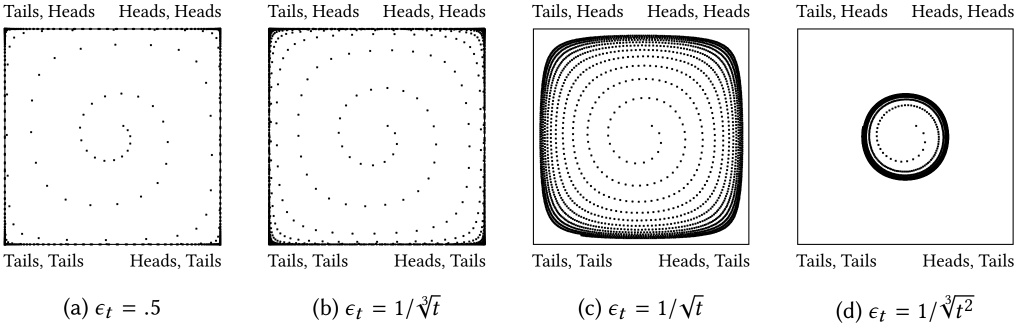
Multiplicative Weights Update in Zero-Sum Games

JAMES P. BAILEY, Singapore University of Technology and Design
GEORGIOS PILIOURAS, Singapore University of Technology and Design

We study the classic setting where two agents compete against each other in a zero-sum game by applying the Multiplicative Weights Update (MWU) algorithm. In a twist of the standard approach of [11], we focus on the K-L divergence from the equilibrium but instead of providing an upper bound about the rate of increase we provide a nonnegative lower bound for games with interior equilibria. This implies movement away from equilibria and towards the boundary. In the case of zero-sum games without interior equilibria convergence to the boundary (and in fact to the minimal product of subsimplexes that contains all Nash equilibria) follows via an orthogonal argument. In that subspace divergence from the set of NE applies for all nonequilibrium initial conditions via the first argument. We argue the robustness of this non-equilibrating behavior by considering the following generalizations:

- **Step size:** Agents may be using different and even decreasing step sizes.
- **Dynamics:** Agents may be using Follow-the-Regularized-Leader algorithms and possibly apply different regularizers (e.g. MWU versus Gradient Descent). We also consider a linearized version of MWU.
- **More than two agents:** Multiple agents can interact via arbitrary networks of zero-sum polymatrix games and their affine variants.

Our results come in stark contrast with the standard interpretation of the behavior of MWU (and more generally regret minimizing dynamics) in zero-sum games, which is typically referred to as "converging to equilibrium". If equilibria are indeed predictive *even for the benchmark class of zero-sum games*, agents in practice must deviate robustly from the axiomatic perspective of optimization driven dynamics as captured by MWU and variants and apply carefully tailored equilibrium-seeking behavioral dynamics.



(a) $\epsilon_t = .5$ (b) $\epsilon_t = 1/\sqrt[3]{t}$ (c) $\epsilon_t = 1/\sqrt{t}$ (d) $\epsilon_t = 1/\sqrt[3]{t^2}$

Fig. 1. Player Strategies Spiraling Outwards in Matching Pennies when Updated with 2500 Iterations of Multiplicative Weights. All Experiments Start with the Same Initial Condition.

Authors' addresses: James P. Bailey, Singapore University of Technology and Design, james_bailey@sutd.edu.sg; Georgios Piliouras, Singapore University of Technology and Design, georgios@sutd.edu.sg.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.
© 2018 Copyright held by the owner/author(s). Publication rights licensed to ACM.
ACM EC'18, June 18–22, 2018, Ithaca, NY, USA. ACM ISBN 978-1-4503-4529-3/18/06…$15.00
https://doi.org/10.1145/3219166.3219235

1 INTRODUCTION

We study arguably the most well known online learning dynamic, Multiplicative Weights Update (MWU) [12, 16] in the most well studied class of games (zero-sum games). Naturally, the prominence of each of these objects in their respective fields, i.e., the centrality of MWU for online learning, optimization and classification and similarly of zero-sum games for game theory is indisputable. In fact, both of these objects as well as the question about how does MWU behave in zero-sum games is a subject matter that is typically studied in undergrad and graduate courses on these subjects.

Our current understanding and interpretation of this setting revolves around a classic work by Freund and Schapire [11]. Their analysis focuses on establishing strong regret guarantees. In the case of a zero-sum game this produces a simple proof of von Neumann's min-max theorem and a provable method of approximately solving a game. Specifically, given that both agents apply (MWU), both the time average of the mixed strategy profiles as well as the utility of both agents converge approximately to their Nash equilibrium values, where the approximation error can become arbitrarily close to zero by choosing a sufficiently small step size. This result extends straightforwardly to general no-regret dynamics, and, in fact, the other desirable properties of the Nash equilibria such as polynomial-time tractability and convexity can be shown to readily follow from this close connection between no-regret learning and equilibration.

In the game theoretic literature, this major result is typically celebrated by the shorthand statement "MWU (no-regret dynamics) converges to equilibria in zero-sum games". Naturally, this statement is not formally accurate as it indicates a much stronger result, i.e., that the actual day-to-day behavior of MWU converges to equilibria. In spite of the classic nature of the problem, we are not aware of any formal analysis of the asymptotic behavior of MWU even for a specific instance of a zero-sum game (e.g., Matching Pennies).

To some extent, this lack of formal theoretical arguments is not surprising, as even a single dimensional discrete time dynamical system can exhibit totally unpredictable, chaotic behavior that is hard to characterize theoretically [15]. To make matters worse, recent results strongly suggest that the behavior of MWU in 2x2 coordination/potential games[1] is impossibly hard to completely characterize and predict. Specifically, [21] has shown that there exist specific instances of 2x2 potential games where the behavior of MWU exhibits bifurcations at critical values of its step size. For small values the system always converges to equilibria whereas if we keep increasing the step size eventually limit cycles emerge. For a different 2x2 potential game it is proven that MWU can exhibit chaotic behavior. Simulations of this system suggest that as we increase the step size of MWU, the system undergoes an infinite number of period doubling phase transitions. This behavior is not universal in 2x2 potential games as there seem to exist instances where it is impossible to induce chaotic behavior no matter how large we choose the step size to be. If the behavior MWU is so complex in 2x2 potential games, why should we expect it to be any simpler in 2x2 zero-sum games, let alone general constant-sum games?

Well, at least in the case of Matching Pennies, the simulations of MWU in Figure 1 seem to suggest an intuitively clear story. Given any fixed step size, e.g., $\epsilon = .5$ the dynamic converges to the boundary with the rate of convergence being an increasing function of ϵ. For decreasing step sizes, e.g., $\epsilon_t = 1/\sqrt[3]{t}$ or $1/\sqrt{t}$, the system moves away from the equilibrium along a spiral. Interestingly enough, we get qualitatively different limit behavior depending on the rate of decrease of the step size with the $1/\sqrt{t}$ rate encoding the boundary between the two. For slower decreasing step sizes we still quickly converge to the boundary, whereas for faster rates the system converges to a closed curve in the interior. With hindsight, this curve is the boundary of a K-L divergence ball.

[1] These are games that lie on the antipode of zero-sum games. In coordination/potential games, the agents' incentives are strongly aligned and all agents act as if they wish to maximize a common potential function.

That is, this set contains all points at a fixed K-L divergence from the equilibrium. For the special case of $1/\sqrt{t}$ we still diverge to the boundary. We will show that this intuitive picture largely carries over to general zero-sum games, and furthermore present several extensions for different games and dynamics.

1.1 Results

We begin by showing that the K-L divergence from any fully mixed Nash equilibrium (NE) to the player strategies is increasing when players update their strategies using the MWU. This implies that the set of Nash equilibria repel player strategies explaining the outward spirals we observe in Figure 1. Building on this result, we establish that player strategies converge to the boundary for any constant ϵ. Importantly, this analysis is robust. Our results extend to constant-sum polymatrix games and even if individual players opt to use different values for ϵ. Moreover, our results apply even to the setting where agents use shrinking step sizes. Formally, player strategies converge to the boundary iff $\sum_{t=1}^{\infty} \epsilon_t^2 = \infty$ perfectly describing the contrast between the different phase portraits in Figure 1.

Otherwise, in the event where the only Nash equilibrium of the game is on the boundary, the K-L divergence may actually decrease in an iteration of MWU. Using different proof techniques we establish that player strategies converge to the smallest face containing the set of Nash equilibria. In that subspace divergence away from the set of NE applies for all nonequilibrium initial conditions via the first argument, completing our picture for the behavior of the MWU dynamics.

Finally we consider other standard regret minimizing dynamics. We show how our results carry over to the more general class of Follow-the-Regularized-Leader algorithms. In addition, our results hold for the linear version of MWU[2] in skew-symmetric games.

1.2 Related Work

The study of learning dynamics in game theory has a long history dating back to the work of Brown and Robinson [2, 26] on fictitious play in zero-sum games, which followed shortly after von Neumann's seminal work on zero-sum games [30, 31]. A good reference book cataloguing these developments is [13]. The classic results about time-average convergence of no-regret dynamics have been successfully generalized to include multiplayer extensions of network constant-sum games [3, 4, 9].

Non-equilibrating dynamics in game theory. Although traditionally research in the area aims at proving results about convergence of dynamics to equilibria, within the algorithmic game theory (AGT) community, there has been a steady stream of results cataloguing interesting non-equilibrium effects. Proving such non-equilibrium effects is typically rather hard, and results in this area typically revolve around specific examples of games with a handful of agents and strategies. [7] showed that MWU does not converge even in a time-average sense in the case of a specific 3x3 game. [14] established non-convergence for a continuous-time variant of MWU, known as the replicator dynamic, for a 2x2x2 game and show that as a result the system social welfare converges to states that dominate all Nash equilibria. Balcan et al. [1] studied MWU in rank-1 games (i.e., games where the summation of the payoff matrices of the two agents results in a rank-1 matrix). This games are in some sense almost constant-sum, and the paper shows that there exist such games where not even the time-average of MWU converges to equilibria. [21] proved the existence of Li-Yorke chaos in MWU dynamics of 2x2 potential games.

[2] This is a well known variant of MWU, where the weights are updated as $w \leftarrow w(1+\epsilon)^u$ they are updated as $w \leftarrow w(1+\epsilon u)$. This variant enjoys strong connections to biological and evolutionary dynamics [5, 18, 19].

Continuous time dynamics in game theory. From the perspective of evolutionary game theory, which typically studies continuous time dynamics, numerous nonconvergence results are known but again typically for small games [27]. In [23, 24], the authors show that replicator dynamics, the continuous time version of MWU exhibit a specific type of near periodic behavior, which is known as Poincaré recurrence. Recently, [20] showed how to generalize this cyclic behavior for replicator to more general continuous time variants of follow the regularized leader (FTRL) dynamics. Finally, [17] showed that this arguments can also be adapted in the case of dynamically evolving games. The papers in this category combine delicate arguments such as volume preservation and the existence of constants of motions for the dynamics to establish cyclic behavior. In the case of discrete time dynamics, such as the MWU, the system trajectories are more "rough" and these arguments which are suited for smooth dynamics are no longer valid. Finally, [22] has put forward a program for a general connection between game theoretic dynamics and concepts in topology that holds the promise of becoming a universal tool for analyzing non-equilibrium dynamics in games.

Fast convergence to low regret states. It is widely known that the time-average of no-regret algorithms converge to the set of coarse correlated equilibria. The "black-box" rate of convergence is $O(1/\sqrt{t})$ and it is achieved by MWU with suitably shrinking step size without making any assumptions about its environment. Recently, several authors have focused instead on obtaining stronger regret guarantees for systems of learning algorithms in games. [6] and [25] developed dynamics with a $O(\log t/t)$ regret minimization rate in two-player zero-sum games. [28] further analyzed a recency biased variant of follow the regularized leader (FTRL) in more general games and showed a $O(t^{-3/4})$ regret minimization rate. The social welfare converges at a rate of $O(1/t)$, a result which was extended to standard versions of FRTL dynamics in [10]. Finally, [8] studies the convergence properties of a specifically tailored dynamic in bilinear saddle problems and show convergence of the daily behavior to equilibrium. Based on this, they develop novel training algorithms for generative adversarial neural networks (GANs).

2 DEFINITIONS

2.1 Normal Form Games

We begin with basic definitions from game theory. A finite normal-form game $\Gamma \equiv \Gamma(\mathcal{N}, \mathcal{S}, u)$ consists of a set of players $\mathcal{N} = \{1, ..., N\}$ where player i may select from a finite set of actions or pure strategies $\mathcal{S}_i$. Each player has a payoff function $u_i : \mathcal{S} \equiv \prod_i \mathcal{S}_i \to \mathbb{R}$ assigning reward $u_i(s)$ to player i. It is common to describe u_i with a payoff tensor $A^{(i)}$ where $u_i(s) = A^{(i)}_s$.

Players are also allowed to use mixed strategies $x_i = (x_{is_i})_{s_i \in \mathcal{S}_i} \in \Delta(\mathcal{S}_i) \equiv \mathcal{X}_i$. The set of mixed strategies is $\mathcal{X} = \prod_i \mathcal{X}_i$. A strategy is fully mixed if $x_{is_i} > 0$ for all $s_i \in \mathcal{S}_i$ and $i \in \mathcal{N}$. Individuals the payoff of a mixed strategy linearly using expectation. Formally,

$$u_i(x) = \sum_{s \in \mathcal{S}} u_i(s) \prod_{i \in \mathcal{N}} x_{is_i}. \tag{1}$$

We also introduce additional notation to express player payouts for brevity in our analysis later. Let $v_{is_i}(x) = u_i(s_i; x_{-i})$[3] denote the reward i receives if i opts to play pure strategy s_i when everyone else commits to their strategies described by x. This results in $u_i(x) = \langle v_i(x), x_i \rangle$. Let $P_i(x) = v_{is_i}(x)$ with probability x_{is_i} be a random variable corresponding to the payout for player i given the mixed profile x. Using this definition, only player i introduces randomness to i's reward and x_{-i} is treated as a deterministic strategy. We can then write $u_i(x) = E[P_i(x)]$.

[3] $(s_i; x_{-i})$ denotes the strategy x after replacing x_i with s_i.

The most commonly used solution concept for games is the *Nash equilibrium.* A Nash equilibrium (NE) is a strategy $x^* \in \mathcal{X}$ where no player can do better by deviating from x_i^*. Formally,

$$u_i(x_i^*; x_{-i}^*) \geq u_i(x_i; x_{-i}^*) \text{ for all } x_i \in \mathcal{X}_i \text{ and } i \in \mathcal{N}. \tag{NE}$$

2.2 2-Player and Polymatrix Constant-Sum Games

The most commonly studied class of games are 2-player zero-sum games. A two-player zero-sum game Γ is such that $\mathcal{N} = \{1, 2\}$ and $u_1 + u_2 = 0$. Letting $u \equiv u_1 = -u_2$, the value of a 2-player zero-sum game is

$$u_\Gamma = \max_{x_1 \in \mathcal{X}_1} \min_{x_2 \in \mathcal{X}_2} u(x_1; x_2) = \min_{x_2 \in \mathcal{X}_2} \max_{x_1 \in \mathcal{X}_1} u(x_1; x_2) \tag{2}$$

where equality comes from von Neuman's min-max theorem [30]. The solutions to (2) are the set of Nash equilibria of the game.

Constant-sum games are closely related to zero-sum games. A constant-sum game with parameter c is such that $u_1 + u_2 = c$. By shifting either players utility function, a two-player constant-sum game can be transformed into a zero-sum game without changing the set of Nash equilibria.

Also of interest in this paper are games featuring a network of competitors. An N-player pairwise constant-sum polymatrix game Γ consists of an interaction graph $\mathcal{G} = \mathcal{G}(\mathcal{N}, \mathcal{E})$ where the set of nodes $\mathcal{N}$ represent of players and where $\{i, j\}$ is an edge in $\mathcal{E}$ only if i and j compete in a constant-sum game with parameter c_{ij}.

Player i's utility, u_i is now expressed as a sum of the utilities gained in the games i plays in the graph $\mathcal{G}$. Formally, let $u_{ij} = c_{ij} - u_{ji}$ be utility gained in i's game against j and

$$u_i(x) = \sum_{j:\{i,j\} \in \mathcal{E}} u_{ij}(x). \tag{3}$$

Without loss of generality we can assume $\mathcal{G}$ is a complete graph by letting $u_{ij}(x) = 0$ for all x if i and j do not compete. Under this assumption, we can write the simpler $u_i(x) = \sum_{j \in \mathcal{N} \setminus \{i\}} u_{ij}(x)$.

2.3 Bregman Divergence from a Nash Equilibrium

Let $x^* \in \mathcal{X}$ be a Nash equilibrium and let $x \in \mathcal{X}$ be an arbitrary strategy profile. Then the Bregman divergence from x^* to x with respect to convex function h is

$$D_h(x^* || x) = h(x^*) - h(x) - \langle \nabla h(x), x^* - x \rangle. \tag{Bregman Divergence}$$

A convex function we will be particularly interested in the course of this paper is the negative entropy function, $h(x) = \sum_{s_i \in \mathcal{S}_i} x_{is_i} \ln(x_{is_i})$. The Bregman divergence for this function is referred to as the Kullback-Leibler (K-L) divergence and is given by

$$D_{KL}(x^* || x) = \sum_{i \in \mathcal{N}} \sum_{s_i \in \mathcal{S}_i} x_{is_i}^* \left(\ln x_{is_i}^* - \ln x_{is_i} \right) \tag{K-L Divergence}$$

2.4 Follow-the-Regularized-Leader

$$x_i^t = \operatorname*{argmax}_{x_i \in \mathcal{X}_i} \eta \sum_{s=0}^{t-1} u_i(x_i; x_{-i}^s) - h(x_i) \tag{FTRL}$$

If the negative entropy function is used as the regularizer, then FTRL results in the exponential weighted update algorithm given by

$$x_{is_i}^t = \frac{x_{is_i}^{t-1} (1+\epsilon)^{v_{is_i}(x^{t-1})}}{\sum_{s_i' \in \mathcal{S}_i} x_{is_i'}^{t-1} (1+\epsilon)^{v_{is_i'}(x^{t-1})}} = \frac{x_{is_i}^t (1+\epsilon)^{v_{is_i}(x^{t-1})}}{E\left[(1+\epsilon)^{P_i(x^{t-1})}\right]} \tag{MWU_e}$$

where $\epsilon = e^{\eta} - 1$.

We will also consider the Linear Multiplicative Weight Update algorithm. While Linear MWU is not a FTRL algorithm, it is a regret minimizing algorithm for decaying values of ϵ. It is given by

$$x^t_{is_i} = \frac{x^{t-1}_{is_i}(1 + \epsilon v_{is_i}(x^{t-1}))}{\sum_{s'_i \in \mathcal{S}_i} x^{t-1}_{is'_i}(1 + \epsilon v_{is'_i}(x^{t-1}))} = \frac{x^{t-1}_{is_i}(1 + \epsilon v_{is_i}(x^{t-1}))}{1 + \epsilon u_i(x^{t-1})} \tag{MWU_ℓ}$$

3 CONVERGENCE TO BOUNDARY IN MULTIPLICATIVE WEIGHTED UPDATE

We start our analysis by proving the repelling property of Nash equilibria.

Theorem 3.1. *The K-L divergence between player strategies and any fully mixed Nash equilibrium is non-decreasing when strategies are updated with (MWU_e) for any 2-player constant-sum game.*

Proof. Plugging (MWU_e) into (K-L Divergence), we obtain

$$\begin{aligned}
D_{KL}(x^*||x^t) &= \sum_{i \in \mathcal{N}} \sum_{s_i \in \mathcal{S}_i} x^*_{is_i} \left(\ln x^*_{is_i} - \ln x^t_{is_i}\right) && (4)\\
&= \sum_{i \in \mathcal{N}} \sum_{s_i \in \mathcal{S}_i} x^*_{is_i} \left(\ln x^*_{is_i} - \ln x^{t-1}_{is_i} - v_{is_i}(x^{t-1}) \ln(1+\epsilon) + \ln E\left[(1+\epsilon)^{P_i(x^{t-1})}\right]\right) && (5)\\
&= D_{KL}(x^*||x^{t-1}) + \sum_{i \in \mathcal{N}} \sum_{s_i \in \mathcal{S}_i} x^*_{is_i} \left(\ln E\left[(1+\epsilon)^{P_i(x^{t-1})}\right] - v_{is_i}(x^{t-1}) \ln(1+\epsilon)\right) && (6)\\
&= D_{KL}(x^*||x^{t-1}) + \sum_{i \in \mathcal{N}} \left(\ln E\left[(1+\epsilon)^{P_i(x^{t-1})}\right] - u_i(x^*_i; x^{t-1}_{-i}) \ln(1+\epsilon)\right) && (7)\\
&= D_{KL}(x^*||x^{t-1}) + \sum_{i \in \mathcal{N}} \left(\ln E\left[(1+\epsilon)^{P_i(x^{t-1})}\right] - u_i(x^{t-1}) \ln(1+\epsilon)\right) && (8)
\end{aligned}$$

The equlibrium x^* is fully mixed implying $u_i(x^*_i, x^{t-1}_{-i}) = u_i(x^*)$ and $\sum_{i \in \mathcal{N}} u_i(x^*_i; x^{t-1}_{-i}) = \sum_{i \in \mathcal{N}} u_i(x^*) = \sum_{i \in \mathcal{N}} u_i(x^{t-1})$ since the game is constant-sum. Therefore in the t^{th} iteration of (MWU_e) the K-L divergence changes by

$$\begin{aligned}
D_{KL}(x^*||x^t) - D_{KL}(x^*||x^{t-1}) &= \sum_{i \in \mathcal{N}} \ln E\left[(1+\epsilon)^{P_i(x^{t-1}) - u_i(x^{t-1})}\right] && (9)\\
&\geq \sum_{i \in \mathcal{N}} E\left[P_i(x^{t-1}) - u_i(x^{t-1})\right] \ln(1+\epsilon) = 0 && (10)
\end{aligned}$$

where the last inequality follows from Jensen's Inequality. □

Definition 3.2. Strategy $s_i \in \mathcal{S}_i$ is *essential* iff there is a Nash equilibrium x^* where $x^*_{is_i} > 0$.

It is straightforward to check that there is no fully mixed Nash equilibrium iff there is a non-essential strategy. The backward direction follows by definition, whereas the forward follows from the convexity of the equilibrium set.

We are now ready to prove our main result.

Theorem 3.3. *For almost every 2-player constant-sum game, there exists $\epsilon_0 > 0$, such that as long as all agents use (MWU_e) with $\epsilon < \epsilon_0$, all non-equilibrium initial conditions converge to the boundary.*

Specifically, in any game that has at least one interior NE, for all ϵ, all non-equilibrium initial conditions converge to the boundary. In any game with no interior NE, as long as it has a unique NE (a genericity assumption), all non-equilibrium initial conditions converge to the minimal subspace that contains all essential strategies.

We present the proof of convergence in two parts. First we show that if there is a fully mixed Nash equilibrium then the K-L divergence between the Nash equilibrium and the player strategies goes to infinity implying convergence to the boundary. This portion of the proof works for any ϵ. In the second part of the proof, for the case of games with non-interior Nash equilibria, we show that under the generic assumption that the zero-sum game in question has a unique equilibrium[4] the probability of playing any non-essential strategy goes to 0 completing the proof of the theorem.

Proof of Theorem 3.3. First suppose there is a fully mixed Nash equilibrium x^*. By Lemma 3.1, $D_{KL}(x^*||x^t)$ is non-decreasing and there is a compact set B such that $x^t \in B$ for all t and $x^* \notin B$.

For contradiction, suppose x^t does not converge to the boundary. This implies there exists $w > 0$ such $D_{KL}(x^*||x^t) \leq w$ for all t since $D_{KL}(x^*||x^t)$ is non-decreasing. Thus we may assume that B also excludes the boundary. (10) is saturated only if $x^t = x^*$ or x^t is on the boundary. Therefore $D_{KL}(x^*||x^t) - D_{KL}(x^*||x^{t-1}) > 0$ for all x^t since $x^t \in B$. $D_{KL}(x^*||x^t) - D_{KL}(x^*||x^{t-1})$ is continuous with respect to x^t and each x^t is in the compact B. Thus, there is a constant $d > 0$ such that $D_{KL}(x^*||x^t) - D_{KL}(x^*||x^{t-1}) \geq d$. This implies

$$\lim_{t\to\infty} D_{KL}(x^*||x^t) = D_{KL}(x^*||x^0) + \sum_{t=1}^{\infty} \left(D_{KL}(x^*||x^t) - D_{KL}(x^*||x^{t-1})\right) \tag{11}$$

$$\geq D_{KL}(x^*||x^0) + \sum_{t=0}^{\infty} d = \infty \tag{12}$$

and thus we reach a contradiction.

Now suppose there is no fully mixed Nash equilibrium and that player i has a non-essential strategy s_i. Iteratively, the update (MWU_e) can be written as

$$x^t_{is_i} = \frac{x^0_{is_i}(1+\epsilon)^{t\cdot v_{is_i}(\bar{x}^{t-1})}}{E\left[(1+\epsilon)^{t\cdot P_i(\bar{x}^{t-1})}\right]} \tag{13}$$

where $\bar{x}^t = \sum_{s=0}^{t} \frac{x^s}{t+1}$ is the time average of player strategies and where the expectation is taken with respect to x_i^{t-1}. Since (MWU_e) has an $O(\epsilon)$ time-average regret, $\bar{x}^t$ converges to a $O(\epsilon)$-approximate Nash equilibrium.[5] Since for $\epsilon \to 0$ the set of $O(\epsilon)$-approximate Nash equilibria converges to the set of Nash equilibria, given any open neighborhood around the set of Nash equilibria we can choose an appropriate ϵ so that all the $O(\epsilon)$-approximate Nash equilibria are contained in that neighborhood.

For this portion of the proof we assume without loss of generality that $u_i(x^*) = 0$. Almost every constant-sum game has a unique Nash equilibrium x^* [29]. Moreover, by [20] we have that that for every non-essential strategy i, $v_{is_i}(x^*) < u_i(x^*) = 0$. Let $\delta = v_{is_i}(x^*) < 0$. By the continuity of payoffs we have that for any mixed strategy profile y in a small enough neighborhood around the equilibrium set, $v_{is_i}(y) < 2\delta/3$ whereas for any strategy s_i' in the equilibrium support $v_{is_i'}(y) > \delta/3$. Let ϵ_0 in the statement of the theorem be such that all $O(\epsilon_0)$-approximate Nash equilibria as strictly

[4] The set of zero-sum games with a unique equilibrium is open and dense in the space of all zero-sum games. Moreover, this slightly stronger result is also true. Within the set of all zero-sum games, the complement of the set of zero-sum games with unique equilibrium is closed and has Lebesgue measure zero [29].

[5] This is a well known fact. By the $O(\epsilon)$-regret property of the second agent, the time average utility of the first agent cannot be larger than $u_\Gamma + O(\epsilon)$. By the $O(\epsilon)$-regret property of the first agent, his utility for deviating to any strategy is at most $O(\epsilon)$ greater that his current time average utility which is at most $u_\Gamma + O(\epsilon)$. Specifically, his expected utility at $\bar{x}^t$ when both agents play their time average strategies is at most $u_\Gamma + O(\epsilon)$. A similar argument from the perspective of the second agent produces the analogous lower bound. This strategy profile is thus an $O(\epsilon)$ equilibrium of the game.

contained in the above neighborhood. We have that $t \cdot \left(v_{is_i}(\bar{x}^t) - v_{is'_i}(\bar{x}^t)\right) \to -\infty$ as $t \to \infty$. As a result, $x^t_{is_i}/x^t_{is'_i} \to 0$ as $t \to \infty$ and since $x^t_{is'_i}$ is upper bounded by 1, $x^t_{is_i} \to 0$ as $t \to \infty$. This completes the proof of the theorem. □

3.1 Exponential MWU Convergence in Polymatrix Games

THEOREM 3.4. *For every polymatrix constant-sum game with a fully mixed Nash equilibrium, every (non-equilibrium) initial strategy converges to the boundary when updated with (MWU_e) for any* $\epsilon > 0$.

Let x denote an arbitrary strategy profile. Observe that

$$\sum_{i\in\mathcal{N}} u_i(x_i^*; x_{-i}) = \sum_{i\in\mathcal{N}} \sum_{j\in\mathcal{N}\setminus\{i\}} u_{ij}(x_i^*; x_j) \tag{14}$$

$$= \sum_{j\in\mathcal{N}} \sum_{i\in\mathcal{N}\setminus\{j\}} u_{ij}(x_i^*; x_j) \tag{15}$$

$$= \sum_{j\in\mathcal{N}} \sum_{i\in\mathcal{N}\setminus\{j\}} u_{ij}(x^*) \tag{16}$$

$$= \sum_{i\in\mathcal{N}} u_i(x^*) \tag{17}$$

$$= \sum_{i\in\mathcal{N}} u_i(x) \tag{18}$$

The proof of Theorem 3.4 then follows analogously to Theorems 3.1 and 3.3.

3.2 Exponential MWU with Different Step Sizes

Players may opt to use a different value of ϵ in each iteration of (MWU_e). For instance, a common selection is $\epsilon_t = \frac{1}{\sqrt{t}}$ which guarantees vanishing regret. We extend our results to this setting.

$$x^t_{is_i} = \frac{x^{t-1}_{is_i}(1+\epsilon_t)^{v_{is_i}(x^{t-1})}}{E\left[(1+\epsilon_t)^{P_i(x^{t-1})}\right]} \tag{$MWU_e^{\epsilon_t}$}$$

THEOREM 3.5. *For every polymatrix constant-sum game with a fully mixed Nash equilibrium, every (non-equilibrium) initial strategy converges to the boundary when updated with ($MWU_e^{\epsilon_t}$) iff* $\sum_{t=1}^{\infty} \epsilon_t^2 = \infty$.

The proof of Theorem 3.5 requires showing that the $K-L$ divergence increases by approximately $\sum_{i\in\mathcal{N}} \frac{Var[P_i(x^t)]}{2}\epsilon_t^2$. The remainder of the proof then follows analogously to Theorem 3.3. The details of this proof appear in Appendix A.

In practice, there is no reason all players update their weights using the same ϵ_t, i.e., players i may instead use weight ϵ_{it}. As we show next, the result holds even if players all use different values for ϵ_t under certain conditions.

THEOREM 3.6. *The statement of Theorem 3.5 holds even if players are allowed to select different values for* ϵ_{it}, *as long as* $\frac{\prod_{k\in\mathcal{N}\setminus\{i\}}\ln(1+\epsilon_{kt})}{\sum_{j\in\mathcal{N}}\ln(1+\epsilon_{jt})}$ *is time invariant for all* $i \in \mathcal{N}$.

The proof of Theorem 3.6 requires examining $\sum_{i\in\mathcal{N}} \frac{\prod_{k\in\mathcal{N}\setminus\{i\}}\ln(1+\epsilon_{kt})}{\sum_{j\in\mathcal{N}}\ln(1+\epsilon_{jt})} D_{KL}(x_i^*||x_i^t) \leq c \cdot D_{KL}(x^*||x^t)$ for some constant $c > 0$ and then follows analogously to Theorems 3.1 and 3.5. As a corollary, for arbitrary, fixed but possibly different step-sizes ϵ_i (MWU_e) converges to the boundary.

While Theorem 3.6 guarantees there are many possible selections of ϵ_{it} that guarantee convergence of the boundary, we also show that ϵ_{it} may be selected so that the K-L divergence between player strategies and any fully mixed Nash equilibrium may actually decrease.

Proposition 3.7. *The K-L divergence between player strategies and any fully mixed Nash equilibrium may decrease when strategies are updated with (MWU_e) if players are allowed to select values for ϵ_{it} arbitrarily.*

The proof of Proposition 3.7 appears in Appendix B.

4 EXTENSIONS

In this section we examine other regret-minimizing dynamics. Specifically, we consider (i) Follow-the-Regularized-Leader algorithms (FTRL) and (ii) the linear version of MWU (MWU_ℓ) in strictly adversarial settings and establish divergence away from Nash equilibria.

4.1 FTRL

Theorem 4.1. *For every constant-sum game with a fully mixed Nash equilibrium and every (non-equilibrium) initial strategy, the Bregman divergence from the NE will strictly increase in every iteration when player strategies are updated with (FTRL) and a strictly convex regularizer, as long as the updated strategies are fully mixed.*

Proof. The *KKT* optimality conditions for (FTRL) are

$$\begin{aligned}
\nabla h_i(x_i^t) &= \eta \sum_{s=0}^{t-1} v_i(x_{-i}^s) - \lambda^t \mathbf{1} + \delta^t \\
\sum_{s_i \in \mathcal{S}_i} x_{is_i}^t &= 1 \\
x_{is_i} &\geq 0 \\
\delta^t &\geq 0 \\
\langle \delta^t, x_i^t \rangle &= 0
\end{aligned} \tag{FTRL KKT Conditions}$$

where $\lambda^t \in \mathbb{R}$ is the variable associated with the constraint $\sum_{s_i \in \mathcal{S}_i} x_{is_i}^t = 1$ and $\delta^t \in \mathbb{R}_{\geq 0}^{|\mathcal{S}_i|}$ is the vector associated with the constraints $x_{is_i} \geq 0$ for all $s_i \in \mathcal{S}_i$. Therefore,

$$\begin{aligned}
D_h(x_i^* || x_i^t) &= h(x_i^*) - h(x_i^t) - \langle \nabla h(x_i^t), x_i^* - x_i^t \rangle && (19)\\
&= h(x_i^*) - h(x_i^t) - \langle \eta \sum_{s=0}^{t-1} v_i(x_{-i}^s) - \lambda^t \mathbf{1} + \delta^t, x_i^* - x_i^t \rangle && (20)\\
&= h(x_i^*) - h(x_i^t) + \eta \sum_{s=0}^{t-1} \left(u_i(x_i^t, x_{-i}^s) - u_i(x_i^*, x_{-i}^s) \right) - \langle \delta^t, x_i^* \rangle && (21)\\
&\leq h(x_i^*) - h(x_i^t) + \eta \sum_{s=0}^{t-1} \left(u_i(x_i^t, x_{-i}^s) - u_i(x^*) \right). && (22)
\end{aligned}$$

Now suppose x_i^t is fully mixed implying $\delta^t = \mathbf{0}$ and (22) holds with equality. Through similar reasoning and by strict convexity of h,

$$h(x_i^{t-1}) > h_i(x_i^t) + \langle \nabla h_i(x_i^t), x_i^{t-1} - x_i^t \rangle \tag{23}$$

$$= h_i(x_i^t) + \eta \sum_{s=0}^{t-1} \left(u_i(x_i^{t-1}, x_{-i}^s) - u_i(x_i^t; x_{-i}^s) \right) \tag{24}$$

Combining (22) and (24), we have

$$D_h(x^*||x^t) - D_h(x^*||x^{t-1}) = \sum_{i \in \mathcal{N}} \left(D_h(x_i^*||x_i^t) - D_h(x_i^*||x_i^{t-1}) \right) \tag{25}$$

$$> \sum_{i \in \mathcal{N}} \eta \left(u_i(x^{t-1}) - u_i(x^*) \right) = 0 \tag{26}$$

completing the proof of the theorem. □

Theorem 4.1 guarantees that when the strategy and its update are fully mixed the update will move further from the Nash equilibrium explaining the outward spirals in Figure 1. Moreover, it guarantees that the Bregman divergence continues to increase when the current strategy isn't fully mixed as long as its update is. Let $r > 0$ be the maximum value such that $D_h(x^*||x) \geq r$ for all x on the boundary. If x^{t-1} is on the boundary, then either (1) x^t is on the boundary and $D_h(x^*||x^t) \geq r$ by definition or (2) x^t is fully mixed and $D_h(x^*||x^t) > D_h(x^*||x^{t-1}) \geq r$ by Theorem 4.1. Thus, once the updated strategies hit the boundary, the strategies will never again enter the smallest ball centered at the Nash equilibrium that intersects with the boundary.

Theorem 4.2. *For every 2-player constant-sum game with a fully mixed equilibrium, if all agents use (FTRL) with fixed η and a strongly convex regularizer, any (non-equilibrium) initial strategies come arbitrarily close to the boundary infinitely often.*

Proof. For contradiction, suppose there exists a T such that that x^t never comes close to boundary and there is a w such that $D_h(x^*||x^t) \leq w$ for all $t \geq T$. This implies $x_{is_i}^t > 0$ and that we can disregard the constraint $x_{is_i} \geq 0$ in (FTRL) for all $t \geq T$. By Theorem 4.1 there exists an $r > 0$ such that $D_h(x^*||x^t) \geq r$ for all t and therefore there is a compact B excluding x^* and the boundary such that $x^t \in B$ for all $t \geq T$.

Let $\hat{h}$, $\hat{u}$, and $\hat{v}$ be the functions obtained by plugging $x_{is_i} = 1 - \sum_{s_i' \in \mathcal{S}_i \setminus \{s_i\}} x_{is_i}$ into h, u, and v respectively for an arbitrarily selected s_i. Once this substitution is made, x^t is determined by

$$x_i^t = \text{argmax}\, \eta \sum_{s=0}^{t-1} \hat{u}_i(x_i; x_{-i}^s) - \hat{h}(x_i) \tag{Unconstrained FTRL}$$

where $x_{is_i}^t$ is assigned $1 - \sum_{s_i' \in \mathcal{S}_i \setminus \{s_i\}} x_{is_i}$. This function has KKT optimality conditions

$$\nabla \hat{h}(x_i^{t+1}) = \eta \sum_{s=0}^{t} \hat{v}_i(x_{-i}^s) = \eta \nabla \hat{h}(x_i^t) + \eta \hat{v}_i(x_{-i}^t). \tag{27}$$

Thus x^{t+1} is uniquely determined from x^t and therefore $f(x^t) = ||x^{t+1} - x^t||_2^2$ is well-defined. Since x^t is fully mixed, $x^t = x^{t+1}$ if and only if $x^t = x^s$ for all $s \geq t$. This only occurs if $x^t = x^*$ which cannot occur since $x^t \in B$ and therefore $f(x^t) > 0$ for all $t \geq T$. By compactness of B, there exists a $d > 0$ such that $f(x^t) \geq d$ for all $t \geq T$.

If h is strongly convex with parameter $m > 0$, then the proof of Theorem 4.1 can readily be modified to show that the divergence increases by at least $m||x^{t+1} - x^t||_2^2$ in iteration t of (FTRL) since x^t is always fully mixed for $t \geq T$. Thus,

$$\lim_{t\to\infty} D_h(x^*||x^t) = D_h(x^*||x^T) + \sum_{t=T}^{\infty}\left(D_h(x^*||x^{t+1}) - D_h(x^*||x^t)\right) \tag{28}$$

$$\geq D_h(x^*||x^T) + \sum_{t=T}^{\infty} m||x^{t+1} - x^t||_2^2 \tag{29}$$

$$\geq D_h(x^*||x^T) + \sum_{t=T}^{\infty} md = \infty \tag{30}$$

a contradiction. □

Theorems 4.1 and 4.2 imply that if the regularizer used in (FTRL) guarantees x^t will be fully mixed for each t then x^t converges to the boundary. An example of such a regularizer is the negative entropy function yielding the update rule (MWU_e). However, not all regularizers come with this guarantee. For instance in Figure 2, we see that for the regularizer $h(x) = ||x||_2^2$ (yielding the Gradient Descent algorithm) that player strategies move outward until they collide with the boundary. After this initial collision however, we see that the strategies may once again become fully mixed until they eventually again hit the boundary. Interestingly, we see that player strategies still appear to converge to the boundary in the absolute sense. However, the proof techniques we established for (MWU_e) cannot capture this simply because the Bregman divergence may decrease when player strategies are on the boundary.

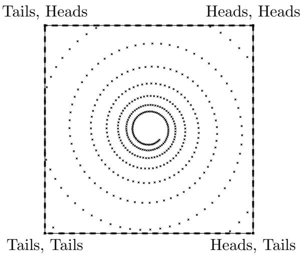


Fig. 2. Matching Pennies updated 1000 times with (FTRL), Regularizer $h(x) = ||x||_2^2$ and $\eta = .08$.

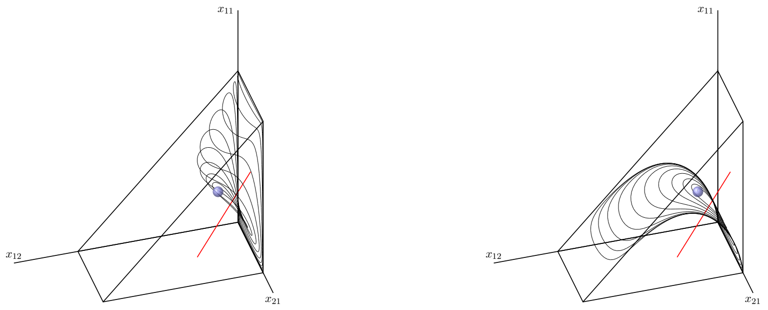
Similar to our results for (MWU_e) Theorem 4.2 holds in polymatrix games even when individuals select different values of η so long as $\frac{\prod_{k\in N\setminus\{i\}}\eta_{kt}}{\sum_{j\in N}\eta_{jt}}$ is time invariant and the selection of η can guarantee arbitrarily low regret. The proof of the results follow analogously to Theorems 3.4, 3.6, and 4.2. In addition, Theorem 4.2 holds even if players use different regularizers. The proof of Theorems 4.1 only makes use of the convexity of the regularizer.

Theorem 4.3. *For almost every 2-player constant-sum game with no fully mixed equilibria, there exists $\eta_0 > 0$, such that as long as all agents use (FTRL) with $\eta < \eta_0$ any (non-equilibrium) initial strategies converge to the boundary.*

This proof follows similarly to the second part of Theorem 3.3 and is deferred to Appendix C.

4.2 Linear MWU

The linear variant of MWU (MWU_ℓ) is similar to the (MWU_e) in that they have the same first order approximation for the update in player strategies and thus we may expect that the two algorithms should behave similarly. However, as shown in Figure 3, the algorithms can result in player strategies evolving in very different ways. Figure 3 suggests that (MWU_ℓ) still implies convergence to the boundary.



(a) Linear Update

(b) Exponential Update

Fig. 3. Player Strategies Spiraling Outwards for the Zero-Sum Game with Player 1 Payoff Matrix $\begin{pmatrix} 1 & -1 \\ 2 & -2 \\ -1 & 1 \end{pmatrix}$ with $\epsilon = .5$.

We explore convergence of (MWU_ℓ) in this section. We establish that the proof techniques used in Theorem 3.3 combined with the K-L divergence are insufficient for showing convergence to the boundary in (MWU_ℓ) – The K-L divergence may actually decrease when players update their strategies with (MWU_ℓ). The proof techniques are however sufficient to guarantee convergence to the boundary in the large class of zero-sum games, skew-symmetric games. A two-player game is *skew-symmetric* if both players have the same payoff matrix. Rock, Paper, Scissors is a classical example of a skew-symmetric game.

Theorem 4.4. *Let Γ be a 2-player constant-sum game that admits an interior Nash equilibrium. If both players update their strategies according to (MWU_ℓ) in iteration t, then the K-L divergence increases by $\sum_{i\in\mathcal{N}}\left(\ln\left(1+\epsilon u_i(x^{t-1})\right)-\sum_{s_i\in\mathcal{S}_i} x_i^*\ln\left(1+\epsilon v_{is_i}(x^{t-1})\right)\right)$.*

The proof of Theorem 4.4 follows analogously to Theorem 3.1.

Proposition 4.5. *There exists a 2-player zero-sum game Γ with an interior point equilibrium and a fully mixed strategy x^0 so that the K-L divergences decreases whenever players update strategies with (MWU_ℓ) with a sufficiently small ϵ.*

Proof. Consider the zero-sum game with payoff matrix $A_1 = \begin{pmatrix} 3 & -2 \\ -1 & 4 \end{pmatrix}$. The unique Nash equilibrium of this game is $x_1^* = (1/2, 1/2)$ and $x_2^* = (3/5, 2/5)$. By Theorem 4.4, the change in KL divergence after one iteration given the initial strategies $x_1^0 = (1/4, 3/4)$ and $x_2^0 = (1/3, 2/3)$ is

$$\ln\left(1+\frac{5}{3}\epsilon\right)+\ln\left(1-\frac{5}{3}\epsilon\right)-\frac{1}{2}\ln\left(1-\frac{1}{3}\epsilon\right)-\frac{1}{2}\ln\left(1+\frac{7}{3}\epsilon\right)-\frac{2}{5}\ln\left(1-\frac{5}{2}\epsilon\right)<0 \tag{31}$$

for $\epsilon \lessapprox 0.1736$. □

Theorem 4.6. *Let Γ be a 2-player skew-symmetric game that admits an interior Nash equilibrium. Further suppose that $x_1^0 = x_2^0$ is a fully mixed non-equilibrium initial strategy and both players update* *their strategies according to* (MWU_ℓ) *then* x^t *converges to the boundary. In the case of time evolving* ϵ_t*, its* ($MWU_\ell^{\epsilon_t}$) *variant converges to the boundary iff* $\sum_{t=1}^{\infty} \epsilon^2 = \infty$.

Since the game is skew-symmetric and since players are playing the same strategies, $u_1(x) = u_2(x) = 0$ for any strategy x. Therefore in an iteration of (MWU_ℓ) the K-L divergence increases by $-2\sum_{s_1 \in \mathcal{S}_1} x_1^* \ln\left(1 + \epsilon v_{1s_1}(x^{t-1})\right)$. The proof then follows analogously to Theorem 3.5 and obtains a similar rate of convergence to the boundary.

5 CONCLUSIONS

Our results suggest that the behavior of MWU (as well as of most classic no-regret dynamics) is more intricate than what is suggested by regret-minimizing algorithms' guarantee of "convergence to equilibrium". Even though the time-average of strategies converge, actual player strategies are repelled away from the equilibrium. If equilibria are predictive and we expect individual strategies to approach equilibria over time then the standard no-regret approach to updating strategies via MWU is inadequate and we must consider other equilibrium-seeking algorithms. Our work opens up the possibility of a much tighter understanding of the true, realized behavior of such dynamics in many contexts and raises interesting questions from a behavioral game theory standpoint. For example, what type of learning dynamics do people apply in practice? Can we exploit our understanding of the shape of these trajectories (e.g. the geometry of the limit cycles) to perform behavioral model fitting?

ACKNOWLEDGEMENTS

We acknowledge SUTD grant SRG ESD 2015 097, MOE AcRF Tier 2 Grant 2016-T2-1-170 and NRF 2018 Fellowship NRF-NRFF2018-07.

REFERENCES

[1] Maria-Florina Balcan, Florin Constantin, and Ruta Mehta. 2012. The Weighted Majority Algorithm does not Converge in Nearly Zero-sum Games. In *ICML Workshop on Markets, Mechanisms and Multi-Agent Models*.

[2] G.W. Brown. 1951. Iterative Solutions of Games by Fictitious Play. *In Activity Analysis of Production and Allocation, T.C. Koopmans (Ed.), New York: Wiley.* (1951).

[3] Yang Cai, Ozan Candogan, Constantinos Daskalakis, and Christos Papadimitriou. 2016. Zero-Sum Polymatrix Games: A Generalization of Minmax. *Mathematics of Operations Research* 41, 2 (2016), 648–655.

[4] Yang Cai and Costantinos Daskalakis. 2011. On Minmax Theorems for Multiplayer Games. In *ACM-SIAM Symposium on Discrete Algorithms (SODA)*. 217–234.

[5] Erick Chastain, Adi Livnat, Christos Papadimitriou, and Umesh Vazirani. 2014. Algorithms, games, and evolution. *Proceedings of the National Academy of Sciences (PNAS)* 111, 29 (2014), 10620–10623.

[6] Constantinos Daskalakis, Alan Deckelbaum, and Anthony Kim. 2011. Near-optimal No-regret Algorithms for Zero-sum Games. In *Proceedings of the Twenty-second Annual ACM-SIAM Symposium on Discrete Algorithms (SODA '11)*. Society for Industrial and Applied Mathematics, Philadelphia, PA, USA, 235–254. http://dl.acm.org/citation.cfm?id=2133036.2133057

[7] C. Daskalakis, R. Frongillo, C. Papadimitriou, G. Pierrakos, and G. Valiant. 2010. On learning algorithms for Nash equilibria. *Symposium on Algorithmic Game Theory (SAGT)* (2010), 114–125.

[8] Constantinos Daskalakis, Andrew Ilyas, Vasilis Syrgkanis, and Haoyang Zeng. 2018. Training GANs with Optimism.

[9] Constantinos Daskalakis and Christos Papadimitriou. 2009. On a Network Generalization of the Minmax Theorem. In *ICALP*. 423–434.

[10] Dylan J Foster, Thodoris Lykouris, Karthik Sridharan, and Eva Tardos. 2016. Learning in games: Robustness of fast convergence. In *Advances in Neural Information Processing Systems*. 4727–4735.

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

[12] Drew Fudenberg and David K Levine. 1995. Consistency and cautious fictitious play. *Journal of Economic Dynamics and Control* 19, 5-7 (1995), 1065–1089.

[13] Drew Fudenberg and David K. Levine. 1998. *The Theory of Learning in Games*. The MIT Press.

[14] R. Kleinberg, K. Ligett, G. Piliouras, and É. Tardos. 2011. Beyond the Nash equilibrium barrier. In *Symposium on Innovations in Computer Science (ICS)*.

[15] Tien-Yien Li and James A. Yorke. 1975. Period Three Implies Chaos. *The American Mathematical Monthly* 82, 10 (1975), 985–992.

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

[17] T. Mai, I. Panageas, W. Ratcliff, V. V. Vazirani, and P. Yunker. 2017. Rock-Paper-Scissors, Differential Games and Biological Diversity. *ArXiv e-prints* (Oct. 2017). arXiv:math.DS/1710.11249

[18] Ruta Mehta, Ioannis Panageas, and Georgios Piliouras. 2015. Natural Selection as an Inhibitor of Genetic Diversity: Multiplicative Weights Updates Algorithm and a Conjecture of Haploid Genetics. In *Innovations in Theoretical Computer Science*.

[19] R. Mehta, I. Panageas, G. Piliouras, and S. Yazdanbod. 2016. The Computational Complexity of Genetic Diversity. *European Symposium on Algorithms (ESA)* (2016).

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

[21] Gerasimos Palaiopanos, Ioannis Panageas, and Georgios Piliouras. 2017. Multiplicative Weights Update with Constant Step-Size in Congestion Games: Convergence, Limit Cycles and Chaos. In *Proceedings of the 30th International Conference on Neural Information Processing Systems (NIPS'17)*.

[22] Christos Papadimitriou and Georgios Piliouras. 2016. From Nash equilibria to chain recurrent sets: Solution concepts and topology. In *ITCS*.

[23] Georgios Piliouras, Carlos Nieto-Granda, Henrik I. Christensen, and Jeff S. Shamma. 2014. Persistent Patterns: Multi-agent Learning Beyond Equilibrium and Utility. In *AAMAS*. 181–188.

[24] Georgios Piliouras and Jeff S Shamma. 2014. Optimization despite chaos: Convex relaxations to complex limit sets via Poincaré recurrence. In *Proceedings of the twenty-fifth annual ACM-SIAM symposium on Discrete algorithms*. SIAM, 861–873.

[25] Sasha Rakhlin and Karthik Sridharan. 2013. Optimization, learning, and games with predictable sequences. In *Advances in Neural Information Processing Systems*. 3066–3074.

[26] J. Robinson. 1951. An Iterative Method of Solving a Game. *Annals of Mathematics* 54 (1951), 296–301.

[27] William H. Sandholm. 2010. *Population Games and Evolutionary Dynamics*. MIT Press.

[28] Vasilis Syrgkanis, Alekh Agarwal, Haipeng Luo, and Robert E. Schapire. 2015. Fast Convergence of Regularized Learning in Games. In *Proceedings of the 28th International Conference on Neural Information Processing Systems (NIPS'15)*. MIT Press, Cambridge, MA, USA, 2989–2997. http://dl.acm.org/citation.cfm?id=2969442.2969573

[29] Eric Van Damme. 1991. *Stability and perfection of Nash equilibria*. Vol. 339. Springer.

[30] John von Neumann. 1928. Zur Theorie der Gesellschaftsspiele. *Math. Ann.* 100 (1928), 295–300.

[31] John von Neumann and Oskar Morgenstern. 1944. *Theory of Games and Economic Behavior*. Princeton University Press.

A PROOF OF THEOREM 3.5

The first proof we present is a bound on the generalized binomial coefficient which will allow us to bound the change in K-L divergence in each iteration of ($MWU_e^{\epsilon_t}$).

Lemma A.1. *The generalized binomial coefficient* $\binom{x}{k} = x(x-1)\cdots(x-k+1)$ *is such that* $\binom{x}{k} = \sum_{j=1}^{k} a_{jk}x^j$ *where* $\sum_{j=1}^{k}|a_{jk}| = k!$.

Proof. Recursively, it follows that $a_{kk} = 1$, $a_{1k} = (1-k)a_{1,k-1}$ and $a_{jk} = (1-k)a_{j,k-1} + a_{j-1,k-1}$. We now proceed by induction. The statement of the lemma trivially holds for $k = 1$. Assume the result hold for $k-1$. Then by the the inductive hypothesis,

$$\sum_{j=1}^{k}|a_{jk}| = |a_{kk}| + |a_{1k}| + \sum_{j=2}^{k-1}|a_{jk}| \tag{32}$$

$$= 1 + (k-1)|a_{1,k-1}| + \sum_{j=2}^{k-1}\Big((k-1)|a_{j,k-1}| + |a_{j-1,k-1}|\Big) \tag{33}$$

$$= 1 + (k-1)\sum_{j=1}^{k-1}|a_{j,k-1}| + \sum_{j=1}^{k-1}|a_{j,k-1}| - |a_{k-1,k-1}| \tag{34}$$

$$= 1 + (k-1)(k-1)! + (k-1)! - 1 = k! \tag{35}$$

completing the proof of the lemma. □

Lemma A.2. *Let* Γ *be a game with a fully mixed Nash equilibrium. Let* $b = \max\left\{1, \max_{s,s'\in\mathcal{S}}\{u_1(s) - u_1(s')\}\right\}$ *and suppose* $\epsilon_t \le \frac{1}{2b}$. *If player's update their strategies according to* ($MWU_e^{\epsilon_t}$) *then* $D_{KL}(x^*||x^t) - D_{KL}(x^*||x^{t-1}) \le \sum_{i\in\mathcal{N}} \frac{Var[P_i(x^t)]}{2}\epsilon_t^2 + 4b^3\epsilon_t^3$.

Proof. Let a_{jk} be as described in Lemma A.1. By Theorem 3.1,

$$D_{KL}(x^*||x^t) - D_{KL}(x^*||x^{t-1}) = \sum_{i\in\mathcal{N}} \ln E\left[(1+\epsilon_t)^{P_i(x^t)-u_i(x^t)}\right] \tag{36}$$

$$= \sum_{i\in\mathcal{N}} \ln E\left[\sum_{k=0}^{\infty}\binom{P_i(x^t)-u_i(x^t)}{k}\epsilon_t^k\right] \tag{37}$$

$$= \sum_{i\in\mathcal{N}} \ln\left(1 + \frac{Var[P_i(x^t)]}{2}\epsilon_t^2 + \sum_{k=3}^{\infty}\frac{\sum_{j=1}^{k} a_{jk}E\left[(P_i(x^t)-u_i(x^t))^j\right]}{k!}\epsilon_t^k\right). \tag{38}$$

Next observe that $E\left[(P_i(x^t)-u_i(x^t))^j\right] \le b^j \le b^k$ for $k \ge j$. Therefore,

$$D_{KL}(x^*||x^t) - D_{KL}(x^*||x^{t-1}) \leq \sum_{i\in\mathcal{N}} \ln\left(1 + \frac{Var[P_i(x^t)]}{2}\epsilon_t^2 + \sum_{k=3}^{\infty} \frac{\sum_{j=1}^{k} |a_{jk}| b^k}{k!}\epsilon_t^k\right) \tag{39}$$

$$= \sum_{i\in\mathcal{N}} \ln\left(1 + \frac{Var[P_i(x^t)]}{2}\epsilon_t^2 + \sum_{k=3}^{\infty} b^k\epsilon_t^k\right) \tag{40}$$

$$= \sum_{i\in\mathcal{N}} \ln\left(1 + \frac{Var[P_i(x^t)]}{2}\epsilon_t^2 + \frac{b^3\epsilon_t^3}{1-b\epsilon_t}\right) \tag{41}$$

$$\leq \sum_{i\in\mathcal{N}} \ln\left(1 + \frac{Var[P_i(x^t)]}{2}\epsilon_t^2 + 2b^3\epsilon_t^3\right) \tag{42}$$

following from the Taylor series expansion of $\frac{y^3}{1-y}$ and since $\epsilon_t \leq \frac{1}{2b}$. Furthermore $\ln(1+y) \leq y$ and

$$D_{KL}(x^*||x^t) - D_{KL}(x^*||x^{t-1}) \leq \sum_{i\in\mathcal{N}} \left(\frac{Var[P_i(x^t)]}{2}\epsilon_t^2 + 2b^3\epsilon_t^3\right) \tag{43}$$

$$= \sum_{i\in\mathcal{N}} \frac{Var[P_i(x^t)]}{2}\epsilon_t^2 + 4b^3\epsilon_t^3 \tag{44}$$

completing the proof of the lemma. □

Lemma A.2 is sufficient to give the "only if" portion of Theorem 3.5. To obtain the "if" portion, we need a similar lower bound.

Lemma A.3. *Let Γ be a game with a fully mixed Nash equilibrium. Let* $b = \max\left\{1, \max_{s,s'\in\mathcal{S}}\{u_1(s) - u_1(s')\}\right\}$ *and suppose* $\epsilon_t \leq \frac{1}{2b}$. *If player's update their strategies according to* ($MWU_e^{\epsilon_t}$) *then* $D_{KL}(x^*||x^t) - D_{KL}(x^*||x^{t-1}) \geq \sum_{i\in\mathcal{N}} \frac{Var[P_i(x^t)]}{4}\epsilon_t^2 - 2b^3\epsilon_t^3$.

Proof. Following symmetrically with the proof of Lemma A.2 for $\epsilon_t \leq \frac{1}{2b}$,

$$D_{KL}(x^*||x^t) - D_{KL}(x^*||x^{t-1}) \geq \sum_{i\in\mathcal{N}} \ln\left(1 + \frac{Var[P_i(x^t)]}{2}\epsilon_t^2 - 2b^3\epsilon_t^3\right) \tag{45}$$

$$\geq \sum_{i\in\mathcal{N}} \frac{Var[P_i(x^t)]}{4}\epsilon_t^2 - 2b^3\epsilon_t^3 \tag{46}$$

since $\ln(1+y) \geq \frac{y}{2}$ completing the proof of the lemma. □

The statement of Lemma A.3 can be tightened to $\frac{Var[P_i(x^t)]}{2c}\epsilon_t^2$ for any $c > 1$ by taking ϵ_t sufficiently small. However, Lemma A.3 as written is sufficient for the proof of Theorem 3.5.

Proof of Theorem 3.5. We show the result only for 2-player games. The extension to polymatrix games follows in the same fashion as Theorem 3.4. We begin by showing player strategies do not converge to the boundary if $\sum_{t=1}^{\infty}\epsilon_t^2 < \infty$. Let b be as defined in Lemma A.2. Since the sum is finite, there must be a T such that $\epsilon_t \leq \frac{1}{2b}$ for all $t \geq T$. Any fully mixed strategy updated by ($MWU_e^{\epsilon_t}$) remains fully mixed and therefore $D_{KL}(x^*||x^{T-1}) < \infty$. Thus,

$$\lim_{t\to\infty} D_{KL}(x^*||x^t) = D_{KL}(x^*||x^{T-1}) + \sum_{t=T}^{\infty}\left(D_{KL}(x^*||x^t) - D_{KL}(x^*||x^{t-1})\right) \tag{47}$$

$$\leq D_{KL}(x^*||x^{T-1}) + \sum_{t=T}^{\infty}\left(\sum_{i\in\mathcal{N}} \frac{Var[P_i(x^t)]}{2}\epsilon_t^2 + 4b^3\epsilon_t^3\right) \tag{48}$$

$$\leq D_{KL}(x^*||x^{T-1}) + \sum_{t=T}^{\infty}\left(\sum_{i\in\mathcal{N}} \frac{b^2}{2}\epsilon_t^2 + 4b^3\epsilon_t^2\right) < \infty \tag{49}$$

by Lemma A.2, since $Var[P_i(x^t)] \leq b^2$, and $\sum_{t=1}^{\infty}\epsilon_t^2 < \infty$. Thus player strategies move outward but do not converge to the boundary completing the first part of the proof.

We now show that player strategies converge to the boundary if $\sum_{t=1}^{\infty}\epsilon_t^2 = \infty$. Let x^* be a fully mixed Nash equilibrium. For contradiction, suppose player strategies do not converge to the boundary and there is a $w > 0$ such that $D_{KL}(x^*||x^t) < w$ for all t. Thus, as in the proof of Theorem 3.3, there exists a compact B such that $x^t \in B$ for all t but B excludes the boundary and x^*. $\sum_{i\in\mathcal{N}} Var[P_i(x)] > 0$ for all $x \in B$ and since B is compact and $Var[P_i(x)]$ is continuous in x, there exists a $d > 0$ such that $\sum_{i\in\mathcal{N}} Var[P_i(x^t)] \geq d$ for all t.

If there are infinitely many t such that $\epsilon_t > \min\{\frac{1}{2b}, \frac{d}{16b^3}\}$ then the proof follows identically to Theorem 3.3 and we may assume there exists a T such that $\epsilon_t \leq \min\{\frac{1}{2b}, \frac{d}{16b^3}\}$ for all $t \geq T$. Finally,

$$\lim_{t\to\infty} D_{KL}(x^*||x^t) = D_{KL}(x^*||x^{T-1}) + \sum_{t=T}^{\infty}\left(D_{KL}(x^*||x^t) - D_{KL}(x^*||x^{t-1})\right) \tag{50}$$

$$\geq D_{KL}(x^*||x^{T-1}) + \sum_{t=T}^{\infty}\left(\sum_{i\in\mathcal{N}} \frac{Var[P_i(x^t)]}{4}\epsilon_t^2 - 2b^3\epsilon_t^3\right) \tag{51}$$

$$\geq D_{KL}(x^*||x^{T-1}) + \sum_{t=T}^{\infty}\left(\frac{d}{4}\epsilon_t^2 - \frac{d}{8}\epsilon_t^2\right) = \infty \tag{52}$$

by Lemma A.3 and since $\epsilon_t \leq \frac{d}{16b^3}$ and $\sum_{t=1}^{\infty}\epsilon_t^2 = \infty$. This is a contradiction, completing the proof of the theorem. □

B PROOF OF PROPOSITION 3.7

To show this proposition, we actually show that for almost every 2x2 constant-sum game with a fully mixed Nash equilibria x^* that there exists $\{\{\epsilon_{it}\}_{i\in\mathcal{N}}\}_{t=1}^{\infty}$ so that $\lim_{t\to\infty} x^t \to x^*$

The process for achieving this is quite simple; it requires at most 4 unique selections for ϵ_{it} to come arbitrarily close to x^*. Moreover if we allow ϵ_{it} to be arbitrarily large we can come arbitrarily close to the Nash equilibrium in 3 steps with 3 unique selections for ϵ_{it}. Rather than give the full proof, we describe selection process used to achieve convergence to x^*.

If $u_1(x^1) > u_1(x^*)$ then player 1 is profiting (relative to the Nash strategy) and selects $\epsilon_{1t} = 0$ so that strategy does not change. The game is constant-sum and $u_2(x^1) < u_2(x^*)$ implying player 2 is losing utility. He assigns $\epsilon_{it} = c$ for some $c > 0$. MWU is myopic and after some number of iterations $u_2(x^T) - u_2(x^*)$ becomes positive. In a 2x2 game $u_i(x^T) = u_i(x^*)$ if and only if at least one player is playing a Nash strategy. Therefore by re-selecting ϵ_{2T} so that $u_2(x^T)$ is arbitrarily close to $u_2(x^*)$, he also ensures that x_2^T is arbitrarily close to x_2^*. Moreover, he can do this so that $u_2(x^T) > u_2(x^*)$. Since $u_2(x^T) > u_2(x^*)$, $u_1(x^T) < u_1(x^*)$ and the players can repeat a symmetric process so that for $T' > T$, $x_1^{T'}$ is arbitrarily close to x_1^*. Moreover, since $\epsilon_{2t} = 0$ for all $t \in \{T+1, ..., T'\}$, $x_2^T = x_2^{T'}$ and both players are submitting a strategy arbitrarily close to the Nash equilibrium.

Since this process requires for $u_i(x^t) \neq u_i(x^*)$, it may require one additional step at the beginning if $x_i^0 = x_i^*$ for either player. However, this process is trivial since almost every selection of ϵ_{i1} will result in $x_i^* \neq x_i^1$ for both i. Moreover, if we allow ϵ_{it} to be arbitrarily large then we can select ϵ_{it} so that $T = 2$ and $T' = 3$.

C PROOF OF THEOREM 4.3

Similar to Theorem 3.3, we show the probability of playing a non-essential strategy goes to 0 as $t \to \infty$. Suppose player i has a non-essential strategy s_i. Without loss of generality we assume $u_i(x^*) = 0$. For almost every zero-sum game there is a unique Nash equilibrium x^*. Suppose x^t is updated according to (FTRL). Let $f_i^t(x_i)$ be the function optimized in (FTRL) to find x_i^t.

$$f_i^t(x_i) = t \cdot \eta \cdot u_i(x_i; \bar{x}_{-i}^{t-1}) - h(x_i) \tag{53}$$

where $\bar{x}^t = \sum_{s=0}^t x^s/t$ is the time-average of the player strategies.

As in the proof of Theorem 3.3, let $\delta = v_{is_i}(x^*) < 0$. For any y sufficiently close to x^*, $v_{is_i}(y) < 2\delta/3$ by continuity of payoffs. Let η_0 in the statement of the theorem be such that the time average of FTRL converges to the interior of the above neighborhood. For any $c > 0$ and any x_i where $x_{is_i} > c$, $f_i^t(x_1) \to -\infty$ as $t \to \infty$. However, $u_i(x_i^*; y_-i) \geq 0$ implying that for sufficiently large t, $f_i^t(x_i^*) > f_i^t(x_i)$. Thus, for arbitrary $c > 0$, $\limsup x_{is_i}^t < c$ implying $x_{is_i}^t \to 0$ for every non-essential s_i.