Convergence of No-Swap-Regret Dynamics in Self-Play

Renato Paes Leme
Google Research

Georgios Piliouras
Google Deepmind

Jon Schneider
Google Research

Abstract

In this paper, we investigate the question of whether no-swap-regret dynamics have stronger convergence properties in repeated games than regular no-external-regret dynamics. We prove that in almost all *symmetric* zero-sum games under *symmetric* initializations of the agents, *no-swap-regret* dynamics in self-play are guaranteed to converge in a strong "frequent-iterate" sense to the Nash equilibrium: in all but a vanishing fraction of the rounds, the players must play a strategy profile close to a symmetric Nash equilibrium. Remarkably, relaxing any of these three constraints, i.e. by allowing either i) asymmetric initial conditions, or ii) an asymmetric game or iii) no-external regret dynamics suffices to destroy this result and lead to complex non-equilibrating or even chaotic behavior.

In a dual type of result, we show that the power of no-swap-regret dynamics comes at a cost of imposing a time-asymmetry on its inputs. While no-external-regret dynamics can be completely determined by the cumulative reward vector received by each player, we show there does not exist any general no-swap-regret dynamics defined on the same state space. In fact, we prove that any no-swap-regret learning algorithm must play a time-asymmetric function over the set of previously observed rewards, ruling out any dynamics based on a symmetric function of the current set of rewards.

1 Introduction

The analysis of learning dynamics in games is a well-established problem situated at the intersection of game theory, online optimization and evolutionary game theory [16, 47, 54]. The significance of this area has been amplified by the emergence of prominent machine learning architectures and applications relying on multi-agent, typically zero-sum, games [42, 32, 46, 52, 14, 49, 10]. Symmetric zero-sum games and their dynamics are actually of particular interest both from a traditional evolutionary perspective [54, 13] as well as from a modern Machine Learning perspective as creating a population of agents that compete against each other in a heads-up fashion to outperform each other ("self-play") has been shown to be a reliable recipe for creating super-humanly capable agents for a wide range of tasks [9, 38].

Despite growing interest in understanding and predicting the long-term behavior of such systems, recent studies have revealed a wide array of negative results, demonstrating the elusiveness of game dynamics. These range from non-convergence results to the establishment of chaotic or even essentially arbitrary behavior [44, 28, 30, 40, 3, 4, 55, 26, 45]. Notably, these instability and chaotic behaviors persist even in the analysis of well-known online optimization algorithms, such as Multiplicative Weights Updates (MWU/Hedge), Online Gradient Descent, Follow-the-Regularized-Leader a.o., even within the narrow but seminal class of (symmetric) zero-sum games [7, 18, 19, 20].

Although it is possible to stabilize learning dynamics in zero-sum games using e.g., optimistic variants of MWU [22, 43, 23], such results leave something to be desired as they presuppose that the agents coordinate to use a specific instantiation of a learning algorithm. Ideally, we would like instead to be able to prove such strong convergence results based on more abstract properties of the dynamics.

38th Conference on Neural Information Processing Systems (NeurIPS 2024).

Swap (internal) regret is once such abstract property of online learning dynamics. Unlike the more permissive case of no-(external) regret, where the algorithm's performance needs to compete against the best fixed action with hindsight, no-swap-regret algorithms need to compete against the best adaptive deviation policy with hindsight (i.e., for each occurrence of action i we consider to the best possible deviating action j with hindsight). Somewhat surprisingly, it is possible to adapt no-regret algorithms to no-internal-regret algorithms efficiently [11, 53] with very recent work providing further efficient such reductions [21, 48]. The stronger nature of swap regret results into numerous applications, such as (multi)-calibration [31, 37, 51, 35, 33], robustness against dynamic strategic behavior [24, 41, 15, 5] and AI Safety [17]. Arguably, however, its most important application is due to its tight connection to correlated equilibria, introduced by Aumann [6], as it is well known that the time-average empirical distribution of play resulting from no-internal/swap regret algorithms is guaranteed to converge to the set of correlated equilibria (CE)[1], a (typically strict) relaxation of the predominant game theoretic solution concept of Nash equilibria. In contrast, no-regret algorithms only guarantee time-average convergence to the even more relaxed solution concept of coarse correlated equilibria (CCE) (see preliminary section for precise definitions). Interestingly, in the special case of zero-sum games, for almost all but pathological zero-measure instances of them, the notions of Nash equilibria and correlated equilibria coincide and are in fact unique (whereas CCE do not). This opens the following tantalizing possibility:

Does no-swap-regret minimization suffice for Nash convergence in (almost) all zero-sum games?

The answer to above question is strongly negative. Even in trivial two strategy zero-sum games, such as Matching Pennies, swap regret minimization does not suffice for convergence. Interestingly, however, a sweeping positive result holds for the case of **symmetric** zero-sum games.

Informal theorem: In almost all symmetric zero-sum games, under arbitrary symmetric initializations for both agents, any no-swap-regret algorithm in self-play is guaranteed to converge to the Nash equilibrium, except[2] for a vanishingly small fraction of iterates.

At a technical level, the result depends on two different arguments. First, even in the case of symmetric zero-sum games it is still possible to show that generically the correlated equilibria remain unique, however, the argument does not follow the analysis of general zero-sum games as symmetric cases are themselves non-generic within the larger class. Secondly, we leverage the symmetry of the trajectories to show that time-averaged convergence of symmetric action profiles to a product distribution implies the desired convergent behavior.

This unexpected connection between swap regret minimization and symmetry in games inspires the investigation of other ways that symmetry can insert itself in the study of online learning itself. For example, it is well understood that most standard no-regret algorithms such as MWU, can be completely determined by the vector of cumulative rewards and thus their outputs remain invariant to any permutation of their history, i.e. they exhibit a strong type of time-symmetry. Interestingly, we show that such time-symmetry is provably at odds with swap regret minimization. At a technical level, this argument is based on a construction that couples the behavior of online learning dynamics to particular classes of card guessing games (e.g., [25]) that enable precise control over the algorithm's optimal expected utility (in particular, showing that the play of any such algorithm must be very close to the Follow-The-Leader algorithm, at least when averaged over certain segments of time).

2 Model and Preliminaries

2.1 Games and learning

We consider a setting where two learners (Alice and Bob) are repeatedly playing a game G for T rounds. We assume the game G has N actions for both Alice and Bob, and Alice and Bob will play mixed strategies belonging to the N-dimensional simplex Δ_N. The game G can be thought of as a pair of bilinear functions (G_A, G_B) describing the payoffs for Alice and Bob: in round t, if Alice plays action $a_t \in \Delta_N$ and Bob plays action $b_t \in \Delta_N$, then Alice receives payoff $G_A(a_t, b_t)$ and Bob receives utility $G_B(a_t, b_t)$. One specific class of games we consider are *zero-sum games*, where

[1] Due to their connections to equilibria, establishing fast swap regret minimization (and variations thereof) for different classes of games is a subject of a lot of recent work (e.g., [1, 2, 50, 56, 27]).

[2] This minor disclaimer is necessary as it is always possible to inject such vanishingly small "noise" in any trajectory without affecting its time-average regret.

$G_B(a_t, b_t) = -G_A(a_t, b_t)$; for such games we will omit the subscript and write $G(a_t, b_t)$ to denote $G_A(a_t, b_t)$.

For our purposes, a learning algorithm for Alice in this repeated game is a function which maps the history of played mixed strategies (e.g., $a_1, b_1, a_2, b_2, \ldots, a_{t-1}, b_{t-1}$) to the mixed strategy that Alice will play next (a_t). Learning algorithms for Bob are defined symmetrically. Note that we operate in the deterministic full-information setting where both Alice and Bob know the game G and can see each other's mixed strategy after each round[3].

We consider two classes of learning algorithms, *no-regret algorithms* and *no-swap-regret algorithms*. Alice's (external) regret is defined via

$$\mathsf{Reg}_A = \sum_{t=1}^{T} G_A(a_t, b_t) - \max_{a^* \in \Delta_N} \sum_{t=1}^{T} G_A(a^*, b_t) \tag{1}$$

(with Bob's external regret Reg_B defined similarly). Alice's swap regret is defined via

$$\mathsf{SwapReg}_A = \sum_{t=1}^{T} G_A(a_t, b_t) - \max_{\pi_A : [N] \to [N]} \sum_{t=1}^{T} G_A(\pi_A(a_t), b_t). \tag{2}$$

In (2), the maximum is over all "swap functions" π mapping the set of N actions to itself. We extend π to act on mixed strategies (elements of Δ_N) in the natural way (i.e., $\pi(x)_i = \sum_j x_j \cdot \mathbf{1}(\pi(j) = i)$). We say Alice's learning algorithm is *no-regret* if it is guaranteed that $\mathsf{Reg}_A = o(T)$. Similarly, we say it is *no-swap-regret* if it is guaranteed that $\mathsf{SwapReg}_A = o(T)$. It is known that both efficient no-regret and no-swap-regret algorithms exist, with regret scaling as $\widetilde{O}(\sqrt{T})$ [11] .

2.1.1 Symmetric games and symmetric learners

In this note we primarily consider symmetric learning dynamics in symmetric games. A game G is *symmetric* if $G_A(a_t, b_t) = G_B(b_t, a_t)$. As with zero-sum games, for symmetric games we will omit subscripts and use $G(a_t, b_t)$ to refer to $G_A(a_t, b_t)$. Some games are both symmetric and zero-sum (e.g., the Rock-Paper-Scissors example we introduce later in Example 1).

In a symmetric game, it's natural to consider the setting where Alice and Bob play identical learning algorithms with identical initialization. This results in completely symmetric learning dynamics for Alice and Bob (i.e., $a_t = b_t$ for all rounds t). We write $x_t \in \Delta_N$ to denote the common strategy that Alice and Bob play at time t.

2.2 Equilibria in games

We are interested in the convergence of various types of learning algorithms to specific equilibria of G. We begin by defining the equilibria of interest.

For a game G, a (not necessarily symmetric) *joint strategy profile* σ is a distribution over all N^2 pairs of pure strategies for Alice and Bob. Note that this is not necessarily a product distribution, and in particular allows for Alice's mixed strategy and Bob's mixed strategy to be correlated. It is convenient to identify the set of joint distributions Δ_{N^2} with the convex subset $\mathcal{S}$ of the tensor product space $\mathbb{R}^N \otimes \mathbb{R}^N$ defined as the convex hull of all elements of the form $a \otimes b$ where $a, b \in \Delta_N$. In particular, given $a, b \in \Delta_N$, the element $a \otimes b$ corresponds to the joint strategy profile where Alice plays mixed strategy a and Bob independently plays mixed strategy b. In other words, the pair (i, j) is played with probability $a_i b_j$.

We consider three different types of equilibria, which are (in increasing order of fineness) coarse-correlated equilibria, correlated equilibria, and Nash equilibria. We define these below:

- A *coarse-correlated equilibrium* is a joint strategy profile where neither Alice nor Bob has an incentive to unilaterally deviate to a single pure action. Formally, σ is a coarse-correlated equilibrium if both

[3] Alternatively, everything we describe also holds in a slightly weaker setting, where the players do not know the game G but instead after each round each player sees the counterfactual payoffs they would have received for each of their possible N actions.

$$\mathop{\mathbb{E}}_{(i,j)\sim\sigma}[G_A(i,j)] \geq \mathop{\mathbb{E}}_{(i,j)\sim\sigma}[G_A(i^*,j)], \forall i^* \in [N]$$

$$\mathop{\mathbb{E}}_{(i,j)\sim\sigma}[G_B(i,j)] \geq \mathop{\mathbb{E}}_{(i,j)\sim\sigma}[G_B(i,j^*)], \forall j^* \in [N]$$

- A *correlated equilibrium* is a joint strategy profile where neither Alice nor Bob has an incentive to deviate from their assigned action, where their deviation may depend on their original action. Formally, σ is a correlated equilibrium if both

$$\mathop{\mathbb{E}}_{(i,j)\sim\sigma}[G_A(i,j)] \geq \mathop{\mathbb{E}}_{(i,j)\sim\sigma}[G_A(\pi_A(i),j)], \forall \pi_A : [N] \to [N]$$

$$\mathop{\mathbb{E}}_{(i,j)\sim\sigma}[G_B(i,j)] \geq \mathop{\mathbb{E}}_{(i,j)\sim\sigma}[G_B(i,\pi_B(j))], \forall \pi_B : [N] \to [N]$$

- Finally, σ is a *Nash equilibrium* if σ is a product distribution $\sigma = a \otimes b$ and neither Alice nor Bob has an incentive to deviate ($G_A(a,b) \geq G_A(a',b)$ and $G_B(a,b) \geq G_B(a,b')$). Note that we can alternatively think of Nash equilibria as the intersection of coarse-correlated (or correlated) equilibria with the set of product distributions.

Example 1. Consider the Rock-Paper-Scissors zero-sum game, defined via:

$$G(i,i) = 0, G(i,(i+1) \bmod 3) = 1, G(i,(i-1) \bmod 3) = -1.$$

The unique correlated equilibrium (and hence unique Nash equilibrium) in this game is the product distribution $(\frac{e_1+e_2+e_3}{3}) \otimes (\frac{e_1+e_2+e_3}{3})$. However, the set of coarse correlated equilibria is much larger – it contains, for example, the element $\frac{1}{3}(e_1 \otimes e_1 + e_2 \otimes e_2 + e_3 \otimes e_3)$. Here there is no incentive to unilaterally deviate, but there is an incentive to deviate based on your choice of action (e.g., whenever you play e_1, you can improve your utility by instead playing e_3).

2.3 Convergence to equilibria

When players run learning algorithms in repeated games, they may over time converge to an equilibrium. There are three senses in which this may happen (that we discuss here). We say that Alice and Bob's strategies have *time-averaged convergence* to a joint strategy profile σ if

$$\lim_{T\to\infty} \left\| \left(\frac{1}{T} \sum_{t=1}^{T} a_t \otimes b_t \right) - \sigma \right\| = 0.$$

In other words, time-averaged convergence means that the average of the joint strategy profiles Alice and Bob play converges over time to σ.

Likewise, we say that Alice and Bob's strategies have *frequent-iterate convergence* to σ if for any $\varepsilon > 0$,

$$\lim_{T\to\infty} \Pr_{t\leq T} [\|(a_t \otimes b_t) - \sigma\| > \varepsilon] = 0.$$

Here the probability is taken over t being drawn uniformly at random from all rounds between 1 and T. In other words, frequent-iterate convergence means that, as time goes on, almost all joint strategies profiles Alice and Bob play will be arbitrarily close to σ.

Finally, we say that Alice and Bob's strategies have *last-iterate convergence* to σ if

$$\lim_{T\to\infty} \|(a_T \otimes b_T) - \sigma\| = 0.$$

Last-iterate convergence means that the sequence of joint action profiles played by Alice and Bob directly converge to σ. Note that last-iterate convergence is a stronger property than frequent-iterate convergence, which in turn is a stronger property than time-averaged convergence.

It is known that if Alice and Bob run certain types of learning algorithms, they will have time-averaged convergence to a certain type of equilibrium. Here are some known facts about learning dynamics []:

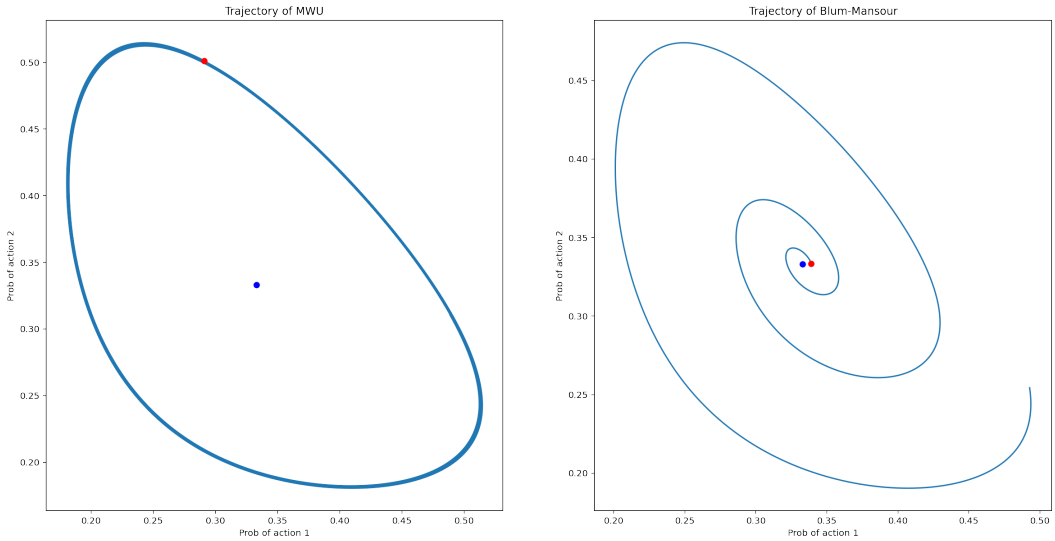

Figure 1: Trajectories of two learning algorithms (Multiplicative Weights and Blum-Mansour) playing Rock-Paper-Scissors. The axis corresponds to the probability of playing Rock and Paper. The blue dot corresponds to the probabilities at Nash equilibrium while the red dot is the last iterate after 10000 steps with learning rate $\eta = 0.001$.

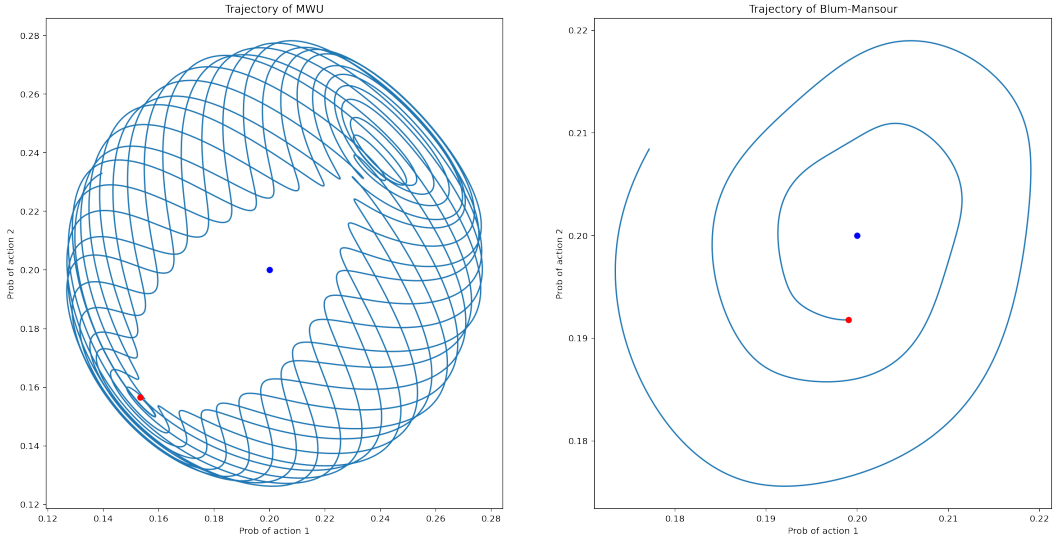

Figure 2: Trajectories of Multiplicative Weights and Blum-Mansour for Rock-Paper-Scissors-Lizard-Spock (a the 5-strategy generalization of Rock-Paper-Scissors).

- If Alice and Bob are running no-regret algorithms in a general game G, their strategies will have time-averaged convergence to a coarse correlated equilibrium of G.
- If Alice and Bob are running no-swap-regret algorithms in a general game G, their strategies will have time-averaged convergence to a correlated equilibrium of G.
- There exist zero-sum games G (including Rock-Paper-Scissors of Example 1) where if Alice and Bob run certain no-regret algorithms (e.g. Multiplicative Weights) Alice and Bob's strategies will not have last-iterate convergence to a coarse correlated equilibrium.

3 Convergence of Symmetric Swap Regret in Symmetric Zero Sum Games

There exist zero-sum games G (including Rock-Paper-Scissors of Example 1) where if Alice and Bob run certain no-regret algorithms (e.g. Multiplicative Weights) Alice and Bob's strategies will not have last-iterate convergence to a equilibrium. In the left side of Figure 1 plot the trajectory of Multiplicative Weights when Alice and Bob have the same initialization. The blue dot in the middle corresponds to the Nash equilibrium. While the strategies played average to the equilibrium, they never converge there.

In this section, we will show that for almost all *symmetric*, *zero-sum* games G, if Alice and Bob employ the same no-swap-regret learning algorithms with symmetric initializations they will have frequent-iterate convergence to a Nash equilibrium of G. In the right side of the same figure, we show the Blum-Mansour trajectory converging to equilibrium. In Appendix A, we show via a simple counter-example that symmetric initializations are necessary to achieve such strong convergence guarantees.

Our main tool is the following technical lemma, which shows that time-averaged convergence of symmetric action profiles to a product distribution implies frequent-iterate convergence to the same distribution.

Lemma 1. *Let $x_1, x_2, \ldots$ be a sequence of elements in Δ_N such that*

$$\lim_{T\to\infty} \left\| \left(\frac{1}{T} \sum_{t=1}^{T} x_t \otimes x_t \right) - (y \otimes y) \right\| = 0$$

for some element y of Δ_N (i.e., the profiles $x_t \otimes x_t$ time-averaged converge to $y \otimes y$). Then it is the case that for any $\varepsilon > 0$,

$$\lim_{T\to\infty} \Pr_{t\le T} [\|(x_t \otimes x_t) - (y \otimes y)\| > \varepsilon] = 0.$$

(i.e., the profiles $x_t \otimes x_t$ frequent-iterate converge to $y \otimes y$).

Proof. For any $y \in \Delta_N$, we will show that there exists a linear functional $L_y : \mathbb{R}^N \otimes \mathbb{R}^N \to \mathbb{R}$ with the property that, among all elements of the form $x \otimes x$ with $x \in \Delta_N$, L_y is uniquely minimized at $y \otimes y$. As a consequence, this implies that that there exists a $\delta > 0$ such that if

$$\|(x \otimes x) - (y \otimes y)\| > \varepsilon,$$

for some element $x \in \Delta_N$, then $L_y(x \otimes x) - L_y(y \otimes y) > \delta$. In particular, if for some T we have that

$$\Pr_{t\le T} [\|(x_t \otimes x_t) - (y \otimes y)\| > \varepsilon] \ge \gamma,$$

then we consequently have that

$$\frac{1}{T} \sum_{t=1}^{T} (L_y(x_t \otimes x_t) - L_y(y \otimes y)) \ge \gamma\delta,$$

and in turn that

$$\left\| \left(\frac{1}{T} \sum_{t=1}^{T} (x_t \otimes x_t) \right) - (y \otimes y) \right\| \ge \gamma\delta \|L_y\|_*,$$

(where $\|L_y\|_*$ is the dual norm of the linear functional L_y and is bounded below by some constant). This directly implies the lemma statement (it is impossible for the second quantity to approach zero as T goes to infinity without the first quantity approaching zero).

We now describe the linear functional L_y. Let $y = (y_1, y_2, \ldots, y_N)$ (and for a general $x \in \Delta_N$, let $x = (x_1, x_2, \ldots, x_N)$). Note that for any linear functional $L_y : \mathbb{R}^N \otimes \mathbb{R}^N \to \mathbb{R}$, the value of

$L_y(x \otimes x)$ will be a homogeneous quadratic polynomial over the x_i; conversely, any such polynomial can be implemented as a linear functional over $\mathbb{R}^N \otimes \mathbb{R}^N$. Moreover, since $\sum x_i = 1$, we can convert any (non-homogeneous) quadratic polynomial to a homogeneous one (e.g. transforming $x_1^2 + x_1 + 1$ to $x_1^2 + x_1 \sum x_i + (\sum x_i)^2$). It therefore suffices to find a quadratic polynomial over the x_i that is minimized when $x = y$.

We begin with the case where all $y_i > 0$. In this case, we claim that the polynomial

$$L_y(x \otimes x) = \sum_{i=1}^{N} \frac{1}{y_i} x_i^2$$

satisfies these constraints. To see this, note that by Cauchy-Schwartz,

$$\sum_{i=1}^{N} \frac{1}{y_i} x_i^2 = \left(\sum_{i=1}^{N} \frac{x_i^2}{y_i}\right)\left(\sum_{i=1}^{N} y_i\right) \geq \left(\sum_{i=1}^{N} x_i\right)^2 = 1,$$

with equality only holding when $x_i^2/y_i = \lambda y_i$ for some fixed λ. Since x and y both belong to Δ_N, this is only possible when $\lambda = 1$ and $x = y$.

What if some of the y_i equal zero? Without loss of generality, assume $y_i > 0$ for $1 \leq i \leq n$ and $y_i = 0$ for $n + 1 \leq i \leq N$. Now, consider the polynomial

$$L_y(x \otimes x) = \left(\sum_{i=1}^{n} \frac{1}{y_i} x_i^2\right) - 2\left(\sum_{i=1}^{n} x_i\right).$$

We begin by minimizing this expression subject to $\sum_{i=1}^{n} x_i = s$, for some $0 \leq s \leq 1$. In this case, the second term in the above polynomial is identically $-2s$, so it suffices to minimize the first term. Again applying Cauchy-Schwartz, we find the minimum of the first term occurs only when $x_i = sy_i$, where it equals s^2. The overall minimum of L_y (subject to this constraint) is therefore $s^2 - 2s$. This is in turn uniquely minimized when $s = 1$ (and $x_i = 0$ for all $i > n$), and therefore this L_y is uniquely minimized at $x = y$. □

Next we show a *'generic'* symmetric zero-sum game has a unique correlated equilibrium and hence coincides with the Nash equilibrium of this game. First, we define what we mean by generic. We can identify the set of zero-sum games with $\mathbb{R}^{N \times N}$, where each element corresponds to an $N \times N$ payoff matrix for Alice. We say that a property $\mathcal{P}$ holds for a generic zero-sum game, if the set of points in $\mathbb{R}^{N \times N}$ for which $\mathcal{P}$ doesn't holds form a measure zero subset of $\mathbb{R}^{N \times N}$.

By this definition, the set of symmetric zero-sum games forms a measure zero subset of the set of zero-sum games, so we need a refined definition to describe a generic symmetric zero-sum game. We can identify the set of symmetric zero-sum games with with $\mathbb{R}^{N(N-1)/2}$ where each element corresponds to a skew-symmetric $N \times N$ payoff matrix for Alice. We say that a property $\mathcal{P}$ holds for a generic symmetric zero-sum game, if the set of points in $\mathbb{R}^{N \times N}$ for which $\mathcal{P}$ doesn't holds form a measure zero subset of $\mathbb{R}^{N(N-1)/2}$.

The uniqueness of a correlated equilibrium for zero-sum games is known for generic zero-sum games by combining results by Forges [29] and Bohnenblust, Karlin and Shapley [12]. But since symmetric zero-sum games are a measure zero subset of zero-sum games, this result does not directly extend. Below, we extend it to symmetric zero-sum games.

Finally, to complete the picture, we further establish that this generic uniqueness of correlated equilibria strongly does not extend to the case of coarse correlated equilibria, which are the limit points of (external) regret-minimizing algorithms.

Lemma 2. *Almost all (i.e., all but a measure zero set of) two-player symmetric zero-sum games have a unique correlated/Nash equilibrium.*

Proof. By [29] a zero-sum game has a unique correlated equilibrium if and only if it has a unique Nash equilibrium, thus it suffices to prove the generic uniqueness of Nash equilibria. Let A be the skew-symmetric ($A^\top = -A$) payoff matrix corresponding to the symmetric zero-sum game. A Nash

equilibrium is called quasi-strict (sometimes also referred to as regular or quasi-strong [34]) if for all agents deviations to strategies outside their support result in a strict decrease of their payoff. By Corollary 3.4 in [36] we have that in any two agent game if all equilibria are quasi-strict then the number of equilibria is finite. Thus, if a symmetric zero-sum game has multiple equilibria, this implies the existence of a non-quasi-strict equilibrium/optimal strategy. Let x be that non-quasi-strict mixed strategy and let S_x denote its support, i.e., the set of strategies played in x with positive probability and let i be the index of the strategy not in S_x such that deviating to that strategy from the symmetric (x, x) Nash equilibrium still results into payoff of zero (since the game value of any zero-sum symmetric game is zero). This implies that both antisymmetric sub-matrices of A defined by its restrictions to the sets S_x and $S_x \cup i$ respectively have a zero eigenvalue where the corresponding eigenvector is the analogously restricted subvector of x. Thus, both of their determinants are equal to zero. However, at least one of them has even dimension. It is well known that the determinant of an even dimension skew-symmetric matrix is a non-trivial polynomial (and in fact is the square of a polynomial in its coefficients, see e.g. [39]). The entries of this submatrix correspond to the vanishing set of a non-trivial polynomial and therefore have Lebesgue measure zero. Thus, the set of all symmetric zero-sum games with unique Nash/correlated equilibrium has zero measure.

□

Combining Lemma 1 and Theorem 2, we arrive at the result mentioned at the beginning of this section.

Theorem 3. *In a generic symmetric zero-sum game G, if Alice and Bob run identical no-swap-regret algorithms with the same initialization to play G repeatedly, their joint strategy profiles will have frequent-iterate convergence to a Nash equilibrium of G. Furthermore, this result is* tight*, i.e., it is not possible to prove (last-iterate) convergence to Nash equilibrium.*

Proof. Since Alice and Bob are both running the same no-swap-regret algorithm, their strategy profiles stay identical and hence their joint strategy profile is of the form $x_t \otimes x_t$. Since no-swap-regret algorithms have time average convergence to correlated equilibrium, then there is a correlated equilibrium σ of G such that $\|\frac{1}{T}\sum_{t=1}^{T} x_t \otimes x_t - \sigma\| \to 0$. By Lemma 2, σ is a Nash equilibrium, so we can write $\sigma = y \otimes y$. Now, we can apply Lemma 1 to obtain frequent iterated convergence to σ.

It is not possible to prove anything stronger, i.e., convergence to Nash, based on (symmetric) no-swap-regret learning because we can take any such no-swap-regret dynamics and interject for a vanishing fraction of the history some arbitrary symmetric play where those payoff inputs are ignored by the learning dynamics (e.g. the Blum-Mansour algorithm does not see these fictitious entries). This rare interleaving of the trajectory with noise does not significantly affect the swap regret analysis which will remain sublinear if the original dynamic is no-swap-regret but at the same time it suffices to destroy any hope of true last iterate convergence. □

3.1 Differences with respect to External Regret

It is useful to consider which parts of the proof break when we move from swap to external regret. Both Lemma 1 and Lemma 2 no longer hold.

Consider for example the executions of MWU and BM in Figure 1. In both cases we have that $\frac{1}{T}(\sum_t x_t) \to x_{\text{Nash}} := (\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$. If we look at the empirical joint distribution $\bar{\sigma}_T = \frac{1}{T}(\sum_t x_t \otimes x_t)$ we obtain the following:

$$\bar{\sigma}_T^{\text{BM}} = \begin{bmatrix} 0.111 & 0.111 & 0.111 \\ 0.111 & 0.111 & 0.111 \\ 0.111 & 0.111 & 0.111 \end{bmatrix} \qquad \bar{\sigma}_T^{\text{MWU}} = \begin{bmatrix} 0.120 & 0.105 & 0.105 \\ 0.105 & 0.120 & 0.105 \\ 0.105 & 0.105 & 0.120 \end{bmatrix}$$

Only in the swap regret algorithm we have $\bar{\sigma} \to x_{\text{Nash}} \otimes x_{\text{Nash}}$. Lemma 1 crucially relies on $\frac{1}{T}(\sum_t x_t \otimes x_t)$ converging a product distribution.

Lemma 2 also breaks when we replace CE (achieved by swap regret minimization) with CCE (achieved by external regret minimization):

Lemma 4. *Given any symmetric zero-sum game as long as its set of optimal strategies does not consist of a single pure (i.e., deterministic) strategy then it has a continuum of coarse correlated equilibria.*

Proof. In any such game there exists an optimal (Nash) strategy that is randomizing amongst at least two strategies. We will define the correlated distribution that applies positive probability only symmetric outcomes of the game (e.g, such as (Rock, Rock) (Paper, Paper) and (Scissors, Scissors) in the RPS game) such as its resulting marginal distribution corresponds to the Nash equilibrium strategy. The expected payoff of each agent in this distribution is equal to zero. Furthermore, any deviating strategy cannot result in positive payoff since the marginal distribution encodes a Nash equilibrium. Thus, the original distribution is a CCE. By taking convex combinations of these CCE and the Nash equilibrium we have that each such game has a continuum of CCE. □

An immediate corollary of Lemma 4 is that the rich, non-equilibrating behavior of MWU and other no-regret dynamics, even under symmetric initializations, shown for Rock-Paper-Scissors and Rock-Paper-Scissors-Lizard-Spock shown in Figures 1,2 should be common for many other symmetric games and regret-minimizing dynamics as well.

3.2 Asymmetric Zero Sum Games

We remark that the results don't generalize to asymmetric zero sum games. One example of such games is matching pennies where no-external regret dynamics like MWU are known not to exhibit last-iterate convergence and in fact diverge chaotically towards the boundary for all interior initializations, including all symmetric (non-Nash) initial conditions [8, 18]. For games with 2 actions per player, any no-external-regret algorithm is also no-swap-regret as we show in the following lemma.

Lemma 5. *For a game with* 2 *actions per player,* $\mathsf{SwapReg} \leq 2\,\mathsf{Reg}$.

Proof. Let $\{0, 1\}$ be Alice's actions in let $a_t, b_t \in \Delta_2$ be a sequence of strategies played by Alice and Bob. There are 3 non-trivial swap strategies $\pi : [2] \to [2]$ for Alice: $s \mapsto 0$, $s \mapsto 1$, $s \mapsto 1 - s$. The regret of using the first two swap strategies is bounded by Reg since they map to a constant action. Let $a_t[s]$ be the s-th component of $a_t \in \Delta_2$. Then, the regret of the third strategy π is bounded by:

$$\begin{aligned}\sum_t G_A(a_t, b_t) - G_A(\pi(a_t), b_t) &= \sum_t a_t[0](G_A(0, b_t) - G_A(1, b_t)) + \sum_t a_t[1](G_A(1, b_t) - G_A(0, b_t)) \\ &= \left[\sum_t G_A(a_t, b_t) - G_A(1, b_t)\right] + \left[\sum_t G_A(a_t, b_t) - G_A(0, b_t)\right] \leq 2\,\mathsf{Reg}\end{aligned}$$

□

4 No-Swap-Regret Algorithms are Time-Asymmetric

One interesting feature of no(-external)-regret algorithms in two-player games is that the action they select at time t can depend entirely on the average historical strategy played by the other player up to time $t - 1$, and not on any other information about how this strategy (or the player's own strategy) evolved over time.

In this section we show that it is impossible for a no-swap-regret learning algorithm to have this property. In fact, we prove the following more general statement.

Theorem 6. *Consider any learning algorithm $\mathcal{A}$ which decides what action to take on behalf of Alice in a game G at round t via a* symmetric *function $A_t(b_1, b_2, \ldots, b_{t-1})$ of Bob's mixed strategies up until $t - 1$ (here symmetric means that the function is unchanged for any permutation of the inputs). Then there exists a game G and a sequence of play for Bob where Alice incurs $\Omega(T)$ swap regret.*

As mentioned, many no-external-regret algorithms (such as multiplicative weights, and follow-the-regularized leader) have the property that each A_t can be written as a function of the average $\frac{1}{t-1}\sum_{s=1}^{t-1} b_s$, and hence are symmetric.

We provide a sketch of the proof of Theorem 6, deferring all details and proofs of lemmas to Appendix B. We will consider the game G given by the three-action generalization of Matching Pennies. Specifically, Alice and Bob will both have 3 actions, and $G_A(i, j)$ equals 1 if $i = j$ and 0 otherwise (since we will specify Bob's sequence of actions adversarially, his payoff G_B is irrelevant).

We will construct an adversarial sequence of actions for Bob where in each round Bob plays one of the three pure actions in Δ_3 (i.e., $b_t \in \{e_1, e_2, e_3\}$). For each $i \in [3]$, let $n_{t,i}$ equal the number

of rounds $s \le t-1$ where Bob played action i. Since A_t is a symmetric function of the b_s for $1 \le s \le t-1$, we can write A_t as a function $A_t(n_{t,1}, n_{t,2}, n_{t,3})$. Moreover, since we must have $n_{t,1} + n_{t,2} + n_{t,3} = t-1$, we can summarize the entire learning algorithm with a single 3-variable function $A(n_1, n_2, n_3) : \mathbb{Z}^3_{\ge 0} \to \Delta_3$ satisfying

$$A(n_1, n_2, n_3) = A_{n_1+n_2+n_3+1}(n_1, n_2, n_3).$$

Our main strategy will be to show that if A has no swap regret (or even no external regret), on average A must be very close to the "follow the leader" strategy, which puts all the weight on action i if n_i is significantly larger than the other n_j. We formalize this in the following lemma (which employs a result of [25] on the number of mistakes necessary in certain card guessing games):

Lemma 7. *Assume that the algorithm $\mathcal{A}$ guarantees that Alice incurs at most sublinear external regret, i.e., $\mathsf{Reg}_A = o(T)$. Then for any $L \ge T/100$ and $n_1, n_2 \le n_3 \le T - L$, we have that*

$$\frac{1}{L} \sum_{m=n_3+1}^{n_3+L} A(n_1, n_2, m)_3 = 1 - o(1).$$

Lemma 7 shows that $\mathcal{A}$ mostly plays the leader (i.e., highest-utility) action in any sufficiently long segment of rounds where Bob is playing the leader action. We would also like to show that $\mathcal{A}$ mostly plays the leader action in stretches where Bob is playing some other fixed action. The following lemma gives a weak form of this claim (but that will be sufficient for proving Theorem 6).

Lemma 8. *Fix any $L, L' \ge T/100$ and let $n_2 \ge \min(n_1, L')$. Then there exists an $n_3 \in [n_2, n_2 + L]$ such that*

$$\frac{1}{L'} \sum_{m'=n_2-L'}^{n_2-1} A(n_1, m', n_3)_3 = 1 - o(1).$$

With Lemmas 7 and 8 (and their symmetric counterparts), we can construct a sequence of play for Bob where Alice incurs high swap regret. Roughly, this sequence of play proceeds as follows.

- First Bob plays action 1 for approximately $T/3$ rounds. By Lemma 7, we can guarantee that Alice plays action 1 for most of these rounds.
- Bob then plays action 2 for approximately $T/3$ rounds. By the guarantee of Lemma 8, Alice will still play action 1 for most of these rounds.
- Bob then plays action 3 for the remaining rounds. Again by applying Lemma 8, we can guarantee that Alice will play action 2 for most of these rounds.

It is straightforward to check that Alice incurs linear swap regret in the above trajectory – Alice would improve her expected utility by $\Omega(T)$ if she played action 3 every time she played action 2. The details of this proof are deferred to Appendix B.

5 Conclusion

In this paper, we study the role of symmetry in the behavior of no-swap regret dynamics. In our first result, we show that no-swap-regret dynamics in self-play in symmetric zero-sum games lead to converge in a strong "frequent-iterate" sense to the Nash equilibrium. Specifically, in all but a vanishing fraction of the rounds, the players must play a strategy profile close to a symmetric Nash equilibrium. Furthermore, we show that the power of no-swap-regret dynamics comes at a cost of imposing a time-asymmetry on its inputs. Specifically, any such algorithm, unlike no-external regret dynamics, must apply a time-asymmetric function over the set of previously observed rewards.

The interplay between symmetry, (external/swap) regret and learning in games emerges as as interesting direction for future work. One particularly interesting such direction would be to explore generalizations of our results beyond two player games.

References

[1] I. Anagnostides, G. Farina, C. Kroer, C.-W. Lee, H. Luo, and T. Sandholm. Uncoupled learning dynamics with o (\log t) swap regret in multiplayer games. *Advances in Neural Information Processing Systems*, 35:3292–3304, 2022.

[2] I. Anagnostides, G. Farina, and T. Sandholm. Near-optimal ϕ-regret learning in extensive-form games. In *International Conference on Machine Learning*, pages 814–839. PMLR, 2023.

[3] G. P. Andrade, R. Frongillo, and G. Piliouras. Learning in matrix games can be arbitrarily complex. In M. Belkin and S. Kpotufe, editors, *Proceedings of Thirty Fourth Conference on Learning Theory*, volume 134 of *Proceedings of Machine Learning Research*, pages 159–185. PMLR, 15–19 Aug 2021.

[4] G. P. Andrade, R. Frongillo, and G. Piliouras. No-regret learning in games is turing complete. *arXiv preprint arXiv:2202.11871*, 2022.

[5] E. R. Arunachaleswaran, N. Collina, and J. Schneider. Pareto-optimal algorithms for learning in games. *arXiv preprint arXiv:2402.09549*, 2024.

[6] R. J. Aumann. Subjectivity and correlation in randomized strategies. *Journal of mathematical Economics*, 1(1):67–96, 1974.

[7] J. P. Bailey and G. Piliouras. Multiplicative weights update in zero-sum games. In *ACM Conference on Economics and Computation*, 2018.

[8] J. P. Bailey and G. Piliouras. Multiplicative weights update in zero-sum games. In *Proceedings of the 2018 ACM Conference on Economics and Computation*, pages 321–338, 2018.

[9] D. Balduzzi, M. Garnelo, Y. Bachrach, W. Czarnecki, J. Perolat, M. Jaderberg, and T. Graepel. Open-ended learning in symmetric zero-sum games. In *International Conference on Machine Learning*, pages 434–443. PMLR, 2019.

[10] A. Bighashdel, Y. Wang, S. McAleer, R. Savani, and F. A. Oliehoek. Policy space response oracles: A survey. *arXiv preprint arXiv:2403.02227*, 2024.

[11] A. Blum and Y. Mansour. From external to internal regret. *Journal of Machine Learning Research*, 8(6), 2007.

[12] H. Bohnenblust, S. Karlin, and L. Shapley. Solutions of discrete, two-person games. *Contributions to the Theory of Games*, 1:51–72, 1950.

[13] V. Boone and G. Piliouras. From Darwin to Poincaré and von Neumann: Recurrence and cycles in evolutionary and algorithmic game theory. In *International Conference on Web and Internet Economics*, pages 85–99. Springer, 2019.

[14] N. Brown and T. Sandholm. Superhuman ai for multiplayer poker. *Science*, 365(6456):885–890, 2019.

[15] W. Brown, J. Schneider, and K. Vodrahalli. Is learning in games good for the learners? *Advances in Neural Information Processing Systems*, 36, 2024.

[16] N. Cesa-Bianchi and G. Lugoisi. *Prediction, Learning, and Games*. Cambridge University Press, 2006.

[17] X. Chen, A. Chen, D. Foster, and E. Hazan. Playing Large Games with Oracles and AI Debate. *arXiv e-prints*, page arXiv:2312.04792, Dec. 2023.

[18] Y. K. Cheung and G. Piliouras. Vortices instead of equilibria in minmax optimization: Chaos and butterfly effects of online learning in zero-sum games. In *Conference on Learning Theory*, pages 807–834. PMLR, 2019.

[19] Y. K. Cheung and G. Piliouras. Chaos, extremism and optimism: Volume analysis of learning in games. In *NeurIPS*, 2020.

[20] Y. K. Cheung, G. Piliouras, and Y. Tao. The evolution of uncertainty of learning in games. In *International Conference on Learning Representations*, 2021.

[21] Y. Dagan, C. Daskalakis, M. Fishelson, and N. Golowich. From external to swap regret 2.0: An efficient reduction and oblivious adversary for large action spaces. *arXiv preprint arXiv:2310.19786*, 2023.

[22] C. Daskalakis, A. Ilyas, V. Syrgkanis, and H. Zeng. Training gans with optimism. In *ICLR*, 2018.

[23] C. Daskalakis and I. Panageas. Last-iterate convergence: Zero-sum games and constrained min-max optimization. *ITCS*, 2019.

[24] Y. Deng, J. Schneider, and B. Sivan. Strategizing against no-regret learners. In *Advances in Neural Information Processing Systems 32: Annual Conference on Neural Information Processing Systems 2019, NeurIPS 2019, 8-14 December 2019, Vancouver, BC, Canada*, pages 1577–1585, 2019.

[25] P. Diaconis and R. Graham. The analysis of sequential experiments with feedback to subjects. *The Annals of Statistics*, 9(1):3–23, 1981.

[26] M. Engel and G. Piliouras. A stochastic variant of replicator dynamics in zero-sum games and its invariant measures. *arXiv preprint arXiv:2302.06969*, 2023.

[27] G. Farina and C. Pipis. Polynomial-time computation of exact phi-equilibria in polyhedral games. *arXiv preprint arXiv:2402.16316*, 2024.

[28] L. Flokas, E.-V. Vlatakis-Gkaragkounis, T. Lianeas, P. Mertikopoulos, and G. Piliouras. No-regret learning and mixed nash equilibria: They do not mix. In *NeurIPS*, 2020.

[29] F. Forges. Correlated equilibrium in two-person zero-sum games. *Econometrica (1986-1998)*, 58(2):515, 1990.

[30] A. Giannou, E. V. Vlatakis-Gkaragkounis, and P. Mertikopoulos. Survival of the strictest: Stable and unstable equilibria under regularized learning with partial information. In M. Belkin and S. Kpotufe, editors, *Proceedings of Thirty Fourth Conference on Learning Theory*, volume 134 of *Proceedings of Machine Learning Research*, pages 2147–2148. PMLR, 15–19 Aug 2021.

[31] I. Globus-Harris, D. Harrison, M. Kearns, A. Roth, and J. Sorrell. Multicalibration as boosting for regression. *arXiv preprint arXiv:2301.13767*, 2023.

[32] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio. Generative adversarial nets. In *Advances in neural information processing systems*, pages 2672–2680, 2014.

[33] N. Haghtalab, C. Podimata, and K. Yang. Calibrated stackelberg games: Learning optimal commitments against calibrated agents. *Advances in Neural Information Processing Systems*, 36, 2024.

[34] J. C. Harsanyi. Oddness of the number of equilibrium points: a new proof. *International Journal of Game Theory*, 2(1):235–250, 1973.

[35] L. Hu and Y. Wu. Calibration error for decision making, 2024.

[36] M. Jansen. Regularity and stability of equilibrium points of bimatrix games. *Mathematics of Operations Research*, 6(4):530–550, 1981.

[37] B. Kleinberg, R. P. Leme, J. Schneider, and Y. Teng. U-calibration: Forecasting for an unknown agent. In *The Thirty Sixth Annual Conference on Learning Theory*, pages 5143–5145. PMLR, 2023.

[38] M. Lanctot, V. Zambaldi, A. Gruslys, A. Lazaridou, K. Tuyls, J. Pérolat, D. Silver, and T. Graepel. A unified game-theoretic approach to multiagent reinforcement learning. *Advances in neural information processing systems*, 30, 2017.

[39] W. Ledermann. A note on skew-symmetric determinants. *Proceedings of the Edinburgh Mathematical Society*, 36(2):335–338, 1993.

[40] A. Letcher. On the impossibility of global convergence in multi-loss optimization, 2021.

[41] Y. Mansour, M. Mohri, J. Schneider, and B. Sivan. Strategizing against learners in bayesian games. In *Conference on Learning Theory*, pages 5221–5252. PMLR, 2022.

[42] H. B. McMahan, G. J. Gordon, and A. Blum. Planning in the presence of cost functions controlled by an adversary. In *Proceedings of the 20th International Conference on Machine Learning (ICML-03)*, pages 536–543, 2003.

[43] P. Mertikopoulos, B. Lecouat, H. Zenati, C.-S. Foo, V. Chandrasekhar, and G. Piliouras. Optimistic mirror descent in saddle-point problems: Going the extra(-gradient) mile. In *ICLR*, 2019.

[44] P. Mertikopoulos, C. Papadimitriou, and G. Piliouras. Cycles in adversarial regularized learning. In *ACM-SIAM Symposium on Discrete Algorithms*, 2018.

[45] J. Milionis, C. Papadimitriou, G. Piliouras, and K. Spendlove. An impossibility theorem in game dynamics. *Proceedings of the National Academy of Sciences*, 120(41):e2305349120, 2023.

[46] M. Moravčík, M. Schmid, N. Burch, V. Lisỳ, D. Morrill, N. Bard, T. Davis, K. Waugh, M. Johanson, and M. Bowling. Deepstack: Expert-level artificial intelligence in heads-up no-limit poker. *Science*, 356(6337):508–513, 2017.

[47] N. Nisan, T. Roughgarden, E. Tardos, and V. V. Vazirani. *Algorithmic Game Theory*. Cambridge University Press, New York, NY, USA, 2007.

[48] B. Peng and A. Rubinstein. Fast swap regret minimization and applications to approximate correlated equilibria. *arXiv preprint arXiv:2310.19647*, 2023.

[49] J. Perolat, B. De Vylder, D. Hennes, E. Tarassov, F. Strub, V. de Boer, P. Muller, J. T. Connor, N. Burch, T. Anthony, et al. Mastering the game of stratego with model-free multiagent reinforcement learning. *Science*, 378(6623):990–996, 2022.

[50] G. Piliouras, M. Rowland, S. Omidshafiei, R. Elie, D. Hennes, J. Connor, and K. Tuyls. Evolutionary dynamics and phi-regret minimization in games. *Journal of Artificial Intelligence Research*, 74:1125–1158, 2022.

[51] A. Roth and M. Shi. Forecasting for swap regret for all downstream agents. *arXiv preprint arXiv:2402.08753*, 2024.

[52] D. Silver, J. Schrittwieser, K. Simonyan, I. Antonoglou, A. Huang, A. Guez, T. Hubert, L. Baker, M. Lai, A. Bolton, et al. Mastering the game of go without human knowledge. *nature*, 550(7676):354–359, 2017.

[53] G. Stoltz and G. Lugosi. Internal regret in on-line portfolio selection. *Machine Learning*, 59:125–159, 2005.

[54] J. W. Weibull. *Evolutionary Game Theory*. MIT Press; Cambridge, MA: Cambridge University Press., 1995.

[55] A. Wibisono, M. Tao, and G. Piliouras. Alternating mirror descent for constrained min-max games. In *Proceedings of the 36nd International Conference on Neural Information Processing Systems*, 2022.

[56] B. H. Zhang, I. Anagnostides, G. Farina, and T. Sandholm. Efficient phi-regret minimization with low-degree swap deviations in extensive-form games. *arXiv preprint arXiv:2402.09670*, 2024.

A No-swap regret dynamics with asymmetric initial conditions do not converge to Nash equilibrium

We show via a simple counter-example that symmetric initializations are necessary for convergence to Nash. Specifically, in Figure 3 we show the trajectory of two players running the Blum-Mansour no-swap-regret algorithm against each other, initialized with asymmetric starting conditions in the symmetric zero-sum game of Rock-Paper-Scissors. Their behavior does not converge to Nash, whereas as we have seen in Figure 1, symmetric initialization in the same game would have led to convergence.

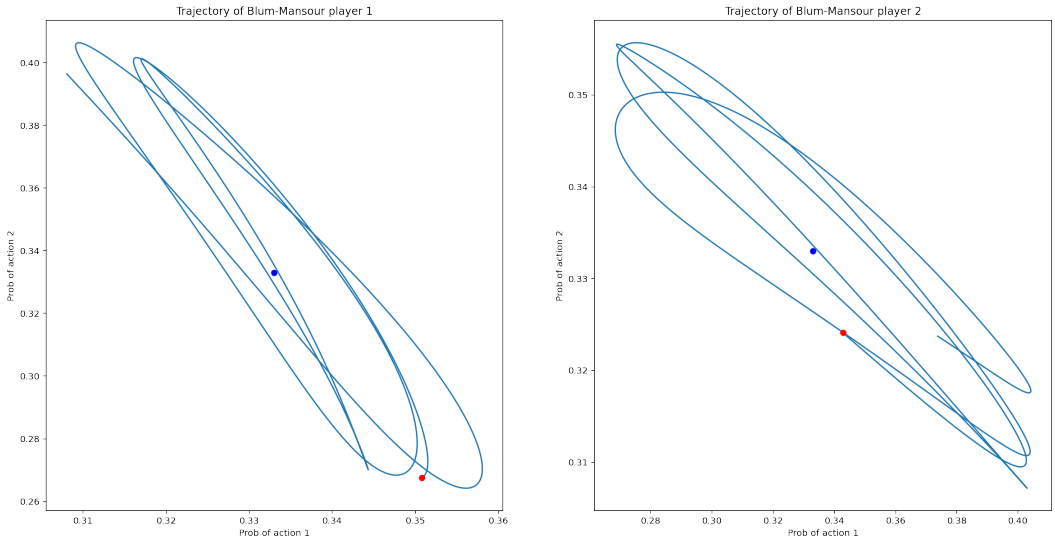

Figure 3: This figure shows the trajectory of two players running the Blum-Mansour no-swap-regret algorithm against each other initialized with asymmetric starting conditions. Unlike the symmetric dynamics, these do not ultimately converge to the unique symmetric Nash equilibrium (the blue point).

B Omitted proofs of Section 4

B.1 Proof of Lemma 7

Proof of Lemma 7. Define $W = \sum_{m=n_3+1}^{n_3+L} A(n_1, n_2, m)_3$. We will present a distribution $\mathcal{D}$ over action sequences of length T for Bob such that the expected regret Reg_A that Alice incurs when playing against a randomly drawn sequence from $\mathcal{D}$ is at least $(L - W) - O(\sqrt{T})$. If the algorithm $\mathcal{A}$ has the guarantee that $\mathsf{Reg}_A = o(T)$, then this implies that we must have $W = L - o(T)$, from which the theorem statement follows.

We will construct the distribution $\mathcal{D}$ as follows:

- For the first $n_1 + n_2 + n_3$ rounds, Bob will play a uniform random sequence that includes action 1 n_1 times, action 2 n_2 times, and action 3 n_3 times.
- In the next L rounds, Bob will always play action 3.
- Finally, in the last $T' = T - (n_1 + n_2 + n_3)$ rounds, Bob will play a uniform random sequence that includes action 1 $T'/3$ times, action 2 $T'/3$ times, and action 3 $T'/3$ times.

Note that for any such sequence, the best action in hindsight for Alice is action 3, which achieves utility exactly $n_3 + L + T'/3$. To compute Alice's expected regret, it suffices to compare Alice's expected utility to this quantity.

To bound Alice's optimal expected utility, we will make use of the following result of Diaconis and Graham [25]. Consider a game where there is a uniformly shuffled deck of $N = n_1 + n_2 + n_3$ cards with n_1 cards labeled 1, n_2 cards labeled 2, and n_3 cards labeled 3. In this game, the player must repeatedly guess the label of the card on the top of the deck, at which point it is revealed to the player and discarded. Then Diaconis and Graham show that the expected number of correct guesses of the player is at most $\max(n_1, n_2, n_3) + O(\sqrt{N})$ (in fact in Theorem 3 of [25] provides bounds that apply to any number of labels, but we only need this specific consequence).

Note that in the first $n_1 + n_2 + n_3$ rounds, Alice is facing essentially this exact game (since she gains utility 1 when she matches the action of Bob, and 0 otherwise). Therefore Alice's expected utility in the first segment of the rounds is at most $n_3 + O(\sqrt{n_1 + n_2 + n_3}) = n_3 + O(\sqrt{T})$. Similarly, in the last segment of rounds, Alice's expected utility is at most $T'/3 + O(\sqrt{T})$. Finally, in the middle segment of rounds, Alice's utility is exactly W (as this is the total weight she places on action 3). It follows that Alice's expected regret is at least $(L - W) - O(\sqrt{T})$, as desired. □

B.2 Proof of Lemma 8

Proof of Lemma 8. By applying Lemma 7 L times, we have that:

$$\frac{1}{L' \cdot L} \sum_{m'=n_2-L'}^{n_2-1} \sum_{m=n_2}^{n_2+L} A(n_1, m', m)_3 = 1 - o(1).$$

By switching the order of summation, this implies that there must exist a fixed value of $m \in [n_2, n_2 + L]$ such that

$$\frac{1}{L'} \sum_{m'=n_2-L'}^{n_2-1} A(n_1, m', m)_3 = 1 - o(1).$$

We can take n_3 to be this value of m. □

With Lemmas 7 and 8 (and their symmetric counterparts), we can construct a sequence of play for Bob where Alice incurs high swap regret.

B.3 Proof of Theorem 6

Proof of Theorem 6. Fix $L = T/100$. Bob will begin by selecting an $n_1 \in [T/3, T/3 + T/1000]$ such that:

$$\frac{1}{T/3} \sum_{m'=1}^{T/3} A(n_1, m', 0)_3 = 1 - o(1).$$

(Such an n_1 is guaranteed to exist by a symmetric variant of Lemma 8). Bob will then play action 1 for n_1 rounds followed by action 2 for $T/3$ rounds.

Bob will then select a value $n_2 \in [T/3 + T/1000, T/3 + 2T/1000]$ such that:

$$\frac{1}{(T/3) - (T/100)} \sum_{m'=1}^{T/3-T/100} A(n_1, n_2, m')_3 = 1 - o(1).$$

(Again, such an n_2 is guaranteed to exist by a symmetric variant of Lemma 8). Bob will then play action 2 for $n_2 - T/3$ rounds, and action 3 for the remaining rounds.

What does Alice do against this sequence of play of Bob? We break this down segment by segment:

- First Bob plays action 1 for n_1 rounds, moving the state from $(0, 0, 0)$ to $(n_1, 0, 0)$. Since $n_1 > T/1000$, Lemma 7 implies that Alice plays action 1 for $1 - o(1)$ of these rounds.

- Bob then plays action 2 for $T/3$ rounds, moving the state to $(n_1, T/3, 0)$. By the guarantee of Lemma 8, Alice will still play action 1 for $1 - o(1)$ of these rounds.
- Bob then plays action 2 for $n_2 - T/3$ more rounds, moving the state to $(n_1, n_2, 0)$. This is at most $T/500$ rounds, which will be negligible in our final swap regret computation.
- Bob then plays action 3 for $(T/3) - (T/100)$ rounds, moving the state to $(n_1, n_2, T/3 - T/100)$. By the guarantee of Lemma 8, Alice will play action 2 for $1 - o(1)$ of these rounds.
- Bob then plays action 3 for the remaining rounds. This is at most $T/100$ rounds, which will be negligible in our final swap regret computation.

The key observation is that Alice would significantly improve her expected utility (by $\Omega(T)$) by playing action 3 every time she played action 2, and therefore Alice has linear swap regret. To see this, note that Alice plays action 2 for at least $(1 - o(1))((T/3) - (T/100)) \geq T/4$ rounds when Bob is playing action 3. On the other hand the number of rounds where both Alice and Bob play action 2 is at most the $T/500$ rounds in the third segment. Therefore the expected gain from switching from action 2 to action 3 is at least $T/4 - T/500 = \Omega(T)$, and therefore the algorithm $\mathcal{A}$ incurs $\Omega(T)$ swap regret. □

NeurIPS Paper Checklist

1. **Claims**

   Question: Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope?

   Answer: [Yes]

   Justification: This is a theoretical paper, and the claims made in the abstract and introduction don't overstate what's theoretically established in the paper.

   Guidelines:

   - The answer NA means that the abstract and introduction do not include the claims made in the paper.
   - The abstract and/or introduction should clearly state the claims made, including the contributions made in the paper and important assumptions and limitations. A No or NA answer to this question will not be perceived well by the reviewers.
   - The claims made should match theoretical and experimental results, and reflect how much the results can be expected to generalize to other settings.
   - It is fine to include aspirational goals as motivation as long as it is clear that these goals are not attained by the paper.

2. **Limitations**

   Question: Does the paper discuss the limitations of the work performed by the authors?

   Answer: [Yes]

   Justification: Conditions and assumptions for all results are clearly stated.

   Guidelines:

   - The answer NA means that the paper has no limitation while the answer No means that the paper has limitations, but those are not discussed in the paper.
   - The authors are encouraged to create a separate "Limitations" section in their paper.
   - The paper should point out any strong assumptions and how robust the results are to violations of these assumptions (e.g., independence assumptions, noiseless settings, model well-specification, asymptotic approximations only holding locally). The authors should reflect on how these assumptions might be violated in practice and what the implications would be.
   - The authors should reflect on the scope of the claims made, e.g., if the approach was only tested on a few datasets or with a few runs. In general, empirical results often depend on implicit assumptions, which should be articulated.
   - The authors should reflect on the factors that influence the performance of the approach. For example, a facial recognition algorithm may perform poorly when image resolution is low or images are taken in low lighting. Or a speech-to-text system might not be used reliably to provide closed captions for online lectures because it fails to handle technical jargon.
   - The authors should discuss the computational efficiency of the proposed algorithms and how they scale with dataset size.
   - If applicable, the authors should discuss possible limitations of their approach to address problems of privacy and fairness.
   - While the authors might fear that complete honesty about limitations might be used by reviewers as grounds for rejection, a worse outcome might be that reviewers discover limitations that aren't acknowledged in the paper. The authors should use their best judgment and recognize that individual actions in favor of transparency play an important role in developing norms that preserve the integrity of the community. Reviewers will be specifically instructed to not penalize honesty concerning limitations.

3. **Theory Assumptions and Proofs**

   Question: For each theoretical result, does the paper provide the full set of assumptions and a complete (and correct) proof?

   Answer: [Yes]

Justification: The proofs of all theorems and supporting lemmas are provided in the main text or the appendix.

Guidelines:

- The answer NA means that the paper does not include theoretical results.
- All the theorems, formulas, and proofs in the paper should be numbered and cross-referenced.
- All assumptions should be clearly stated or referenced in the statement of any theorems.
- The proofs can either appear in the main paper or the supplemental material, but if they appear in the supplemental material, the authors are encouraged to provide a short proof sketch to provide intuition.
- Inversely, any informal proof provided in the core of the paper should be complemented by formal proofs provided in appendix or supplemental material.
- Theorems and Lemmas that the proof relies upon should be properly referenced.

4. **Experimental Result Reproducibility**

   Question: Does the paper fully disclose all the information needed to reproduce the main experimental results of the paper to the extent that it affects the main claims and/or conclusions of the paper (regardless of whether the code and data are provided or not)?

   Answer: [NA]

   Justification: The paper does not include experiments. (We have produced some helpful illustrations with code, but these serve entirely to help illustrate the mathematical theorems in the text).

   Guidelines:

   - The answer NA means that the paper does not include experiments.
   - If the paper includes experiments, a No answer to this question will not be perceived well by the reviewers: Making the paper reproducible is important, regardless of whether the code and data are provided or not.
   - If the contribution is a dataset and/or model, the authors should describe the steps taken to make their results reproducible or verifiable.
   - Depending on the contribution, reproducibility can be accomplished in various ways. For example, if the contribution is a novel architecture, describing the architecture fully might suffice, or if the contribution is a specific model and empirical evaluation, it may be necessary to either make it possible for others to replicate the model with the same dataset, or provide access to the model. In general. releasing code and data is often one good way to accomplish this, but reproducibility can also be provided via detailed instructions for how to replicate the results, access to a hosted model (e.g., in the case of a large language model), releasing of a model checkpoint, or other means that are appropriate to the research performed.
   - While NeurIPS does not require releasing code, the conference does require all submissions to provide some reasonable avenue for reproducibility, which may depend on the nature of the contribution. For example
     (a) If the contribution is primarily a new algorithm, the paper should make it clear how to reproduce that algorithm.
     (b) If the contribution is primarily a new model architecture, the paper should describe the architecture clearly and fully.
     (c) If the contribution is a new model (e.g., a large language model), then there should either be a way to access this model for reproducing the results or a way to reproduce the model (e.g., with an open-source dataset or instructions for how to construct the dataset).
     (d) We recognize that reproducibility may be tricky in some cases, in which case authors are welcome to describe the particular way they provide for reproducibility. In the case of closed-source models, it may be that access to the model is limited in some way (e.g., to registered users), but it should be possible for other researchers to have some path to reproducing or verifying the results.

5. **Open access to data and code**

Question: Does the paper provide open access to the data and code, with sufficient instructions to faithfully reproduce the main experimental results, as described in supplemental material?

Answer: [NA]

Justification: The paper does not include experiments requiring code.

Guidelines:

- The answer NA means that paper does not include experiments requiring code.
- Please see the NeurIPS code and data submission guidelines (`https://nips.cc/public/guides/CodeSubmissionPolicy`) for more details.
- While we encourage the release of code and data, we understand that this might not be possible, so "No" is an acceptable answer. Papers cannot be rejected simply for not including code, unless this is central to the contribution (e.g., for a new open-source benchmark).
- The instructions should contain the exact command and environment needed to run to reproduce the results. See the NeurIPS code and data submission guidelines (`https://nips.cc/public/guides/CodeSubmissionPolicy`) for more details.
- The authors should provide instructions on data access and preparation, including how to access the raw data, preprocessed data, intermediate data, and generated data, etc.
- The authors should provide scripts to reproduce all experimental results for the new proposed method and baselines. If only a subset of experiments are reproducible, they should state which ones are omitted from the script and why.
- At submission time, to preserve anonymity, the authors should release anonymized versions (if applicable).
- Providing as much information as possible in supplemental material (appended to the paper) is recommended, but including URLs to data and code is permitted.

6. **Experimental Setting/Details**

   Question: Does the paper specify all the training and test details (e.g., data splits, hyperparameters, how they were chosen, type of optimizer, etc.) necessary to understand the results?

   Answer: [NA]

   Justification: The paper does not include experiments.

   Guidelines:

   - The answer NA means that the paper does not include experiments.
   - The experimental setting should be presented in the core of the paper to a level of detail that is necessary to appreciate the results and make sense of them.
   - The full details can be provided either with the code, in appendix, or as supplemental material.

7. **Experiment Statistical Significance**

   Question: Does the paper report error bars suitably and correctly defined or other appropriate information about the statistical significance of the experiments?

   Answer: [NA]

   Justification: The paper does not include experiments.

   Guidelines:

   - The answer NA means that the paper does not include experiments.
   - The authors should answer "Yes" if the results are accompanied by error bars, confidence intervals, or statistical significance tests, at least for the experiments that support the main claims of the paper.
   - The factors of variability that the error bars are capturing should be clearly stated (for example, train/test split, initialization, random drawing of some parameter, or overall run with given experimental conditions).
   - The method for calculating the error bars should be explained (closed form formula, call to a library function, bootstrap, etc.)

- The assumptions made should be given (e.g., Normally distributed errors).
- It should be clear whether the error bar is the standard deviation or the standard error of the mean.
- It is OK to report 1-sigma error bars, but one should state it. The authors should preferably report a 2-sigma error bar than state that they have a 96% CI, if the hypothesis of Normality of errors is not verified.
- For asymmetric distributions, the authors should be careful not to show in tables or figures symmetric error bars that would yield results that are out of range (e.g. negative error rates).
- If error bars are reported in tables or plots, The authors should explain in the text how they were calculated and reference the corresponding figures or tables in the text.

8. **Experiments Compute Resources**

   Question: For each experiment, does the paper provide sufficient information on the computer resources (type of compute workers, memory, time of execution) needed to reproduce the experiments?

   Answer: [NA]

   Justification: The paper does not include experiments.

   Guidelines:

   - The answer NA means that the paper does not include experiments.
   - The paper should indicate the type of compute workers CPU or GPU, internal cluster, or cloud provider, including relevant memory and storage.
   - The paper should provide the amount of compute required for each of the individual experimental runs as well as estimate the total compute.
   - The paper should disclose whether the full research project required more compute than the experiments reported in the paper (e.g., preliminary or failed experiments that didn't make it into the paper).

9. **Code Of Ethics**

   Question: Does the research conducted in the paper conform, in every respect, with the NeurIPS Code of Ethics `https://neurips.cc/public/EthicsGuidelines`?

   Answer: [Yes]

   Justification: The research conforms to the NeurIPS Code of Ethics.

   Guidelines:

   - The answer NA means that the authors have not reviewed the NeurIPS Code of Ethics.
   - If the authors answer No, they should explain the special circumstances that require a deviation from the Code of Ethics.
   - The authors should make sure to preserve anonymity (e.g., if there is a special consideration due to laws or regulations in their jurisdiction).

10. **Broader Impacts**

    Question: Does the paper discuss both potential positive societal impacts and negative societal impacts of the work performed?

    Answer: [NA]

    Justification: There is no societal impact of the work performed (it focuses on a fundamental theoretical question).

    Guidelines:

    - The answer NA means that there is no societal impact of the work performed.
    - If the authors answer NA or No, they should explain why their work has no societal impact or why the paper does not address societal impact.
    - Examples of negative societal impacts include potential malicious or unintended uses (e.g., disinformation, generating fake profiles, surveillance), fairness considerations (e.g., deployment of technologies that could make decisions that unfairly impact specific groups), privacy considerations, and security considerations.

- The conference expects that many papers will be foundational research and not tied to particular applications, let alone deployments. However, if there is a direct path to any negative applications, the authors should point it out. For example, it is legitimate to point out that an improvement in the quality of generative models could be used to generate deepfakes for disinformation. On the other hand, it is not needed to point out that a generic algorithm for optimizing neural networks could enable people to train models that generate Deepfakes faster.
- The authors should consider possible harms that could arise when the technology is being used as intended and functioning correctly, harms that could arise when the technology is being used as intended but gives incorrect results, and harms following from (intentional or unintentional) misuse of the technology.
- If there are negative societal impacts, the authors could also discuss possible mitigation strategies (e.g., gated release of models, providing defenses in addition to attacks, mechanisms for monitoring misuse, mechanisms to monitor how a system learns from feedback over time, improving the efficiency and accessibility of ML).

11. **Safeguards**

    Question: Does the paper describe safeguards that have been put in place for responsible release of data or models that have a high risk for misuse (e.g., pretrained language models, image generators, or scraped datasets)?

    Answer: [NA]

    Justification: The paper does not release models nor datasets.

    Guidelines:

    - The answer NA means that the paper poses no such risks.
    - Released models that have a high risk for misuse or dual-use should be released with necessary safeguards to allow for controlled use of the model, for example by requiring that users adhere to usage guidelines or restrictions to access the model or implementing safety filters.
    - Datasets that have been scraped from the Internet could pose safety risks. The authors should describe how they avoided releasing unsafe images.
    - We recognize that providing effective safeguards is challenging, and many papers do not require this, but we encourage authors to take this into account and make a best faith effort.

12. **Licenses for existing assets**

    Question: Are the creators or original owners of assets (e.g., code, data, models), used in the paper, properly credited and are the license and terms of use explicitly mentioned and properly respected?

    Answer: [NA]

    Justification: The paper does not use existing assets

    Guidelines:

    - The answer NA means that the paper does not use existing assets.
    - The authors should cite the original paper that produced the code package or dataset.
    - The authors should state which version of the asset is used and, if possible, include a URL.
    - The name of the license (e.g., CC-BY 4.0) should be included for each asset.
    - For scraped data from a particular source (e.g., website), the copyright and terms of service of that source should be provided.
    - If assets are released, the license, copyright information, and terms of use in the package should be provided. For popular datasets, `paperswithcode.com/datasets` has curated licenses for some datasets. Their licensing guide can help determine the license of a dataset.
    - For existing datasets that are re-packaged, both the original license and the license of the derived asset (if it has changed) should be provided.

- If this information is not available online, the authors are encouraged to reach out to the asset's creators.

13. **New Assets**

    Question: Are new assets introduced in the paper well documented and is the documentation provided alongside the assets?

    Answer: [NA]

    Justification: The paper does not release new assets

    Guidelines:

    - The answer NA means that the paper does not release new assets.
    - Researchers should communicate the details of the dataset/code/model as part of their submissions via structured templates. This includes details about training, license, limitations, etc.
    - The paper should discuss whether and how consent was obtained from people whose asset is used.
    - At submission time, remember to anonymize your assets (if applicable). You can either create an anonymized URL or include an anonymized zip file.

14. **Crowdsourcing and Research with Human Subjects**

    Question: For crowdsourcing experiments and research with human subjects, does the paper include the full text of instructions given to participants and screenshots, if applicable, as well as details about compensation (if any)?

    Answer: [NA]

    Justification: The paper does not involve crowdsourcing nor research with human subjects

    Guidelines:

    - The answer NA means that the paper does not involve crowdsourcing nor research with human subjects.
    - Including this information in the supplemental material is fine, but if the main contribution of the paper involves human subjects, then as much detail as possible should be included in the main paper.
    - According to the NeurIPS Code of Ethics, workers involved in data collection, curation, or other labor should be paid at least the minimum wage in the country of the data collector.

15. **Institutional Review Board (IRB) Approvals or Equivalent for Research with Human Subjects**

    Question: Does the paper describe potential risks incurred by study participants, whether such risks were disclosed to the subjects, and whether Institutional Review Board (IRB) approvals (or an equivalent approval/review based on the requirements of your country or institution) were obtained?

    Answer: [NA]

    Justification: The paper does not involve crowdsourcing nor research with human subjects

    Guidelines:

    - The answer NA means that the paper does not involve crowdsourcing nor research with human subjects.
    - Depending on the country in which research is conducted, IRB approval (or equivalent) may be required for any human subjects research. If you obtained IRB approval, you should clearly state this in the paper.
    - We recognize that the procedures for this may vary significantly between institutions and locations, and we expect authors to adhere to the NeurIPS Code of Ethics and the guidelines for their institution.
    - For initial submissions, do not include any information that would break anonymity (if applicable), such as the institution conducting the review.