \infty} x_t = x^*.$$

**REMARK 2.** *The pointwise convergence results can be extended to general decomposable time-varying games where there exists a matrix game  $A$  satisfying  $\mathcal{X}_A^* = \mathcal{X}^*$ .*

## 4.2 Asymptotic Zero-Sum Games: An Extension to General Matrix Games

Recent work [13, 14] investigated the last-iterate convergence separation between EG and OGDA in time-varying zero-sum games. Their results demonstrate that OGDA fails to achieve last-iterate convergence in such games. This limitation stems from OGDA's reliance on payoff matrix information from two preceding iterations:

when the matrices change rapidly, the algorithm cannot obtain reliable needed information.

In contrast, our heterogeneous alternating learning dynamics (MD-BR) permit sequential strategy updates where players may use different matrix information ( $C_t$  for MD player and  $B_t$  for Best Response player) within the same round  $t$ . We show that vanishing step sizes  $1/\lambda_t$  can effectively mitigate the impact of such matrix discrepancies on dynamic convergence.

Consider the following learning process: Starting from an arbitrary initial strategy  $x_0 \in \Delta(I)$ , the strategy updates for  $t \geq 0$  are given by:

$$\begin{aligned} y_t &\in \text{BR}(x_t; B_t), \\ x_{t+1} &= \arg \min_{x \in \Delta(I)} \{ \langle C_t y_t, x - x_t \rangle + \lambda_t D_\psi(x, x_t) \}, \end{aligned} \quad (5)$$

The following theorem establishes that when  $\{C_t - B_t\}$  converges to zero matrix at an appropriate rate (hence referred to as **asymptotic zero-sum games**), Player X can still extract gradient information about  $V_t(x_t) = V_{C_t}(x_t)$  through the vanishing step size and the sequence  $\{x_t\}$  generated by (5) converges to Nash policy set of  $C_t$ .

**THEOREM 7.** *Let  $\{x_t\}$  be the strategy sequence generated by the dynamics (5). Under Assumption 1 for  $C_t$  and  $\{\lambda_t\}$  as in Theorem 2, if  $\{B_t\}$  satisfies:*

$$\sum_{t \geq 1} \frac{1}{\lambda_t} \|C_t - B_t\|_1 < \infty,$$

*then the value of strategy sequence achieves:*

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

**REMARK 3.** *Duvocelle et al. [12] and Chen and Yu [8] both analyze homogeneous learning convergence in time-varying monotone games but are limited to the convergent perturbed games. Crucially, Duvocelle et al. [12] requires convergence to strongly monotone games while Chen and Yu [8] imposes the BAP condition on vanishing terms  $\mathcal{E}_t$ . In contrast, our results require only asymptotic convergence to zero-sum games, a non-strongly-monotone subclass of monotone games, without BPA conditions, thereby substantially expanding the class of time-varying games admitting last-iterate convergence.*

## 5 SIMULATIONS

In this section, we design four simulations to verify our theoretical results and evaluate empirical performance. Our simulations are lightweight and do not require any specialized hardware. All experiments were run on a personal laptop without GPU acceleration.

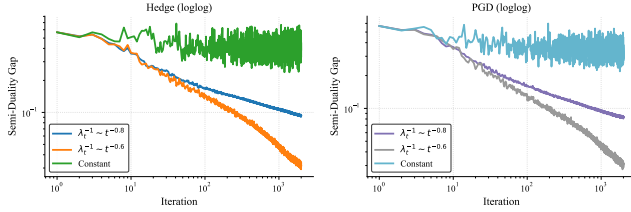
### 5.1 Convergence of Heterogeneous Learning in Convergent Perturbed Games

In this subsection, we empirically verify the convergence of heterogeneous learning dynamics to varying equilibria in convergent perturbed games. Moreover, we show how the decay rate of the step size sequence influences the convergence behavior.

We generate a random  $100 \times 200$  payoff matrix  $C$  with entries from  $(-5, 5)$  as the limiting game. The game runs for 2000 rounds, with the stage payoff at time  $t$  given by  $C + \mathcal{E}_t$ , where  $\mathcal{E}_t$  is a

perturbation matrix with entries in  $(-1/t^\varepsilon, 1/t^\varepsilon)$ . We set the decay rate  $\varepsilon = 0.7$  to ensure smooth convergence to the limiting game.

We focus on two classic MD algorithms: Hedge and PGD. For each algorithm, we experiment with three different decay rates for the step size: (1) Large:  $\lambda_t^{-1} \sim t^{-0.8}$  (2) Small:  $\lambda_t^{-1} \sim t^{-0.6}$  (3) Constant: 0.1 for Hedge and 0.001 for PGD. The semi-duality gap  $V_t(x_t) - v_t^*$  is used to evaluate the quality of the strategy  $x_t$ .



**Figure 1: Convergence under different step-size decay rates.**

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, properly decaying rates lead to convergence, and larger rate results in slower convergence rates.

## 5.2 Convergence Efficiency of PGD Compared with EG and OGDA

In this subsection, the convergence efficiency of the heterogeneous learning method induced by projected gradient descent (PGD) is compared with that of extra-gradient (EG) and optimistic gradient descent ascent (OGDA) in approaching the NE. The three aforementioned algorithms are tested with a common step size of  $\lambda_t^{-1} \sim t^{-0.6}$ . Additionally, EG and OGDA are tested with a constant step size 0.04. The time-varying payoff matrix is defined as:

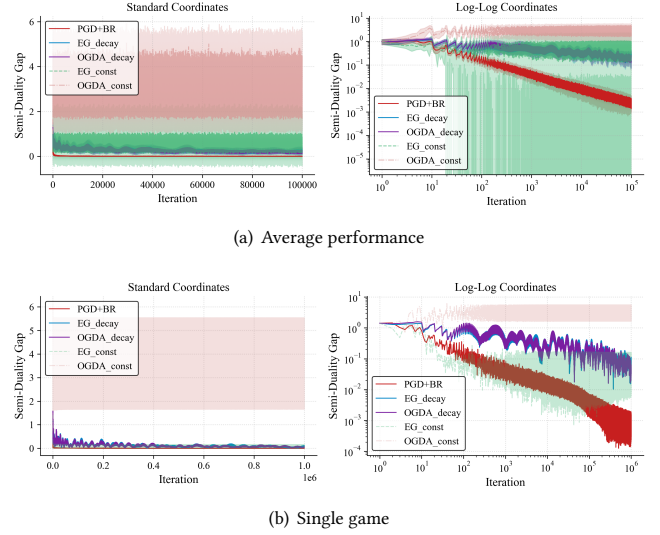
$$C_t = (p_t/10 + 1)A + \mathcal{E}_t,$$

where  $p_t = t \bmod 10$ ,  $A$  is a  $20 \times 30$  matrix with uniformly random entries sampled from  $(-5, 5)$ , and  $\mathcal{E}_t$  is a decaying perturbation satisfying  $\lim_{t \rightarrow \infty} \mathcal{E}_t = \mathbf{O}$  with decay rate 0.5. The time horizon is set to 100000 for evaluating average performance over 10 randomly generated matrices  $A$  (Figure 2(a)) and to 1000000 for a specific instance of  $A$  (Figure 2(b)).

Figure 2 demonstrates that under the decay step size, the PGD-induced heterogeneous learning method achieves faster and more stable convergence than both EG and OGDA. Although the constant step-size EG exhibits competitive performance during certain iterations, it displays significant instability.

## 5.3 Behavior in Periodic Games with Shared vs. Distinct Equilibria

In this subsection, we investigate the behavior of heterogeneous learning in periodic games. These experiments reveal how the equilibrium location and the matrix magnitude affect the learning trajectory, while the theoretical analysis is left for future work.



**Figure 2: Convergence performance of heterogeneous learning, compared with EG and OGDA using both constant (\_const) and decaying step sizes (\_decay). Figure 2(a) shows the average performance over ten distinct time-varying games across 100000 rounds, whereas Figure 2(b) shows the performance of a single time-varying game over 1000000 rounds.**

*Case 1: Periodic Games with a Shared Equilibrium.* We first consider a periodic game of period 3 with following stage matrices:

$$A1 : \begin{pmatrix} -1 & 1 \\ 1 & -1 \end{pmatrix} \quad A2 : \begin{pmatrix} 2 & -1 \\ 0 & 3 \end{pmatrix} \quad A3 : \begin{pmatrix} 0 & 3 \\ 2 & -1 \end{pmatrix}$$

In all three games, Player X's Nash policy is  $(1/2, 1/2)$ , while Player Y's equilibrium strategy differs across matrices. We apply both Hedge and PGD with a step size  $\lambda_t^{-1} \sim t^{-0.5}$  and use  $(0.1, 0.9)$  as the initial strategy for Player X. The game repeats 200 times.

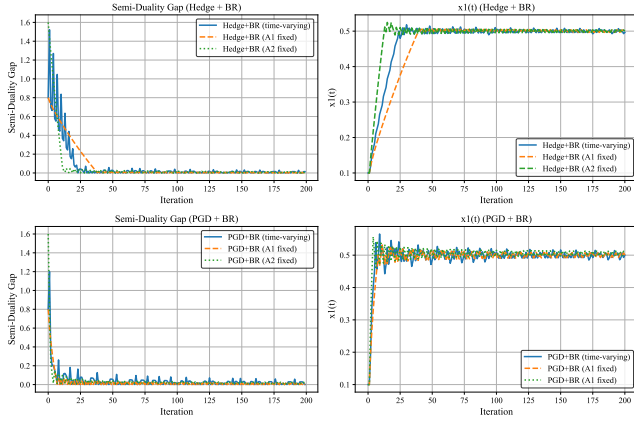
To examine the effect of varying payoff matrices, we compare three cases: the above periodic game, and two time-invariant games using  $A1$  and  $A2$ , respectively.

From Figure 3, we observe that in all three cases, the semi-duality gap converges to 0, confirming convergence to the Nash policy. However, in the periodic game, the gap exhibits oscillations due to variation of the games. Meanwhile, the right subplots show that the first component of  $x_t$  converges smoothly without oscillations, demonstrating that the shared NE serves as a strong attractor.

Interestingly, the convergence rate of the periodic game lies between those of the  $A1$ - and  $A2$ -based static games. This suggests that while the equilibrium location determines the direction of dynamics, the magnitude affects the speed.

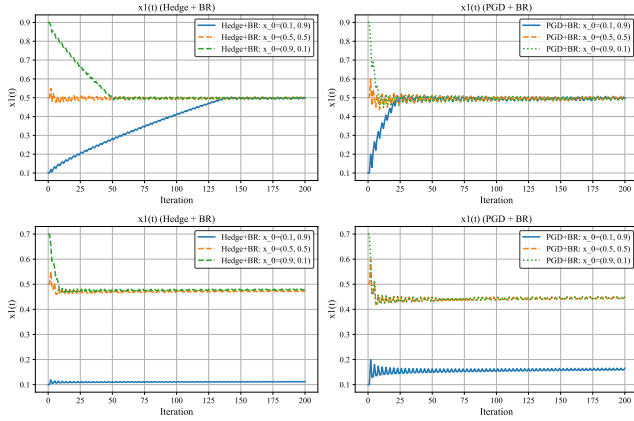
*Case 2: Periodic Games with Distinct Equilibria.* We now consider another periodic game of period 3, with each stage game having a different equilibrium:

$$A1 : \begin{pmatrix} -1 & 1 \\ 1 & -1 \end{pmatrix} \quad A2 : \begin{pmatrix} 1 & 2 \\ -1 & -2 \end{pmatrix} \quad A3 : \begin{pmatrix} 0 & -1 \\ 1 & 2 \end{pmatrix}$$



**Figure 3: The performance of PGD and Hedge under heterogeneous learning in periodic and time-invariant games. The left subplots display the semi-duality gap of the strategy sequence  $\{x_t\}$ ; the right subplots track the evolution of the first component of  $x_t$ .**

The corresponding Nash policies for Player X are  $(1/2, 1/2)$ ,  $(0, 1)$ , and  $(1, 0)$ , respectively. Other settings remain the same.



**Figure 4: The limiting behavior of PGD and Hedge under heterogeneous learning in periodic games with distinct equilibria. The upper subplots show the first coordinate of  $x_t$  from different initializations. The lower subplots show results when A3 is modified.**

The upper subplots of Figure 4 show the behavior of the first component of  $x_t$  under different initializations. For both Hedge and PGD, the trajectories fluctuate around the average, rather than converging to a single point. This indicates that multiple NE targets may compete and pull the trajectory in different directions.

To probe further, we modify matrix A3 to reduce the strength of its NE while maintaining  $(1, 0)$  as the Nash policy for Player X:

$$A3 : \begin{pmatrix} 0 & -0.01 \\ 0.01 & 0.02 \end{pmatrix}.$$

As shown in the lower subplots of Figure 4, the trajectories now tend to deviate more significantly from  $(1, 0)$  and lean toward  $(0, 1)$ . This confirms that both the location of NE and the matrix magnitude influence the learning dynamics. The NE location dictates the direction of convergence, while the matrix magnitude controls the size of each update step. When the pull of a certain NE weakens due to smaller payoff magnitudes, the trajectory may be attracted by stronger NE in other stage games.

## 6 CONCLUSION

This paper defines a new class of time-varying zero-sum games and establishes convergence guarantees for heterogeneous learning dynamics in these games, where one player employs MD while the opponent adopts Best Response strategies. We prove that when the game variations satisfy certain convergence rate conditions, appropriate vanishing step sizes for the MD player ensure convergence to Nash policy sets. Our theoretical results can be naturally extended to asymptotic zero-sum games. Simulations demonstrate that this adversarial learning approach achieves faster convergence speed and more stable performance compared to EG and OGD, suggesting new directions for developing effective learning algorithms.

These findings open promising research avenues for extending the heterogeneous learning framework to general convex-concave games and potentially non-convex-non-concave settings, thereby enabling novel algorithmic developments for large-scale min-max optimization problems including GANs training and preference-based alignment of LLMs. Moreover, several open questions raised in the simulation section are worthy of further investigation.

## APPENDIX: MISSING PROOFS

**PROOF OF LEMMA 1.** Based on the definition of the Bregman divergence:

$$\begin{aligned} D_\psi(p, x) &= \psi(p) - \psi(x) - \langle \nabla \psi(x), p - x \rangle, \\ D_\psi(p, x') &= \psi(p) - \psi(x') - \langle \nabla \psi(x'), p - x' \rangle, \\ D_\psi(x, x') &= \psi(x) - \psi(x') - \langle \nabla \psi(x'), x - x' \rangle, \end{aligned}$$

we directly have

$$D_\psi(p, x) = D_\psi(p, x') + D_\psi(x', x) + \langle \nabla \psi(x) - \nabla \psi(x'), x' - p \rangle. \quad (6)$$

By the optimality of  $x'$ , we obtain that for any  $p \in \mathcal{X}$ ,

$$\langle g(x) + \lambda(\nabla \psi(x') - \nabla \psi(x)), p - x' \rangle \geq 0 \quad (7)$$

Combining Equations (6) and (7) leads to the famous **three-point identity**:

$$\begin{aligned} D_\psi(p, x') &= D_\psi(p, x) - D_\psi(x', x) + \langle \nabla \psi(x) - \nabla \psi(x'), p - x' \rangle \\ &\leq D_\psi(p, x) - D_\psi(x', x) + \frac{1}{\lambda} \langle g(x), p - x' \rangle \\ &= D_\psi(p, x) - D_\psi(x', x) + \frac{1}{\lambda} \langle g(x), p - x \rangle + \frac{1}{\lambda} \langle g(x), x - x' \rangle. \end{aligned}$$

Besides, by Young's inequality, we have

$$\frac{1}{\lambda} \langle g(x), x - x' \rangle \leq \frac{\|g(x)\|^2}{2\rho\lambda^2} + \frac{\rho\|x - x'\|^2}{2}$$

Together with the strong convexity of  $\psi$ , we have

$$\begin{aligned} D_\psi(p, x') &\leq D_\psi(p, x) + \frac{1}{\lambda} \langle g(x), p - x \rangle + \frac{\|g(x)\|^2}{2\rho\lambda^2} + \\ &\quad \frac{\rho\|x - x'\|^2}{2} - D_\psi(x', x) \\ &\leq D_\psi(p, x) + \frac{1}{\lambda} \langle g(x), p - x \rangle + \frac{\|g(x)\|^2}{2\rho\lambda^2}. \end{aligned}$$

This completes the proof.  $\square$

PROOF OF LEMMA 3. From Condition 1 in Lemma 3, for all  $t \geq 0$ , we have:

$$0 \leq X_{t+1} + \sum_{k=0}^t Z_k + \sum_{k=t+1}^{\infty} Y_k \leq X_t + \sum_{k=0}^{t-1} Z_k + \sum_{k=t}^{\infty} Y_k,$$

where we let  $\sum_{k=0}^{-1} Z_k = 0$  for convenience.

This implies the existence of a constant  $W_\infty > 0$  such that:

$$X_t + \sum_{k=0}^{t-1} Z_k + \sum_{k=t}^{\infty} Y_k \rightarrow W_\infty, \quad \text{as } t \rightarrow \infty.$$

By Condition 2, the sequence  $X_t + \sum_{k=0}^{t-1} Z_k$  converges. Since  $X_t \geq 0$ , the sums  $\sum_{k=0}^{t-1} Z_k$  form a non-decreasing sequence with an upper bound  $W_\infty$ , which implies  $\sum_{t=0}^{\infty} Z_t < \infty$  by the monotone convergence theorem.

Finally, the convergence of  $X_t$  follows immediately from these results.  $\square$

PROOF OF THEOREM 6. In the proof of Theorem 2, by lemma 3, for any  $x^* \in \mathcal{X}_A^*$ , the sequence  $D_\psi(x^*, x_t)$  converges to some limit  $D_{x^*}$ .

From inequality (4), we derive:

$$\begin{aligned} &\langle C_t y_t, x^* - x_t \rangle \\ &\leq \langle C_t y_t, x^* \rangle - \langle C_t \text{BR}(x_t; A), x_t \rangle \\ &\leq \langle A_t \text{BR}(x^*; A_t), x^* \rangle + \langle \mathcal{E}_t y_t, x^* \rangle \\ &\quad - \langle A_t \text{BR}(x_t; A), x_t \rangle - \langle \mathcal{E}_t \text{BR}(x_t; A), x_t \rangle \\ &\leq v_A^* - V_A(x_t) + 2\|\mathcal{E}_t\|_1, \end{aligned}$$

where  $V_A(\cdot)$  is defined in Corollary 5.

Following the proof of Theorem 2, we obtain  $V_A(x_t) \rightarrow v_A^*$ . The continuity of  $V_A(x)$  implies  $x_t \rightarrow x_A^*$ . Since  $\mathcal{X}_A^*$  is a closed convex set, there exists a Nash strategy  $x^* \in \mathcal{X}_A^*$  and a subsequence  $\{x_{t_k}\}$  of  $\{x_t\}$  such that  $x_{t_k} \rightarrow x^*$ .

The Bregman reciprocity property yields  $D_\psi(x^*, x_{t_k}) \rightarrow 0$ , which implies  $D_{x^*} = 0$ . Consequently,  $D_\psi(x^*, x_t) \rightarrow 0$  for the entire sequence, establishing the convergence  $x_t \rightarrow x^* \in \mathcal{X}_A^*$ .  $\square$

PROOF OF THEOREM 7. Following the proof of Theorem 2, we re-estimate the key term  $\langle C_t \text{BR}(x_t; B_t), x^* - x_t \rangle$ :

$$\begin{aligned} &\langle C_t \text{BR}(x_t; B_t), x^* - x_t \rangle \\ &= \langle C_t \text{BR}(x_t; B_t), x^* \rangle - \langle C_t \text{BR}(x_t; C_t), x_t^* \rangle \\ &\quad + \langle C_t \text{BR}(x_t; C_t), x_t \rangle - \langle C_t \text{BR}(x_t; B_t), x_t \rangle + v_t^* - V_t(x_t) \\ &\leq \langle A_t \text{BR}(x_t; A_t), x^* - \bar{x}_t \rangle + \langle \mathcal{E}_t \text{BR}(x_t; B_t), x^* \rangle \\ &\quad + \langle A_t \text{BR}(x_t; A_t), \bar{x}_t \rangle - \langle C_t \text{BR}(x_t^*; A_t), x_t^* \rangle \\ &\quad + \langle (C_t - B_t) \text{BR}(x_t; C_t), x_t \rangle + \langle B_t \text{BR}(x_t; B_t), x_t \rangle \\ &\quad - \langle C_t \text{BR}(x_t; B_t), x_t \rangle + v_t^* - V_t(x_t) \\ &\leq M \text{dist}(x^*, \bar{x}_t) + 2\|\mathcal{E}_t\|_1 + 2\|C_t - B_t\|_1 + v_t^* - V_t(x_t). \end{aligned}$$

The conclusion follows from the assumptions and Lemma 3.  $\square$

## 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* (2025).
- [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*. 321–338.
- [3] Jakub Bielawski, Thiparat Chotibut, Fryderyk Falniowski, Michał Misiurewicz, and Georgios Piliouras. 2025. Heterogeneity, reinforcement learning, and chaos in population games. *Proceedings of the National Academy of Sciences* 122, 25 (2025), e2319929121.
- [4] 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 preprint arXiv:2204.09228* (2022).
- [5] Adrian Rivera Cardoso, Jacob Abernethy, He Wang, and Huan Xu. 2019. Competing against nash equilibria in adversarially changing zero-sum games. In *International Conference on Machine Learning*. PMLR, 921–930.
- [6] Gong Chen and Marc Teboulle. 1993. Convergence analysis of a proximal-like minimization algorithm using Bregman functions. *SIAM Journal on Optimization* 3, 3 (1993), 538–543.
- [7] Han-Fu Chen. 2002. *Stochastic Approximation and Its Applications*. Springer New York, NY.
- [8] Yanzheng Chen and Jun Yu. 2025. Classic but Everlasting: Traditional Gradient-Based Algorithms Converge Fast Even in Time-Varying Multi-Player Games. In *The Thirteenth International Conference on Learning Representations*.
- [9] Constantinos Daskalakis, Andrew Ilyas, Vasilis Syrgkanis, and Haoyang Zeng. 2017. Training gans with optimism. *arXiv preprint arXiv:1711.00141* (2017).
- [10] Constantinos Daskalakis and Ioannis Panageas. 2018. Last-iterate convergence: Zero-sum games and constrained min-max optimization. *arXiv preprint arXiv:1807.04252* (2018).
- [11] Manh Hong Duong, Hoang Minh Tran, et al. 2019. On the expected number of internal equilibria in random evolutionary games with correlated payoff matrix. *Dynamic Games and Applications* 9, 2 (2019), 458–485.
- [12] 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
- [13] 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 *Advances in Neural Information Processing Systems*, A. Oh, T. Naumann, A. Globerson, K. Saenko, M. Hardt, and S. Levine (Eds.), Vol. 36. Curran Associates, Inc., 21933–21944. [https://proceedings.neurips.cc/paper\\_files/paper/2023/file/457ab261562014550e53351422f69834-Paper-Conference.pdf](https://proceedings.neurips.cc/paper_files/paper/2023/file/457ab261562014550e53351422f69834-Paper-Conference.pdf)
- [14] Yi Feng, Ping Li, Ioannis Panageas, and Xiao Wang. 2024. Last-iterate convergence separation between extra-gradient and optimism in constrained periodic games. In *Proceedings of the Fortieth Conference on Uncertainty in Artificial Intelligence* (Barcelona, Spain) (UAI '24). JMLR.org, Article 63, 32 pages.
- [15] 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.
- [16] 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 AAAI Conference on Artificial Intelligence*, Vol. 39. 23194–23202.
- [17] Ian J Goodfellow, Jonathon Shlens, and Christian Szegedy. 2014. Explaining and harnessing adversarial examples. *arXiv preprint arXiv:1412.6572* (2014).

- [18] Eduard Gorbunov, Adrien Taylor, and Gauthier Gidel. 2022. Last-iterate convergence of optimistic gradient method for monotone variational inequalities. *Advances in neural information processing systems* 35 (2022), 21858–21870.
- [19] 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* (2025), 1–25.
- [20] Martin Heusel, Hubert Ramsauer, Thomas Unterthiner, Bernhard Nessler, and Sepp Hochreiter. 2017. Gans trained by a two time-scale update rule converge to a local nash equilibrium. *Advances in neural information processing systems* 30 (2017).
- [21] Michael Johanson, Nolan Bard, Neil Burch, and Michael Bowling. 2012. Finding optimal abstract strategies in extensive-form games. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 26. 1371–1379.
- [22] Frank Kelly, Aman Kumar Maulloo, and D K H Tan. [n.d.]. Rate Control for Communication Networks: Shadow Prices, Proportional Fairness and Stability. 49, 3 ([n. d.]), 237–252. <https://doi.org/10.1057/palgrave.jors.2600523>
- [23] Tianyi Lin, Chi Jin, and Michael I Jordan. 2025. Two-timescale gradient descent ascent algorithms for nonconvex minimax optimization. *Journal of Machine Learning Research* 26, 11 (2025), 1–45.
- [24] 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 Twenty-Eighth International Joint Conference on Artificial Intelligence, IJCAI-19*. International Joint Conferences on Artificial Intelligence Organization, 464–470.
- [25] Aleksander Madry, Aleksandar Makelov, Ludwig Schmidt, Dimitris Tsipras, and Adrian Vladu. 2017. Towards deep learning models resistant to adversarial attacks. *arXiv preprint arXiv:1706.06083* (2017).
- [26] 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*. 339–350.
- [27] 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>
- [28] 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 preprint arXiv:1807.02629* (2018).
- [29] 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*. SIAM, 2703–2717.
- [30] A. S. Nemirovsky and D. B. Yudin. 1984. *Problem Complexity and Method Efficiency in Optimization*. Wiley, New York.
- [31] Alexander Rakhlin and Karthik Sridharan. 2013. Online learning with predictable sequences. In *Conference on Learning Theory*. PMLR, 993–1019.
- [32] 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.
- [33] Dimitris Tsipras, Shibani Santurkar, Logan Engstrom, Alexander Turner, and Aleksander Madry. 2018. Robustness may be at odds with accuracy. *arXiv preprint arXiv:1805.12152* (2018).
- [34] 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.
- [35] Oriol Vinyals, Igor Babuschkin, Wojciech M. Czarnecki, Michaël Mathieu, Andrew Dudzik, Junyoung Chung, David H. Choi, Richard Powell, Timo Ewalds, Petko Georgiev, Junhyuk Oh, Dan Horgan, Manuel Kroiss, Ivo Danihelka, Aja Huang, Laurent Sifre, Trevor Cai, John P. Agapiou, Max Jaderberg, Alexander S. Vezhnevets, Rémi Leblond, Tobias Pohlen, Valentin Dalibard, David Budden, Yury Sulsky, James Molloy, Tom L. Paine, Caglar Gulcehre, Ziyu Wang, Tobias Pfaff, Yuhuai Wu, Roman Ring, Dani Yogatama, Dario Wünsch, Katrina McKinney, Oliver Smith, Tom Schaul, Timothy Lillicrap, Koray Kavukcuoglu, Demis Hassabis, Chris Apps, and David Silver. 2019. Grandmaster Level in StarCraft II Using Multi-Agent Reinforcement Learning. *Nature* 575, 7782 (Nov. 2019), 350–354. <https://doi.org/10.1038/s41586-019-1724-z>
- [36] John Von Neumann and Oskar Morgenstern. 2007. *Theory of games and economic behavior (60th Anniversary Commemorative Edition)*. Princeton university press.
- [37] Chen-Yu Wei, Chung-Wei Lee, Mengxiao Zhang, and Haipeng Luo. 2020. Linear last-iterate convergence in constrained saddle-point optimization. *arXiv preprint arXiv:2006.09517* (2020).