LEARNING IN TIME-VARYING ZERO-SUM GAMES: INFORMATION-THEORETIC AND COMPUTATIONAL LIMITS

Anonymous authors
Paper under double-blind review

ABSTRACT

In many multi-agent learning settings, the underlying game evolves over time, raising the question of whether learning remains possible in a changing environment. In this setting, Zhang et al. (2022) introduced dynamic Nash equilibrium (NE) regret as a performance measure for learning in time-varying zero-sum games. We study the information-theoretic and computational limits of achieving small dynamic NE regret, with the nonstationarity of the environment measured by the second-order path-length variation of the payoff matrices, denoted by V_T. On the information-theoretic side, we show that sublinear V_T is sufficient to achieve vanishing average dynamic NE regret, and complement this result with a variation-dependent lower bound. On the computational side, we show that, unless $\mathsf{P} = \mathsf{PPAD}$, stable learners with n actions cannot achieve vanishing average dynamic NE regret within $\mathsf{poly}(n)$ rounds. Our results reveal a separation between information-theoretic and computational learnability in time-varying zero-sum games.

1 INTRODUCTION

Learning in games (Fudenberg & Levine, 1998) asks how adaptive dynamics drive repeated play toward equilibrium. No-regret learning provides a well-studied route to this goal, with classical results connecting regret minimization to correlated equilibrium in general-sum games and to Nash equilibrium in two-player zero-sum games (Hart & Mas-Colell, 2000; Cesa-Bianchi et al., 2006; Freund & Schapire, 1999). Much of this classical theory, however, considers a fixed game whose payoff structure remains unchanged over time. Such a stationary model misses an important feature of many applications, where payoffs evolve as market conditions shift, populations change, or other agents adapt, making the underlying game inherently dynamic.

This limitation motivates the study of time-varying games (Lykouris et al., 2016; Duvocelle et al., 2023), in which the payoff structure evolves over the course of repeated play. Learners must therefore adapt not only to one another, but also to shifts in the game itself. In this work, we focus on two-player time-varying zero-sum games (Cardoso et al., 2019; Zhang et al., 2022), a canonical setting in which the underlying games are individually tractable, yet learning must contend with a nonstationary environment. In this setting, the equilibrium benchmark may itself evolve over time, so learning performance must be measured relative to a changing sequence of games rather than a single fixed one. To capture this, Zhang et al. (2022) introduced *dynamic Nash equilibrium (NE) regret* and showed that vanishing dynamic NE regret is achievable when the payoff sequence satisfies suitable regularity conditions.

A central question is therefore which regularity conditions on the payoff sequence are sufficient to guarantee equilibrium learning. A natural candidate is the second-order payoff variation

$$V_T := \sum_{t=2}^{T} \|\mathbf{A}_t - \mathbf{A}_{t-1}\|_{\max}^2 ,$$

which directly captures how much the payoff matrix changes from one round to the next. Zhang et al. (2022) obtained vanishing dynamic NE regret under the condition that $Q_T = V_T + \min\{P_T, W_T\}$

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

is sublinear, where P_T measures the variation of the equilibrium sequence and W_T measures the deviation of the payoff matrices from their temporal average. This leaves open a natural question:

> *Is sublinear V_T alone sufficient for vanishing dynamic NE regret, and, if so, can the same guarantee be attained efficiently by natural learning dynamics?*

1.1 Our results

Our results reveal a sharp separation between the information-theoretic and computational limits of learning in time-varying zero-sum games. We first establish a positive information-theoretic result: the payoff variation V_T alone is sufficient for equilibrium learning. Specifically, even for the vector feedback setting, we obtain a vanishing average dynamic NE regret whenever V_T is sublinear.

Theorem 1.1 (Informal; Formal version in Theorem 3.3)**.** *Consider a time-varying two-player zero-sum game with second-order payoff variation V_T. There exists a learning algorithm such that, when employed by both players, it achieves*

$$\frac{\text{DynNE-Reg}_{[T]}}{T} \le O\left(V_T^{1/4} T^{-1/4}\right)$$

after T rounds under vector feedback.

In contrast to the parameter-free guarantee of Zhang et al. (2022), which depends on the combined quantity $Q_T = V_T + \min\{P_T, W_T\}$, our bound depends only on V_T, at the cost of requiring an a priori upper bound on the variation V_T. This raises the question of how strongly the learning performance must depend on V_T. We answer this with a complementary information-theoretic lower bound that quantifies the unavoidable dependence on V_T.

Theorem 1.2 (Informal; Formal version in Theorem 3.6)**.** *For any pair of online learners, there exists a time-varying zero-sum game such that, after T rounds,*

$$\frac{\text{DynNE-Reg}_{[T]}}{T} \ge \Omega(V_T^{1/2} T^{-1/2}).$$

Notably, this lower bound holds even with full-matrix feedback. Together with Observation 3.2, which achieves an average duality gap of order $O(\sqrt{V_T/T})$ under the same feedback model, it characterizes the optimal dependence on V_T in the full-information setting.

Taken together, Theorem 1.1 and Theorem 1.2 delineate the information-theoretic limits for time-varying zero-sum games with controlled payoff variation. In particular, when $V_T = O(T^{1-\delta})$ for some constant $\delta > 0$, any inverse-polynomial accuracy $\epsilon = \mathsf{poly}(1/m, 1/n)$ can be reached within $T = \mathsf{poly}(m, n)$ rounds. The remaining challenge is computational: whether efficient learning dynamics can attain the same inverse-polynomial accuracy in the same regime where V_T is strongly sublinear. We show that this task is PPAD-hard for *strongly stable* learners, namely, learners whose strategy sequence satisfies $\sum_{t=2}^{T} \|\boldsymbol{x}_t - \boldsymbol{x}_{t-1}\|_1 = O(T^{1-\delta})$ for some constant $\delta > 0$.

Theorem 1.3 (Informal; Formal version in Theorem 4.3)**.** *For every strongly stable learning algorithm, there exists some time-varying zero-sum games with strongly sublinear V_T and some accuracy $\epsilon = \mathsf{poly}(1/n)$ such that it is* PPAD*-hard for the learning algorithm to achieve*

$$\frac{\text{DynNE-Reg}_{[T]}}{T} \le \epsilon.$$

Theorem 1.3 builds upon the recent PPAD-hardness of computing approximate Nash equilibria in 2 vs. 2 team zero-sum games (Bernasconi et al., 2026; Anagnostides et al., 2026). Starting from such a four-player game, we construct four time-varying two-player zero-sum games, one associated with each player in the original game. In each induced game, the strategies of two players are encoded into the payoff matrix, while the remaining players interact through an ordinary two-player zero-sum game. The construction is designed so that the dynamic NE regret of each induced game captures the unilateral deviation incentive of the corresponding player in the original team game. Consequently, small dynamic NE regret in the four induced games recovers an approximate Nash equilibrium of the original team zero-sum game, which is PPAD-hard. The strong stability assumption ensures that

the induced payoff sequences have strongly sublinear second-order variation V_T, so the reduction remains within the same regime considered in Theorem 1.1. Consequently, unless P = PPAD, no polynomial-time strongly stable learner can attain inverse-polynomial average dynamic NE regret, even if V_T is also strongly sublinear.

Why strongly stable learners? Strong stability captures a natural structural property of many standard online learning dynamics. Stability has long played an important role in regret analysis (Ross & Bagnell, 2011; Saha et al., 2012), and many commonly used methods—including projected gradient descent, mirror descent, follow-the-regularized-leader, and their optimistic variants—produce slowly varying strategy sequences under vanishing learning rates. This connection is particularly explicit in the Regret bounded by Variation in Utilities (RVU) framework (Syrgkanis et al., 2015; Anagnostides et al., 2022; 2023). A similar idea underlies the DRVU framework of Zhang et al. (2022), where controlling the variation of the generated strategies is a central ingredient in the analysis of time-varying games.

1.2 Related work

A growing literature studies online learning when the underlying game evolves over time. Cardoso et al. (2019); Fiez et al. (2021) consider zero-sum games with evolving payoff matrices while Zhang et al. (2022) systematically study online learning in time-varying two-player zero-sum games, introducing dynamic NE regret and deriving guarantees for regret and duality gap. More recently, Anagnostides et al. (2023) analyze optimistic gradient descent dynamics in time-varying games, while Feng et al. (2023); Zhang et al. (2026) focus on the last-iterate convergence of the extra-gradient method. Related researches also arise in the broader dynamic-regret literature, where performance is characterized in terms of temporal variation of the objectives or comparator sequence (Jadbabaie et al., 2015; Besbes et al., 2015; Yang et al., 2016; Perdomo et al., 2020; Campolongo & Orabona, 2021).

2 Preliminaries

Notation We use bold lowercase letters, such as $\boldsymbol{x}$ to denote vectors in $\mathbb{R}^d$, and bold uppercase letters $\mathbf{A}$ to denote matrices in $\mathbb{R}^{m\times n}$. We denote the Euclidean, ℓ_1, and ℓ_∞ norms by $\|\cdot\|_2$, $\|\cdot\|_1$, and $\|\cdot\|_\infty$, respectively. Unless specified otherwise, sometimes we use $\|\cdot\|$ to represent $\|\cdot\|_2$. For a matrix $\mathbf{A}$, we define the entry-wise maximum norm by $\|\mathbf{A}\|_{\max} \coloneqq \max_{i,j} |\mathbf{A}_{i,j}|$. For an arbitrary positive integer $k \in \mathbb{Z}_+$, let $[k] \coloneqq \{1,\dots,k\}$. We use the standard asymptotic notation $O(\cdot)$ and $\Omega(\cdot)$ to suppress constant factors. Moreover, we use $o(\cdot)$ to denote a strictly smaller asymptotic order.

Definition 2.1 (Nash, 1950)**.** A joint strategy $(\boldsymbol{x}_1,\dots,\boldsymbol{x}_N) \in \Delta(\mathcal{A}_1)\times\cdots\times\Delta(\mathcal{A}_N)$ is an ϵ-*approximate Nash equilibrium* if for any player $i\in[N]$, we have

$$u_i(\boldsymbol{x}_i,\boldsymbol{x}_{-i}) \geq u_i(\boldsymbol{x}_i',\boldsymbol{x}_{-i}) - \epsilon \qquad \forall \boldsymbol{x}_i' \in \Delta(\mathcal{A}_i),$$

where $\boldsymbol{x}_{-i} \coloneqq (\boldsymbol{x}_1,\dots,\boldsymbol{x}_{i-1},\boldsymbol{x}_{i+1},\dots,\boldsymbol{x}_N)$ is the joint strategy profile of all players other than player i, and $u_i(\cdot)$ denotes the utility function of player i.

For a two-player zero-sum game with payoff matrix $\mathbf{A}$, it admits the min-max formulation: $\min_{\boldsymbol{x}\in\Delta_m}\max_{\boldsymbol{y}\in\Delta_n}\boldsymbol{x}^\top\mathbf{A}\boldsymbol{y}$. Then a joint strategy $(\boldsymbol{x},\boldsymbol{y})$ is an ϵ-approximate Nash equilibrium if the following condition holds:

$$\boldsymbol{x}^\top\mathbf{A}\boldsymbol{y} - \min_{\boldsymbol{x}'\in\Delta_m}\boldsymbol{x}'^\top\mathbf{A}\boldsymbol{y} \leq \epsilon \quad \text{and} \quad \max_{\boldsymbol{y}'\in\Delta_n}\boldsymbol{x}^\top\mathbf{A}\boldsymbol{y}' - \boldsymbol{x}^\top\mathbf{A}\boldsymbol{y} \leq \epsilon.$$

Next, we define the duality gap, which is the sum of these two unilateral deviation gaps.

Definition 2.2. For a two-player zero-sum game with payoff matrix $\mathbf{A}\in\mathbb{R}^{m\times n}$ and a strategy pair $(\boldsymbol{x},\boldsymbol{y})\in\Delta_m\times\Delta_n$, the *duality gap* of $(\boldsymbol{x},\boldsymbol{y})$ is

$$\text{Dual-Gap}_{\mathbf{A}}(\boldsymbol{x},\boldsymbol{y}) \coloneqq \max_{\boldsymbol{y}'\in\Delta_n}\boldsymbol{x}^\top\mathbf{A}\boldsymbol{y}' - \min_{\boldsymbol{x}'\in\Delta_m}\boldsymbol{x}'^\top\mathbf{A}\boldsymbol{y}.$$

When the payoff matrix and the strategy profile are clear from context, we suppress the argument and write Dual-Gap instead.

Time-varying zero-sum games An instance of a *two-player time-varying zero-sum games* consists of a sequence of two-player zero-sum games with payoff matrices $\mathbf{A}_1, \dots, \mathbf{A}_T \in [-1, 1]^{m \times n}$. At each round $t \in [T]$, the environment selects the payoff matrix $\mathbf{A}_t$. Without observing $\mathbf{A}_t$, the minimizer and the maximizer choose $\boldsymbol{x}_t \in \Delta_m$ and $\boldsymbol{y}_t \in \Delta_n$, respectively, yielding the realized payoff $\boldsymbol{x}_t^\top \mathbf{A}_t \boldsymbol{y}_t$. After the round, the players receive feedback about $\mathbf{A}_t$. Under *vector feedback*, player $\boldsymbol{x}$ observes $\mathbf{A}_t \boldsymbol{y}_t$, while player $\boldsymbol{y}$ observes $\mathbf{A}_t^\top \boldsymbol{x}_t$. Under *full-matrix feedback*, both players observe the entire matrix $\mathbf{A}_t$. For a sequence of payoff matrices $\{\mathbf{A}_t\}_{t=1}^T$, we define the *cumulative duality gap* as

$$\text{Dual-Gap}_{[T]} := \sum_{t=1}^{T} \text{Dual-Gap}_{\mathbf{A}_t}(\boldsymbol{x}_t, \boldsymbol{y}_t). \tag{1}$$

We also define the *cumulative dynamic NE regret* (Zhang et al., 2022), which compares the cumulative payoff generated by the strategy sequence $\{(\boldsymbol{x}_t, \boldsymbol{y}_t)\}_{t=1}^T$ with the cumulative Nash equilibrium values. In particular,

$$\text{DynNE-Reg}_{[T]} := \left| \sum_{t=1}^{T} \left(\boldsymbol{x}_t^\top \mathbf{A}_t \boldsymbol{y}_t - \min_{\boldsymbol{x} \in \Delta_m} \max_{\boldsymbol{y} \in \Delta_n} \boldsymbol{x}^\top \mathbf{A}_t \boldsymbol{y} \right) \right|. \tag{2}$$

Payoff variation. For a sequence of payoff matrices $\{\mathbf{A}_t\}_{t=1}^T$, define its second-order path-length variation by

$$V_T(\mathbf{A}) := \sum_{t=2}^{T} \|\mathbf{A}_t - \mathbf{A}_{t-1}\|_{\max}^2. \tag{3}$$

When the payoff matrix sequence is clear from context, we simply write V_T.

3 Information-Theoretic Bounds

In this paper, we focus on time-varying zero-sum games with sublinear second-order payoff path-length variation, i.e., $V_T = o(T)$. Zhang et al. (2022) established positive learning guarantees under additional bounds on P_T and W_T, where P_T measures the variation of Nash equilibria along the sequence and W_T measures the deviation of the payoff matrices from their average. This raises a natural question: is sublinear V_T alone sufficient to guarantee vanishing average duality gap and dynamic NE regret? We address this question from both information-theoretic and algorithmic perspectives. In Section 3.1, we provide learning algorithms showing that whenever $V_T = o(T)$, both the average duality gap and the average dynamic NE regret vanish over time. In Section 3.2, we complement these positive results with an information-theoretic lower bound that captures the unavoidable dependence on the payoff variation.

We begin this section with the following proposition, which will be used repeatedly.

Proposition 3.1 (Zhang et al., 2022)**.** *For any sequence of time-varying two-player zero-sum games, it holds that* $\text{DynNE-Reg}_{[T]} \le \text{Dual-Gap}_{[T]}$.

3.1 Information-theoretic upper bound

We first consider the full-matrix feedback setting, where both players observe the entire payoff matrix $\mathbf{A}_t$ after each round. In the single-agent setting, Jadbabaie et al. (2015) showed that playing a minimizer of the previous round's loss yields a V_T-dependent bound on dynamic regret. Motivated by this idea, we consider the analogous strategy for time-varying zero-sum games: at each round, both players play a Nash equilibrium of the previous payoff matrix.

Observation 3.2. *Consider a sequence of two-player zero-sum games* $\{\mathbf{A}_t\}_{t=1}^T \subseteq [-1, 1]^{m \times n}$. *Under full-matrix feedback, there exists a learning algorithm such that, when applying to both players, achieves*

$$\frac{\text{Dual-Gap}_{[T]}}{T} \le \frac{2}{T} + 2\sqrt{\frac{V_T}{T}}.$$

In particular, if V_T is sublinear in T, then the average duality gap vanishes.

Algorithm 1

Require: Horizon T, variation budget $V \geq V_T$, number of actions m, n

Choose block length $H \leftarrow \sqrt{m+n}\left(\frac{T}{V}\right)^{1/4}$.
Partition $[T]$ into $B = \lceil \frac{T}{H} \rceil$ consecutive blocks of length H.
for each block $b \in [B]$ **do**
  Probing phase:
  $t_0 \leftarrow (b-1)H$
  for $i = 1, \ldots, m+n$ **do**
    $t_i \leftarrow t_0 + i$
    Minimizer plays $\boldsymbol{x}_t = \boldsymbol{e}_{(i \bmod m)}$.
    Maximizer player plays $\boldsymbol{y}_t = \boldsymbol{e}_{(i \bmod n)}$.
    if $i \leq n$ **then**
      Minimizer construct the i-th column $\widehat{\mathbf{A}}_b^x(\cdot, i) = \mathbf{A}_{t_i}\boldsymbol{y}_{t_i}$
    end if
    if $i \leq m$ **then**
      Maximizer construct the i-th row $\widehat{\mathbf{A}}_b^y(i, \cdot) = \boldsymbol{x}_{t_i}^\top \mathbf{A}_{t_i}$
    end if
  end for
  End probing phase
  Minimizer computes $\widehat{\boldsymbol{x}}_b \in \operatorname{argmin}_{\boldsymbol{x}\in\Delta_m} \max_{\boldsymbol{y}\in\Delta_n} \boldsymbol{x}^\top \widehat{\mathbf{A}}_b^x \boldsymbol{y}$.
  Maximizer computes $\widehat{\boldsymbol{y}}_b \in \operatorname{argmax}_{\boldsymbol{y}\in\Delta_n} \min_{\boldsymbol{x}\in\Delta_m} \boldsymbol{x}^\top \widehat{\mathbf{A}}_b^y \boldsymbol{y}$.
  for $t = t_0 + m + n + 1, \ldots, bH$ **do**
    Play $\boldsymbol{x}_t = \widehat{\boldsymbol{x}}_b$ and $\boldsymbol{y}_t = \widehat{\boldsymbol{y}}_b$.
  end for
end for

Proof sketch. Because $(\boldsymbol{x}_t, \boldsymbol{y}_t)$ is a Nash equilibrium of $\mathbf{A}_{t-1}$, its duality gap with respect to $\mathbf{A}_{t-1}$ is zero. Also, duality gap under the current matrix $\mathbf{A}_t$ can increase by at most $2\|\mathbf{A}_t - \mathbf{A}_{t-1}\|_{\max}$. Lastly, summing over rounds and applying Cauchy–Schwarz gives $\text{Dual-Gap}_{[T]} \leq 2 + 2\sqrt{TV_T}$, which yields the stated average bound after dividing by T. The complete proof is in Appendix B.1. □

From Proposition 3.1, dynamic NE regret is upper bounded by the cumulative duality gap, so Observation 3.2 also implies the same upper bound for dynamic NE regret.

We now move to the settings where each player only observes the vector feedback (Zhang et al., 2022). At each round t, after the players choose $\boldsymbol{x}_t \in \Delta_m$ and $\boldsymbol{y}_t \in \Delta_n$, the minimizing player observes $\mathbf{A}_t\boldsymbol{y}_t$, while the maximizing player observes $\mathbf{A}_t^\top\boldsymbol{x}_t$. In this setting, having the two players playing Nash equilibrium from the previous round is no longer applicable as each agent does not directly observe the payoff matrix. Instead, we design Algorithm 1 that partitions the horizon into several blocks and probes the payoff matrix for each block by playing pure actions.

Specifically, the algorithm divides the horizon into consecutive blocks of length $H = \sqrt{m+n}(T/V)^{1/4}$ for a known budget $V_T \leq V$, where the last block may be smaller. Consequently, the number of blocks is $B = \lceil \frac{T}{H} \rceil$. At the beginning of each block, the players cycle through pure actions for $m+n$ rounds. Under vector feedback, these probes reveal the columns of the payoff matrix to the minimizer and the rows to the maximizer, allowing each player to construct its own matrix estimate. The minimizer then plays a minimax strategy of its estimate, while the maximizer plays a maximin strategy of its estimate for the remainder of the block. The probing schedule is fixed in advance, so the players need not communicate.

Theorem 3.3. *Consider a time-varying two-player zero-sum game with payoff matrices $\{\mathbf{A}_t\}_{t=1}^T \subseteq [-1, 1]^{m\times n}$ and second-order payoff variation $V_T = \sum_{t=2}^T \|\mathbf{A}_t - \mathbf{A}_{t-1}\|_{\max}^2$. Suppose that V_T is known to both players. Then there exists a learning algorithm (Algorithm 1) such that, when applying to both players, achieves*

$$\frac{\text{Dual-Gap}_{[T]}}{T} \leq O\left(\frac{m+n}{T} + \sqrt{m+n}\left(\frac{V_T}{T}\right)^{1/4}\right),$$

Consequently, the same bound holds for the average dynamic NE regret.

Proof sketch. For each block b, let V_b denote the payoff variation within the block. Since every row and column used to construct the blockwise matrix estimates $\widehat{\mathbf{A}}_b^x$ and $\widehat{\mathbf{A}}_b^y$ is observed during that same block, telescoping the matrix changes and applying Cauchy–Schwarz gives $\|\mathbf{A}_t - \widehat{\mathbf{A}}_b^x\|_{\max},\ \|\mathbf{A}_t - \widehat{\mathbf{A}}_b^y\|_{\max} \le \sqrt{HV_b}$ for every round t after the probing phase. Moreoever, due to the Lipschitz continuity of equilibrium value, the resulting duality gap on these rounds is at most $O(\sqrt{HV_b})$ per round. Summing over blocks and using $\sum_b V_b \le V$ gives a total duality-gap contribution of $O(H\sqrt{TV})$ from the non-probing rounds. Detailed proof is presented in Section B.2.

The probing rounds contribute at most $2(m+n)$ duality gap per block, hence $O((m+n)T/H+m+n)$ in total. Combining these two contributions, dividing by T, and setting $H = \sqrt{m+n}(T/V)^{1/4}$ yields the stated bound. The complete proof is given in Appendix B.2. □

Remark 3.4. Compared with the parameter-free algorithm developed by Zhang et al. (2022), Algorithm 1 requires the players to know an upper bound on the payoff variation V_T in advance. In return, our guarantee depends only on V_T, without requiring additional control on the variation of equilibrium or strategy sequences. Similar prior knowledge assumptions appear in the dynamic-regret literature when seeking guarantees that depend only on the variation of the underlying objective functions (Besbes et al., 2015).

Consider a time-varying zero-sum game with strongly sublinear second-order payoff variation V_T, for any inverse-polynomial accuracy, Theorem 3.3 implies the following corollary.

Corollary 3.5. *Suppose that the second-order payoff variation satisfies $V_T = O(T^{1-\delta})$ for some constant $\delta > 0$. Let the accuracy $\epsilon = \mathsf{poly}(1/m, 1/n)$ be inverse-polynomial, in $T = \mathsf{poly}(m, n)$ rounds, Algorithm 1 applying to both players can achieve $\text{Dual-Gap}_{[T]}/T \le \epsilon$ and $\text{DynNE-Reg}_{[T]}/T \le \epsilon$.*

3.2 Information-theoretic lower bound

We now derive an information-theoretic lower bound relating the payoff variation V_T to the dynamic NE regret incurred by any pair of online learners. The lower bound holds even for time-varying zero-sum games with only two actions per player. In particular, we show that every learning algorithm must incur $\Omega(\sqrt{V_T/T})$ average dynamic NE regret on some time-varying zero-sum game. This matches the rate established in Observation 3.2 under full-matrix feedback, showing that the dependence on V_T is information-theoretically tight.

Theorem 3.6. *For any pair of online learners, there exists a deterministic sequence of zero-sum payoff matrices $\mathbf{A}_1, \dots, \mathbf{A}_T \in [-1, 1]^{2\times 2}$ with sublinear $V_T = o(T)$ such that $\text{DynNE-Reg}_{[T]}/T \ge \Omega\left(\sqrt{V_T/T}\right)$. Consequently, the same bound holds for the average duality gap.*

Proof sketch. Following the random-perturbation approach of Yang et al. (2016), we construct a sequence of 2×2 time-varying zero-sum games whose payoffs have magnitude α and whose signs change independently across rounds. Because the round t strategy is chosen before the current sign is revealed, it forces expected dynamic NE regret to be $\Omega(\alpha)$ per round and hence $\Omega(\alpha T)$ cumulatively. Each sign change alters the payoff matrix by only $O(\alpha)$, so the resulting second-order variation satisfies $V_T = O(\alpha^2 T)$. An averaging argument then fixes a deterministic sign sequence with the same lower-bound order. Eliminating α between these two bounds yields $\text{DynNE-Reg}_{[T]} = \Omega(\sqrt{TV_T})$. Choosing $\alpha = o(1)$ ensures $V_T = o(T)$. The complete proof is given in Appendix B.3. □

The preceding upper and lower bounds together characterize the information-theoretic complexity of the problem up to polynomial factors. When $V_T = O(T^{1-\delta})$ for a fixed constant $\delta > 0$, the upper bound shows that inverse-polynomial average accuracy, $\text{Dual-Gap}_{[T]}/T \le \epsilon$ and $\text{DynNE-Reg}_{[T]}/T \le \epsilon$, with $\epsilon = \mathsf{poly}(1/n, 1/m)$, is achievable within a polynomial number of rounds $T = \mathsf{poly}(m, n)$. On the other hand, the information-theoretic lower bound shows that polynomially many rounds are necessary: for example, achieving average error at most ϵ over the

corresponding slowly varying class requires $T = \Omega(\epsilon^{-2/\delta})$, which remains polynomial when ϵ is inverse-polynomial and $\delta > 0$ is a fixed constant.

4 DynNE-Reg REDUCTION AND COMPUTATIONAL HARDNESS

In this section, we address the computational question by reducing a 2 vs. 2 team zero-sum game to a collection of two-player time-varying zero-sum games. The reduction allows us to transfer the hardness of computing approximate Nash equilibria in team games to the problem of achieving small dynamic NE regret. In particular, our reduction relies on the following result of Bernasconi et al. (2026); Anagnostides et al. (2026).

Proposition 4.1. *Let n denote the number of actions in the game. For some inverse-polynomial accuracy parameter $\epsilon = \mathsf{poly}(1/n)$, finding an ϵ-approximate Nash equilibrium of a 2 vs. 2 team zero-sum game is* PPAD*-complete.*

4.1 FOUR-GAME REDUCTION

Given a two-team game, we construct *induced time-varying payoff matrices* that preserve the payoff of the original game. In particular, in each induced matrix game, we select one player from each team as the two explicit players, while the strategies of the remaining two players are encoded into the entries of the payoff matrix. As these strategies evolve over time, the corresponding induced payoff matrix becomes time-varying.

Specifically, consider a 2 vs. 2 team zero-sum game with minimizing players $\boldsymbol{x}_1, \boldsymbol{x}_2$ and maximizing players $\boldsymbol{y}_1, \boldsymbol{y}_2$. Player $\boldsymbol{x}_i$ has strategy space $\Delta(\mathcal{A}_i)$ and player $\boldsymbol{y}_i$ has strategy space $\Delta(\mathcal{B}_i)$, where $m_i := |\mathcal{A}_i|$ and $n_i := |\mathcal{B}_i|$ for $i \in \{1, 2\}$. For a joint strategy profile $\boldsymbol{z} = (\boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{y}_1, \boldsymbol{y}_2)$, define the expected payoff to the maximizing team by

$$f(\boldsymbol{z}) = \sum_{a_1,a_2,b_1,b_2} \mathbf{U}_{a_1,a_2,b_1,b_2} \boldsymbol{x}_1(a_1)\boldsymbol{x}_2(a_2)\boldsymbol{y}_1(b_1)\boldsymbol{y}_2(b_2),$$

where $\mathbf{U} \in [-1, 1]^{m_1 \times m_2 \times n_1 \times n_2}$ is the four-dimensional payoff tensor and $\mathbf{U}_{a_1,a_2,b_1,b_2}$ denotes the payoff to the maximizing team at pure-action profile (a_1, a_2, b_1, b_2). The other team receives $-f(\boldsymbol{z})$.

For any strategies $\boldsymbol{x}_2 \in \Delta_{m_2}$ and $\boldsymbol{y}_2 \in \Delta_{n_2}$, we define each entry of the induced payoff matrix $\mathbf{A}^{(1)}(\boldsymbol{x}_2, \boldsymbol{y}_2) \in [-1, 1]^{m_1 \times n_1}$ by $\mathbf{A}^{(1)}(\boldsymbol{x}_2, \boldsymbol{y}_2)_{a_1,b_1} := \sum_{a_2,b_2} \boldsymbol{x}_2(a_2)\boldsymbol{y}_2(b_2)\mathbf{U}_{a_1,a_2,b_1,b_2}$. It is then straightforward to verify that $\boldsymbol{x}_1^\top \mathbf{A}^{(1)}(\boldsymbol{x}_2, \boldsymbol{y}_2)\boldsymbol{y}_1 = f(\boldsymbol{z})$. By using the strategies of players $\boldsymbol{x}_2$ and $\boldsymbol{y}_2$ to define the induced payoff matrix $\mathbf{A}^{(1)}(\boldsymbol{x}_2, \boldsymbol{y}_2)$, the payoff of the original team zero-sum game can be represented as the payoff of a two-player matrix game between $\boldsymbol{x}_1$ and $\boldsymbol{y}_1$. Similarly, we define the entries of the symmetric induced matrix $\mathbf{A}^{(2)}(\boldsymbol{x}_1, \boldsymbol{y}_1) \in [-1, 1]^{m_2 \times n_2}$ as $\mathbf{A}^{(2)}(\boldsymbol{x}_1, \boldsymbol{y}_1)_{a_2,b_2} := \sum_{a_1,b_1} \boldsymbol{x}_1(a_1)\boldsymbol{y}_1(b_1)\mathbf{U}_{a_1,a_2,b_1,b_2}$, so that $\boldsymbol{x}_2^\top \mathbf{A}^{(2)}(\boldsymbol{x}_1, \boldsymbol{y}_1)\boldsymbol{y}_2 = f(\boldsymbol{z})$.

We now consider the induced construction over some time horizon T. For a joint strategy profile $\boldsymbol{z}_t = (\boldsymbol{x}_{1,t}, \boldsymbol{x}_{2,t}, \boldsymbol{y}_{1,t}, \boldsymbol{y}_{2,t})$ at round $t \in [T]$, let the induced matrices for that round be $\mathbf{A}_t^{(1)} := \mathbf{A}^{(1)}(\boldsymbol{x}_{2,t}, \boldsymbol{y}_{2,t})$, and $\mathbf{A}_t^{(2)} := \mathbf{A}^{(2)}(\boldsymbol{x}_{1,t}, \boldsymbol{y}_{1,t})$.

The two induced games preserve the payoff of the original team game, but this alone is not enough to relate dynamic NE regret in the two-player games to approximate equilibrium in the team game. Therefore, based on $\mathbf{A}_t^{(1)}$ and $\mathbf{A}_t^{(2)}$, we design four time-varying zero-sum games such that together their dynamic NE regret quantities can recover the Nash equilibrium gap of the original team game.

Specifically, define the four payoff matrices to be

$$\mathbf{C}_t^{\boldsymbol{x}_1} := \left(\mathbf{A}_t^{(1)}\boldsymbol{y}_{1,t}\right)\mathbf{1}_{n_1}^\top \in [-1, 1]^{m_1 \times n_1}, \qquad \mathbf{C}_t^{\boldsymbol{x}_2} := \left(\mathbf{A}_t^{(2)}\boldsymbol{y}_{2,t}\right)\mathbf{1}_{n_2}^\top \in [-1, 1]^{m_2 \times n_2},$$
$$\mathbf{C}_t^{\boldsymbol{y}_1} := \mathbf{1}_{m_1}\left(\boldsymbol{x}_{1,t}^\top \mathbf{A}_t^{(1)}\right) \in [-1, 1]^{m_1 \times n_1}, \qquad \mathbf{C}_t^{\boldsymbol{y}_2} := \mathbf{1}_{m_2}\left(\boldsymbol{x}_{2,t}^\top \mathbf{A}_t^{(2)}\right) \in [-1, 1]^{m_2 \times n_2}.$$

Each payoff matrix is associated with a single team game player, as indicated by its superscript. The matrices associated with $\boldsymbol{x}_i$ have identical columns, while those associated with $\boldsymbol{y}_i$ have identical rows, making one player's strategic choice irrelevant in each game. This allows the corresponding

dynamic NE regret quantity to isolate the unilateral deviation incentive of the associated team game player. Summing the four resulting quantities recovers the Nash equilibrium gap of the original team game. We formalize this mechanism in Proposition C.4 of Appendix C.1.

4.2 From stability to computational hardness

We first introduce the definition of stable learners, which plays a central role in the reductions to bound V_T of the constructed games.

Definition 4.2. Let $\{\boldsymbol{x}_t\}_{t=1}^T$ be the strategy sequence produced by an online learner, with $\boldsymbol{x}_t \in \Delta$ for each round t. We define the learner's strategy path-length by $S_T(\boldsymbol{x}) := \sum_{t=2}^T \|\boldsymbol{x}_t - \boldsymbol{x}_{t-1}\|_1$. We say that the learner is *strongly stable* if $S_T(\boldsymbol{x}) = O(T^{1-\delta})$ for some constant $\delta > 0$.

The induced payoff matrices $\mathbf{A}_t^{(1)}$ and $\mathbf{A}_t^{(2)}$ depend Lipschitz-continuously on the underlying player strategies. Consequently, for our four-game construction, strongly stable strategy sequences generate payoff sequences with strongly sublinear matrix variation. We formalize this stability-to-variation argument in Proposition C.6 of Appendix C.2.

Together with Corollary 3.5, this shows that the reduction operates within a variation regime where inverse-polynomial average dynamic NE regret is information-theoretically achievable. Thus, we can proceed to the computational complexity result which proves that achieving the same guarantee is PPAD-hard, establishing a separation between information-theoretic and computational learnability.

Theorem 4.3. *Consider any time-varying two-player zero-sum payoff sequence with strongly sublinear second-order payoff path-length variation. There exists an inverse-polynomial accuracy $\epsilon = \mathsf{poly}(1/n)$ such that it is PPAD-hard to design an online learner that simultaneously satisfies the following two properties: i) $S_T(\boldsymbol{x}) = O(T^{1-\delta})$ with some constant $\delta > 0$ for every strategy sequence $\{\boldsymbol{x}_t\}_{t=1}^T$ generated by the learner; and ii) $\text{DynNE-Reg}_{[T]}/T \le \epsilon$. Consequently, the same hardness result applies to the average duality gap.*

Proof sketch. Let G be a 2 vs. 2 team zero-sum game whose inverse-polynomial approximate Nash equilibrium problem is PPAD-hard. Applying the strongly stable learner on the four induced games of Section 4.1 yields four payoff sequences with strongly sublinear variation. Recall that each game is constructed so that its dynamic NE regret recovers the unilateral deviation incentive of one player of the original game. Therefore, by Proposition C.4, small average dynamic NE regret across these games implies that some profile $\boldsymbol{z}_{t^*}$ is an ϵ-approximate Nash equilibrium of G. The complete proof is given in Appendix C.3. □

Corollary 4.4. *There exists some inverse-polynomial accuracy ϵ such that, unless $\mathsf{P} = \mathsf{PPAD}$, there is no polynomial-time learning algorithm with learner stability $S_T = O(T^{1-\delta})$ for some constant $\delta > 0$ that can achieve $\text{Dual-Gap}_{[T]}/T \le \epsilon$ or $\text{DynNE-Reg}_{[T]}/T \le \epsilon$ for every time-varying two-player zero-sum payoff sequence satisfying payoff matrix variation $V_T = O(T^{1-\delta})$.*

5 Experiment

We support our theoretical results with a simple experiment demonstrating that strongly stable learners can perform poorly in time-varying zero-sum games, even when the payoff variation is strongly sublinear. Specifically, we construct a sequence with $V_T = O(\sqrt{T})$ and evaluate several standard no-regret algorithms whose strategy path lengths satisfy $S_T = O(\sqrt{T})$. Our experiments show that both the average duality gap and the average dynamic NE regret remain bounded away from zero over the tested horizons.

5.1 Game description

We define a time-varying zero-sum game where the minimizer $\boldsymbol{x}$ and maximizer $\boldsymbol{y}$ each have two actions, denoted by H and T. Fix $\epsilon = 0.1$ and consider the following payoff matrices

$$\mathbf{A}^{\mathsf{H}} = \begin{pmatrix} 0 & -\epsilon \\ 1 & 1 \end{pmatrix}, \qquad \mathbf{A}^{\mathsf{T}} = \begin{pmatrix} 1 & 1 \\ -\epsilon & 0 \end{pmatrix}.$$

The unique Nash equilibrium of $\mathbf{A}^{\mathsf{H}}$ is (H,H), while the unique Nash equilibrium of $\mathbf{A}^{\mathsf{T}}$ is (T,T). Both games have value zero. Thus, alternating between the two payoff matrices moves the Nash equilibrium between (H,H) and (T,T), while leaving the equilibrium value unchanged.

For a fixed horizon T, let $B_T = \lceil\sqrt{T}\rceil$. The environment plays $\mathbf{A}^{\mathsf{H}}$ for B_T consecutive rounds, then $\mathbf{A}^{\mathsf{T}}$ for B_T rounds, and continues alternating in this way for the remainder of the horizon. Every switch satisfies $\left\|\mathbf{A}^{\mathsf{H}} - \mathbf{A}^{\mathsf{T}}\right\|_{\max} = 1+\epsilon$, and the number of switches is $\lfloor \frac{T-1}{B_T} \rfloor$. Therefore $V_T = \left\lfloor \frac{T-1}{B_T} \right\rfloor (1+\epsilon)^2 = \Theta\left(\sqrt{T}\right)$.

5.2 Learning Results

We consider four standard no-regret learning algorithms: OMWU, PGD, FTRL, and OGDA. All four algorithms use the common horizon-dependent learning rate $\eta_T = T^{-1/2}$. Under this choice, each of the learning algorithms is strongly stable, with $S_T(\boldsymbol{x}) = S_T(\boldsymbol{y}) = O(\sqrt{T})$. We defer the verification of the stability to Appendix D.1. The resulting average V_t, average duality gap, and average dynamic NE regret are reported in Figure 1. Additional experiments over longer horizons are provided in Appendix D.

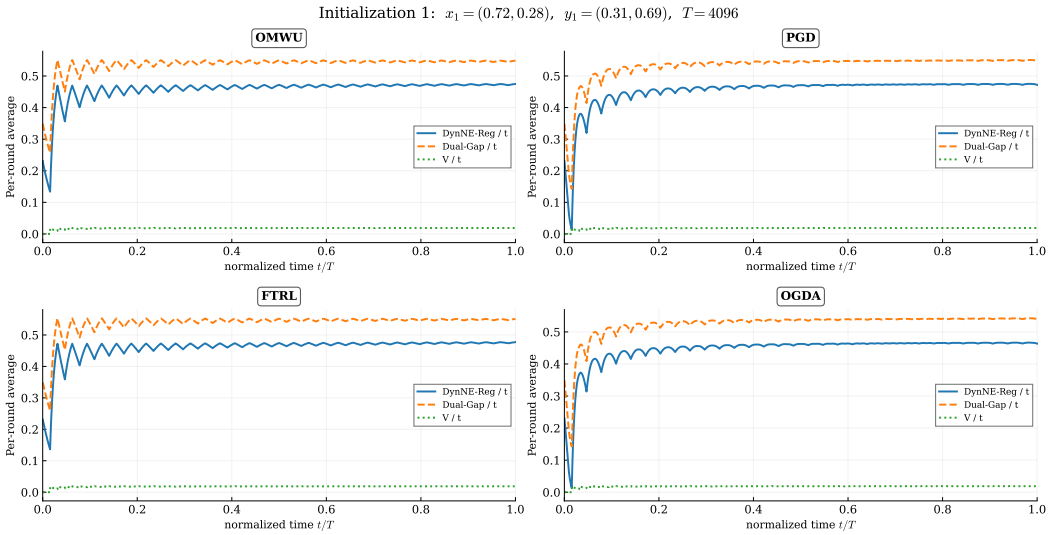


Figure 1: Results for the initialization $\boldsymbol{x}_1 = (0.72, 0.28)$ and $\boldsymbol{y}_1 = (0.31, 0.69)$ at horizon $T = 4096$.

6 Conclusion

In this paper, we study when learning remains possible in time-varying zero-sum games. Our results show that slowly varying games can be learnable in principle: sublinear payoff variation V_T is sufficient for vanishing average dynamic NE regret, and we give a matching variation-dependent lower bound up to polynomial factors. At the same time, we show that this information-theoretic possibility does not translate into efficient learning. For strongly stable learners, achieving inverse-polynomial average dynamic NE regret is PPAD-hard. Together, these results reveal a separation between information-theoretic and computational learnability in time-varying environments.

Several directions remain open. On the algorithmic side, it would be useful to close the gap between our upper and lower bounds under vector feedback and to remove the payoff variation parameter required in Algorithm 1. On the computational side, an important question is whether the hardness result can be extended beyond strongly stable learners. More broadly, it would be interesting to understand whether similar information–computation separations arise in richer classes of time-varying games.

AI DISCLOSURE STATEMENT

We use GPT Pro to assist with editing and polishing the writing of this paper, including improving clarity, concision, and organization. All AI-assisted text was reviewed and revised by the authors, who take full responsibility for the final content of the paper.

REFERENCES

Ioannis Anagnostides, Ioannis Panageas, Gabriele Farina, and Tuomas Sandholm. On last-iterate convergence beyond zero-sum games. In *International Conference on Machine Learning*, pp. 536–581. PMLR, 2022.

Ioannis Anagnostides, Ioannis Panageas, Gabriele Farina, and Tuomas Sandholm. On the convergence of no-regret learning dynamics in time-varying games. *Advances in Neural Information Processing Systems*, 36:16367–16405, 2023.

Ioannis Anagnostides, Ioannis Panageas, Tuomas Sandholm, and Jingming Yan. The computational complexity of team zero-sum games. *arXiv preprint arXiv:2606.16139*, 2026.

Martino Bernasconi and Matteo Castiglioni. The complexity of min-max optimization with product constraints. In *Proceedings of the 58th Annual ACM Symposium on Theory of Computing*, pp. 489–499, 2026.

Martino Bernasconi, Matteo Castiglioni, Andrea Celli, and Alexandros Hollender. The complexity of min-max optimization for quadratic polynomials, 2026. URL `https://arxiv.org/abs/2606.17000`.

Omar Besbes, Yonatan Gur, and Assaf Zeevi. Non-stationary stochastic optimization. *Operations research*, 63(5):1227–1244, 2015.

Yang Cai and Constantinos Daskalakis. On minmax theorems for multiplayer games. In *Proceedings of the twenty-second annual ACM-SIAM symposium on Discrete algorithms*, pp. 217–234. SIAM, 2011.

Yang Cai, Ozan Candogan, Constantinos Daskalakis, and Christos Papadimitriou. Zero-sum polymatrix games: A generalization of minmax. *Mathematics of Operations Research*, 41(2):648–655, 2016.

Nicolo Campolongo and Francesco Orabona. A closer look at temporal variability in dynamic online learning. *arXiv preprint arXiv:2102.07666*, 2021.

Adrian Rivera Cardoso, Jacob Abernethy, He Wang, and Huan Xu. Competing against equilibria in zero-sum games with evolving payoffs. *arXiv preprint arXiv:1907.07723*, 2019.

Nicolo Cesa-Bianchi, Gábor Lugosi, et al. *Prediction, learning, and games*, volume 1. Cambridge university press Cambridge, 2006.

Xi Chen, Xiaotie Deng, and Shang-Hua Teng. Settling the complexity of computing two-player Nash equilibria. *Journal of the ACM*, 56(3):14:1–14:57, 2009.

Constantinos Daskalakis, Paul W. Goldberg, and Christos H. Papadimitriou. The complexity of computing a Nash equilibrium. *SIAM Journal of Computing*, 39(1):195–259, 2009.

Benoit Duvocelle, Panayotis Mertikopoulos, Mathias Staudigl, and Dries Vermeulen. Multiagent online learning in time-varying games. *Mathematics of Operations Research*, 48(2):914–941, 2023.

Yi Feng, Hu Fu, Qun Hu, Ping Li, Ioannis Panageas, Xiao Wang, et al. On the last-iterate convergence in time-varying zero-sum games: Extra gradient succeeds where optimism fails. *Advances in Neural Information Processing Systems*, 36:21933–21944, 2023.

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

Yoav Freund and Robert E Schapire. Adaptive game playing using multiplicative weights. *Games and Economic Behavior*, 29(1-2):79–103, 1999.

Drew Fudenberg and David K Levine. *The theory of learning in games*, volume 2. MIT press, 1998.

Sergiu Hart and Andreu Mas-Colell. A simple adaptive procedure leading to correlated equilibrium. *Econometrica*, 68(5):1127–1150, 2000.

Alexandros Hollender, Gilbert Maystre, and Sai Ganesh Nagarajan. The complexity of two-team polymatrix games with independent adversaries. In *International Conference on Learning Representations*, volume 2025, pp. 30456–30471, 2025.

Ali Jadbabaie, Alexander Rakhlin, Shahin Shahrampour, and Karthik Sridharan. Online optimization: Competing with dynamic comparators. In *Artificial Intelligence and Statistics*, pp. 398–406. PMLR, 2015.

Nick Littlestone and Manfred K Warmuth. The weighted majority algorithm. *Information and computation*, 108(2):212–261, 1994.

Thodoris Lykouris, Vasilis Syrgkanis, and Éva Tardos. Learning and efficiency in games with dynamic population. In *Proceedings of the twenty-seventh annual ACM-SIAM symposium on Discrete algorithms*, pp. 120–129. SIAM, 2016.

John Nash. *Non-cooperative games*. PhD thesis, Princeton University, 1950.

Juan Perdomo, Tijana Zrnic, Celestine Mendler-Dünner, and Moritz Hardt. Performative prediction. In *International conference on machine learning*, pp. 7599–7609. PMLR, 2020.

Alexander Rakhlin and Karthik Sridharan. Online learning with predictable sequences. In *Conference on Learning Theory*, pp. 993–1019. PMLR, 2013a.

Sasha Rakhlin and Karthik Sridharan. Optimization, learning, and games with predictable sequences. *Advances in Neural Information Processing Systems*, 26, 2013b.

Stéphane Ross and J Andrew Bagnell. Stability conditions for online learnability. *arXiv preprint arXiv:1108.3154*, 2011.

Ankan Saha, Prateek Jain, and Ambuj Tewari. The interplay between stability and regret in online learning. *arXiv preprint arXiv:1211.6158*, 2012.

Shai Shalev-Shwartz. Online learning and online convex optimization. *Foundations and Trends® in Machine Learning*, 4(2):107–194, 2025.

Vasilis Syrgkanis, Alekh Agarwal, Haipeng Luo, and Robert E Schapire. Fast convergence of regularized learning in games. *Advances in Neural Information Processing Systems*, 28, 2015.

Tianbao Yang, Lijun Zhang, Rong Jin, and Jinfeng Yi. Tracking slowly moving clairvoyant: Optimal dynamic regret of online learning with true and noisy gradient. In *International Conference on Machine Learning*, pp. 449–457. PMLR, 2016.

Junyue Zhang, Xinxiang Guo, Yi Feng, Xiao Wang, and Yifen Mu. 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*, pp. 3280–3282, 2026.

Mengxiao Zhang, Peng Zhao, Haipeng Luo, and Zhi-Hua Zhou. No-regret learning in time-varying zero-sum games. In *Proceedings of the 39th International Conference on Machine Learning*, volume 162 of *Proceedings of Machine Learning Research*, pp. 26772–26808, 2022.

Martin Zinkevich. Online convex programming and generalized infinitesimal gradient ascent. In *Proceedings of the 20th international conference on machine learning (icml-03)*, pp. 928–936, 2003.

A Further related work

Computational complexity of equilibrium computation. Nash equilibria in two-player zero-sum games can be computed in polynomial time by linear programming, in sharp contrast to general-sum games, where equilibrium computation is PPAD-complete (Daskalakis et al., 2009; Chen et al., 2009). This tractability extends to certain multiplayer zero-sum structures, including zero-sum polymatrix games (Cai & Daskalakis, 2011; Cai et al., 2016). Recent work established CLS-hardness for equilibrium computation in two-team polymatrix games (Hollender et al., 2025). Building upon recent breakthroughs in the complexity of min-max optimization (Bernasconi & Castiglioni, 2026), Bernasconi et al. (2026); Anagnostides et al. (2026) further establish PPAD-hardness for team zero-sum games, even when each team consists of only two players, the game is polymatrix, and the target accuracy is inverse-polynomial.

Online learning and no-regret dynamics. No-regret learning is a central paradigm in online decision-making and repeated games (Littlestone & Warmuth, 1994; Freund & Schapire, 1999). In online convex optimization, classic algorithms such as PGD, mirror descent, and FTRL provide no-regret guarantees (Zinkevich, 2003; Shalev-Shwartz, 2025). A later line of work exploits predictability and optimism to obtain sharper guarantees in benign environments, leading to optimistic mirror-descent and related dynamics (Rakhlin & Sridharan, 2013a;b). In games, these ideas underpin faster convergence and refined regret analyses for optimistic learning dynamics (Syrgkanis et al., 2015; Anagnostides et al., 2022). Our work is particularly related to this literature through the role of stability and path-length control in understanding the behavior of online learners.

B Omitted Proofs of Section 3

B.1 Proof of Observation 3.2

Here we provide the complete proof of the information-theoretic upper bound for full-matrix feedback.

Observation 3.2. *Consider a sequence of two-player zero-sum games $\{\mathbf{A}_t\}_{t=1}^T \subseteq [-1,1]^{m\times n}$. Under full-matrix feedback, there exists a learning algorithm such that, when applying to both players, achieves*

$$\frac{\text{Dual-Gap}_{[T]}}{T} \le \frac{2}{T} + 2\sqrt{\frac{V_T}{T}}.$$

In particular, if V_T is sublinear in T, then the average duality gap vanishes.

Proof. At the first round, the players choose arbitrary strategies $\boldsymbol{x}_1 \in \Delta_m$ and $\boldsymbol{y}_1 \in \Delta_n$. For every round $t \ge 2$, after observing $\mathbf{A}_{t-1}$ from the previous round, the players compute a Nash equilibrium $(\boldsymbol{x}_t, \boldsymbol{y}_t)$ of the zero-sum game with payoff matrix $\mathbf{A}_{t-1}$ and play this strategy pair at round t. Specifically, the players choose

$$\boldsymbol{x}_t \in \operatorname*{argmin}_{\boldsymbol{x}\in\Delta_m} \max_{\boldsymbol{y}\in\Delta_n} \boldsymbol{x}^\top \mathbf{A}_{t-1}\boldsymbol{y} \qquad \boldsymbol{y}_t \in \operatorname*{argmax}_{\boldsymbol{y}\in\Delta_n} \min_{\boldsymbol{x}\in\Delta_m} \boldsymbol{x}^\top \mathbf{A}_{t-1}\boldsymbol{y}.$$

We note that the choice of $\boldsymbol{x}_t$ and $\boldsymbol{y}_t$ can be computed in polynomial time through linear programs. Since Nash equilibrium is interchangeable in two-player zero-sum games, $(\boldsymbol{x}_t, \boldsymbol{y}_t)$ is a Nash equilibrium of $\mathbf{A}_{t-1}$. Therefore, it holds that

$$\max_{\boldsymbol{y}\in\Delta_n} \boldsymbol{x}_t^\top \mathbf{A}_{t-1}\boldsymbol{y} = \min_{\boldsymbol{x}\in\Delta_m} \boldsymbol{x}^\top \mathbf{A}_{t-1}\boldsymbol{y}_t.$$

Moreover, since $\mathbf{A}_t \in [-1,1]^{m\times n}$, for any $\boldsymbol{x} \in \Delta_m$ and $\boldsymbol{y} \in \Delta_n$, we have

$$\left|\boldsymbol{x}^\top(\mathbf{A}_t - \mathbf{A}_{t-1})\boldsymbol{y}\right| \le \|\mathbf{A}_t - \mathbf{A}_{t-1}\|_{\max}.$$

Therefore,

$$\max_{\boldsymbol{y}\in\Delta_n} \boldsymbol{x}_t^\top \mathbf{A}_t\boldsymbol{y} \le \max_{\boldsymbol{y}\in\Delta_n} \boldsymbol{x}_t^\top \mathbf{A}_{t-1}\boldsymbol{y} + \|\mathbf{A}_t - \mathbf{A}_{t-1}\|_{\max},$$

and similarly

$$\min_{\boldsymbol{x}\in\Delta_m} \boldsymbol{x}^\top \mathbf{A}_t \boldsymbol{y}_t \geq \min_{\boldsymbol{x}\in\Delta_m} \boldsymbol{x}^\top \mathbf{A}_{t-1} \boldsymbol{y}_t - \|\mathbf{A}_t - \mathbf{A}_{t-1}\|_{\max}.$$

Since $(\boldsymbol{x}_t, \boldsymbol{y}_t)$ is a Nash equilibrium of $\mathbf{A}_{t-1}$, we have $\min_{\boldsymbol{x}\in\Delta_m} \boldsymbol{x}^\top \mathbf{A}_{t-1}\boldsymbol{y}_t = \max_{\boldsymbol{y}\in\Delta_n} \boldsymbol{x}_t^\top \mathbf{A}_{t-1}\boldsymbol{y}$. Thus at round t, the duality gap at round t satisfies

$$\text{Dual-Gap}_t = \max_{\boldsymbol{y}\in\Delta_n} \boldsymbol{x}_t^\top \mathbf{A}_t \boldsymbol{y} - \min_{\boldsymbol{x}\in\Delta_m} \boldsymbol{x}^\top \mathbf{A}_t \boldsymbol{y}_t \leq 2\|\mathbf{A}_t - \mathbf{A}_{t-1}\|_{\max}.$$

Every entry of $\mathbf{A}_1$ lies in $[-1, 1]$, the duality gap for the first round is at most 2. Summing over all rounds gives

$$\text{Dual-Gap}_{[T]} = \sum_{t=1}^{T} \text{Dual-Gap}_t \leq 2 + 2\sum_{t=2}^{T} \|\mathbf{A}_t - \mathbf{A}_{t-1}\|_{\max}.$$

By the Cauchy–Schwarz inequality, we have

$$\sum_{t=2}^{T} \|\mathbf{A}_t - \mathbf{A}_{t-1}\|_{\max} \leq \sqrt{(T-1)\sum_{t=2}^{T} \|\mathbf{A}_t - \mathbf{A}_{t-1}\|_{\max}^2} \leq \sqrt{TV_T}.$$

Dividing by T yields

$$\frac{\text{Dual-Gap}_{[T]}}{T} \leq \frac{2}{T} + \frac{2\sqrt{TV_T}}{T} \leq \frac{2}{T} + 2\sqrt{\frac{V_T}{T}}.$$

This completes the proof. □

B.2 Proof of Theorem 3.3

Next, for the upper bound on vector feedback, we prove Proposition B.1. By setting $V = V_T$, we recover Theorem 3.3.

Proposition B.1. *Consider a sequence of two-player zero-sum games $\{\mathbf{A}_t\}_{t=1}^T \subseteq [-1, 1]^{m\times n}$ with $V_T \leq V$. There exists a learning algorithm (Algorithm 1) such that, when applying to both players, achieves*

$$\frac{\text{Dual-Gap}_{[T]}}{T} \leq O\left(\frac{m+n}{T} + \sqrt{m+n}\left(\frac{V}{T}\right)^{1/4}\right).$$

Consequently, the same bound holds for the average dynamic NE regret.

Proof. We first assume $m + n \leq H \leq T$, the case where $H > T$ or $H < m + n$ will be handled separately. For any block $b \in [B]$, let $I_b = [(b-1)H + 1, \min\{bH, T\}]$ denote the interval of rounds that block b covers and define the variation within the block by

$$V_b := \sum_{t\in I_b} \|\mathbf{A}_t - \mathbf{A}_{t-1}\|_{\max}^2.$$

Clearly, it holds that

$$\sum_{b\in[B]} V_b \leq V_T \leq V.$$

Consider a round $t \in I_b$ after the probing phase. For each column j, let $t_j \in I_b$ be the probing round at which the minimizing player $\boldsymbol{x}$ observes column j. From Algorithm 1, we have

$$\widehat{\mathbf{A}}_b^x(\cdot, j) = \mathbf{A}_{t_j}(\cdot, j).$$

Therefore, for any round $t \in I_b$ after the probing phase, any column $j \in [n]$, and for every entry (i, j),

$$\begin{aligned}
\left|\mathbf{A}_t(i,j) - \widehat{\mathbf{A}}_b^x(i,j)\right| &= \left|\mathbf{A}_t(i,j) - \mathbf{A}_{t_j}(i,j)\right| \\
&\leq \sum_{r=t_j+1}^{t} \|\mathbf{A}_r - \mathbf{A}_{r-1}\|_{\max} \\
&\leq \sqrt{(t-t_j)\sum_{r=t_j+1}^{t} \|\mathbf{A}_r - \mathbf{A}_{r-1}\|_{\max}^2} \\
&\leq \sqrt{HV_b},
\end{aligned}$$

where the second inequality follows from Cauchy–Schwarz. Hence for any round $t \in I_b$ after the probing phase, we have

$$\left\|\mathbf{A}_t - \widehat{\mathbf{A}}_b^x\right\|_{\max} \le \sqrt{HV_b}. \tag{4}$$

For $\widehat{\mathbf{A}}_b^y$, the same argument yields

$$\left\|\mathbf{A}_t - \widehat{\mathbf{A}}_b^y\right\|_{\max} \le \sqrt{HV_b}. \tag{5}$$

We next translate the matrix-estimation error in (4) and (5) into the cumulative duality gap. For any payoff matrix $\mathbf{M}$, denote the zero-sum equilibrium value to be

$$v(\mathbf{M}) = \min_{\boldsymbol{x}\in\Delta_m} \max_{\boldsymbol{y}\in\Delta_n} \boldsymbol{x}^\top \mathbf{M} \boldsymbol{y}.$$

It is straightforward to show that value of a zero-sum game is 1-Lipschitz with respect to the $\|\cdot\|_{\max}$, that is,

$$|v(\mathbf{M}) - v(\mathbf{N})| \le \|\mathbf{M} - \mathbf{N}\|_{\max}.$$

Since $\widehat{\boldsymbol{x}}_b$ is a minimax strategy of $\widehat{\mathbf{A}}_b^x$,

$$\begin{aligned}\max_{\boldsymbol{y}\in\Delta_n} \widehat{\boldsymbol{x}}_b^\top \mathbf{A}_t \boldsymbol{y} &\le \max_{\boldsymbol{y}\in\Delta_n} \widehat{\boldsymbol{x}}_b^\top \widehat{\mathbf{A}}_b^x \boldsymbol{y} + \left\|\mathbf{A}_t - \widehat{\mathbf{A}}_b^x\right\|_{\max} \\ &= v(\widehat{\mathbf{A}}_b^x) + \left\|\mathbf{A}_t - \widehat{\mathbf{A}}_b^x\right\|_{\max}.\end{aligned}$$

Similarly, since $\widehat{\boldsymbol{y}}_b$ is a maximin strategy of $\widehat{\mathbf{A}}_b^y$, we have

$$\begin{aligned}\min_{x\in\Delta_m} \boldsymbol{x}^\top \mathbf{A}_t \widehat{\boldsymbol{y}}_b &\ge \min_{\boldsymbol{x}\in\Delta_m} \boldsymbol{x}^\top \widehat{\mathbf{A}}_b^y \widehat{\boldsymbol{y}}_b - \left\|\mathbf{A}_t - \widehat{\mathbf{A}}_b^y\right\|_{\max} \\ &= v(\widehat{\mathbf{A}}_b^y) - \left\|\mathbf{A}_t - \widehat{\mathbf{A}}_b^y\right\|_{\max}.\end{aligned}$$

Therefore, at any round $t \in I_b$ after the probing phase, it holds that

$$\begin{aligned}\text{Dual-Gap}_t &= \max_{\boldsymbol{y}\in\Delta_n} \widehat{\boldsymbol{x}}_b^\top \mathbf{A}_t \boldsymbol{y} - \min_{\boldsymbol{x}\in\Delta_m} \boldsymbol{x}^\top \mathbf{A}_t \widehat{\boldsymbol{y}}_b \\ &\le v(\widehat{\mathbf{A}}_b^x) - v(\widehat{\mathbf{A}}_b^y) + \left\|\mathbf{A}_t - \widehat{\mathbf{A}}_b^x\right\|_{\max} + \left\|\mathbf{A}_t - \widehat{\mathbf{A}}_b^y\right\|_{\max}.\end{aligned}$$

Moreover, the Lipschitz continuity of $v(\cdot)$ yields

$$\begin{aligned}v(\widehat{\mathbf{A}}_b^x) - v(\widehat{\mathbf{A}}_b^y) &\le \left|v(\widehat{\mathbf{A}}_b^x) - v(\mathbf{A}_t)\right| + \left|v(\mathbf{A}_t) - v(\widehat{\mathbf{A}}_b^y)\right| \\ &\le \left\|\mathbf{A}_t - \widehat{\mathbf{A}}_b^x\right\|_{\max} + \left\|\mathbf{A}_t - \widehat{\mathbf{A}}_b^y\right\|_{\max}.\end{aligned}$$

Combining the above inequalities with (4) and (5) yields

$$\text{Dual-Gap}_t \le 2\left(\left\|\mathbf{A}_t - \widehat{\mathbf{A}}_b^x\right\|_{\max} + \left\|\mathbf{A}_t - \widehat{\mathbf{A}}_b^y\right\|_{\max}\right) \le 4\sqrt{HV_b}.$$

There are at most H rounds in block b, so the cumulative duality gap incurred after probing phase in this block is at most

$$\sum_{t\notin\text{probing phase}} \text{Dual-Gap}_t \le 4H^{\frac{3}{2}}\sqrt{V_b}.$$

Summing over the blocks and applying Cauchy–Schwarz,

$$\begin{aligned}\sum_{b=1}^{B} 4H^{\frac{3}{2}}\sqrt{V_b} &\le 4H^{\frac{3}{2}}\sqrt{B\sum_{b=1}^{B} V_b} &(6)\\ &\le 4H^{\frac{3}{2}}\sqrt{BV} &(7)\\ &\le 8H\sqrt{TV}, &(8)\end{aligned}$$

where the last step follows from $B \le \frac{T}{H} + 1 \le \frac{2T}{H}$.

It remains to bound the cost of probing. Since $\mathbf{A}_t \in [-1,1]^{m\times n}$, the duality gap at any round is at most 2. There are at most $m+n$ probing rounds per block, and therefore the cumulative probing cost is at most

$$2(m+n)B \le 2(m+n)\left(\frac{T}{H}+1\right).$$

Combining with (8) and dividing by T, we have

$$\frac{\text{Dual-Gap}_{[T]}}{T} \le \frac{2(m+n)}{H} + \frac{2(m+n)}{T} + 8\,H\sqrt{\frac{V}{T}}. \tag{9}$$

Plugging in the value of $H = \sqrt{m+n}(T/V)^{1/4}$, we get

$$\frac{\text{Dual-Gap}_{[T]}}{T} \le O\left(\frac{(m+n)}{T} + \sqrt{m+n}\left(\frac{V}{T}\right)^{\frac{1}{4}}\right).$$

For completeness, if $H > T$, we choose $H = T$. In this case $V < (m+n)^2T^{-3}$, and (9) gives

$$\frac{\text{Dual-Gap}_{[T]}}{T} = O\left(\frac{m+n}{T}\right).$$

On the other hand, if $H < m+n$, then

$$\frac{m+n}{H} = \sqrt{m+n}\left(\frac{V}{T}\right)^{\frac{1}{4}} > 1, \tag{10}$$

while the average duality gap is always at most 2, which is trivially bounded by (10). We therefore conclude that in all cases,

$$\frac{\text{Dual-Gap}_{[T]}}{T} \le O\left(\frac{m+n}{T} + \sqrt{m+n}\left(\frac{V}{T}\right)^{\frac{1}{4}}\right).$$

Finally, from Proposition 3.1, the dynamic NE regret is always upper bounded by the duality gap, so the proof is complete. □

B.3 Proof of Theorem 3.6

And lastly for the information-theoretic section, we provide the complete proof of the complementary lower bound.

Theorem 3.6. *For any pair of online learners, there exists a deterministic sequence of zero-sum payoff matrices $\mathbf{A}_1,\dots,\mathbf{A}_T \in [-1,1]^{2\times 2}$ with sublinear $V_T = o(T)$ such that $\text{DynNE-Reg}_{[T]}/T \ge \Omega\left(\sqrt{V_T/T}\right)$. Consequently, the same bound holds for the average duality gap.*

Proof. For an arbitrary pair of online learners, consider a payoff-scaling parameter $\alpha \in (0,1]$ and independent random signs $\sigma_1,\dots,\sigma_T \in \{-1,+1\}$, each uniform on $\{-1,+1\}$, and set the payoff matrix at round t

$$\mathbf{A}_t = \alpha\sigma_t\begin{pmatrix}1 & 1\\ -1 & -1\end{pmatrix}. \tag{11}$$

Because $\alpha \le 1$, every entry of every payoff matrix lies in $[-1,1]$.

At round t, write the minimizing player's mixed strategy as

$$\boldsymbol{x}_t = (p_t, 1-p_t),$$

where $p_t \in [0,1]$ denotes the probability to play the first action. Since the two columns of $\mathbf{A}_t$ are identical, the strategy for player $\boldsymbol{y}$ does not affect the realized payoff. Thus, for every $\boldsymbol{y} \in \Delta_2$,

$$\boldsymbol{x}_t^\top \mathbf{A}_t \boldsymbol{y} = \alpha\sigma_t p_t - \alpha\sigma_t(1-p_t) = \alpha\sigma_t(2p_t-1).$$

Next, consider the minimax value of the game at round t. Regardless of the sign of σ_t, one of the two rows of $\mathbf{A}_t$ achieves payoff $-\alpha$ against every column strategy. Therefore,

$$\min_{\boldsymbol{x}\in\Delta_2}\max_{\boldsymbol{y}\in\Delta_2} \boldsymbol{x}^\top \mathbf{A}_t \boldsymbol{y} = -\alpha.$$

Hence the dynamic regret at round t is

$$\text{DynNE-Reg}_t = \boldsymbol{x}_t^\top \mathbf{A}_t \boldsymbol{y}_t - \min_{\boldsymbol{x}\in\Delta_2} \max_{\boldsymbol{y}\in\Delta_2} \boldsymbol{x}^\top \mathbf{A}_t \boldsymbol{y} = \alpha\,(1 + \sigma_t(2p_t - 1))\,.$$

Since $p_t \in [0,1]$, we have $2p_t - 1 \in [-1,1]$. Thus

$$1 + \sigma_t(2p_t - 1) \geq 0.$$

Therefore we conclude that the dynamic NE regret is nonnegative at every round, so the absolute value in the definition of dynamic NE regret introduces no cancellation. Consequently,

$$\text{DynNE-Reg}_{[T]} = \sum_{t=1}^{T} \alpha\,(1 + \sigma_t(2p_t - 1))\,.$$

We note that the strategies chosen by the learners at round t are independent of the random sign σ_t. Since $\mathbb{E}[\sigma_t] = 0$, taking expectation over the random sign sequence, and summing over all rounds, gives

$$\mathbb{E}\left[\text{DynNE-Reg}_{[T]}\right] = \alpha T. \tag{12}$$

Notice that by setting $\alpha = o(1)$, the payoff variation V_T becomes sublinear.

To retrieve a bound in terms of V_T, we next relate the payoff scale α to the payoff variation. Using the definition of $\mathbf{A}_t$ in (11), we have

$$\mathbf{A}_t - \mathbf{A}_{t-1} = \alpha(\sigma_t - \sigma_{t-1}) \begin{pmatrix} 1 & 1 \\ -1 & -1 \end{pmatrix}.$$

Since $|\sigma_t - \sigma_{t-1}| \leq 2$, we have

$$\|\mathbf{A}_t - \mathbf{A}_{t-1}\|_{\max}^2 \leq 4\alpha^2.$$

Therefore, every realization of the random sequence $\sigma = (\sigma_1, \ldots, \sigma_T)$ satisfies

$$V_T = \sum_{t=2}^{T} \|\mathbf{A}_t - \mathbf{A}_{t-1}\|_{\max}^2 \leq 4\alpha^2(T-1).$$

In order to remove the randomness from our current payoff sequence, we must choose a fixed sign sequence $\boldsymbol{\sigma}^* = (\sigma_1^*, \ldots, \sigma_T^*)$. Since the variation bound holds for every realization of σ, we have

$$V_T(\mathbf{A}^{\boldsymbol{\sigma}^*}) \leq 4\alpha^2(T-1), \tag{13}$$

where $\{\mathbf{A}_t^{\boldsymbol{\sigma}^*}\}_{t=1}^T$ denotes the payoff matrix sequence obtained by using some sequence $\boldsymbol{\sigma}^*$ in (11). Since (12) gives an average expected dynamic regret over the random sign sequence, there exists a fixed sequence $\boldsymbol{\sigma}^*$ such that

$$\text{DynNE-Reg}_{[T]} \geq \alpha T. \tag{14}$$

To recover our final lower bound form, we solve for α in terms of $V_T(\mathbf{A}^{\boldsymbol{\sigma}^*})$ in (13), yielding

$$\alpha \geq \frac{1}{2}\sqrt{\frac{V_T(\mathbf{A}^{\boldsymbol{\sigma}^*})}{T-1}}.$$

And substituting back into our bound (14) gives

$$\text{DynNE-Reg}_{[T]} \geq \alpha T \geq \frac{T}{2}\sqrt{\frac{V_T(\mathbf{A}^{\boldsymbol{\sigma}^*})}{T-1}} = \Omega\left(\sqrt{T V_T(\mathbf{A}^{\boldsymbol{\sigma}^*})}\right).$$

Thus, for every pair of online learners, there exists a deterministic payoff sequence with realized variation V_T for which the learners must incur average dynamic NE regret of order $\Omega(\sqrt{V_T/T})$.

Finally, by Proposition 3.1, the duality gap is always lower bounded by the dynamic NE regret, so the proof is complete. □

Although random signs are used in the argument, the averaging step fixes a deterministic sequence $\boldsymbol{\sigma}^*$ independent of the learners' realized actions, so the final hard sequence can be chosen before play begins.

C Omitted Proofs of Section 4

Team game deviation gaps Recall the 2 vs. 2 team zero-sum game and joint strategy profile $\boldsymbol{z} = (\boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{y}_1, \boldsymbol{y}_2)$ introduced in Section 4.1. We now define the unilateral deviation gaps used in the analysis of this section. For $i \in \{1, 2\}$, let

$$g_{\boldsymbol{x}_i}(\boldsymbol{z}) := f(\boldsymbol{z}) - \min_{\boldsymbol{x}_i' \in \Delta(\mathcal{A}_i)} f(\boldsymbol{x}_i', \boldsymbol{x}_{-i}, \boldsymbol{y}_1, \boldsymbol{y}_2), g_{\boldsymbol{y}_i}(\boldsymbol{z}) := \max_{\boldsymbol{y}_i' \in \Delta(\mathcal{B}_i)} f(\boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{y}_i', \boldsymbol{y}_{-i}) - f(\boldsymbol{z}),$$

where $\boldsymbol{x}_{-i}$ and $\boldsymbol{y}_{-i}$ denote the other player's strategy within the corresponding team.

Observe that each of these quantities is nonnegative. Moreover, we define

$$\text{Nash-Gap}_{\max}(\boldsymbol{z}) := \max\{g_{\boldsymbol{x}_1}(\boldsymbol{z}), g_{\boldsymbol{x}_2}(\boldsymbol{z}), g_{\boldsymbol{y}_1}(\boldsymbol{z}), g_{\boldsymbol{y}_2}(\boldsymbol{z})\},$$

to be the maximum unilateral deviation improvement among the four players at the strategy profile $\boldsymbol{z}$. Denote

$$\text{Nash-Gap}_{\Sigma}(\boldsymbol{z}) := g_{\boldsymbol{x}_1}(\boldsymbol{z}) + g_{\boldsymbol{x}_2}(\boldsymbol{z}) + g_{\boldsymbol{y}_1}(\boldsymbol{z}) + g_{\boldsymbol{y}_2}(\boldsymbol{z}).$$

From the definition of Nash equilibrium in Definition 2.1, a strategy profile $\boldsymbol{z}$ is an ϵ-approximate Nash equilibrium if and only if $\text{Nash-Gap}_{\max}(\boldsymbol{z}) \le \epsilon$.

C.1 DynNE-Reg and Nash-Gap recovery

To relate the DynNE-Reg quantities of our four-game reduction in Section 4.1 to the Nash-Gap of a team zero-sum game, we first consider the game $\mathbf{C}_t^{\boldsymbol{x}_1}$ associated with $\boldsymbol{x}_1$.

Lemma C.1. *For any column strategy $\tilde{\boldsymbol{y}}_1 \in \Delta_{n_1}$,*

$$\boldsymbol{x}_{1,t}^\top \mathbf{C}_t^{\boldsymbol{x}_1} \tilde{\boldsymbol{y}}_1 - \min_{\boldsymbol{x}_1'} \max_{\boldsymbol{y}_1'} \boldsymbol{x}_1'^\top \mathbf{C}_t^{\boldsymbol{x}_1} \boldsymbol{y}_1' = g_{\boldsymbol{x}_1}(\boldsymbol{z}_t).$$

Consequently, the cumulative dynamic NE regret of strategy sequence $\{(\boldsymbol{x}_{1,t}, \tilde{\boldsymbol{y}}_1)\}_{t=1}^T$ with respect to the game sequence $\{\mathbf{C}_t^{\boldsymbol{x}_1}\}_{t=1}^T$ is

$$\text{DynNE-Reg}_{[T]}^{\boldsymbol{x}_1} = \sum_{t=1}^T g_{\boldsymbol{x}_1}(\boldsymbol{z}_t).$$

Proof. Every column of $\mathbf{C}_t^{\boldsymbol{x}_1}$ is equal to the vector $\mathbf{A}_t^{(1)} \boldsymbol{y}_{1,t}$. Since $\mathbf{1}_{n_1}^\top \tilde{\boldsymbol{y}}_1 = 1$,

$$\begin{aligned} \boldsymbol{x}_{1,t}^\top \mathbf{C}_t^{\boldsymbol{x}_1} \tilde{\boldsymbol{y}}_1 &= \boldsymbol{x}_{1,t}^\top \left(\mathbf{A}_t^{(1)} \boldsymbol{y}_{1,t} \right) \mathbf{1}_{n_1}^\top \tilde{\boldsymbol{y}}_1 \\ &= \boldsymbol{x}_{1,t}^\top \mathbf{A}_t^{(1)} \boldsymbol{y}_{1,t} \\ &= f(\boldsymbol{z}_t). \end{aligned}$$

Likewise, for every $\boldsymbol{x}_1' \in \Delta_{m_1}$ and $\boldsymbol{y}_1' \in \Delta_{n_1}$,

$$\boldsymbol{x}_1'^\top \mathbf{C}_t^{\boldsymbol{x}_1} \boldsymbol{y}_1' = \boldsymbol{x}_1'^\top \mathbf{A}_t^{(1)} \boldsymbol{y}_{1,t},$$

so the maximizing player's choice has no effect. Hence

$$\begin{aligned} \min_{\boldsymbol{x}_1'} \max_{\boldsymbol{y}_1'} \boldsymbol{x}_1'^\top \mathbf{C}_t^{\boldsymbol{x}_1} \boldsymbol{y}_1' &= \min_{\boldsymbol{x}_1'} \boldsymbol{x}_1'^\top \mathbf{A}_t^{(1)} \boldsymbol{y}_{1,t} \\ &= \min_{\boldsymbol{x}_1'} f(\boldsymbol{x}_1', \boldsymbol{x}_{2,t}, \boldsymbol{y}_{1,t}, \boldsymbol{y}_{2,t}). \end{aligned}$$

Subtracting gives

$$f(\boldsymbol{z}_t) - \min_{\boldsymbol{x}_1'} f(\boldsymbol{x}_1', \boldsymbol{x}_{2,t}, \boldsymbol{y}_{1,t}, \boldsymbol{y}_{2,t}) = g_{\boldsymbol{x}_1}(\boldsymbol{z}_t),$$

which proves the first statement of the lemma. Since $g_{\boldsymbol{x}_1}(\boldsymbol{z}_t) \ge 0$ for every t, the absolute value in (2) introduces no cancellation, and therefore

$$\text{DynNE-Reg}_{[T]}^{\boldsymbol{x}_1} = \left| \sum_{t=1}^T g_{\boldsymbol{x}_1}(\boldsymbol{z}_t) \right| = \sum_{t=1}^T g_{\boldsymbol{x}_1}(\boldsymbol{z}_t).$$

□

Now for the game $\mathbf{C}_t^{\boldsymbol{y}_1}$ associated with player $\boldsymbol{y}_1$, we have the analogous lemma.

Lemma C.2. *For any row strategy $\tilde{\boldsymbol{x}}_1 \in \Delta_{m_1}$,*

$$\tilde{\boldsymbol{x}}_1^\top \mathbf{C}_t^{\boldsymbol{y}_1} \boldsymbol{y}_{1,t} - \min_{\boldsymbol{x}_1'} \max_{\boldsymbol{y}_1'} \boldsymbol{x}_1'^\top \mathbf{C}_t^{\boldsymbol{y}_1} \boldsymbol{y}_1' = -g_{\boldsymbol{y}_1}(\boldsymbol{z}_t).$$

Consequently, the cumulative dynamic NE regret of strategy sequence $\{(\tilde{\boldsymbol{x}}_1, \boldsymbol{y}_{1,t})\}_{t=1}^T$ with respect to the game sequence $\{\mathbf{C}_t^{\boldsymbol{y}_1}\}_{t=1}^T$ is

$$\text{DynNE-Reg}_{[T]}^{\boldsymbol{y}_1} = \sum_{t=1}^T g_{\boldsymbol{y}_1}(\boldsymbol{z}_t).$$

Proof. Every row of $\mathbf{C}_t^{\boldsymbol{y}_1}$ is identical. Since $\tilde{\boldsymbol{x}}_1^\top \mathbf{1}_{m_1} = 1$,

$$\tilde{\boldsymbol{x}}_1^\top \mathbf{C}_t^{\boldsymbol{y}_1} \boldsymbol{y}_{1,t} = \boldsymbol{x}_{1,t}^\top \mathbf{A}_t^{(1)} \boldsymbol{y}_{1,t} = f(\boldsymbol{z}_t).$$

Moreover, the minimizing player's choice has no effect, so

$$\begin{aligned}\min_{\boldsymbol{x}_1'} \max_{\boldsymbol{y}_1'} \boldsymbol{x}_1'^\top \mathbf{C}_t^{\boldsymbol{y}_1} \boldsymbol{y}_1' &= \max_{\boldsymbol{y}_1'} \boldsymbol{x}_{1,t}^\top \mathbf{A}_t^{(1)} \boldsymbol{y}_1' \\ &= \max_{\boldsymbol{y}_1'} f(\boldsymbol{x}_{1,t}, \boldsymbol{x}_{2,t}, \boldsymbol{y}_1', \boldsymbol{y}_{2,t}).\end{aligned}$$

Thus the per-round residual is

$$f(\boldsymbol{z}_t) - \max_{\boldsymbol{y}_1'} f(\boldsymbol{x}_{1,t}, \boldsymbol{x}_{2,t}, \boldsymbol{y}_1', \boldsymbol{y}_{2,t}) = -g_{\boldsymbol{y}_1}(\boldsymbol{z}_t).$$

Every such residual is nonpositive. Hence the outer absolute value in dynamic NE regret gives

$$\text{DynNE-Reg}_{[T]}^{\boldsymbol{y}_1} = \left| -\sum_{t=1}^T g_{\boldsymbol{y}_1}(\boldsymbol{z}_t) \right| = \sum_{t=1}^T g_{\boldsymbol{y}_1}(\boldsymbol{z}_t).$$

□

And for the games $\mathbf{C}_t^{\boldsymbol{x}_2}$ and $\mathbf{C}_t^{\boldsymbol{y}_2}$, we utilize the previous lemmas.

Lemma C.3. *For the game sequences $\{\mathbf{C}_t^{\boldsymbol{x}_2}\}_{t=1}^T$ and $\{\mathbf{C}_t^{\boldsymbol{y}_2}\}_{t=1}^T$, the dynamic NE regret of the respective strategy sequences $\{(\boldsymbol{x}_{2,t}, \tilde{\boldsymbol{y}}_2)\}_{t=1}^T$ and $\{(\tilde{\boldsymbol{x}}_2, \boldsymbol{y}_{2,t})\}_{t=1}^T$ satisfy*

$$\text{DynNE-Reg}_{[T]}^{\boldsymbol{x}_2} = \sum_{t=1}^T g_{\boldsymbol{x}_2}(\boldsymbol{z}_t), \qquad \text{DynNE-Reg}_{[T]}^{\boldsymbol{y}_2} = \sum_{t=1}^T g_{\boldsymbol{y}_2}(\boldsymbol{z}_t),$$

where $\tilde{\boldsymbol{y}}_2 \in \Delta_{n_2}$ is any column strategy, and $\tilde{\boldsymbol{x}}_2 \in \Delta_{m_2}$ is any row strategy.

Proof. The proof for $\text{DynNE-Reg}_{[T]}^{\boldsymbol{x}_2}$ is analogous to Lemma C.1 and $\text{DynNE-Reg}_{[T]}^{\boldsymbol{y}_2}$ is analogous to Lemma C.2, both using $\boldsymbol{x}_{2,t}^\top \mathbf{A}_t^{(2)} \boldsymbol{y}_{2,t} = f(\boldsymbol{z}_t)$. □

Having shown that each of the four constructed game sequences exactly recovers one player's cumulative unilateral deviation gap, we can now combine these four identities to recover the cumulative Nash gap of the original team zero-sum game.

Proposition C.4. *For the four game sequences $\{\mathbf{C}_t^{\boldsymbol{x}_1}\}_{t=1}^T, \{\mathbf{C}_t^{\boldsymbol{x}_2}\}_{t=1}^T, \{\mathbf{C}_t^{\boldsymbol{y}_1}\}_{t=1}^T$, and $\{\mathbf{C}_t^{\boldsymbol{y}_2}\}_{t=1}^T$, and the respective strategy sequences $\{(\boldsymbol{x}_{1,t}, \tilde{\boldsymbol{y}}_1)\}_{t=1}^T, \{(\boldsymbol{x}_{2,t}, \tilde{\boldsymbol{y}}_2)\}_{t=1}^T, \{(\tilde{\boldsymbol{x}}_1, \boldsymbol{y}_{1,t})\}_{t=1}^T$, and $\{(\tilde{\boldsymbol{x}}_2, \boldsymbol{y}_{2,t})\}_{t=1}^T$, we have*

$$\text{DynNE-Reg}_{[T]}^{\boldsymbol{x}_1} + \text{DynNE-Reg}_{[T]}^{\boldsymbol{x}_2} + \text{DynNE-Reg}_{[T]}^{\boldsymbol{y}_1} + \text{DynNE-Reg}_{[T]}^{\boldsymbol{y}_2} = \sum_{t=1}^T \text{Nash-Gap}_\Sigma(\boldsymbol{z}_t)$$

where the strategies $(\tilde{\boldsymbol{x}}_1, \tilde{\boldsymbol{x}}_2, \tilde{\boldsymbol{y}}_1, \tilde{\boldsymbol{y}}_2) \in \Delta_{m_1} \times \Delta_{m_2} \times \Delta_{n_1} \times \Delta_{n_2}$ are arbitrary.

Proof. By Lemma C.1, Lemma C.2, and Lemma C.3, we have

$$\begin{aligned}
&\text{DynNE-Reg}^{\boldsymbol{x}_1}_{[T]} + \text{DynNE-Reg}^{\boldsymbol{y}_1}_{[T]} + \text{DynNE-Reg}^{\boldsymbol{x}_2}_{[T]} + \text{DynNE-Reg}^{\boldsymbol{y}_2}_{[T]} \\
&= \sum_{t=1}^{T} \left(g_{\boldsymbol{x}_1}(\boldsymbol{z}_t) + g_{\boldsymbol{y}_1}(\boldsymbol{z}_t) + g_{\boldsymbol{x}_2}(\boldsymbol{z}_t) + g_{\boldsymbol{y}_2}(\boldsymbol{z}_t) \right) \\
&= \sum_{t=1}^{T} \text{Nash-Gap}_{\Sigma}(\boldsymbol{z}_t).
\end{aligned}$$

□

C.2 Stability to variation bound

To prove slow variation of the games in our reduction, we first show how the induced games $\mathbf{A}^{(1)}(\boldsymbol{x}_2, \boldsymbol{y}_2)$ and $\mathbf{A}^{(2)}(\boldsymbol{x}_1, \boldsymbol{y}_1)$ are controlled by the strategies that define them.

Lemma C.5. *For any strategy pairs $(\boldsymbol{x}_2, \boldsymbol{y}_2)$ and $(\boldsymbol{x}_2', \boldsymbol{y}_2')$, it holds that*

$$\left\| \mathbf{A}^{(1)}(\boldsymbol{x}_2, \boldsymbol{y}_2) - \mathbf{A}^{(1)}(\boldsymbol{x}_2', \boldsymbol{y}_2') \right\|_{\max} \leq \|\boldsymbol{x}_2 - \boldsymbol{x}_2'\|_1 + \|\boldsymbol{y}_2 - \boldsymbol{y}_2'\|_1 .$$

Similarly, for any strategy pairs $(\boldsymbol{x}_1, \boldsymbol{y}_1)$ and $(\boldsymbol{x}_1', \boldsymbol{y}_1')$, we have

$$\left\| \mathbf{A}^{(2)}(\boldsymbol{x}_1, \boldsymbol{y}_1) - \mathbf{A}^{(2)}(\boldsymbol{x}_1', \boldsymbol{y}_1') \right\|_{\max} \leq \|\boldsymbol{x}_1 - \boldsymbol{x}_1'\|_1 + \|\boldsymbol{y}_1 - \boldsymbol{y}_1'\|_1 .$$

Proof. For arbitrary actions $a_1 \in \mathcal{A}_1$ and $b_1 \in \mathcal{B}_1$, the corresponding entry of the induced payoff matrix is given by

$$\mathbf{A}^{(1)}(\boldsymbol{x}_2, \boldsymbol{y}_2)_{a_1, b_1} = \sum_{a_2 \in \mathcal{A}_2} \sum_{b_2 \in \mathcal{B}_2} \boldsymbol{x}_2(a_2) \boldsymbol{y}_2(b_2) \mathbf{U}_{a_1, a_2, b_1, b_2}.$$

Since the team zero-sum game payoffs are normalized such that $\|\mathbf{U}\|_{\max} \leq 1$, for any two strategy pairs $(\boldsymbol{x}_2, \boldsymbol{y}_2)$ and $(\boldsymbol{x}_2', \boldsymbol{y}_2')$,

$$\begin{aligned}
&\left| \mathbf{A}^{(1)}(\boldsymbol{x}_2, \boldsymbol{y}_2)_{a_1, b_1} - \mathbf{A}^{(1)}(\boldsymbol{x}_2', \boldsymbol{y}_2')_{a_1, b_1} \right| \\
&\leq \sum_{a_2, b_2} \left| \boldsymbol{x}_2(a_2) \boldsymbol{y}_2(b_2) - \boldsymbol{x}_2'(a_2) \boldsymbol{y}_2'(b_2) \right| \\
&= \sum_{a_2, b_2} \left| \big(\boldsymbol{x}_2(a_2) - \boldsymbol{x}_2'(a_2)\big) \boldsymbol{y}_2(b_2) + \boldsymbol{x}_2'(a_2) \big(\boldsymbol{y}_2(b_2) - \boldsymbol{y}_2'(b_2)\big) \right| \\
&\leq \sum_{a_2, b_2} \left| \boldsymbol{x}_2(a_2) - \boldsymbol{x}_2'(a_2) \right| \boldsymbol{y}_2(b_2) + \sum_{a_2, b_2} \boldsymbol{x}_2'(a_2) \left| \boldsymbol{y}_2(b_2) - \boldsymbol{y}_2'(b_2) \right| \\
&= \left(\sum_{a_2} \left| \boldsymbol{x}_2(a_2) - \boldsymbol{x}_2'(a_2) \right| \right) \left(\sum_{b_2} \boldsymbol{y}_2(b_2) \right) + \left(\sum_{a_2} \boldsymbol{x}_2'(a_2) \right) \left(\sum_{b_2} \left| \boldsymbol{y}_2(b_2) - \boldsymbol{y}_2'(b_2) \right| \right) \\
&\leq \|\boldsymbol{x}_2 - \boldsymbol{x}_2'\|_1 + \|\boldsymbol{y}_2 - \boldsymbol{y}_2'\|_1 .
\end{aligned}$$

The final inequality follows from $\boldsymbol{x}_2' \in \Delta$ and $\boldsymbol{y}_2 \in \Delta$. Since this bound holds entrywise, we have

$$\left\| \mathbf{A}^{(1)}(\boldsymbol{x}_2, \boldsymbol{y}_2) - \mathbf{A}^{(1)}(\boldsymbol{x}_2', \boldsymbol{y}_2') \right\|_{\max} \leq \|\boldsymbol{x}_2 - \boldsymbol{x}_2'\|_1 + \|\boldsymbol{y}_2 - \boldsymbol{y}_2'\|_1 .$$

The same argument applies to $\mathbf{A}^{(2)}$. The proof is complete. □

Using strongly stable strategies together with Lemma C.5, we can prove the bound on strongly sublinear matrix variation for all four game sequences in our reduction.

Proposition C.6. *For the four-game construction, suppose the four strategy sequences $\{\boldsymbol{x}_{1,t}\}_{t=1}^T, \{\boldsymbol{x}_{2,t}\}_{t=1}^T, \{\boldsymbol{y}_{1,t}\}_{t=1}^T$, and $\{\boldsymbol{y}_{2,t}\}_{t=1}^T$ are strongly stable, i.e.*

$$S_T(\boldsymbol{x}_1),\ S_T(\boldsymbol{x}_2),\ S_T(\boldsymbol{y}_1),\ S_T(\boldsymbol{y}_2) = O(T^{1-\delta})$$

for some constant $\delta > 0$. Then each of the four payoff matrix sequences satisfies strongly sublinear variation:

$$V_T(\mathbf{C}^{\boldsymbol{x}_1}), V_T(\mathbf{C}^{\boldsymbol{y}_1}), V_T(\mathbf{C}^{\boldsymbol{x}_2}), V_T(\mathbf{C}^{\boldsymbol{y}_2}) = O(T^{1-\delta}).$$

Proof. We first consider the game associated with player $\boldsymbol{x}_1$. Since every column of $\mathbf{C}_t^{\boldsymbol{x}_1}$ is identical,

$$\begin{aligned}\|\mathbf{C}_t^{\boldsymbol{x}_1} - \mathbf{C}_{t-1}^{\boldsymbol{x}_1}\|_{\max} &= \left\|\mathbf{A}_t^{(1)}\boldsymbol{y}_{1,t} - \mathbf{A}_{t-1}^{(1)}\boldsymbol{y}_{1,t-1}\right\|_\infty \\ &\le \left\|(\mathbf{A}_t^{(1)} - \mathbf{A}_{t-1}^{(1)})\boldsymbol{y}_{1,t}\right\|_\infty + \left\|\mathbf{A}_{t-1}^{(1)}(\boldsymbol{y}_{1,t} - \boldsymbol{y}_{1,t-1})\right\|_\infty .\end{aligned}$$

Because $\boldsymbol{y}_{1,t}$ is a probability distribution and the induced matrices have entries bounded in absolute value by 1, this gives

$$\|\mathbf{C}_t^{\boldsymbol{x}_1} - \mathbf{C}_{t-1}^{\boldsymbol{x}_1}\|_{\max} \le \left\|\mathbf{A}_t^{(1)} - \mathbf{A}_{t-1}^{(1)}\right\|_{\max} + \|\boldsymbol{y}_{1,t} - \boldsymbol{y}_{1,t-1}\|_1 .$$

Applying Lemma C.5 to $\mathbf{A}^{(1)}$ yields

$$\|\mathbf{C}_t^{\boldsymbol{x}_1} - \mathbf{C}_{t-1}^{\boldsymbol{x}_1}\|_{\max} \le \|\boldsymbol{x}_{2,t} - \boldsymbol{x}_{2,t-1}\|_1 + \|\boldsymbol{y}_{2,t} - \boldsymbol{y}_{2,t-1}\|_1 + \|\boldsymbol{y}_{1,t} - \boldsymbol{y}_{1,t-1}\|_1 .$$

Each ℓ_1 distance between two probability distributions is at most 2, so the quantity on the right-hand side is at most 6. Therefore,

$$\begin{aligned}V_T(\mathbf{C}^{\boldsymbol{x}_1}) &\le \sum_{t=2}^{T}\Big(\|\boldsymbol{x}_{2,t} - \boldsymbol{x}_{2,t-1}\|_1 + \|\boldsymbol{y}_{2,t} - \boldsymbol{y}_{2,t-1}\|_1 + \|\boldsymbol{y}_{1,t} - \boldsymbol{y}_{1,t-1}\|_1\Big)^2 \\ &\le 6\sum_{t=2}^{T}\Big(\|\boldsymbol{x}_{2,t} - \boldsymbol{x}_{2,t-1}\|_1 + \|\boldsymbol{y}_{2,t} - \boldsymbol{y}_{2,t-1}\|_1 + \|\boldsymbol{y}_{1,t} - \boldsymbol{y}_{1,t-1}\|_1\Big) \\ &= 6\Big(S_T(\boldsymbol{x}_2) + S_T(\boldsymbol{y}_2) + S_T(\boldsymbol{y}_1)\Big) \\ &= O(T^{1-\delta}).\end{aligned}$$

The remaining three games satisfy analogous bounds. In particular,

$$\begin{aligned}V_T(\mathbf{C}^{\boldsymbol{y}_1}) &\le 6\Big(S_T(\boldsymbol{x}_2) + S_T(\boldsymbol{y}_2) + S_T(\boldsymbol{x}_1)\Big), \\ V_T(\mathbf{C}^{\boldsymbol{x}_2}) &\le 6\Big(S_T(\boldsymbol{x}_1) + S_T(\boldsymbol{y}_1) + S_T(\boldsymbol{y}_2)\Big), \\ V_T(\mathbf{C}^{\boldsymbol{y}_2}) &\le 6\Big(S_T(\boldsymbol{x}_1) + S_T(\boldsymbol{y}_1) + S_T(\boldsymbol{x}_2)\Big).\end{aligned}$$

Since all four strategy sequences are strongly stable, each right-hand side is $O(T^{1-\delta})$. Hence all four payoff sequences have strongly sublinear matrix variation.

Notice that the arbitrary strategies do not appear in these bounds. As established in the four-game reduction, they serve only to complete valid strategy profiles in the games and have no effect on the corresponding payoff matrices. □

C.3 Proof of Theorem 4.3

We now combine the four-game reduction of Section 4.1 which follows the four-gap identity of Proposition C.4, and the stability-to-variation bound from Proposition C.6, to obtain the corresponding computational hardness result for dynamic NE regret.

Theorem 4.3. *Consider any time-varying two-player zero-sum payoff sequence with strongly sublinear second-order payoff path-length variation. There exists an inverse-polynomial accuracy $\epsilon = \mathsf{poly}(1/n)$ such that it is* PPAD*-hard to design an online learner that simultaneously satisfies the following two properties: i) $S_T(\boldsymbol{x}) = O(T^{1-\delta})$ with some constant $\delta > 0$ for every strategy sequence $\{\boldsymbol{x}_t\}_{t=1}^T$ generated by the learner; and ii) $\mathrm{DynNE\text{-}Reg}_{[T]}/T \le \epsilon$. Consequently, the same hardness result applies to the average duality gap.*

Proof. Let G be a 2 vs. 2 team zero-sum game of input size n. By Proposition 4.1, there exists an inverse-polynomial accuracy parameter $\epsilon_G = 1/\mathsf{poly}(n)$ such that computing an ϵ_G-approximate Nash equilibrium of G is PPAD-hard.

Following the four-game reduction in Section 4.1, we run one time-varying two-player zero-sum process for each of the four players $\boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{y}_1, \boldsymbol{y}_2$. Under the assumed stability guarantee, Proposition C.6 implies that each of the four resulting payoff sequences has strongly sublinear variation,

$$V_T(\mathbf{C}^{\boldsymbol{x}_1}), V_T(\mathbf{C}^{\boldsymbol{y}_1}), V_T(\mathbf{C}^{\boldsymbol{x}_2}), V_T(\mathbf{C}^{\boldsymbol{y}_2}) = O(T^{1-\delta})$$

for some constant $\delta > 0$. Hence the assumed dynamic NE regret guarantee applies to every copy, giving

$$\text{DynNE-Reg}^{v}_{[T]} \le \epsilon T \qquad \text{for each } v \in \{\boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{y}_1, \boldsymbol{y}_2\}.$$

By the four-game identity established in Proposition C.4,

$$\sum_{t=1}^{T} \text{Nash-Gap}_{\Sigma}(\boldsymbol{z}_t) = \text{DynNE-Reg}^{\boldsymbol{x}_1}_{[T]} + \text{DynNE-Reg}^{\boldsymbol{x}_2}_{[T]} + \text{DynNE-Reg}^{\boldsymbol{y}_1}_{[T]} + \text{DynNE-Reg}^{\boldsymbol{y}_2}_{[T]}.$$

Therefore,

$$\sum_{t=1}^{T} \text{Nash-Gap}_{\Sigma}(\boldsymbol{z}_t) \le 4\epsilon T.$$

Dividing by T yields

$$\frac{1}{T}\sum_{t=1}^{T} \text{Nash-Gap}_{\Sigma}(\boldsymbol{z}_t) \le 4\epsilon.$$

Since every Nash gap is nonnegative, there must exist some round $t^* \in [T]$ such that

$$\text{Nash-Gap}_{\Sigma}(\boldsymbol{z}_{t^*}) \le 4\epsilon.$$

Choose $\epsilon \le \epsilon_G/4$. Because $\epsilon_G = 1/\text{poly}(n)$, this choice of ϵ is also inverse-polynomial in n. It follows that

$$\text{Nash-Gap}_{\Sigma}(\boldsymbol{z}_{t^*}) \le \epsilon_G.$$

Since

$$\text{Nash-Gap}_{\max}(\boldsymbol{z}) \le \text{Nash-Gap}_{\Sigma}(\boldsymbol{z}),$$

the profile $\boldsymbol{z}_{t^*}$ is an ϵ_G-approximate Nash equilibrium of G.

At this point we know such a round t^* exists; in order to recover one algorithmically, we can evaluate $\text{Nash-Gap}_{\Sigma}(\boldsymbol{z}_t)$ for all $t \in [T]$ and simply output the minimizing strategy profile. For a polynomially bounded horizon $T = \text{poly}(n)$, this requires only polynomially many evaluations. Each of the four payoff matrices is computable in polynomial time, and each round of the assumed deterministic learner also runs in polynomial time. Thus the complete four-game construction and the extraction of $\boldsymbol{z}_{t^*}$ run in polynomial time. From Proposition 4.1, we conclude that the dynamic NE regret learning problem is PPAD-hard.

Finally, from Proposition 3.1, the duality gap is always lower bounded by the dynamic NE regret, so the duality gap learning problem is also PPAD-hard. □

D Additional Experiment Details

This appendix provides additional details omitted from Section 5. We first emphasize that these experiments are intended only as an illustration of the phenomenon captured by our theoretical results. They do not establish the general lower bound, since our hardness results apply to the full class of strongly stable algorithms rather than only the particular learning dynamics considered here. Instead, the goal is to show that several commonly used no-regret algorithms can have strongly sublinear matrix variation and strongly sublinear strategy movement while their equilibrium-tracking error remains bounded away from zero.

We begin by verifying the strong-stability claim for the four learning dynamics, then report results for a second initialization and for horizons beyond the one shown in the main text.

D.1 Stability of the Learning Dynamics

Recall that all four algorithms use the horizon-dependent learning rate $\eta_T = T^{-1/2}$. Because each payoff matrix lies in $[-1, 1]^{2\times 2}$ and each player chooses a mixed strategy in Δ_2, all feedback vectors are uniformly bounded.

For PGD, nonexpansiveness of Euclidean projection gives $\|\boldsymbol{x}_{t+1} - \boldsymbol{x}_t\|_2 \le \eta_T \|\mathbf{A}_t \boldsymbol{y}_t\|_2 = O(\eta_T)$. Since the dimension is fixed, this also implies $\|\boldsymbol{x}_{t+1} - \boldsymbol{x}_t\|_1 = O(\eta_T)$. The same argument applies

to OGDA, since its optimistic gradient $2\mathbf{A}_t\boldsymbol{y}_t - \mathbf{A}_{t-1}\boldsymbol{y}_{t-1}$ remains uniformly bounded whenever the individual feedback vectors are bounded.

For entropic FTRL, strong convexity of the negative-entropy regularizer implies that a bounded one-round change in the cumulative score changes the resulting mixed strategy by $O(\eta_T)$ in ℓ_1. OMWU is the corresponding optimistic entropic update, and its score changes by a bounded combination of consecutive feedback vectors, yielding the same order of one-step movement.

Thus, for each of the four learning dynamics, $\|\boldsymbol{x}_{t+1} - \boldsymbol{x}_t\|_1 = O(\eta_T)$, and $\|\boldsymbol{y}_{t+1} - \boldsymbol{y}_t\|_1 = O(\eta_T)$. Summing over the horizon gives

$$S_T(\boldsymbol{x}) = O(\eta_T T) = O(\sqrt{T}), \qquad S_T(\boldsymbol{y}) = O(\sqrt{T}).$$

Hence all four algorithms satisfy the strong-stability condition with $\delta = 1/2$ under the tuning used in the experiment.

D.2 Additional Initialization

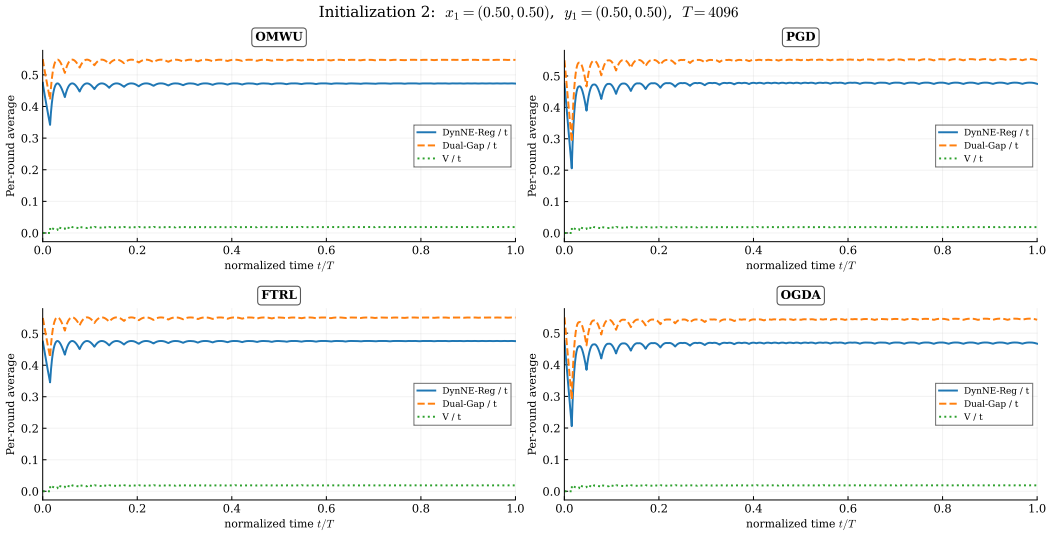


Figure 2: Results for the symmetric initialization $\boldsymbol{x}_1 = \boldsymbol{y}_1 = (0.50, 0.50)$ at horizon $T = 4096$.

The main text reports results for the asymmetric initialization $\boldsymbol{x}_1 = (0.72, 0.28)$, and $\boldsymbol{y}_1 = (0.31, 0.69)$. To check that the observed behavior is not specific to this starting point, we repeat the experiment from the symmetric initialization $\boldsymbol{x} = \boldsymbol{y} = (0.50, 0.50)$, and view the results in Figure 2. The horizon, payoff sequence, learning rate, and switching period are otherwise unchanged: $T = 4096, \eta_T = T^{-1/2}$, and $B_T = \lceil\sqrt{T}\rceil$.

Across both initializations and all four learning dynamics, V_t/t becomes small while the average duality gap and average dynamic NE regret remain bounded away from zero over the plotted horizon. Thus the behavior shown in Figure 1 is not specific to the choice of initial strategies.

D.3 Scaling Beyond the Plotted Horizon

The figures use the horizon $T = 4096 = 2^{12}$. To examine whether the same behavior persists for larger horizons, we repeat the experiment at $T \in \{2^{13}, 2^{14}, 2^{15}\}$. At each horizon, we retune the learning rate and switching period according to $\eta_T = T^{-1/2}$, and $B_T = \lceil\sqrt{T}\rceil$.

Table 1 reports $\mathrm{DynNE\text{-}Reg}_{[T]}/T$, averaged over the two initializations. The payoff variation is common to all four algorithms.

As the horizon increases from 2^{13} to 2^{15}, V_T/T decreases from approximately 0.0133 to 0.00665, consistent with $V_T/T = O(T^{-1/2})$. In contrast, the average dynamic NE regret remains nearly un-

Table 1: Scaling beyond the horizon shown in the figures.

T	V_T/T	OMWU	PGD	FTRL	OGDA
2^{13}	0.013293	0.473789	0.474873	0.476207	0.468941
2^{14}	0.009379	0.472921	0.475911	0.474711	0.471676
2^{15}	0.006647	0.473191	0.474642	0.474428	0.471651

changed: all reported values lie between approximately 0.469 and 0.476. Thus the behavior observed in the figures continues beyond the plotted horizon rather than appearing only at $T = 2^{12}$.