Fast Last-Iterate Convergence of Learning in Games Requires Forgetful Algorithms

Yang Cai
Yale
yang.cai@yale.edu

Gabriele Farina
MIT
gfarina@mit.edu

Julien Grand-Clément
HEC Paris
grand-clement@hec.fr

Christian Kroer
Columbia
ck2945@columbia.edu

Chung-Wei Lee
USC
leechung@usc.edu

Haipeng Luo
USC
haipengl@usc.edu

Weiqiang Zheng
Yale
weiqiang.zheng@yale.edu

Abstract

Self-play via online learning is one of the premier ways to solve large-scale two-player zero-sum games, both in theory and practice. Particularly popular algorithms include optimistic multiplicative weights update (OMWU) and optimistic gradient-descent-ascent (OGDA). While both algorithms enjoy $O(1/T)$ ergodic convergence to Nash equilibrium in two-player zero-sum games, OMWU offers several advantages including logarithmic dependence on the size of the payoff matrix and $\tilde{O}(1/T)$ convergence to coarse correlated equilibria even in general-sum games. However, in terms of last-iterate convergence in two-player zero-sum games, an increasingly popular topic in this area, OGDA guarantees that the duality gap shrinks at a rate of $(1/\sqrt{T})$, while the best existing last-iterate convergence for OMWU depends on some game-dependent constant that could be arbitrarily large. This begs the question: is this potentially slow last-iterate convergence an inherent disadvantage of OMWU, or is the current analysis too loose? Somewhat surprisingly, we show that the former is true. More generally, we prove that a broad class of algorithms that do not forget the past quickly all suffer the same issue: for any arbitrarily small $\delta > 0$, there exists a 2×2 matrix game such that the algorithm admits a constant duality gap even after $1/\delta$ rounds. This class of algorithms includes OMWU and other standard optimistic follow-the-regularized-leader algorithms.

1 Introduction

Self-play via online learning is one of the premier ways to solve large-scale two-player zero-sum games. Major examples include super-human AIs for Go, Poker [Brown and Sandholm, 2018], and human-level AI for Stratego [Perolat et al., 2022] and alignment of large language models [Munos et al., 2023]. In particular, Optimistic Multiplicative Weights Update (OMWU) and Optimistic Gradient Descent-Ascent (OGDA) are two of the most well-known online learning algorithms. When applied to learning a two-player zero-sum game via self-play for T rounds, the *average* iterates of both algorithms are known to be an $O(1/T)$-approximate Nash equilibrium [Rakhlin and Sridharan, 2013, Syrgkanis et al., 2015], while other algorithms, such as vanilla Multiplicative Weights Update

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

https://doi.org/10.52202/079017-0737

(MWU) and vanilla Gradient Descent-Ascent (GDA), have a slower ergodic convergence rate of $O(1/\sqrt{T})$.

For multiple practical reasons, there is growing interest in studying the *last-iterate* convergence of these learning dynamics [Daskalakis and Panageas, 2019, Golowich et al., 2020b, Wei et al., 2021, Lee et al., 2021]. In this regard, existing results seemingly exhibit a gap between OGDA and OMWU — the duality gap of the last iterate of OGDA is known to decrease at a rate of $O(1/\sqrt{T})$ [Cai et al., 2022, Gorbunov et al., 2022], with no dependence on constants beyond the dimension and the smoothness of the players' utility functions of the game.[1] In contrast, the existing convergence rate for OMWU depends on some game-dependent constant that could be arbitrarily large, even after fixing the dimension and the smoothness constant of the game [Wei et al., 2021].[2] Given the fundamental role of OMWU in online learning and its other advantages over OGDA (such as its logarithmic dependence on the number of actions), it is natural to ask the following question:

Is the potentially slow last-iterate convergence an inherent disadvantage of OMWU? (*)

Main Results. In this work, we show that the answer to this question is yes, contrary to a common belief that better analysis and better last-iterate convergence results similar to those of OGDA are possible for OMWU. More specifically, we show the following.

Theorem (Informal)**.** *For OMWU with constant step size, there is no function f such that the corresponding learning dynamics $\{(x^t, y^t)\}_{t\geq 1}$ in two-player zero-sum games $[0,1]^{d_1\times d_2}$ has a last-iterate convergence rate of $f(d_1, d_2, T)$.*[3] *More specifically, no function f can satisfy*

1. $\mathrm{DualityGap}(x^T, y^T) \leq f(d_1, d_2, T)$ *for all matrices* $[0,1]^{d_1\times d_2}$ *and* $T \geq 1$.
2. $\lim_{T\to\infty} f(d_1, d_2, T) \to 0$.

Our findings show that, despite the significantly superior *regret* properties of OMWU compared to OGDA, its *last-iterate convergence* properties are remarkably worse. In turn, this counters the viewpoint that "Follow-the-Regularized-Leader (FTRL) is better than Online Mirror Descent (OMD)" [van Erven, 2021]: crucially, while OMWU is an instance of (optimistic) FTRL, OGDA is an instance of optimistic OMD that cannot be expressed in the FTRL formalism.

We further show that similar negative results extend to several other standard online learning algorithms, including a close variant of OGDA. More concretely, our main results are as follows.

- We identify a broad family of Optimistic FTRL (OFTRL) algorithms that do not forget about the past quickly. We prove that, for any sufficiently small $\delta > 0$, there exists a 2×2 two-player zero-sum game such that, even after $1/\delta$ iterations, the duality gap of the iterate output by these algorithms is still a constant (Theorem 1). This excludes the possibility of showing a game-independent last-iterate convergence rate similar to that of OGDA.
- We prove that many standard online learning algorithms, such as OFTRL with the entropy regularizer (equivalently, OMWU), the Tsallis entropy family of regularizers, the log regularizer, and the squared Euclidean norm regularizer, all fall into this family of non-forgetful algorithms and thus all suffer from the same slow convergence. Also note that Optimistic OMD (OOMD), another well-known family of algorithms, is equivalent to OFTRL when given a Legendre regularizer. Therefore, OOMD with the entropy, Tsallis entropy, and log regularizer also suffer the same issue.[4]
- Finally, we also generalize our negative results from 2×2 games to $2n \times 2n$ games for any positive integer n, strengthening our message that forgetfulness is generally needed in order to achieve fast last-iterate convergence.

[1] In finite two-player zero-sum games, the dependence is polynomial in the number of actions and the largest absolute value in the payoff matrix.

[2] We note that there are also linear-rate last-iterate results for OGDA when we allow dependence on such constants; see [Wei et al., 2021].

[3] Under the same condition, OGDA has a last-iterate convergence rate of $\frac{\mathrm{poly}(d_1,d_2)}{\sqrt{T}}$.

[4] We focus on optimistic variants of these algorithms since it is well-known that their vanilla version does not converge in the last iterate at all, see e.g. [Mertikopoulos et al., 2018, Daskalakis and Panageas, 2018, Bailey and Piliouras, 2018, Cheung and Piliouras, 2019].

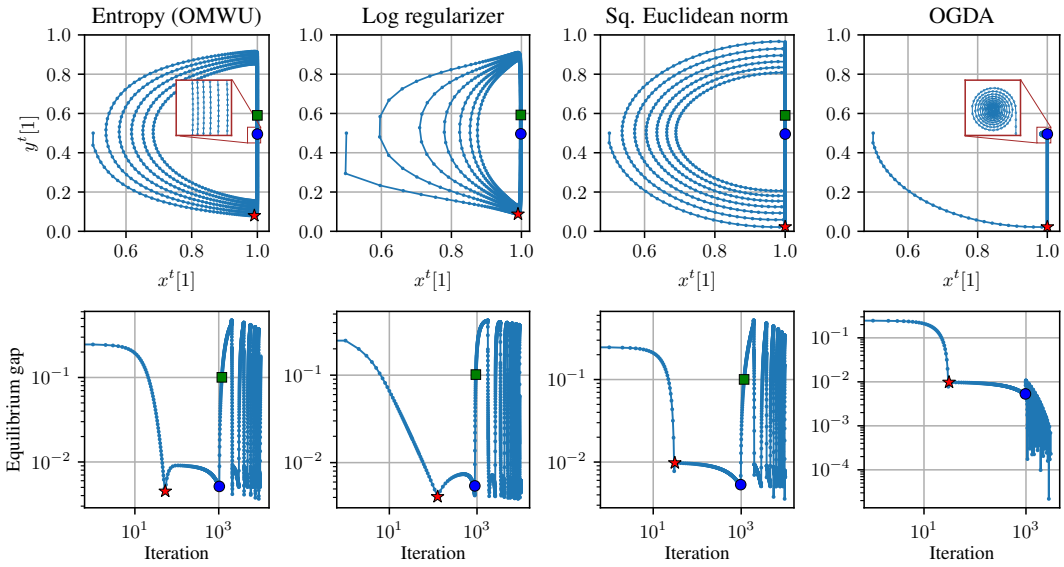

Figure 1: Comparison of the dynamics produced by three variants of OFTRL with different regularizers (negative entropy, logarithmic regularizer, and squared Euclidean norm) and OGDA in the same game A_δ defined in (2) for $\delta := 10^{-2}$. The bottom row shows the duality gap achieved by the last iterates. The OFTRL variants exhibit poor performance due to their lack of *forgetfulness*, while OGDA converges quickly to the Nash equilibrium. Since the regularizers in the first two plots are Legendre, the dynamics are equivalent to the ones produced by optimistic OMD with the respective Bregman divergences. In the plot for OMWU we observe that $x^t[1]$ can get extremely close to the boundary (*e.g.,* in the range $1 - e^{-50} < x^t[1] < 1$). To correctly simulate the dynamics, we used 1000 digits of precision. The red star, blue dot, and green square illustrate the key times T_1, T_2, T_3 defined in our analysis in Section 3.

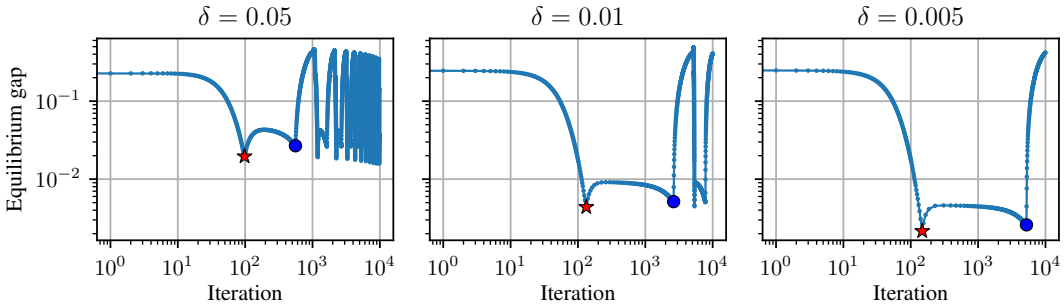

Figure 2: Performance of OMWU on the game A_δ defined in eq. (2) for three choices of δ. In all plots, the learning rate was set to $\eta = 0.1$. As predicted by our analysis, the length of the "flat region" between iteration T_1 (red star) and T_2 (blue dot) scales inversely proportionally with δ.

Main Ideas. Intuitively, we trace the poor last-iterate convergence properties of OFTRL to its *lack of forgetfulness*. The high-level idea of our hard 2×2 game instance, parametrized by $\delta > 0$, is as follows. First, it has a unique Nash equilibrium at which one player is $O(\delta)$ close to the boundary of the simplex. We refer to the first row of plots in Figure 1, where the equilibrium is noted by a blue dot (note that we can plot only $x[1], y[1]$ for each player, since $x[2] = 1 - x[1]$ and $y[2] = 1 - y[2]$). As can be seen, the iterates of OGDA and all three OFTRL variants initially have a two-phase structure. In the first phase, they converge to the lower-right area denoted by a red star in Figure 1. Then, from there all algorithms start moving towards the equilibrium. In particular, $y[1]$ increases. However, once they enter the vicinity of the equilibrium, the behavior depends on the algorithms. For OGDA, the dynamics start spiraling closer and closer to the equilibrium. On the other hand, for the OFTRL algorithm, the x player has built up a lot of "memory" of $x[1]$ being better than $x[2]$, and for this reason, $x[1]$ will stay very close to 1 for a long time. During the time when $x[1]$ is close to 1, $y[1]$ keeps increasing since the y player receives gradients that indicates $y[1]$ is better than $y[2]$. As a

https://doi.org/10.52202/079017-0737

result, the dynamics cannot "stop" near the equilibrium but start to move away from the equilibrium. The dynamics reach a point (denoted by a green square) whose duality gap is a constant and enter a new cycle where they move out towards the starting point of the learning process. This cycle repeats in smaller and smaller semi-ellipses that slowly converge to equilibrium. Note that the semi-ellipses correspond to the seesaw pattern in the equilibrium gap (second row of plots). OFTRL overshoots the equilibrium as it has built up a lot of "memory" of $x[1]$ being better than $x[2]$ along the phase from the red star to the blue circle, and it requires many iterations to "forget" this fact. We show that as we make δ, the parameter defining the nearness to the boundary, smaller and smaller, it takes longer and longer for these semi-ellipses to get close to the equilibrium along the entire path, as illustrated in Figure 2.

Our results are related to numerical observations made in the literature on solving large-scale extensive-form games. There, algorithms based on the regret-matching$^+$ (RM$^+$) algorithm [Tammelin et al., 2015], combined with the counterfactual regret minimization [Zinkevich et al., 2007], perform by far the best in practice. In contrast, the classical regret matching algorithm [Hart and Mas-Colell, 2000] performs much worse, in spite of similar regret guarantees. It was later discovered that RM$^+$ corresponds to OGD, while RM corresponds to FTRL [Farina et al., 2021, Flaspohler et al., 2021]. It was hypothesized that RM builds up too much negative regret at times, and thus is slow to adapt to changes in the learning dynamics related to the strategy of the other player. These numerical results and the hypothesis are consistent with our theoretical findings: FTRL (and thus RM) is not able to "forget," whereas OGD and OGDA can forget and thereby quickly adapt to changes in which actions should be played.

1.1 Related Work

The literature on last-iterate convergence of online learning methods in games is vast. In this section, we will cover key contributions focusing on the case of interest for this paper: discrete-time dynamics for two-player zero-sum normal-form games.

Convergence of OGDA. Average-iterate convergence of OGDA has been studied for minimax optimization problems in both the unconstrained [Mokhtari et al., 2020] and constrained settings [Hsieh et al., 2019]. Last-iterate convergence of OGDA in *unconstrained* saddle-point problems has been shown in [Daskalakis et al., 2018, Golowich et al., 2020a]. In the (constrained) game setting, Wei et al. [2021], Anagnostides et al. [2022] showed *best*-iterate convergence to the set of Nash equilibria in any two-player zero-sum game with payoff matrix A at a rate of $O(\text{poly}(d_1, d_2, \max_{i,j} |A_{i,j}|)/\sqrt{T})$ using constant learning rate, where d_1 and d_2 are the number of actions of the players. A stronger result was shown by Cai et al. [2022], who showed that the same rate applies to the *last* iterate.

Convergence of OMWU. Optimistic multiplicative weights update (also known as optimistic hedge) is often regarded as the premier algorithm for learning in games. Unlike OGDA, it guarantees sublinear regret with a *logarithmic* dependence on the number of actions, and it is known to guarantee only polylogarithmic regret per player when used in self play even for general-sum games [Daskalakis et al., 2021]. It can be applied with similar strong properties beyond normal-form games in several important combinatorial settings [Takimoto and Warmuth, 2003, Koolen et al., 2010, Farina et al., 2022]. The work by Daskalakis and Panageas [2019] established *asymptotic* last-iterate convergence for OMWU in games using a small learning rate under the assumption of a unique Nash equilibrium. Similar asymptotic results without the unique equilibrium assumption were also given by Mertikopoulos et al. [2019], Hsieh et al. [2021]. Wei et al. [2021] were the first to provide *nonasymptotic* learning rates for OMWU. Specifically, they showed a linear rate of convergence in games with a unique equilibrium, albeit with a dependence on a condition number-like quantity that could be arbitrarily large given fixed d_1, d_2, and $\max_{i,j} |A_{i,j}|$.This result was later extended by Lee et al. [2021] to extensive-form games. Unlike OGDA, no last-iterate convergence result for OMWU with a polynomial dependence on only the natural parameters of the game (*i.e.*, d_1, d_2, and $\max_{i,j} |A_{i,j}|$) is known. As we show in this paper, perhaps surprisingly, this is no coincidence: in general, OMWU does not exhibit a last-iterate convergence rate that solely depends on these parameters, whether polynomial or not.

FTRL vs. OMD. While the last-iterate convergence of instantiations of Optimistic Online Mirror Descent has been observed before, the properties of Follow-the-Regularized-Leader dynamics remain mostly elusive. The present paper partly explains this vacuum: all standard instantiations of optimistic FTRL *cannot hope* to converge in iterates with only a polynomial dependence on the natural parameters of the game, unlike optimistic OMD. Complications in obtaining last-iterate convergence

results for continuous-time FTRL instantiations were already reported by Vlatakis-Gkaragkounis et al. [2020], who showed the necessity of *strict* Nash equilibria.

Exploiting a no-regret learner. The forgetfulness property that we identify is closely related to the concept of *mean-based* learning algorithms from Braverman et al. [2018]. Intuitively, mean-based algorithms are ones such that if the mean reward for action a is significantly greater than the mean reward for action b, then the algorithm selects b with negligible probability. They show that MWU is mean-based, along with Follow-the-Perturbed-Leader and the Exp3 bandit algorithm. Braverman et al. [2018] shows that "mean-based" algorithms are exploitable when learning to bid in first-price auctions, whereas Kumar et al. [2024] shows that OGD does not suffer from this exploitability issue.

2 Preliminaries and Problem Setup

We consider the standard setting of no-regret learning in a zero-sum game $A \in [0,1]^{d_1 \times d_2}$. In each iteration $t \geq 1$, the x-player chooses $x^t \in \mathcal{X} := \Delta^{d_1}$ while the y-player chooses $y^t \in \mathcal{Y} := \Delta^{d_2}$. Then the x-player receives loss vector $\ell_x^t = Ay^t$ while the y-player receives loss vector $\ell_y^t = -A^\top x^t$. The goal is to find or approximate a *Nash equilibrium* (x^*, y^*) to the game such that $x^* \in \operatorname{argmin}_{x \in \mathcal{X}} \max_{y \in \mathcal{Y}} x^\top A y$ and $y^* \in \operatorname{argmax}_{y \in \mathcal{Y}} \min_{x \in \mathcal{X}} x^\top A y$. The approximation error of a strategy pair (x, y) is measured by its duality gap, defined as $\mathrm{DualityGap}(x, y) = \max_{y' \in \mathcal{Y}} x^\top A y' - \min_{x' \in \mathcal{X}} x'^\top A y$, which is always non-negative.

Popular no-regret algorithms for solving the game include the Optimistic Follow-the-Regularized-Leader (OFTRL) algorithm and the Optimistic Online Mirror Descent (OOMD) algorithm, both defined in terms of a certain regularizer $R : \Delta^d \to \mathbb{R}$ (for some general dimension d). The corresponding Bregman divergence of R is $D_R(x, x') = R(x) - R(x') - \langle \nabla R(x'), x - x' \rangle$, and the regularizer is 1-strongly convex if $D_R(x, x') \geq \frac{1}{2}\|x - x'\|_2^2$ for all $x, x' \in \Delta^d$.

Optimistic Online Mirror Descent (OOMD) Starting from an initial point $(x^1, y^1) = (\widehat{x}^1, \widehat{y}^1)$, the OOMD algorithm with regularizer R and steps size $\eta > 0$ updates in each iteration $t \geq 2$,

$$\begin{aligned} \widehat{x}^t &= \operatorname*{argmin}_{x \in \mathcal{X}} \{\eta \langle x, \ell_x^{t-1} \rangle + D_R(x, \widehat{x}^{t-1})\}, \quad x^t = \operatorname*{argmin}_{x \in \mathcal{X}} \{\eta \langle x, \ell_x^{t-1} \rangle + D_R(x, \widehat{x}^t)\}, \\ \widehat{y}^t &= \operatorname*{argmin}_{y \in \mathcal{Y}} \{\eta \langle y, \ell_y^{t-1} \rangle + D_R(y, \widehat{y}^{t-1})\}, \quad y^t = \operatorname*{argmin}_{y \in \mathcal{Y}} \{\eta \langle y, \ell_y^{t-1} \rangle + D_R(y, \widehat{y}^t)\}. \end{aligned} \tag{OOMD}$$

In particular, we call OOMD with a squared Euclidean norm regularizer, that is, $R(x) = \frac{1}{2}\sum_{i=1}^d x[i]^2$ *optimistic gradient-descent-ascent* (OGDA). When R is the negative entropy, that is, $R(x) = \sum_{i=1}^d x[i] \log x[i]$, we call the resulting OOMD algorithm *optimistic multiplicative weights update* (OMWU). OGDA and OMWU have been extensively studied in the literature regarding their last-iterate convergence properties in zero-sum games. Specifically, both OMWU and OGDA guarantee that (x^t, y^t) approaches to a Nash equilibrium as $t \to \infty$.

Optimistic Follow-the-Regularized-Leader (OFTRL) Define the cumulative loss vectors $L_x^t := \sum_{k=1}^t \ell_x^k$ and $L_y^t := \sum_{k=1}^t \ell_y^k$. The update rule of OFTRL with regularizer R is for each $t \geq 1$,

$$\begin{aligned} x^t &= \operatorname*{argmin}_{x \in \mathcal{X}} \left\{ \left\langle x, L_x^{t-1} + \ell_x^{t-1} \right\rangle + \frac{1}{\eta} R(x) \right\}, \\ y^t &= \operatorname*{argmin}_{y \in \mathcal{Y}} \left\{ \left\langle y, L_y^{t-1} + \ell_y^{t-1} \right\rangle + \frac{1}{\eta} R(y) \right\}. \end{aligned} \tag{OFTRL}$$

Throughout the paper, we consider the following regularizers:

- Negative entropy ($R(x) = \sum_{i=1}^d x[i] \log x[i]$): the resulting OFTRL algorithm coincides with OMWU defined by the OOMD framework previously.
- Squared Euclidean norm ($R(x) = \frac{1}{2}\sum_{i=1}^d x[i]^2$): note that the resulting algorithm is different from OGDA since the squared Euclidean norm is not a Legendre regularizer. As we will show, the two algorithms behave very differently in terms of last-iterate convergence.
- Log barrier ($R(x) = \sum_{i=1}^d -\log(x[i])$): we also call it the log regularizer.

https://doi.org/10.52202/079017-0737

- Negative Tsallis entropy regularizers ($R(x) = \frac{1-\sum_{i=1}^{d}(x[i])^{\beta}}{1-\beta}$ parameterized by $\beta \in (0,1)$).

The 2-dimension case We denote $x \in \mathbb{R}^2$ as $x = [x[1], x[2]]^\top$. For $d_1 = 2$, finding x^t of OFTRL reduces to the following 1-dimensional optimization problem:

$$x^t[1] = \operatorname*{argmin}_{x\in[0,1]} \left\{ x \cdot (L_x^{t-1}[1] + \ell_x^{t-1}[1] - L_x^{t-1}[2] - \ell_x^{t-1}[2]) + \frac{1}{\eta} R(x) \right\}, \quad x^t[2] = 1 - x^t[1],$$

where we slightly abuse the notation and denote $R(x) = R([x, 1-x])$ for $x \in [0,1]$. We introduce two notations (the case for the y-player is similar): let $e_x^t = \ell_x^t[1] - \ell_x^t[2]$ be the difference between the losses of the two actions, and $E_x^t = \sum_{k=1}^{t} e_x^k$ be the cumulative difference between the losses of the two actions. For OFTRL, it is clear that the update of x^t only depends on the differences E_x^{t-1}, e_x^{t-1}, the step size η, and the regularizer R. For this reason, we define $F_{\eta,R} : \mathbb{R} \to [0,1]$ as follows:

$$F_{\eta,R}(e) := \operatorname*{argmin}_{x\in[0,1]} \left\{ x \cdot e + \frac{1}{\eta} R(x) \right\}. \tag{1}$$

We assume the function $F_{\eta,R}$ is well-defined, *i.e.*, the above optimization problem admits a unique solution in $[0,1]$. This is a condition easily satisfied, for example, when the regularizer R is strongly convex. Then the OFTRL algorithm can be written as

$$x^t[1] = F_{\eta,R}(E_x^{t-1} + e_x^{t-1}), \quad x^t[2] = 1 - x^t[1].$$

The following lemma shows that the function $F_{\eta,R}$ is non-increasing (we defer missing proofs in the section to Appendix A).

Lemma 1 (Monotonicity of $F_{\eta,R}$)**.** *The function $F_{\eta,R}(\cdot) : \mathbb{R} \to [0,1]$ defined in (1) is non-increasing.*

We present some blanket assumptions on the regularizer, which are satisfied by all the regularizers introduced before.

Assumption 1. *We assume that the regularizer R satisfies the following properties: the function $F_{\eta,R} : \mathbb{R} \to [0,1]$ defined in (1) is,*

1. ***Unbiased:*** $F_{\eta,R}(0) = \frac{1}{2}$.
2. ***Rational:*** $\lim_{E\to-\infty} F_{\eta,R}(E) = 1$ *and* $\lim_{E\to+\infty} F_{\eta,R}(E) = 0$.
3. ***Lipschitz continuous:*** *There exists $L \geq 0$ such that $F_{1,R}$ is L-Lipschitz.*

Item 1 in Assumption 1 shows that the initial strategy is the uniform distribution over the two actions, which is standard in practice. The rational assumption (item 2 in Assumption 1) is natural since otherwise, the algorithm could not even converge to a pure Nash equilibrium. The Lipschitzness (item 3 in Assumption 1) is implied when the regularizer is strongly convex over $[0,1]^2$ (see Lemma 4), and it further implies Lipschitzness of $F_{\eta,R}$ for any η as shown in the following proposition.

Proposition 1. *The function $F_{\eta,R}$ satisfies $F_{\eta,R}(E/\eta) = F_{1,R}(E)$. If $F_{1,R}$ is L-Lipschitz, then $F_{\eta,R}$ is ηL-Lipschitz for any $\eta > 0$.*

3 Slow Convergence of OFTRL: A Hard Game Instance

We give negative results on the last-iterate convergence properties of OFTRL by studying its behavior on a surprisingly simple 2×2 two-player zero-sum games. The game's loss matrix A_δ is parameterized by $\delta \in (0, \frac{1}{2})$ and is defined as follows:

$$A_\delta := \begin{bmatrix} \frac{1}{2}+\delta & \frac{1}{2} \\ 0 & 1 \end{bmatrix}. \tag{2}$$

3.1 Basic Properties

We summarize some useful properties of A_δ in the following proposition (missing proofs of this section can be found in Appendix B).

Proposition 2. *The matrix game A_δ satisfies:*

1. *A_δ has a unique Nash equilibrium $x^* = [\frac{1}{1+\delta}, \frac{\delta}{1+\delta}]$ and $y^* = [\frac{1}{2(1+\delta)}, \frac{1+2\delta}{2(1+\delta)}]$.*
2. *For a strategy pair (x^t, y^t), the loss vectors (i.e., gradients) for the two players are respectively:*

$$\ell_x^t = A_\delta y^t = \begin{bmatrix} \frac{1}{2} + \delta y^t[1] \\ 1 - y^t[1] \end{bmatrix} \quad \ell_y^t = -A_\delta^\top x^t = -\begin{bmatrix} (\frac{1}{2} + \delta) x^t[1] \\ 1 - \frac{1}{2} x^t[1] \end{bmatrix}. \tag{3}$$

Moreover,

$$e_x^t = \ell_x^t[1] - \ell_x^t[2] = -\frac{1}{2} + (1+\delta) y^t[1] \in [-\frac{1}{2}, \frac{1}{2} + \delta]$$
$$e_y^t = \ell_y^t[1] - \ell_y^t[2] = 1 - (1+\delta) x^t[1] \in [-\delta, 1].$$

In particular, we notice that $e_y^t \geq -\delta$. It implies that if the cumulative differences between the losses of the two actions E_y^t is large, then it takes $\Omega(\frac{1}{\delta})$ iterations to make E_y^t small (close to 0). This has important implications for non-forgetful algorithms like OFTRL that look at the whole history of losses. Since OFTRL chooses the strategy y^t based on E_y^t, it could be trapped in a bad action for a long time even if the current gradients suggest that the other action is better. This is the key observation for our main negative results on the slow last-iterate convergence rates of OFTRL.

The following lemma shows that in a particular region of (x, y), the duality gap is a constant.

Lemma 2. *Let $\delta, \epsilon \in (0, \frac{1}{2})$. For any $x, y \in \Delta^2$ such that $x[1] \geq \frac{1}{1+\delta}$ and $y[1] \geq \frac{1}{2} + \epsilon$, the duality gap of (x, y) for game A_δ (defined in (2)) satisfies* $\mathrm{DualityGap}(x, y) \geq \epsilon$.

3.2 Slow Last-Iterate Convergence

We further require the following assumption on the regularizer R (and thus the function $F_{1,R}$).

Assumption 2. *Let L be the Lipschitness constant of $F_{1,R}$ in Assumption 1. Denote constant $c_1 = \frac{1}{2} - F_{1,R}(\frac{1}{20L})$. There exist universal constants $\delta', c_2 > 0$ and $c_3 \in (0, \frac{1}{2}]$ such that for any $0 < \delta \leq \delta'$,*

1. *For any E that satisfies $F_{1,R}(E) \geq \frac{1}{1+\delta}$, we have $F_{1,R}(-\frac{c_1^2}{30L\delta} + E) \geq \frac{1+c_3}{1+c_3+\delta}$*
2. *For any E that satisfies $F_{1,R}(E) \geq \frac{1}{2(1+\delta)}$, we have $F_{1,R}(-\frac{c_3 c_1^2}{120L} + \frac{\delta}{4L} + E) \geq \frac{1}{2} + c_2$.*

Although Assumption 2 is technical, the idea is simple. Item 1 in Assumption 2 states that if a loss difference $E < 0$ already makes $F_{1,R}(E) \geq \frac{1}{1+\delta}$, then the loss difference $E' = E - \Omega(\frac{1}{\delta})$ is able to make $F_{1,R}(E')$ greater than $F_{1,R}(E)$ by a margin of $\Omega(\delta)$. Item 2 in Assumption 2 states that if a loss difference E already makes $F_{1,R}(E) \geq \frac{1}{2(1+\delta)} \approx \frac{1}{2}$, then the loss difference $E' = E - \Omega(1)$ is able to make $F_{1,R}(E')$ greater than $\frac{1}{2}$ by a constant margin c_2. In Appendix C, we verify that Assumption 2 holds for the negative entropy, squared Euclidean norm, the log barrier, and the negative Tsallis entropy regularizers.

Now we present the main result of the section showing that even after $\Omega(1/\delta)$ iterations, the duality gap of the iterate output by OFTRL is still a constant.

Theorem 1. *Assume the regularizer R satisfies Assumption 1 and Assumption 2. For any $\delta \in (0, \hat{\delta})$, where $\hat{\delta}$ is a constant depending only on the constants c_1 and δ' defined in Assumption 2, the OFTRL dynamics on A_δ (defined in (2)) with any step size $\eta \leq \frac{1}{4L}$ satisfies the following: there exists an iteration $t \geq \frac{c_1}{3\eta L \delta}$ with a duality gap of at least c_2, a strictly positive constant defined in Assumption 2.*

Proof Sketch: We decompose the analysis into three stages as illustrated in Figure 3. We describe the three stages and the high-level ideas of our proof below and defer the full proof to Appendix B.2.

https://doi.org/10.52202/079017-0737

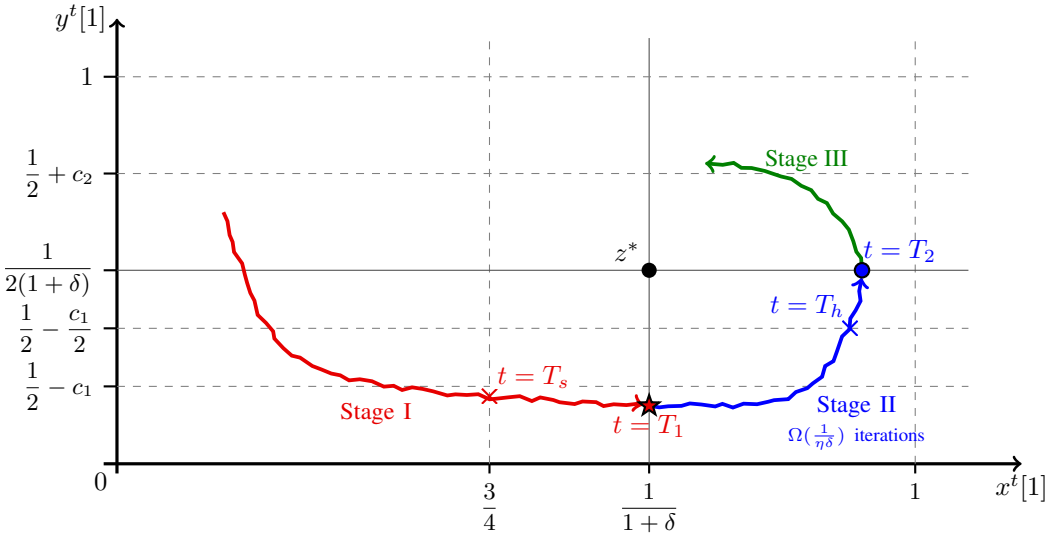

Figure 3: Pictorial depiction of the three stages incurred by the OFTRL dynamics in the game A_δ defined in (2). The point z^* denotes the unique Nash equilibrium. The times T_1 and T_2 are shown for concrete instantiations of OFTRL in Figure 1 by a red star and a blue dot, respectively. The times T_s and T_h are defined in the proof of Theorem 1 in Appendix B.2.

- **Stage I** $[1, T_1 - 1]$**:** Recall that $x^1[1] = y^1[1] = \frac{1}{2}$ by Assumption 1. We show that $x^t[1]$ increases and denote T_1 the first iteration that $x^{T_1}[1] \geq \frac{1}{1+\delta}$. During the time $[1, T_1 - 1]$, since $x^t[1]$ is always smaller than $\frac{1}{1+\delta}$, we know from Proposition 2 action 1 has larger loss than action 2 for the y-player, i.e., $e_y^t = \ell_y^t[1] - \ell_y^t[2] \geq 0$. Thus $y^t[1]$ decreases during stage I and we show that $y^{T_1}[1] \leq \frac{1}{2} - c_1$ with c_1 defined in Assumption 2.
- **Stage II** $[T_1, T_2]$**:** Recall that $y^{T_1}[1] \leq \frac{1}{2} - c_1$. We define $T_2 > T_1$ as the first iteration where $y^{T_2}[1] \geq \frac{1}{2(1+\delta)} > \frac{1}{2} - c_1$. We remark that for $y^t[1]$ to increase, the loss vector must satisfy $e_y^t < 0$. However, the game matrix A_δ guarantees that $e_y^t \geq -\delta$ no matter what the x-player's strategy is (Proposition 2). Thus by the ηL-Lipschitzness of $F_{\eta,R}$ (Proposition 1), the per-iteration increase in $y^t[1]$ is at most $\eta L \delta$. Therefore, we know $T_2 - T_1 = \Omega(\frac{c_1}{\eta L \delta})$. As a result, $e_x^t < 0$ during $[T_1, T_2]$ and the cumulative loss for the x-player decreases to $E_x^{T_2} \leq E_x^{T_1} - \Omega(\frac{1}{\eta L \delta})$. Recall $x^{T_1}[1] \geq \frac{1}{1+\delta}$. Thus $x^{T_2}[1] > x^{T_1}[1]$ is much closer to 1.
- **Stage III** $[T_2, T_3]$**:** Recall that $y^{T_2}[1] \geq \frac{1}{2(1+\delta)}$. Moreover, $y^t[1]$ could keep increasing if $x^t[1] \geq \frac{1}{1+\delta}$ since that implies $e_y^t \leq 0$. Now the question is how long would the x-player stay close to the boundary, i.e, $x^t[1] \geq \frac{1}{1+\delta}$. Since OFTRL-type algorithms are not forgetful, this happens only when $E_x^t \geq E_x^{T_1}$ (recall $x^{T_1}[1] \geq \frac{1}{1+\delta}$). But we have at the end of stage II, $E_x^{T_2} \leq E_x^{T_1} - \Omega(\frac{1}{\eta L \delta})$. Since the per-iteration loss is bounded by 1, it requires at least $\Omega(\frac{1}{\eta L \delta})$ iterations to cancel the cumulative loss of $\Omega(\frac{1}{\eta L \delta})$. Define $T_3 = T_2 + \Omega(\frac{1}{\eta L \delta})$. During $[T_2, T_3]$, the y-player always receives loss such that $e_y^t \leq 0$ and we prove that in the end $y^{T_3}[1] \geq \frac{1}{2} + c_2$ for some constant c_2.
- **Conclusion:** Finally, we get one iteration $T_3 \geq \Omega(\frac{1}{\eta L \delta})$ with $x^{T_3}[1] \geq \frac{1}{1+\delta}$ and $y^{T_3}[1] \geq \frac{1}{2} + c_2$. Using Lemma 2, the duality gap of (x^{T_3}, y^{T_3}) is at least $c_2 > 0$.

Theorem 1 immediately implies the following (proof deferred to Appendix B.3).

Theorem 2. *For optimistic FTRL with any regularizer satisfying Assumption 1 and Assumption 2 and constant steps size $\eta \leq \frac{1}{4L}$ (L is defined in Assumption 1), there is no function f such that the corresponding learning dynamics $\{(x^t, y^t)\}_{t \geq 1}$ in two-player zero-sum games $[0,1]^{d_1 \times d_2}$ has a last-iterate convergence rate of $f(d_1, d_2, T)$. More specifically, no function f can satisfy*

1. $\text{DualityGap}(x^T, y^T) \leq f(d_1, d_2, T)$ *for all* $[0,1]^{d_1 \times d_2}$ *and for all* $T \geq 1$.

2. $\lim_{T\to\infty} f(d_1, d_2, T) \to 0$.

Theorem 1 and Theorem 2 provide impossibility results for getting a last-iterate convergence rate for OFTRL that solely depends on the bounded parameters, even in two-player zero-sum games. Moreover, they show the necessity of forgetfulness for fast last-iterate convergence in games since OGDA has a last-iterate convergence rate of $O(\frac{\text{poly}(d_1,d_2)}{\sqrt{T}})$ [Cai et al., 2022, Gorbunov et al., 2022].

4 Extension to Higher Dimensions

In this section, we extend our negative results from 2×2 matrix games to games with higher dimensions. We start by showing an equivalence result for a single player (say, the first player). We assume that a decision maker is using OFTRL with a 1-strongly convex (w.r.t. the ℓ_2 norm) and separable regularizer $R(x) = R_1(x_1) + R_2(x_2)$ to choose decisions. At a given time time t, they see a loss $\ell^t \in [0, 1]^2$.

Now consider the following $2n$-dimensional decision problem: The player uses OFTRL using the regularizer $\hat{R}(\hat{x}) = \sum_{i=1}^{n} R_1(\hat{x}_i) + \sum_{i=n+1}^{2n} R_2(\hat{x}_i)$, *i.e.*, they use R_1 on the first half of actions, and R_2 on the second half. This is again a 1-strongly convex regularizer (w.r.t. the ℓ_2 norm). Suppose the decision maker sees the rescaled and *duplicated* version of the losses $\ell^1, \dots, \ell^T$ from the 2-dimensional case: $\hat{\ell}_i^t = \frac{1}{n^\alpha}\ell_1^t$ if $i \le n$, and $\hat{\ell}_i^t = \frac{1}{n^\alpha}\ell_2^t$ if $i > n$. The parameter α will be chosen later based on the regularizer.

Now we wish to show that by choosing α in the right way, we get that the decisions for the 2-dimensional and $2n$-dimensional OFTRL algorithms are equivalent. Let $x^1, \dots, x^T$ be the 2-dimensional OFTRL decisions, and let $\hat{x}^1, \dots, \hat{x}^T$ be the $2n$-dimensional OFTRL decisions. Then, we want to show that $\sum_{i=1}^{n} \hat{x}_i^t = x^t[1]$ and $\sum_{i=n+1}^{2n} \hat{x}_i^t = x^t[2]$ for all t.

Lemma 3. *Let the losses $\hat{\ell}^1, \dots, \hat{\ell}^T$ satisfy the duplication procedure given in the preceding paragraph. Then for any time t, we have $\hat{x}_1^t = \cdots = \hat{x}_n^t$ and $\hat{x}_{n+1}^t = \cdots = \hat{x}_{2n}^t$.*

Proof. Suppose not and let $\hat{x}^t$ be the corresponding solution. Then the optimal solution is such that $\hat{x}_i^t \neq \hat{x}_k^t$ for some i, k both less than n, or both greater than n. But then, by symmetry, we have that there is more than one optimal solution to the OFTRL optimization problem at time t: the objective is exactly the same if we create a new solution where we swap the values of $\hat{x}_i^t$ and $\hat{x}_k^t$. This is a contradiction due to strong convexity. □

From lemma 3, we have that the OFTRL decision problem in $2n$ dimensions can equivalently be written as a 2-dimensional decision problem: Since the first n entries must be the same, we can simply optimize over that one shared value, say $x^t[1]$, which we use for all n entries, and similarly we use $x^t[2]$ for the second half of the entries. Let $\text{Dupl} : \Delta^2 \to \Delta^{2n}$ be a function that maps the two-dimensional solution into the corresponding duplicated $2n$-dimensional solution. The equivalent 2-dimensional problem is then:

$$\begin{aligned}
\hat{x}^t &= \text{Dupl}\left[\operatorname*{argmin}_{x\in\frac{1}{n}\cdot\Delta^2}\left\{\frac{n}{n^\alpha}\left\langle x, \sum_{\tau=1}^{t-1}\ell^\tau + \ell^{t-1}\right\rangle + \frac{n}{\eta}R_1(x[1]) + \frac{n}{\eta}R_2(x[2])\right\}\right] \\
&= \text{Dupl}\left[\frac{1}{n}\cdot\operatorname*{argmin}_{x\in\Delta^2}\left\{\frac{n}{n^\alpha}\left\langle \frac{1}{n}x, \sum_{\tau=1}^{t-1}\ell^\tau + \ell^{t-1}\right\rangle + \frac{n}{\eta}R(x/n)\right\}\right] \\
&= \text{Dupl}\left[\frac{1}{n}\cdot\operatorname*{argmin}_{x\in\Delta^2}\left\{\left\langle x, \sum_{\tau=1}^{t-1}\ell^\tau + \ell^{t-1}\right\rangle + \frac{n^{\alpha+1}}{\eta}R(x/n)\right\}\right].
\end{aligned}$$

The next theorem shows that we can choose α for different regularizers and construct $2n \times 2n$ loss matrices whose learning dynamics are equivalent to the learning dynamics in 2×2 games given in the preceding sections. We defer the proof to Appendix D.

Theorem 3. *For any loss matrix $A \in [0, 1]^{2\times 2}$, there exists a loss matrix $\hat{A} \in [0, n^{-\alpha}]^{2n\times 2n}$ such that for the Euclidean ($\alpha = 1$), entropy ($\alpha = 0$), Tsallis ($\beta \in (0, 1)$ and $\alpha = -1 + \beta$), and log ($\alpha = -1$) regularizers, the resulting OFTRL learning dynamics are equivalent in the two games.*

https://doi.org/10.52202/079017-0737

Combining Theorem 1 and Theorem 3, we have the following:

Corollary 1. *In the same setup as Theorem 3, under Assumption 1 and Assumption 2, there exists a game matrix* $\hat{A}_\delta \in [0, n^{-\alpha}]^{2n \times 2n}$ *such that the OFTRL learning dynamics with any step size* $\eta \leq \frac{1}{4L}$ *satisfies the following: there exists an iteration* $t \geq \frac{c_1}{3\eta L \delta}$ *with a duality gap at least* $c_2 n^{-\alpha}$.

Since $\alpha = 0$ for the entropy regularizer, the same results hold more generally for games where one player has more actions than the other. In particular, we can create a $2n \times 2m$ game such that the resulting dynamics are equivalent to those in a 2×2 game. This does not work for the Euclidean and log regularizers because the rescaling factors would be different for the row and column players.

5 Conclusion and Discussions

In this paper, we study last-iterate convergence rates of OFTRL algorithms with various popular regularizers, including the popular OMWU algorithm. Our main results show that even in simple 2×2 two-player zero-sum games parametrized by $\delta > 0$, the lack of forgetfulness of OFTRL leads to the duality gap remaining constant even after $1/\delta$ iterations (Theorem 1). As a corollary, we show that the last-iterate convergence rate of OFTRL must depend on a problem-dependent constant that can be arbitrarily bad (Theorem 2). This highlights a stark contrast with OOMD algorithms: while OGDA with constant step size achieves a $O(\frac{1}{\sqrt{T}})$ last-iterate convergence rate, such a guarantee is impossible for OMWU or more generally OFTRL. We now discuss several interesting questions regarding the convergence guarantees of learning in games and leave them as future directions.

Best-Iterate Convergence Rates While we focus on the last-iterate (*i.e.*, $\mathrm{DualityGap}(x^T, y^T)$), the weaker notion of best-iterate (*i.e.*, $\min_{t \in [T]} \mathrm{DualityGap}(x^t, y^t)$) is also of both practical and theoretical interest. By definition, we know the best-iterate convergence rate is at least as good as the last-iterate convergence rate and could be much faster. This raises the following question:

What is the best-iterate convergence rate of OMWU/OFTRL?

To our knowledge, there are no concrete results on the best-iterate convergence rates of OMWU or other OFTRL algorithms. It is thus interesting to extend our negative results to the best-iterate convergence rates or develop fast best-iterate convergence rates of OMWU/OFTRL.

Dynamic Step Sizes Our negative results hold for OFTRL with *fixed* step sizes. We conjecture that the slow last-iterate convergence of OFTRL persists even with *dynamic* step sizes. In particular, we believe our counterexamples still work for OFTRL with decreasing step sizes. This is because decreasing the step size makes the players move even slower, and they may be trapped in the wrong direction for a longer time due to the lack of forgetfulness. In Appendix E, we include numerical results for OMWU with adaptive stepsize akin to Adagrad [Duchi et al., 2011], which supports our intuition. We observe the same cycling behavior as for fixed step size. While the cycle is smaller than that of fixed step sizes, the dynamics take more steps to finish each cycle. Investigating the effect of dynamic step sizes on last-iterate convergence rates is an interesting future direction.

Slow Convergence due to Lack of Forgetfulness Our work shows that various OFTRL-type algorithms do not have fast last-iterate convergence rates for learning in games. Our proof and hard game instance build on the intuition that these algorithms lack forgetfulness: they do not forget the past quickly. This intuition is also utilized in [Panageas et al., 2023]. In particular, they give an $d \times d$ potential game where the last-iterate convergence rate of the Fictitious Play algorithm, which is equivalent to the Follow-the-Leader (FTL) algorithm, suffers exponential dependence in the dimension d. One natural future direction is to formalize the intuition of non-forgetfulness further and give a general condition for algorithms under which they suffer slow last-iterate convergence. It is also interesting to show other lower-bound results for learning in games.

Acknowledgements

We thank the anonymous reviewers for their constructive comments on improving the paper. Yang Cai was supported by the NSF Awards CCF-1942583 (CAREER) and CCF-2342642. Christian Kroer was supported by the Office of Naval Research awards N00014-22-1-2530 and N00014-23-1-2374, and the National Science Foundation awards IIS-2147361 and IIS-2238960. Julien Grand-Clément was supported by Hi! Paris and Agence Nationale de la Recherche (Grant 11-LABX-0047). Haipeng Luo was supported by NSF award IIS-1943607. Weiqiang Zheng was supported by the NSF Awards CCF-1942583 (CAREER), CCF-2342642, and a Research Fellowship from the Center for Algorithms, Data, and Market Design at Yale (CADMY).

References

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

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

Mark Braverman, Jieming Mao, Jon Schneider, and Matt Weinberg. Selling to a no-regret buyer. In *Proceedings of the 2018 ACM Conference on Economics and Computation*, pages 523–538, 2018.

Noam Brown and Tuomas Sandholm. Superhuman AI for heads-up no-limit poker: Libratus beats top professionals. *Science*, 359(6374):418–424, 2018.

Yang Cai, Argyris Oikonomou, and Weiqiang Zheng. Finite-time last-iterate convergence for learning in multi-player games. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2022.

Yun Kuen Cheung and Georgios 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.

Constantinos Daskalakis and Ioannis Panageas. The limit points of (optimistic) gradient descent in min-max optimization. *Advances in neural information processing systems (NeurIPS)*, 2018.

Constantinos Daskalakis and Ioannis Panageas. Last-iterate convergence: Zero-sum games and constrained min-max optimization. In *10th Innovations in Theoretical Computer Science Conference (ITCS)*, 2019.

Constantinos Daskalakis, Andrew Ilyas, Vasilis Syrgkanis, and Haoyang Zeng. Training gans with optimism. In *International Conference on Learning Representations (ICLR)*, 2018.

Constantinos Daskalakis, Maxwell Fishelson, and Noah Golowich. Near-optimal no-regret learning in general games. *Advances in Neural Information Processing Systems (NeurIPS)*, 2021.

John Duchi, Elad Hazan, and Yoram Singer. Adaptive subgradient methods for online learning and stochastic optimization. *Journal of machine learning research*, 12(7), 2011.

Gabriele Farina, Christian Kroer, and Tuomas Sandholm. Faster game solving via predictive blackwell approachability: Connecting regret matching and mirror descent. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 35, pages 5363–5371, 2021.

Gabriele Farina, Chung-Wei Lee, Haipeng Luo, and Christian Kroer. Kernelized multiplicative weights for 0/1-polyhedral games: Bridging the gap between learning in extensive-form and normal-form games. In *International Conference on Machine Learning (ICML)*, pages 6337–6357, 2022.

Genevieve E Flaspohler, Francesco Orabona, Judah Cohen, Soukayna Mouatadid, Miruna Oprescu, Paulo Orenstein, and Lester Mackey. Online learning with optimism and delay. In *International Conference on Machine Learning*, pages 3363–3373. PMLR, 2021.

 https://doi.org/10.52202/079017-0737

Noah Golowich, Sarath Pattathil, and Constantinos Daskalakis. Tight last-iterate convergence rates for no-regret learning in multi-player games. *Advances in neural information processing systems (NeurIPS)*, 2020a.

Noah Golowich, Sarath Pattathil, Constantinos Daskalakis, and Asuman Ozdaglar. Last iterate is slower than averaged iterate in smooth convex-concave saddle point problems. In *Conference on Learning Theory (COLT)*, 2020b.

Eduard Gorbunov, Adrien Taylor, and Gauthier Gidel. Last-iterate convergence of optimistic gradient method for monotone variational inequalities. In *Advances in Neural Information Processing Systems*, 2022.

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

Yu-Guan Hsieh, Franck Iutzeler, Jérôme Malick, and Panayotis Mertikopoulos. On the convergence of single-call stochastic extra-gradient methods. *Advances in Neural Information Processing Systems*, 32, 2019.

Yu-Guan Hsieh, Kimon Antonakopoulos, and Panayotis Mertikopoulos. Adaptive learning in continuous games: Optimal regret bounds and convergence to nash equilibrium. In *Conference on Learning Theory*, pages 2388–2422. PMLR, 2021.

Wouter M Koolen, Manfred K Warmuth, Jyrki Kivinen, et al. Hedging structured concepts. In *COLT*, pages 93–105. Citeseer, 2010.

Rachitesh Kumar, Jon Schneider, and Balasubramanian Sivan. Strategically-robust learning algorithms for bidding in first-price auctions. In *Proceedings of the 2024 ACM Conference on Economics and Computation*, 2024.

Chung-Wei Lee, Christian Kroer, and Haipeng Luo. Last-iterate convergence in extensive-form games. *Advances in Neural Information Processing Systems*, 34:14293–14305, 2021.

Haipeng Luo. Lecture note 2, Introduction to Online Learning. 2022. URL https://haipeng-luo.net/courses/CSCI659/2022_fall/lectures/lecture2.pdf.

Panayotis Mertikopoulos, Christos Papadimitriou, and Georgios Piliouras. Cycles in adversarial regularized learning. In *Proceedings of the twenty-ninth annual ACM-SIAM symposium on discrete algorithms*, pages 2703–2717. SIAM, 2018.

Panayotis Mertikopoulos, Bruno Lecouat, Houssam Zenati, Chuan-Sheng Foo, Vijay Chandrasekhar, and Georgios Piliouras. Optimistic mirror descent in saddle-point problems: Going the extra (gradient) mile. In *International Conference on Learning Representations (ICLR)*, 2019.

Aryan Mokhtari, Asuman E Ozdaglar, and Sarath Pattathil. Convergence rate of $\mathcal{O}(1/k)$ for optimistic gradient and extragradient methods in smooth convex-concave saddle point problems. *SIAM Journal on Optimization*, 30(4):3230–3251, 2020.

Rémi Munos, Michal Valko, Daniele Calandriello, Mohammad Gheshlaghi Azar, Mark Rowland, Zhaohan Daniel Guo, Yunhao Tang, Matthieu Geist, Thomas Mesnard, Andrea Michi, et al. Nash learning from human feedback. *arXiv preprint arXiv:2312.00886*, 2023.

Ioannis Panageas, Nikolas Patris, Stratis Skoulakis, and Volkan Cevher. Exponential lower bounds for fictitious play in potential games. In *Thirty-seventh Conference on Neural Information Processing Systems*, 2023. URL https://openreview.net/forum?id=tkenkPYkxj.

Julien Perolat, Bart De Vylder, Daniel Hennes, Eugene Tarassov, Florian Strub, Vincent de Boer, Paul Muller, Jerome T Connor, Neil Burch, Thomas Anthony, et al. Mastering the game of stratego with model-free multiagent reinforcement learning. *Science*, 378(6623):990–996, 2022.

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

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

Eiji Takimoto and Manfred K Warmuth. Path kernels and multiplicative updates. *The Journal of Machine Learning Research*, 4:773–818, 2003.

Oskari Tammelin, Neil Burch, Michael Johanson, and Michael Bowling. Solving heads-up limit texas hold'em. In *Twenty-fourth international joint conference on artificial intelligence*, 2015.

Tim van Erven. Why FTRL is better than online mirror descent. https://www.timvanerven.nl/blog/ftrl-vs-omd/, 2021. Accessed: 2024-05-22.

Emmanouil-Vasileios Vlatakis-Gkaragkounis, Lampros Flokas, Thanasis Lianeas, Panayotis Mertikopoulos, and Georgios Piliouras. No-regret learning and mixed nash equilibria: They do not mix. *Advances in Neural Information Processing Systems*, 33:1380–1391, 2020.

Chen-Yu Wei, Chung-Wei Lee, Mengxiao Zhang, and Haipeng Luo. Linear last-iterate convergence in constrained saddle-point optimization. In *International Conference on Learning Representations (ICLR)*, 2021.

Martin Zinkevich, Michael Johanson, Michael Bowling, and Carmelo Piccione. Regret minimization in games with incomplete information. *Advances in neural information processing systems*, 20, 2007.

Contents

 https://doi.org/10.52202/079017-0737

A Missing Proofs in Section 2

A.1 Proof of Lemma 1

Proof. Let $e_1 < e_2$. Denote $x_1 = F_{\eta,R}(e_1)$ and $x_2 = F_{\eta,R}(e_2)$. By definition, we have

$$e_2(x_2 - x_1) \le \frac{1}{\eta}(R(x_1) - R(x_2)) \le e_1(x_2 - x_1).$$

Since $e_1 < e_2$, we have $x_2 \le x_1$. □

A.2 Proof of Proposition 1

Proof. By definition,

$$F_{\eta,R}\left(\frac{E}{\eta}\right) = \operatorname*{argmin}_{x\in[0,1]}\left\{x \cdot \frac{E}{\eta} + \frac{1}{\eta}R(x)\right\} = \operatorname*{argmin}_{x\in[0,1]}\{x \cdot E + R(x)\} = F_{1,R}(E).$$

The second claim on the Lipschitzness follows directly. □

B Missing Proofs in Section 3

B.1 Proof of Lemma 2

Proof. We have

$$\begin{aligned}
\text{DualityGap}(x,y) &= \max_{\tilde{y}\in\Delta^2} x^\top A_\delta \tilde{y} - \min_{\tilde{x}\in\Delta^2} \tilde{x}^\top A_\delta y \\
&= \max_{i\in\{1,2\}} (A_\delta^\top x)[i] - \min_{i\in\{1,2\}} (Ay_\delta)[i] \\
&= (\frac{1}{2} + \delta)x[1] - (1 - y[1]) && (x[1] \ge \tfrac{1}{1+\delta}, \epsilon > 0) \\
&\ge \frac{1}{2}\frac{1+2\delta}{1+\delta} - \frac{1}{2} + \epsilon \\
&\ge \epsilon.
\end{aligned}$$

□

B.2 Proof of Theorem 1

Proof. Recall that $c_1 = \frac{1}{2} - F_{1,R}(\frac{1}{20L})$ defined in Assumption 2. We fix any $\delta < \min\{\frac{1}{15}, \frac{c_1}{6}, \frac{c_1^2}{300}, \delta'\}$. Since $\delta < \delta'$, Assumption 2 holds. We will prove that there exists an iteration $t \ge \frac{c_1}{3\eta L\delta}$ with duality gap c_2.

Proof Plan: We decompose the analysis into three stages as shown in Figure 3. Below, we describe the three stages and the high-level ideas in our proof.

- **Stage I:** Recall that $x^1[1] = y^1[1] = \frac{1}{2}$. In Stage I, we show that $x^t[1]$ will increase and denote $T_1 \ge 1$ the first iteration where $x^t[1] \ge \frac{1}{1+\delta}$. The existence of T_1 can be proved by contradiction (Claim 1). Since before the end of Stage I, $x^t[1] < \frac{1}{1+\delta}$, the loss vector for the y-player satisfies $e_y^t = \ell_y^t[1] - \ell_y^t[2] \ge 0$ meaning action 1 is worse than action 2. We will prove that finally $y^{T_1}[1] \le \frac{1}{2} - c_1$.

- **Stage II:** Now we have that $y^{T_1}[1] \le \frac{1}{2} - c_1$, we denote $T_2 > T_1$ the first iteration where $y^{T_2}[1] \ge \frac{1}{2(1+\delta)} > \frac{1}{2} - c_1$. We remark that in order to increase $y^t[1]$, the loss vector must satisfy $e_y^t < 0$. However, the game matrix A_δ guarantees that $e_y^t \ge -\delta$ no matter what the x-player is playing. Thus by the ηL-Lipschitzness of $F_{\eta,R}$ (Lemma 4), the increase in $y^t[1]$ is at most $\eta L\delta$. Therefore, we know $T_2 - T_1 = \Omega(\frac{c_1}{\eta L\delta})$. But during $[T_1, T_2]$, for the x-player, we have $e_x^t < 0$ which implies its cumulative loss $E_x^{T_2} \le E_x^{T_1} - \Omega(\frac{1}{\eta L\delta})$. In other words, $x^t[1]$ is very close to 1 and the cumulative loss for action 1 is much smaller than that of action 2.
- **Stage III:** Now we have $y^{T_2}[1] \ge \frac{1}{2(1+\delta)}$ and that $y^t[1]$ could keep increasing if $x^t[1] \ge \frac{1}{1+\delta}$ since then the loss satisfies $e_y^t \le 0$. Now the question is how long would the x-player stay close to the boundary, i.e, $x^t[1] \ge \frac{1}{1+\delta}$. Since OFTRL-type algorithms are not forgetful, this happens only when $E_x^t \ge E_x^{T_1}$ (recall $x^{T_1}[1] \ge \frac{1}{1+\delta}$). But we have at the end of stage II, $E_x^{T_2} \le E_x^{T_1} - \Omega(\frac{1}{\eta L\delta})$. Since$e_x^t$ is bounded by a constant, we know $x^t[1] \ge \frac{1}{1+\delta}$ even after $\Omega(\frac{1}{\eta L\delta})$ iterations. Define $T_3 = T_2 + \Omega(\frac{1}{\eta L\delta})$. During $[T_2, T_3]$, the y-player always receives loss such that $e_y^t \le 0$ and we prove that $y^{T_3}[1] \ge \frac{1}{2} + c_2$ for some constant c_2.
- **Conclusion** Finally we get one iteration $T_3 \ge \Omega(\frac{1}{\eta L\delta})$ with $x^{T_3}[1] \ge \frac{1}{1+\delta}$ and $y^{T_3}[1] \ge \frac{1}{2} + c_2$, Using Lemma 2, the duality gap of (x^{T_3}, y^{T_3}) is at least c_2.

Stage I: We know $x^1[1] = y^1[1] = \frac{1}{2}$. We define (i) $T_s > 1$ to be the smallest iteration such that $x^{T_s}[1] \ge \frac{3}{4}$ and (ii) $T_1 > T_s$ to be the smallest iteration such that $x^{T_1}[1] \ge \frac{1}{1+\delta}$. Both T_s and T_1 must exist, and the reason will become clear in the following analysis. We postpone the proof of this fact in Claim 1 at the end of this paragraph.

Notice from Proposition 2, the difference e_x^t is lower bounded: $e_x^t \ge -\frac{1}{2}$ for any t. Thus $E_x^{t-1} + e_x^{t-1} \ge -\frac{t}{2}$ for any $t \ge 1$. Since $x^{T_s}[1] \ge \frac{3}{4} > \frac{1}{2}$, we know that $E_x^{T_s-1} + e_x^{T_s-1} < 0$. As $F_{\eta,R}$ is ηL-Lipschitz,

$$\frac{1}{4} \le x^{T_s}[1] - x^1[1] \le \eta L \cdot \left|E_x^{T_s-1} + e_x^{T_s-1}\right| \le \frac{L\eta T_s}{2}.$$

This implies

$$T_s \ge \frac{1}{2\eta L}.$$

Since $x^t[1] < \frac{3}{4}$ for all $1 \le t \le T_s - 1$, we know that $e_y^t = \ell_y^t[1] - \ell_y^t[2] = 1 - (1+\delta)x^t[1] > \frac{1-3\delta}{4} \ge \frac{1}{5}$ (as $\delta \le \frac{1}{15}$) for all $1 \le t \le T_s - 1$. Moreover, for all $1 \le t \le T_1 - 1$, we know that $e_y^t \ge 0$ as $x^t[1] \le \frac{1}{1+\delta}$. Since the difference e_y^t is at least $1/5$ for all $t \le T_s - 1$ and remains non-negative for all $t \in [T_s, T_1 - 1]$, we can conclude that for all $T_s \le t \le T_1$

$$y^t[1] = F_{\eta,R}(E_y^{t-1} + e_y^{t-1}) \le F_{\eta,R}(E_y^{t-1}),$$

and moreover

$$\begin{aligned} F_{\eta,R}(E_y^{t-1}) &\le F_{\eta,R}\left(\frac{T_s - 1}{5}\right) \\ &\le F_{\eta,R}\left(\frac{1}{20L\eta}\right) && (T_s - 1 \ge \tfrac{1}{2\eta L} - 1 \ge \tfrac{1}{4L\eta}) \\ &= \frac{1}{2} - c_1. \end{aligned}$$

This completes the proof of Stage I, where $x^{T_1}[1] \ge \frac{1}{1+\delta}$ and $y^{T_1}[1] \le \frac{1}{2} - c_1$. Before we proceed to the next stage, we prove the existence of T_s and T_1.

Claim 1. *T_s and T_1 exist.*

Proof. It suffices to prove that T_1 exists as it implies the existence of T_s. Assume for the sake of contradiction that T_1 does not exist, *i.e.*, $x^t[1] < \frac{1}{1+\delta}$ for all $t \ge 1$. By the same analysis as for Stage I, we get $y^t[1] \le \frac{1}{2} - c_1$ for all $t \ge \frac{1}{2\eta L}$. This implies $e_x^t = -\frac{1}{2} + (1+\delta)y^t[1] \le \frac{\delta}{2} - c_1 \le -\frac{c_1}{2}$ for

https://doi.org/10.52202/079017-0737

all $t \geq \frac{1}{2\eta L}$. Then $E_x^t + e_x^t \to -\infty$ as $t \to +\infty$. As a consequence, $x^t[1] = F_{\eta,R}(E_x^{t-1} + e_x^{t-1}) \to 1$ as $t \to +\infty$ by item 2 in Assumption 1. But this contradicts with the assumption that $x^t[1] < \frac{1}{1+\delta}$ for all $t \geq 1$. This completes the proof. □

Stage II We define

$$T := \left\lfloor \frac{c_1}{2L\eta\delta} \right\rfloor \in \left[\frac{c_1}{3L\eta\delta}, \frac{c_1}{2L\eta\delta} \right], \tag{4}$$

where the lower bound on T holds since $\frac{c_1}{6L\eta\delta} \geq \frac{c_1}{6\delta} \geq 1$. We note that $T = \Omega(\frac{1}{\delta})$ since $\eta L \leq \frac{1}{4}$.

In Stage I, we have proved that $y^{T_1}[1] \leq \frac{1}{2} - c_1$. Define $T_h = T_1 + T$. We claim that for all $t \in [T_1, T_h - 1]$, $y^t[1] \leq \frac{1}{2} - \frac{c_1}{2}$. To prove the claim, we first notice that $-\delta \leq e_y^t \leq 1$ for all $t \geq 1$. Then by the monotonicity and the ηL-Lipschitzness of $F_{\eta,R}$ (Lemma 1 and Lemma 4), we get for all $t \in [T_1, T_h - 1]$,

$$\begin{aligned}
y^t[1] &\leq F_{\eta,R}(E_y^{T_1-1}) + \eta L \max\left\{ E_y^{T_1-1} - E_y^{t-1} - e_y^{t-1}, 0 \right\} \\
&\leq \frac{1}{2} - c_1 + \eta L \cdot (t - T_1 + 1)\delta \\
&\leq \frac{1}{2} - c_1 + \eta L T \delta \\
&\leq \frac{1}{2} - \frac{c_1}{2},
\end{aligned}$$

where, in the second-to-last inequality, we use $t - T_1 + 1 \leq T \leq \frac{c_1}{2\eta L\delta}$ by Equation (4).

Now we denote $T_2 \geq T_h$ the smallest iteration when $y^{T_2}[1] \geq \frac{1}{2(1+\delta)}$. The existence of T_2 will become clear in the following analysis, and we postpone the proof to Claim 2 at the end of the discussion. Then for all $t \in [T_s, T_2 - 1]$, we have $y^t[1] \leq \frac{1}{2(1+\delta)}$, which implies $e_x^t \leq 0$. Moreover, for all $t \in [T_s, T_1 + T - 1]$, since $y^t[1] \leq \frac{1}{2} - \frac{c_1}{2}$, we have

$$\begin{aligned}
e_x^t &= \ell_x^t[1] - \ell_x^t[2] \\
&= -\frac{1}{2} + (1+\delta) y^t[1] \\
&\leq \frac{-1 + (1+\delta)(1-c_1)}{2} \\
&\leq \frac{\delta - c_1}{2} \\
&\leq -\frac{c_1}{4}. && (\delta \leq \tfrac{c_1}{2})
\end{aligned}$$

Then for any $T_1 + T \leq t \leq T_2$, we have

$$\begin{aligned}
x^t[1] &= F_{\eta,R}(E_x^{t-1} + e_x^{t-1}) \\
&\geq F_{\eta,R}(E_x^{T_1+T-1}) && (e_x^{t-1} \leq 0 \text{ for all } t \in [T_1 + T, T_2]) \\
&\geq F_{\eta,R}\left(-\frac{c_1 T}{4} + E_x^{T_1-1} \right) \\
&\geq F_{\eta,R}\left(-\frac{c_1 T}{5} + E_x^{T_1-1} + e_x^{T_1-1} \right),
\end{aligned}$$

where in the last inequality, we use the fact that $\frac{c_1 T}{20} \geq \frac{c_1^2}{60\eta L\delta} \geq 1$.

Claim 2. *T_2 exists.*

Proof. Assume for the sake of contradiction that T_2 does not exist, *i.e.*, $y^t[1] < \frac{1}{2(1+\delta)}$ for all $t \geq T_1$ (since we know $y^t[1] \leq \frac{1}{2} - \frac{c_1}{2}$ for all $t \in [T_1, T_1 + T - 1]$). Then by the analysis of Stage II and Equation (5), we have $x^t[1] \geq \frac{4}{4+\delta}$ for all $t \geq T_1$. This implies $e_y^t \leq -\frac{3\delta}{5}$ for all $t \geq T_1$. As a result, we have $E_y^{t-1} + e_y^{t-1} \to -\infty$ as $t \to \infty$. By item 2 in Assumption 1, we get $y^t[1] = F_{\eta,R}(E_y^{t-1} + e_y^{t-1}) \geq \frac{1}{2}$ as $t \to \infty$. But this contradicts with the assumption that $y^t[1] < \frac{1}{2(1+\delta)}$ for all $t \geq T_1$. This completes the proof. □

Stage III Recall that we have argued in State I that $F_{\eta,R}(E_x^{T_1-1}+e_x^{T_1-1}) = F_{1,R}(\eta(E_x^{T_1-1}+e_x^{T_1-1})) = x^{T_1}[1] \geq \frac{1}{1+\delta}$. By item 1 in Assumption 2, we have that

$$\begin{aligned} F_{\eta,R}\left(-\frac{c_1 T}{10}+E_x^{T_1-1}+e_x^{T_1-1})\right) &\geq F_{\eta,R}\left(-\frac{c_1^2}{30L\eta\delta}+E_x^{T_1-1}+e_x^{T_1-1})\right) \\ &= F_{1,R}\left(-\frac{c_1^2}{30L\delta}+\eta(E_x^{T_1-1}+e_x^{T_1-1}))\right) \\ &\geq \frac{1+c_3}{1+c_3+\delta}, \end{aligned} \tag{5}$$

where the first inequality follows from the definition of T and the monotonicity of $F_{\eta,R}$ (Lemma 1).

Now denote $T_3 = T_2 + \lfloor \frac{c_1 T}{10} \rfloor - 2$. For any $T_2 \leq t \leq T_3$, we know that

$$\begin{aligned} x^t[1] &= F_{\eta,R}(E_x^{t-1}+e_x^{t-1}) \\ &= F_{\eta,R}(E_x^{T_2-1}+e_x^{T_2-1}+\sum_{k=T_2}^{t-1} e_x^k + e_x^{t-1} - e_x^{T_2-1}) \\ &\geq F_{\eta,R}(-\frac{c_1 T}{5}+E_x^{T_1-1}+e_x^{T_1-1}+\sum_{k=T_2}^{t-1} e_x^k + e_x^{t-1} - e_x^{T_2-1}) \\ &\geq F_{\eta,R}(-\frac{c_1 T}{5}+E_x^{T_1-1}+e_x^{T_1-1}+\frac{c_1 T}{10}-2+2) \\ &\geq F_{\eta,R}(-\frac{c_1 T}{10}+E_x^{T_1-1}+e_x^{T_1-1})) \\ &\geq \frac{1+c_3}{1+c_3+\delta}. \qquad \text{(by (5))} \end{aligned}$$

Note that $1+c_3+\delta \leq 2$. This implies $e_y^t = 1-(1+\delta)x^t[1] = -\frac{c_3\delta}{1+c_3+\delta} \leq -\frac{c_3\delta}{2}$ for all $T_2 \leq t \leq T_3$. Moreover, we know that $e_y^t \geq -\delta$ for any t. Then

$$\begin{aligned} y^{T_3}[1] &= F_{\eta,R}(E_y^{T_3-1}+e_y^{T_3-1})) \\ &\geq F_{\eta,R}(E_y^{T_2-1}+e_y^{T_2-1}+\sum_{k=T_2}^{T_3-1} e_y^k + e_y^{T_3-1} - e_y^{T_2-1})) \\ &\geq F_{\eta,R}(E_y^{T_2-1}+e_y^{T_2-1}-\frac{c_3\delta(T_3-T_2)}{2}+\delta) \\ &\geq F_{\eta,R}(E_y^{T_2-1}+e_y^{T_2-1}-\frac{c_3\delta c_1 T}{40}+\delta) \qquad (T_3-T_2 = \lfloor \tfrac{c_1 T}{10} \rfloor - 2 \geq \tfrac{c_1 T}{20}) \\ &\geq F_{\eta,R}(E_y^{T_2-1}+e_y^{T_2-1}-\frac{c_3 c_1^2}{120\eta L}+\delta) \qquad (T \geq \tfrac{c_1}{3\eta L\delta}) \\ &= F_{1,R}(\eta(E_y^{T_2-1}+e_y^{T_2-1})-\frac{c_3 c_1^2}{120L}+\eta\delta) \\ &\geq F_{1,R}(\eta(E_y^{T_2-1}+e_y^{T_2-1})-\frac{c_3 c_1^2}{120L}+\frac{\delta}{4L}). \qquad (\eta \leq \tfrac{1}{4L}) \end{aligned}$$

Recall that $F_{1,R}(\eta(E_y^{T_2-1}+e_y^{T_2-1})) = F_{\eta,R}(E_y^{T_2-1}+e_y^{T_2-1}) = y^{T_2}[1] \geq \frac{1}{2(1+\delta)}$. By item 2 in Assumption 2, we have $F_{1,R}(\eta(E_y^{T_2-1}+e_y^{T_2-1})-\frac{c_1^2}{120L}+\frac{\delta}{4L}) \geq \frac{1}{2}+c_2$ for some absolute constant $c_2 > 0$. Thus, we have $y^{T_3}[1] \geq \frac{1}{2}+c_2$. Recall that $x^{T_3}[1] \geq \frac{1+c_3}{1+c_3+\delta} \geq \frac{1}{1+\delta}$. Then by Lemma 2 we can conclude that the duality gap of (x^{T_3}, y^{T_3}) is at least $c_2 > 0$. This completes the proof as $T_3 \geq T_2 \geq T \geq \frac{c_1}{3\eta L\delta}$. □

B.3 Proof of Theorem 2

Proof. Assume for the sake of contradiction that there is a function that satisfies both conditions. Then for any $A \in [0,1]^{2\times 2}$, we have the OFTRL learning dynamics over A satisfies

https://doi.org/10.52202/079017-0737

1. $\text{DualityGap}(x^T, y^T) \le f(2,2,T)$ for all T.
2. $\lim_{T\to\infty} f(2,2,T) \to 0$

Since $\lim_{T\to\infty} f(2,2,T) \to 0$, we know there exists $T_0 > 0$ such that for any $t \ge T_0$, $\text{DualityGap}(x^t, y^t) \le f(2,2,t) < c_2$. Now let $\delta \le \min\{\hat{\delta}, \frac{c_1}{3\eta L T_0}\}$. Then by Theorem 1, we know there exists an iteration $t \ge \frac{c_1}{3\eta L\delta} \ge T_0$ such that $\text{DualityGap}(x^t, y^t) \ge c_2$. This completes the proof. □

C Verifying Assumption 2 for Different Regularizers

Lemma 4. *If the regularizer R is 1-strongly convex, then $F_{1,R}$ is $\frac{1}{2}$-Lipschitz.*

Proof. Notice that $R(x) + R(1-x)$ is 2-strongly convex. Thus by standard analysis (see e.g., Luo [2022, Lemma 4]) we know $F_{1,R}$ is $\frac{1}{2}$-Lipschitz. □

By Lemma 4, we can choose $L = \frac{1}{2}$ for any 1-strongly convex regularizer in Assumption 1.

C.1 Negative Entropy

Lemma 5 (Assumption 2 holds for the entropy regularizer)**.** *Consider the negative entropy regularizer R defined as $R(x) = x\log x + (1-x)\log(1-x)$. Then $F_{1,R}$ is $L = \frac{1}{2}$-Lipschitz. We have c_1 and Assumption 2 holds with $\delta' = \frac{c_1^2}{480L}$, $c_2 = F_{1,R}(-\frac{c_1^2}{480L}) - \frac{1}{2}$, and $c_3 = \frac{1}{2}$.*

Proof. It is easy to verify that $F_{1,R}(x)$ has a closed-form representation

$$F_{1,R}(E) = \frac{1}{1+\exp(E)}.$$

Thus $L = \frac{1}{2}$ and $c_1 = \frac{1}{2} - F_{1,R}(\frac{1}{20L})$ is a universal constant. We also choose $c_3 = \frac{1}{2}$.

If $F_{1,R}(E) \ge \frac{1}{1+\delta} \ge \frac{1}{1+\delta}$, then we have $E \le -\log(1/\delta)$. We note that

$$\exp\left(-\frac{c_1^2}{30L\delta}\right) \le \frac{1}{1+c_3} \Rightarrow \frac{1}{1+\exp\left(-\frac{c_1^2}{30L\delta} - \log(1/\delta)\right)} \ge \frac{1+c_3}{1+c_3+\delta}.$$

Thus $\delta \le \delta_1 = \frac{c_1^2}{30L\log(1+c_3))} = \frac{c_1^2}{30\log(\frac{3}{2}))L}$ suffices for item 1 in Assumption 2.

If $F_{1,R}(E) \ge \frac{1}{2(1+\delta)} = \frac{1}{1+1+2\delta}$, we have $E \le \log(1+2\delta)$. Note that since $\log(1+2y) \le 2y$ for $y > 0$, we have

$$\begin{aligned}\delta \le \frac{c_3c_1^2}{480L} &\Rightarrow -\frac{c_3c_1^2}{120L} + \log(1+2\delta) < -\frac{c_3c_1^2}{240L} \\ &\Rightarrow F_{1,R}\left(-\frac{c_3c_1^2}{120L} + E\right) > F_{1,R}\left(-\frac{c_3c_1^2}{240L}\right).\end{aligned}$$

Thus item 2 in Assumption 2 holds for any $\delta \le \delta_2 = \frac{c_3c_1^2}{480L} = \frac{c_1^2}{960L}$ and $c_2 = F_{1,R}(-\frac{c_3c_1^2}{240L}) - \frac{1}{2} = F_{1,R}(-\frac{c_1^2}{480L}) - \frac{1}{2}$.

Combining the above, we know Assumption 2 holds for the negative entropy regularizer with $\delta' = \frac{c_1^2}{960L}$ and $c_2 = F_{1,R}(-\frac{c_1^2}{480L}) - \frac{1}{2}$. □

C.2 Squared Euclidean Norm Regularizer

Lemma 6 (Assumption 2 holds for the Euclidean regularizer)**.** *Consider the Euclidean regularizer R defined as $R(x) = \frac{1}{2}(x^2 + (1-x)^2)$. We have $L = \frac{1}{2}$ and $c_1 = \frac{1}{20}$. We also have Assumption 2 holds with $\delta' = \frac{c_1^2}{480L}$, $c_2 = \frac{c_1^2}{960L}$, and $c_3 = \frac{1}{2}$.*

Proof. It is easy to verify that $F_{1,R}(x)$ has a closed-form representation

$$F_{1,R}(x) = \begin{cases} 1 & \text{if } x \le -1 \\ \frac{1-x}{2} & \text{if } x \in (-1,1) \\ 0 & \text{if } x \ge 1 \end{cases}$$

Thus $F_{1,R}$ is L-Lipschitz with $L = \frac{1}{2}$. Moreover, $c_1 = \frac{1}{2} - F_{1,R}(\frac{1}{20L}) = \frac{1}{20}$. We choose $c_3 = \frac{1}{2}$.

Fix any E such that $F_{1,R}(E) \ge \frac{1}{1+\delta}$. We have $E \le -\frac{1-\delta}{1+\delta} < 0$. We note that for any $\delta \le \frac{c_1^2}{30L} = \frac{c_1^2}{15}$,

$$F_{1,R}\left(-\frac{c_1^2}{30L\delta} + E\right) \ge F_{1,R}(-1) = 1.$$

Thus $\delta \le \delta_1 = \frac{c_1^2}{30L}$ suffices for item 1 in Assumption 2.

Fix any E such that $F_{1,R}(E) \ge \frac{1}{2(1+\delta)} = \frac{1}{2(1+\delta)}$. We have $E \le \frac{\delta}{1+\delta} \le \delta$. The for any $\delta \le \frac{c_3 c_1^2}{240L}$, we have

$$F_{1,R}\left(-\frac{c_3 c_1^2}{120L} + E\right) \ge F_{1,R}\left(-\frac{c_3 c_1^2}{240L}\right) = \frac{1}{2} + \frac{c_3 c_1^2}{480L}$$

Thus item 2 in Assumption 2 holds for any $\delta \le \delta_2 = \frac{c_1^2}{480L}$ and $c_2 = \frac{c_1^2}{960L}$.

Combining the above, we know Assumption 2 holds for the negative entropy regularizer with $\delta' = \min\{\delta_1, \delta_2\} = \frac{c_1^2}{480L}$ and $c_2 = \frac{c_1^2}{960L}$. □

C.3 Log Barrier

Lemma 7 (Assumption 2 holds for the log barrier)**.** *Consider the log barrier regularizer R defined as $R(x) = -\log(x) - \log(1-x)$. Then Assumption 2 holds with the following choices of constants:*

1. $c_1 = \sqrt{\frac{1}{4} + 400L^2} - 20L > 0$.
2. $c_3 = \frac{c_1^2}{60L}$.
3. $c_2 = \sqrt{\frac{1}{4} + (\frac{c_3 c_1^2}{240L})^2} - \frac{c_3 c_1^2}{240L} > 0$.
4. $\delta' = \frac{c_3 c_1^2}{2160L}$.

Proof. By setting the gradient of $x \cdot E + R(x)$ to 0, we get a closed-form expression of $F_{1,R}$:

$$F_{1,R}(E) = \begin{cases} \frac{1}{2} + \frac{1}{E} - \sqrt{\frac{1}{4} + \frac{1}{E^2}} & \text{if } E > 0 \\ \frac{1}{2} & \text{if } E = 0 \\ \frac{1}{2} + \frac{1}{E} + \sqrt{\frac{1}{4} + \frac{1}{E^2}} & \text{if } E < 0. \end{cases}$$

For $x \in (0,1)$, the $F_{1,R}$ function admits an inverse function defined as

$$F_{1,R}^{-1}(x) = \frac{2x-1}{x^2 - x}.$$

https://doi.org/10.52202/079017-0737

Thus we know $E_0 := F_{1,R}^{-1}(\frac{1}{1+\delta}) = -\frac{1-\delta^2}{\delta}$ satisfies $F_{1,R}(E_0) = \frac{1}{1+\delta}$. Moreover, we can calculate

$$\begin{aligned} F_{1,R}^{-1}\left(\frac{1+c_3}{1+c_3+\delta}\right) &= -\frac{(1+c_3)^2-\delta^2}{(1+c_3)\delta} \\ &= -\frac{1+c_3}{\delta} + \frac{\delta}{1+c_3} \\ &= E_0 - \frac{c_3}{\delta} - \frac{c_3\delta}{1+c_3}. \end{aligned}$$

Thus we can choose $c_3 = \frac{c_1^2}{60L}$ so that

$$\begin{aligned} & E_0 - \frac{c_1^2}{30L\delta} \\ &= E_0 - \frac{c_3}{\delta} - \frac{c_3}{\delta} \\ &\le E_0 - \frac{c_3}{\delta} - \frac{c_3\delta}{1+c_3}. && \text{(since } \delta < 1/2 \text{ and } c_3 > 0\text{)} \end{aligned}$$

Thus we have $F_{1,R}(E_0 - \frac{c_1^2}{30L\delta}) \ge F_{1,R}(E_0 - \frac{c_3}{\delta} - \frac{c_3\delta}{1+c_3}) \ge \frac{1+c_3}{1+c_3+\delta}$.

We calculate $E_1 := F_{1,R}^{-1}(\frac{1}{2(1+\delta)}) = \frac{4(\delta+\delta^2)}{1+2\delta} \le 8\delta$. Then we can choose $\delta \le \delta' := \frac{c_3c_1^2}{2160L}$. Then we have

$$\begin{aligned} & F_{1,R}(-\frac{c_3c_1^2}{120L} + \frac{\delta}{4L} + E_1) \\ &\ge F_{1,R}(-\frac{c_3c_1^2}{120L} + 9\delta) \\ &\ge F_{1,R}(-\frac{c_3c_1^2}{240L}) \\ &= \frac{1}{2} + c_2, \end{aligned}$$

where $c_2 = \sqrt{\frac{1}{4} + (\frac{c_3c_1^2}{240L})^2} - \frac{c_3c_1^2}{240L} > 0$ by the closed-form expression of $F_{1,R}$. □

C.4 Negative Tsallis Entropy

For $x \in [0,1]$, the negative Tsallis entropy is a family of regularizers parameterized by $\beta \in (0,1)$:

$$R(x) = \frac{1-x^\beta}{1-\beta}. \tag{6}$$

The corresponding $F_{1,R}$ is defined as

$$F_{1,R}(E) = \operatorname*{argmin}_{x\in(0,1)} \left\{ x\cdot E + \frac{1-x^\beta}{1-\beta} + \frac{1-(1-x)^\beta}{1-\beta} \right\}.$$

For $x \in (0,1)$, we note that $F_{1,R}$ has an inverse function

$$F_{1,R}^{-1}(x) = \frac{\beta}{1-\beta}\big(x^{\beta-1} - (1-x)^{\beta-1}\big).$$

Lemma 8 (Assumption 2 holds for Tsallis entropy)**.** *Consider Tsallis entropy parameterized by $\beta \in (0,1)$. Then $L = \frac{1}{2\beta}$ and Assumption 2 holds with the following choices of constants:*

1. $c_1 = \frac{1}{2} - F_{1,R}(\frac{1}{20L}) > 0$.
2. $c_3 = \frac{1}{2}$.
3. $c_2 = F_{1,R}(-\frac{c_3c_1^2}{240L}) - \frac{1}{2} > 0$.

4. $\delta' = \min\{(\frac{c_1^2(1-\beta)}{120L\beta c_3^{1-\beta}})^{\frac{1}{\beta}}, \frac{c_3 c_1^2}{120}, \frac{1-\beta}{8\beta}\cdot\frac{c_3 c_1^2}{480L}\}$.

Proof. We choose $c_3 = \frac{1}{2}$. We have $c_1 = \frac{1}{2} - F_{1,R}(\frac{1}{20L})$ is a constant.

We note that

$$E_0 := F_{1,R}^{-1}(\frac{1}{1+\delta}) = \frac{\beta}{1-\beta}\left((1+\delta)^{1-\beta} - \left(\frac{1+\delta}{\delta}\right)^{1-\beta}\right)$$

satisfies $F_{1,R}(E_0) = \frac{1}{1+\delta}$. Similarly, we calculate

$$\begin{aligned}
E_1 &:= F_{1,R}^{-1}(\frac{1+c_3}{1+c_3+\delta}) \\
&= \frac{\beta}{1-\beta}\left(\left(\frac{1+c_3+\delta}{1+c_3}\right)^{1-\beta} - \left(\frac{1+c_3+\delta}{\delta}\right)^{1-\beta}\right) \\
&\geq \frac{\beta}{1-\beta}\left((1+\delta)^{1-\beta} - 2 - \left(\frac{1+c_3+\delta}{\delta}\right)^{1-\beta}\right) \\
&\geq \frac{\beta}{1-\beta}\left((1+\delta)^{1-\beta} - \left(\frac{1+\delta}{\delta}\right)^{1-\beta} - \left(\frac{c_3}{\delta}\right)^{1-\beta} - 2\right) \\
&= E_0 - \frac{\beta}{1-\beta}\left(\left(\frac{c_3}{\delta}\right)^{1-\beta} + 2\right)
\end{aligned}$$

where in the first inequality we use the fact that $(1+\delta)^{1-\beta} \leq 2$ since $\delta \leq 1$; the second inequality we use the inequality $(x+y)^{1-\beta} \leq x^{1-\beta} + y^{1-\beta}$. We note that

$$\delta \leq \delta_1 := \left(\frac{c_1^2(1-\beta)}{120L\beta c_3^{1-\beta}}\right)^{\frac{1}{\beta}} \Rightarrow -\frac{c_1^2}{30L\delta} \leq -\frac{\beta}{1-\beta}\left(\left(\frac{c_3}{\delta}\right)^{1-\beta} + 2\right). \tag{7}$$

Thus for any $\delta \leq \delta_1$, we have for any E such that $F_{1,R}(E) \geq \frac{1}{1+\delta}$,

$$-\frac{c_1^2}{30L\delta} + E \leq -\frac{c_1^2}{30L\delta} + E_0 \leq E_0 - \frac{\beta}{1-\beta}\left(\left(\frac{c_3}{\delta}\right)^{1-\beta} + 2\right) \leq E_1.$$

The above implies $F_{1,R}(-\frac{c_1^2}{30L\delta} + E) \geq \frac{1+c_3}{1+c_3+\delta}$ and the first item in Assumption 2 is satisfied.

We define E_2

$$\begin{aligned}
E_2 := F_{1,R}^{-1}\left(\frac{1}{2(1+\delta)}\right) &= \frac{\beta}{1-\beta}\left((2+2\delta)^{1-\beta} - \left(\frac{2+2\delta}{1+2\delta}\right)^{1-\beta}\right) \\
&= \frac{\beta}{1-\beta}(2+2\delta)^{1-\beta}\cdot\left(1 - \left(\frac{1}{1+2\delta}\right)^{1-\beta}\right) \\
&\leq \frac{4\beta}{1-\beta}\cdot\left(1 - \left(1 - \frac{2\delta}{1+2\delta}\right)^{1-\beta}\right) \\
&\leq \frac{4\beta}{1-\beta}\cdot\frac{2\delta}{1+\delta} \\
&= \frac{8\beta\delta}{(1-\beta)(1+\delta)}
\end{aligned}$$

where in the first inequality we use $(2+2\delta)^{1-\beta} \leq 4$ since $0 \leq \delta \leq 1$ and $\beta \in (0,1)$; in the second inequality we use the basic inequality $(1-x)^r \leq 1-x$ for $r, x \in (0,1)$. We define

$$\delta_2 := \min\{\frac{c_3 c_1^2}{120}, \frac{1-\beta}{8\beta}\cdot\frac{c_3 c_1^2}{480L}\}$$

https://doi.org/10.52202/079017-0737

Then for any $\delta \le \delta_2$ and E such that $F_{1,R}[E] \ge \frac{1}{2(1+\delta)}$, we have

$$\begin{aligned}
-\frac{c_3 c_1^2}{120L} + \frac{\delta}{4L} + E &\le -\frac{c_3 c_1^2}{120L} + \frac{\delta}{4L} + E_2 \\
&\le -\frac{c_3 c_1^2}{120L} + \frac{c_3 c_1^2}{480L} + \frac{8\beta\delta}{(1-\beta)(1+\delta)} \\
&\le -\frac{c_3 c_1^2}{120L} + \frac{c_3 c_1^2}{480L} + + \frac{c_3 c_1^2}{480L} \\
&= -\frac{c_3 c_1^2}{240L}.
\end{aligned}$$

Thus we know $F_{1,R}(-\frac{c_3 c_1^2}{120L} + \frac{\delta}{4L} + E) \ge F_{1,R}(-\frac{c_3 c_1^2}{240L})$ and item 2 in Assumption 2 is satisfied by $c_2 = F_{1,R}(-\frac{c_3 c_1^2}{240L}) - \frac{1}{2} > 0$.

Combining the above, we can choose $\hat{\delta} = \min\{\delta_1, \delta_2\}$ so that both items in Assumption 2 hold for $\delta \le \hat{\delta}$. □

D Proof of Theorem 3

Recall the equivalent 2-dimensional problem:

$$\begin{aligned}
\hat{x}^t &= \mathrm{Dupl}\left[\operatorname*{argmin}_{x \in \frac{1}{n}\cdot\Delta^2} \left\{ \frac{n}{n^\alpha} \left\langle x, \sum_{\tau=1}^{t-1} \ell^\tau + \ell^{t-1} \right\rangle + \frac{n}{\eta} R_1(x[1]) + \frac{n}{\eta} R_2(x[2]) \right\}\right] \\
&= \mathrm{Dupl}\left[\frac{1}{n} \cdot \operatorname*{argmin}_{x \in \Delta^2} \left\{ \frac{n}{n^\alpha} \left\langle \frac{1}{n} x, \sum_{\tau=1}^{t-1} \ell^\tau + \ell^{t-1} \right\rangle + \frac{n}{\eta} R(x/n) \right\}\right] \\
&= \mathrm{Dupl}\left[\frac{1}{n} \cdot \operatorname*{argmin}_{x \in \Delta^2} \left\{ \left\langle x, \sum_{\tau=1}^{t-1} \ell^\tau + \ell^{t-1} \right\rangle + \frac{n^{\alpha+1}}{\eta} R(x/n) \right\}\right].
\end{aligned}$$

Euclidean regularizer: this regularizer is homogeneous of degree two. Choosing $\alpha = 1$, the inner minimization problem is exactly the same as the one solved by OFTRL in two dimensions.

Entropy regularizer: we set $\alpha = 0$ to get equivalence:

$$nR(x/n) = \sum_{i=1}^{2} x[i] \log(x[i]/n) = \sum_{i=1}^{2} x[i] \log x[i] - \sum_{i=1}^{2} x[i] \log n = \sum_{i=1}^{2} x[i] \log x[i] - \log n.$$

Now we have equivalence because the last term is a constant that does not affect the argmin.

Log regularizer: we set $\alpha = -1$ to get equivalence, using similar logic as for entropy:

$$R(x/n) = \sum_{i=1}^{2} -\log(x[i]/n) = 2\log n + \sum_{i=1}^{2} -\log x[i].$$

Tsallis entropy regularizer: we set $\alpha = -1 + \beta$ to get equivalence, using similar logic as for entropy:

$$n^\beta R(x/n) = n^\beta \cdot \frac{1 - \sum_{i=1}^{2} (\frac{x[i]}{n})^\beta}{1-\beta} = \frac{n^\beta - 1}{1-\beta} + \frac{1 - \sum_{i=1}^{2} x[i]^\beta}{1-\beta}.$$

E Numerical Experiments with Adaptive Stepsizes

In this section we present our numerical results when OFTRL and OOMD are instantiated with adaptive stepsize [Duchi et al., 2011]: $\eta_t = 1/\sqrt{\epsilon + \sum_{k=1}^{t-1} \|\ell_k\|_k^2}$ with some constant $\epsilon > 0$. We present our numerical experiments in Figure 4, where we choose $\epsilon = 0.1$.

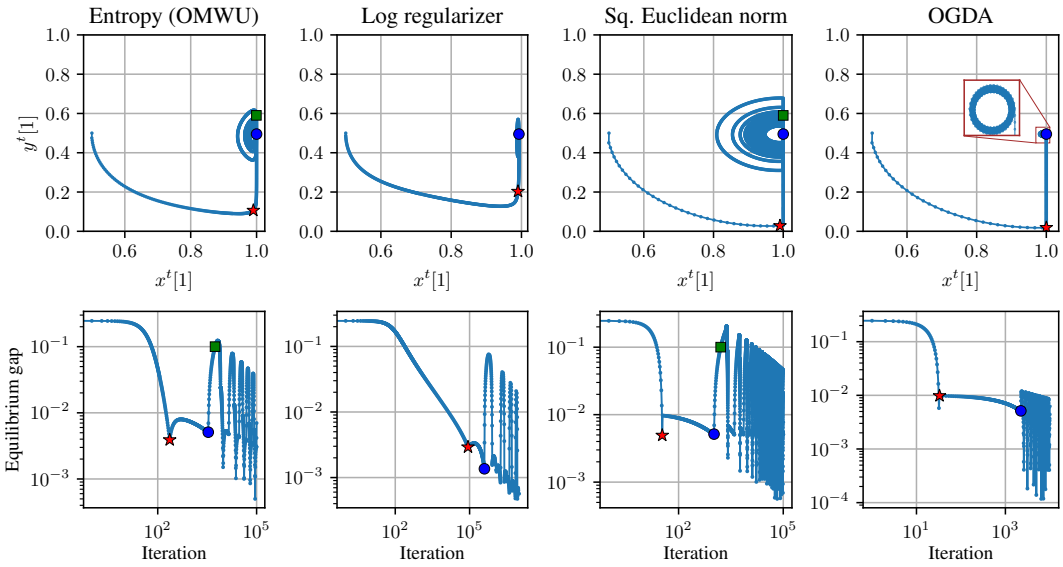

Figure 4: Comparison of the dynamics produced by three variants of OFTRL with different regularizers (negative entropy, logarithmic regularizer, and squared Euclidean norm) and OGDA in the same game A_δ defined in (2) for $\delta := 10^{-2}$ and adaptive step size with $\epsilon = 0.1$. The bottom row shows the duality gap achieved by the iterates.

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: The claims made in the abstract and introduction match both the theoretical and experimental results of 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: The extension to the general $m \times n$ game does not work for the Euclidean and log regularizers because the rescaling factors would be different for the row and column players.

   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.

https://doi.org/10.52202/079017-0737

- 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: See Section 2, Section 3 and 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: [Yes]

   Justification: See Figure 1 and Python codes in the supplementary materials.

   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: [Yes]

   Justification: See Python code in the supplementary materials.

   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?

https://doi.org/10.52202/079017-0737

Answer: [Yes]

Justification: See Python code in the supplementary materials.

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: There is no randomness in our 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: [Yes]

Justification: Our simulations were done within a few hours on an average consumer laptop.

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 authors have reviewed the NeurIPS Code of Ethics. The research conducted in this paper conforms with it in every respect.

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: This paper is mostly theoretical.

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: This paper poses no such risks.

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.

https://doi.org/10.52202/079017-0737

- 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: This 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: This 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: This 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: This 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.

https://doi.org/10.52202/079017-0737