Last-iterate Convergence in Extensive-Form Games

Anonymous Author(s)
Affiliation
Address
email

Abstract

Regret-based algorithms are highly efficient at finding approximate Nash equilibria in sequential games such as poker games. However, most regret-based algorithms, including counterfactual regret minimization (CFR) and its variants, rely on iterate averaging to achieve convergence. Inspired by recent advances on *last-iterate* convergence of optimistic algorithms in zero-sum normal-form games, we study this phenomenon in sequential games, and provide a comprehensive study of last-iterate convergence for zero-sum extensive-form games with perfect recall (EFGs), using various optimistic regret-minimization algorithms over treeplexes. This includes algorithms using the vanilla entropy or squared Euclidean norm regularizers, as well as their dilated versions which admit more efficient implementation. In contrast to CFR, we show that all of these algorithms enjoy last-iterate convergence, with some of them even converging *exponentially* fast. We also provide experiments to further support our theoretical results.

1 Introduction

Extensive-form games (EFGs) are an important class of games in game theory and artificial intelligence which can model imperfect information and sequential interactions. EFGs are typically solved by finding or approximating a *Nash equilibrium*. Regret-minimization algorithms are among the most popular approaches to approximate Nash equilibria. The motivation comes from a classical result which says that in two-player zero-sum games, when both players use *no-regret* algorithms, *the average strategy* converges to Nash equilibrium [18, 24, 49]. Counterfactual Regret Minimization (CFR) [49] and it variants such as CFR+ [42] are based on this motivation.

However, due to their ergodic convergence guarantee, theoretical convergence rates of regret-minimization algorithms are typically limited to $\mathcal{O}(1/\sqrt{T})$ or $\mathcal{O}(1/T)$ for T rounds, and this is also the case in practice [5, 11]. In contrast, it is known that linear convergence rates are achievable for certain other first-order algorithms [43, 22]. Additionally, the averaging procedure can create complications. It not only increases the computational and memory overhead [2], but also makes things difficult when incorporating neural networks in the solution process, where averaging is usually not possible. Indeed, to address this issue, Brown et al. [8] create a separate neural network to approximate the average strategy in their Deep CFR model.

Therefore, a natural idea is to design regret-minimization algorithms whose last strategy converges (we call this *last-iterate convergence*), ideally at a faster rate than the average iterate. Unfortunately, many regret minimization algorithms such as regret matching, regret matching+, and hedge, are known not to satisfy this property empirically and theoretically even for normal-form games. Although Bowling et al. [2] find that in Heads-Up Limit Hold'em poker the last strategy of CFR+ is better than the average strategy, and Farina et al. [16] observe in some experiments the last-iterate of optimistic OMD and FTRL converge fast, a theoretical understanding of this phenomenon is still absent for EFGs.

Submitted to 35th Conference on Neural Information Processing Systems (NeurIPS 2021). Do not distribute.

In this work, inspired by recent results on last-iterate convergence in normal-form games [46], we greatly extend the theoretical understanding of last-iterate convergence of regret-minimization algorithms in two-player zero-sum extensive-form games with perfect recall, and open up many interesting directions both in theory and practice. First, we show that any optimistic online mirror-descent algorithm instantiated with a strongly convex regularizer on the EFG strategy spaces provably enjoys last-iterate convergence, while CFR with either regret matching or regret matching+ fails to converge. Moreover, for some of the optimistic algorithms, we further show explicit convergence rates. In particular, we prove that optimistic mirror descent instantiated with the 1-strongly-convex dilated entropy regularizer [30], which we refer to as *Dilated Optimistic Multiplicative Weights Update* (DOMWU), has a *linear* convergence rate under the assumption that there is a unique Nash equilibrium; this assumption was also made by [13, 46] in order to achieve similar results for normal-form games.

2 Related Work

Extensive-form Games Here we focus on work related to two-player zero-sum perfect-recall games. Although there are many game-solving techniques such as abstraction [27, 19, 7], endgame solving [10, 20], and subgame solving [37, 6], these methods all rely on scalable methods for computing approximate Nash equilibria. There are several classes of algorithms for computing approximate Nash equilibria, such as double-oracle methods [34], fictitious play [3, 25], first-order methods [26, 30], and CFR methods [49, 32, 42]. Notably, variants of the CFR approach have achieved significant success in poker games [2, 38, 4]. Underlying the first-order and CFR approaches is the sequence-form representation [45], which allows the problem to be represented as a bilinear saddle-point problem. This leads to algorithms based on smoothing techniques and other first-order methods [22, 28, 21], and enables CFR via the theorem connecting no-regret guarantees to Nash equilibrium.

Online Convex Optimization and Optimistic Regret Minimization Online convex optimization [48] is a framework for repeated decision making where the goal is to minimize regret. When applied to repeated two-player zero-sum games, it is known that the average strategy converges to Nash Equilibria at the rate of $\mathcal{O}(1/\sqrt{T})$ when both players apply regret-minimization algorithms whose regret grows on the order of $\mathcal{O}(\sqrt{T})$ [18, 24, 49]. Moreover, when the players use optimistic regret-minimization algorithms, the convergence rate is improved to $\mathcal{O}(1/T)$ [39, 41]. Recent works have applied optimism ideas to EFGs, such as optimistic algorithms with dilated regularizers [30, 16], CFR-like local optimistic algorithms [15], and optimistic CFR algorithms [9, 5, 17]. However, the theoretical results in all these existing papers consider the average strategy, while we are the first to consider last-iterate convergence in EFGs.

Last-iterate Convergence in Saddle-point Optimization As mentioned previously, two-player zero-sum games can be formulated as saddle-point optimization problems. Saddle-point problems have recently gained a lot of attention due to their applications in machine learning, for example in generative adversarial networks [23]. Basic algorithms, including gradient descent ascent and multiplicative weights update, diverge even in simple instances [35, 1]. In contrast, their optimistic versions, optimistic gradient descent ascent (OGDA) [14, 36, 46] and optimistic multiplicative weights update (OMWU) [13, 33, 46] have been shown to enjoy attractive last-iterate convergence guarantees. However, almost none of these results apply to the case of EFGs: Wei et al. [46] show a result that implies linear convergence of vanilla OGDA in EFGs (see Corollary 5), but no results are known for vanilla OMWU or more importantly for algorithms instantiated with *dilated* regularizers which lead to fast iterate updates in EFGs. In this work we extend the existing results on normal-form games to EFGs, including the practically-important dilated regularizers.

3 Problem Setup

We start with some basic notation. For a vector $\boldsymbol{z}$, we use z_i to denote its i-th coordinate and $\|\boldsymbol{z}\|_p$ to denote its p-norm (with $\|\boldsymbol{z}\|$ being a shorthand for $\|\boldsymbol{z}\|_2$). For a convex function ψ, the associated Bregman divergence is define as $D_\psi(\boldsymbol{u}, \boldsymbol{v}) = \psi(\boldsymbol{u}) - \psi(\boldsymbol{v}) - \langle \nabla\psi(\boldsymbol{v}), \boldsymbol{u} - \boldsymbol{v}\rangle$, and ψ is called κ-strongly convex with respect to the p-norm if $D_\psi(\boldsymbol{u}, \boldsymbol{v}) \geq \frac{\kappa}{2}\|\boldsymbol{u} - \boldsymbol{v}\|_p^2$ holds for all $\boldsymbol{u}$ and $\boldsymbol{v}$ in

the domain. The Kullback-Leibler divergence, which is the Bregman divergence with respect to the entropy function, is denoted by $\mathrm{KL}(\cdot,\cdot)$. Finally, we use Δ_P to denote the $(P-1)$-dimensional simplex and $[N]$ to denote the set $\{1, 2 \ldots, N\}$ for some positive integer N.

Extensive-form Games as Bilinear Saddle-point Optimization We consider the problem of finding a Nash equilibrium of a two-player zero-sum extensive-form game (EFG) with perfect recall. Instead of formally introducing the definition of an EFG, for the purpose of this work, it suffices to consider an equivalent formulation, which casts the problem as a simple bilinear saddle-point optimization [45]:

$$\min_{\boldsymbol{x}\in\mathcal{X}} \max_{y\in\mathcal{Y}} \boldsymbol{x}^\top \boldsymbol{G}\boldsymbol{y} = \max_{y\in\mathcal{Y}} \min_{\boldsymbol{x}\in\mathcal{X}} \boldsymbol{x}^\top \boldsymbol{G}\boldsymbol{y}, \tag{1}$$

where $\boldsymbol{G} \in [-1,+1]^{M\times N}$ is a known matrix, and $\mathcal{X} \subset \mathbb{R}^M$ and $\mathcal{Y} \subset \mathbb{R}^N$ are two polytopes called *treeplexes* (to be defined soon). The set of Nash equilibria is then defined as $\mathcal{Z}^* = \mathcal{X}^* \times \mathcal{Y}^*$, where $\mathcal{X}^* = \operatorname{argmin}_{\boldsymbol{x}\in\mathcal{X}} \max_{y\in\mathcal{Y}} \boldsymbol{x}^\top \boldsymbol{G}\boldsymbol{y}$ and $\mathcal{Y}^* = \operatorname{argmax}_{y\in\mathcal{Y}} \min_{\boldsymbol{x}\in\mathcal{X}} \boldsymbol{x}^\top \boldsymbol{G}\boldsymbol{y}$. Our goal is to find a point $z \in \mathcal{Z} = \mathcal{X} \times \mathcal{Y}$ that is close to the set of Nash equilibria $\mathcal{Z}^*$, and we use the Bregman divergence (of some function ψ) between $\boldsymbol{z}$ and the closest point in $\mathcal{Z}^*$ to measure the closeness, that is, $\min_{\boldsymbol{z}^*\in\mathcal{Z}^*} D_\psi(\boldsymbol{z}^*, \boldsymbol{z})$.

For notational convenience, we let $P = M + N$ and $F(\boldsymbol{z}) = (\boldsymbol{G}\boldsymbol{y}, -\boldsymbol{G}^\top \boldsymbol{x})$ for any $\boldsymbol{z} = (\boldsymbol{x}, \boldsymbol{y}) \in \mathcal{Z} \subset \mathbb{R}^P$. Without loss of generality, we assume $\|F(\boldsymbol{z})\|_\infty \le 1$ for all $\boldsymbol{z} \in \mathcal{Z}$ (which can always be ensured by normalizing the entries of $\boldsymbol{G}$ accordingly).

Treeplexes The structure of the EFG is implicitly captured by the treeplexes $\mathcal{X}$ and $\mathcal{Y}$, which are generalizations of simplexes that capture the sequential structure of an EFG. The formal definition is as follows.

Definition 1 (Hoda et al. [26])**.** *A treeplex is recursively constructed via the following three operations:*

1. *Every probability simplex is a treeplex.*
2. *Given treeplexes $\mathcal{Z}_1, \ldots, \mathcal{Z}_K$, the Cartesian product $\mathcal{Z}_1 \times \cdots \times \mathcal{Z}_K$ is a treeplex.*
3. *(Branching) Given treeplexes $\mathcal{Z}_1 \subset \mathbb{R}^M$ and $\mathcal{Z}_2 \subset \mathbb{R}^N$, and any $i \in [M]$,*
$$\mathcal{Z}_1 \boxed{i} \mathcal{Z}_2 = \left\{(\boldsymbol{u}, u_i \cdot \boldsymbol{v}) \in \mathbb{R}^{M+N} : \boldsymbol{u} \in \mathcal{Z}_1,\ \boldsymbol{v} \in \mathcal{Z}_2\right\}$$
is a treeplex.

By definition, a treeplex is a tree-like structure built with simplexes, which intuitively represents the tree-like decision space of a single player, and an element in the treeplex represents a strategy for the player. Let $\mathcal{H}^{\mathcal{Z}}$ denote the collection of all the simplexes in treeplex $\mathcal{Z}$, which following Definition 1 can be recursively defined as: $\mathcal{H}^{\mathcal{Z}} = \{\mathcal{Z}\}$ if $\mathcal{Z}$ is a simplex; $\mathcal{H}^{\mathcal{Z}} = \bigcup_{k=1}^K \mathcal{H}^{\mathcal{Z}_i}$ if $\mathcal{Z}$ is a Cartesian product $\mathcal{Z}_1 \times \cdots \times \mathcal{Z}_K$; and $\mathcal{H}^{\mathcal{Z}} = \mathcal{H}^{\mathcal{Z}_1} \cup \mathcal{H}^{\mathcal{Z}_2}$ if $\mathcal{Z} = \mathcal{Z}_1 \boxed{i} \mathcal{Z}_2$. In EFG terminology, $\mathcal{H}^{\mathcal{X}}$ and $\mathcal{H}^{\mathcal{Y}}$ are the collections of *information sets* for player $\boldsymbol{x}$ and player $\boldsymbol{y}$ respectively, which are the decision points for the players, at which they select an action within the simplex. For any $h \in \mathcal{H}^{\mathcal{Z}}$, we let Ω_h denote the set of indices belonging to h, and for any $\boldsymbol{z} \in \mathcal{Z}$, we let $\boldsymbol{z}_h$ be the slice of $\boldsymbol{z}$ whose indices are in Ω_h. For each index i, we also let $h(i)$ be the simplex i falls into, that is, $i \in \Omega_{h(i)}$.

In Definition 1, the last branching operation naturally introduces the concept of a *parent variable* for each $h \in \mathcal{H}^{\mathcal{Z}}$, which can again be recursively defined as: if $\mathcal{Z}$ is a simplex, then it has no parent; if $\mathcal{Z}$ is a Cartesian product $\mathcal{Z}_1 \times \cdots \times \mathcal{Z}_K$, then the parent of $h \in \mathcal{H}^{\mathcal{Z}}$ is the same as the parent of h in the treeplex $\mathcal{Z}_k$ that h belongs to (that is, $h \in \mathcal{H}^{\mathcal{Z}_k}$); finally, if $\mathcal{Z} = \{(\boldsymbol{u}, u_i \cdot \boldsymbol{v}) : \boldsymbol{u} \in \mathcal{Z}_1,\ \boldsymbol{v} \in \mathcal{Z}_2\}$, then for all $h \in \mathcal{H}^{\mathcal{Z}_2}$ without a parent, their parent in $\mathcal{Z}$ is u_i, and for all other h, their parents remain the same as in $\mathcal{Z}_1$ or $\mathcal{Z}_2$. We denote by $\sigma(h)$ the index of the parent variable of h, and let it be 0 if h has no parent. For convenience, we let $z_0 = 1$ for all $z \in \mathcal{Z}$ (so that $z_{\sigma(h)}$ is always well-defined). Also define $\mathcal{H}_i = \{h \in \mathcal{H}^{\mathcal{Z}} : \sigma(h) = i\}$ to be the collection of simplexes whose parent index is i.

Similarly, for an index i, its parent index is defined as $p_i = \sigma(h(i))$, and i is called a *terminal index* if it is not a parent index (that is, $i \neq p_j$ for all j). Finally, for an element $\boldsymbol{z} \in \mathcal{Z}$ and an index i, we define $q_i = z_i / z_{p_i}$. In EFG terminology, q_i specifies the probability of selecting action i in the information set $h(i)$ according to strategy $\boldsymbol{z}$.

4 Optimistic Regret-minimization Algorithms

There are many different algorithms for solving bilinear saddle-point problems over general constrained sets. We focus specifically on a family of regret-minimization algorithms, called Optimistic Online Mirror Descent (OOMD) [39], which are known to be highly efficient. In contrast to the CFR algorithm and its variants, which minimize a local regret notion at each information set (which upper bounds global regret), the algorithms we consider explicitly minimize global regret. As our main results in the next section show, these global regret-minimization algorithms enjoy last-iterate convergence, while CFR and its variants provably diverge.

Specifically, given a step size $\eta > 0$ and a convex function ψ (called a *regularizer*), OOMD sequentially performs the following update for $t = 1, 2, \ldots,$

$$\boldsymbol{x}_t = \operatorname*{argmin}_{\boldsymbol{x}\in\mathcal{X}} \left\{\eta\langle \boldsymbol{x}, \boldsymbol{G}\boldsymbol{y}_{t-1}\rangle + D_\psi(\boldsymbol{x}, \widehat{\boldsymbol{x}}_t)\right\}, \qquad \widehat{\boldsymbol{x}}_{t+1} = \operatorname*{argmin}_{\boldsymbol{x}\in\mathcal{X}} \left\{\eta\langle \boldsymbol{x}, \boldsymbol{G}\boldsymbol{y}_{t}\rangle + D_\psi(\boldsymbol{x}, \widehat{\boldsymbol{x}}_t)\right\},$$

$$\boldsymbol{y}_t = \operatorname*{argmin}_{\boldsymbol{y}\in\mathcal{Y}} \left\{\eta\langle \boldsymbol{y}, -\boldsymbol{G}^\top\boldsymbol{x}_{t-1}\rangle + D_\psi(\boldsymbol{y}, \widehat{\boldsymbol{y}}_t)\right\}, \quad \widehat{\boldsymbol{y}}_{t+1} = \operatorname*{argmin}_{\boldsymbol{y}\in\mathcal{Y}} \left\{\eta\langle \boldsymbol{y}, -\boldsymbol{G}^\top\boldsymbol{x}_{t}\rangle + D_\psi(\boldsymbol{y}, \widehat{\boldsymbol{y}}_t)\right\},$$

with $(\widehat{\boldsymbol{x}}_1, \widehat{\boldsymbol{y}}_1) = (\boldsymbol{x}_0, \boldsymbol{y}_0) \in \mathcal{Z}$ being arbitrary. Using shorthands $\boldsymbol{z}_t = (\boldsymbol{x}_t, \boldsymbol{y}_t)$, $\widehat{\boldsymbol{z}}_t = (\widehat{\boldsymbol{x}}_t, \widehat{\boldsymbol{y}}_t)$, $\psi(\boldsymbol{z}) = \psi(\boldsymbol{x}) + \psi(\boldsymbol{y})$ and recalling the notation $F(\boldsymbol{z}) = (\boldsymbol{G}\boldsymbol{y}, -\boldsymbol{G}^\top\boldsymbol{x})$, the updates above can be compactly written as OOMD with regularizer ψ over treeplex $\mathcal{Z}$:

$$\boldsymbol{z}_t = \operatorname*{argmin}_{\boldsymbol{z}\in\mathcal{Z}} \left\{\eta\langle \boldsymbol{z}, F(\boldsymbol{z}_{t-1})\rangle + D_\psi(\boldsymbol{z}, \widehat{\boldsymbol{z}}_t)\right\}, \ \widehat{\boldsymbol{z}}_{t+1} = \operatorname*{argmin}_{\boldsymbol{z}\in\mathcal{Z}} \left\{\eta\langle \boldsymbol{z}, F(\boldsymbol{z}_{t})\rangle + D_\psi(\boldsymbol{z}, \widehat{\boldsymbol{z}}_t)\right\}. \quad (2)$$

Below, we discuss four different regularizers and their resulting algorithms (throughout, we use notations Φ for regularizers based on Euclidean norm and Ψ for regularizers based on entropy).

Vanilla Optimistic Gradient Descent Ascent (VOGDA) Define the vanilla squared Euclidean norm regularizer as $\Phi^{\text{van}}(\boldsymbol{z}) = \frac{1}{2}\sum_i z_i^2$. We call OOMD instantiated with $\psi = \Phi^{\text{van}}$ Vanilla Optimistic Gradient Descent Ascent (VOGDA). In this case, the Bregman divergence is $D_{\Phi^{\text{van}}}(\boldsymbol{z}, \boldsymbol{z}') = \frac{1}{2}\|\boldsymbol{z} - \boldsymbol{z}'\|^2$ (by definition Φ^{van} is thus 1-strongly convex with respect to the 2-norm), and the updates simply become projected gradient descent. For VOGDA there is no closed-form for Eq. (2), since projection onto the treeplex $\mathcal{Z}$ is required. Nevertheless, the solution can still be computed in $\mathcal{O}(P^2 \log P)$ time (recall that P is the dimension of $\mathcal{Z}$) [22].

Vanilla Optimistic Multiplicative Weight Update (VOMWU) Define the vanilla entropy regularizer as $\Psi^{\text{van}}(\boldsymbol{z}) = \sum_i z_i \ln z_i$. We call OOMD with $\psi = \Psi^{\text{van}}$ Vanilla Optimistic Multiplicative Weights Update (VOMWU). The Bregman divergence in this case is the generalized KL divergence: $D_{\Psi^{\text{van}}}(\boldsymbol{z}, \boldsymbol{z}') = \sum_i z_i \ln(z_i/z_i') - z_i + z_i'$. Although it is well-known that Ψ^{van} is 1-strongly convex with respect to the 1-norm for the special case when $\mathcal{Z}$ is a simplex, this is not true generally on a treeplex. Nevertheless, it can still be shown that Ψ^{van} is 1-strongly convex with respect to the 2-norm; see Appendix B.

The name "Multiplicative Weights Update" is inherited from case when $\mathcal{X}$ and $\mathcal{Y}$ are simplexes, in which case the updates in Eq. (2) have a simple multiplicative form. We emphasize, however, that in general VOMWU does not admit a closed-form update. Instead, to solve Eq. (2), one can equivalently solve a simpler dual optimization problem; see [47, Proposition 1].

The two regularizers mentioned above ignore the structure of the treeplex. *Dilated Regularizers* [26], on the other hand, take the structure into account and allow one to decompose the update into simpler updates at each information set. Specifically, given any convex function ψ defined over the simplex and a weight parameter $\boldsymbol{\alpha} \in \mathbb{R}_+^{\mathcal{H}^{\mathcal{Z}}}$, the dilated version of ψ defined over $\mathcal{Z}$ is:

$$\psi_{\boldsymbol{\alpha}}^{\text{dil}}(\boldsymbol{z}) = \sum_{h\in\mathcal{H}^{\mathcal{Z}}} \alpha_h \cdot z_{\sigma(h)} \cdot \psi\left(\frac{\boldsymbol{z}_h}{z_{\sigma(h)}}\right). \quad (3)$$

This is well-defined since $\boldsymbol{z}_h/z_{\sigma(h)}$ is indeed a distribution within the simplex h (with q_i for $i \in \Omega_h$ being its entries). It can also be shown that $\psi_{\boldsymbol{\alpha}}^{\text{dil}}$ is always convex in $\boldsymbol{z}$ [26]. Intuitively, $\psi_{\boldsymbol{\alpha}}^{\text{dil}}$ applies the base regularizer ψ to the action distribution in each information set and then scales the value by its parent variable and the weight α_h. By picking different base regularizers, we obtain the following two algorithms.

Dilated Optimistic Gradient Descent Ascent (DOGDA) Define the dilated squared Euclidean norm regularizer $\Phi^{\text{dil}}_{\boldsymbol{\alpha}}$ as Eq. (3) with ψ being the vanilla squared Euclidean norm $\psi(\boldsymbol{z}) = \frac{1}{2}\sum_i z_i^2$. Direct calculation shows $\Phi^{\text{dil}}_{\boldsymbol{\alpha}}(\boldsymbol{z}) = \frac{1}{2}\sum_i \alpha_{h(i)} z_i q_i$. We call OOMD with regularizer $\Phi^{\text{dil}}_{\boldsymbol{\alpha}}$ the Dilated Optimistic Gradient Descent Ascent algorithm (DOGDA). It is known that there exists an $\boldsymbol{\alpha}$ such that $\Phi^{\text{dil}}_{\boldsymbol{\alpha}}$ is 1-strongly convex with respect to the 2-norm [16]. Importantly, DOGDA decomposes the update Eq. (2) into simpler gradient descent-style updates at each information set, as shown below.

Lemma 1 (Hoda et al. [26]). *If $\boldsymbol{z}' = \operatorname{argmin}_{\boldsymbol{z}\in\mathcal{Z}} \left\{\eta\langle \boldsymbol{z}, \boldsymbol{f}\rangle + D_{\Phi^{dil}_{\boldsymbol{\alpha}}}(\boldsymbol{z}, \widehat{\boldsymbol{z}})\right\}$, then for every $h \in \mathcal{H}^{\mathcal{Z}}$, the corresponding vector $\boldsymbol{q}'_h = \frac{\boldsymbol{z}'_h}{z'_{\sigma(h)}}$ can be computed by:*

$$\boldsymbol{q}'_h = \operatorname*{argmin}_{\boldsymbol{q}_h \in \Delta_{|\Omega_h|}} \left\{\eta\langle \boldsymbol{q}_h, \boldsymbol{L}_h\rangle + \frac{\alpha_h}{2}\|\boldsymbol{q}_h - \widehat{\boldsymbol{q}}_h\|^2\right\}, \tag{4}$$

where $\widehat{\boldsymbol{q}}_h = \frac{\widehat{\boldsymbol{z}}_h}{\widehat{\boldsymbol{z}}_{\sigma(h)}}$, $\boldsymbol{L}_h$ is the slice of $\boldsymbol{L}$ whose entries are in Ω_h, and $\boldsymbol{L}$ is defined through:

$$L_i = f_i + \sum_{h\in\mathcal{H}_i}\left(\langle \boldsymbol{q}'_h, \boldsymbol{L}_h\rangle + \frac{\alpha_h}{2\eta}\|\boldsymbol{q}'_h - \widehat{\boldsymbol{q}}_h\|^2\right).$$

While the definitions of $\boldsymbol{q}'_h$ and $\boldsymbol{L}$ are seemingly recursive, one can verify that they can in fact be computed in a "bottom-up" manner, starting with the terminal indices. Although Eq. (4) still does not admit closed-form solution, it only requires projection onto a simplex, which can be solved efficiently, see e.g. [12]. Finally, with $\boldsymbol{q}'_h$ computed for all h, $\boldsymbol{z}'$ can be calculated in a "top-down" manner by definition.

Dilated Optimistic Multiplicative Weight Update (DOMWU) Finally, define the dilated entropy regularizer $\Psi^{\text{dil}}_{\boldsymbol{\alpha}}$ as Eq. (3) with ψ being the vanilla entropy $\psi(\boldsymbol{z}) = \sum_i z_i \ln z_i$. Direct calculation shows $\Psi^{\text{dil}}_{\boldsymbol{\alpha}}(\boldsymbol{z}) = \sum_i \alpha_{h(i)} z_i \ln q_i$. We call OOMD with regularizer $\Psi^{\text{dil}}_{\boldsymbol{\alpha}}$ the Dilated Optimistic Multiplicative Weights Update algorithm (DOMWU). Similar to DOGDA, there exists an $\boldsymbol{\alpha}$ such that $\Psi^{\text{dil}}_{\boldsymbol{\alpha}}$ is 1-strongly convex with respect to the 2-norm [30].[1] Moreover, in contrast to all the three algorithms mentioned above, the update of DOMWU has a closed-form solution:

Lemma 2 (Hoda et al. [26]). *Suppose $\boldsymbol{z}' = \operatorname{argmin}_{\boldsymbol{z}\in\mathcal{Z}} \left\{\eta\langle \boldsymbol{z}, \boldsymbol{f}\rangle + D_{\Psi^{dil}_{\boldsymbol{\alpha}}}(\boldsymbol{z}, \widehat{\boldsymbol{z}})\right\}$. Similarly to the notation q_i, define $q'_i = z'_i/z'_{p_i}$ and $\widehat{q}_i = \widehat{z}_i/\widehat{z}_{p_i}$. Then we have*

$$q'_i \propto \widehat{q}_i \exp\left(-\eta L_i/\alpha_{h(i)}\right), \text{ where } L_i = f_i - \sum_{h\in\mathcal{H}_i}\frac{\alpha_h}{\eta}\ln\left(\sum_{j\in\Omega_h}\widehat{q}_j \exp\left(-\eta L_j/\alpha_h\right)\right).$$

This lemma again implies that we can compute q'_i bottom-up, and then $\boldsymbol{z}'$ can be computed top-down. This is similar to DOGDA, except that all updates have a closed-form.

5 Last-iterate Convergence Results

In this section, we present our main last-iterate convergence results for the global regret-minimization algorithms discussed in Section 4. Before doing so, we point out again that the sequence produced by the well-known CFR algorithm or its variants, including CFR+ [42] and their optimistic versions [17], may diverge (even if the average converges to a Nash equilibrium). Indeed, this can happen even for a simple normal-form game, as formally shown below.

Theorem 3. *In the rock-paper-scissors game, CFR, CFR+ (with simultaneous updates), and their optimistic versions (with some particular initialization) all produce a diverging sequence.*

On the contrary, every algorithm from the OOMD family given by Eq. (2) ensures last-iterate convergence, as long as the regularizer is strongly convex.

Theorem 4. *Consider the update rules in Eq. (2). Suppose that ψ is 1-strongly convex with respect to the 2-norm and $\eta \le \frac{1}{8P}$. Then $\boldsymbol{z}_t$ converges to a Nash equilibrium as $t \to \infty$.*

[1] Kroer et al. [30] also show a better strong convexity result with respect to the 1-norm. We focus on the 2-norm here for consistency with our other results, but our analysis can be applied to the 1-norm case as well.

As mentioned, Φ^{van} and Ψ^{van} are both 1-strongly convex with respect to 2-norm, so are $\Phi^{\text{dil}}_{\boldsymbol{\alpha}}$ and $\Psi^{\text{dil}}_{\boldsymbol{\alpha}}$ under some specific choice of $\boldsymbol{\alpha}$ (in the rest of the paper, we fix this choice of $\boldsymbol{\alpha}$). Therefore, all the four algorithms we consider enjoy last-iterate convergence. However, Theorem 4 only provides an asymptotic convergence result, whereas obtaining concrete convergence rates for these algorithms may be a more challenging task. Below, we establish such convergence rates for three of them.

First of all, note that [46, Theorem 5, Theorem 8] already provide a general last-iterate convergence rate for VOGDA over polytopes. Since treeplexes are polytopes, we can directly apply their results and obtain the following corollary.

Corollary 5. *Define* $\text{dist}^2(\boldsymbol{z}, \mathcal{Z}^*) = \min_{\boldsymbol{z}^* \in \mathcal{Z}^*} \|\boldsymbol{z} - \boldsymbol{z}^*\|^2$. *For* $\eta \le \frac{1}{8P}$, VOGDA *guarantees*

$$\text{dist}^2(\boldsymbol{z}_t, \mathcal{Z}^*) \le 64\text{dist}^2(\widehat{\boldsymbol{z}}_1, \mathcal{Z}^*)(1 + C_1)^{-t},$$

where $C_1 > 0$ *is some constant that depends on the game and* η.

However, the results for VOMWU in [46, Theorem 3] is very specific to normal-form game (that is, when $\mathcal{X}$ and $\mathcal{Y}$ are simplexes) and thus cannot be applied here. Nevertheless, we are able to extend their analysis to get the following result.

Theorem 6. *If the EFG has a unique Nash equilibrium* $\boldsymbol{z}^*$, *then* VOMWU *with step size* $\eta \le \frac{1}{8P}$ *guarantees* $\frac{1}{2}\|\widehat{\boldsymbol{z}}_t - \boldsymbol{z}^*\|^2 \le D_{\Psi^{\text{van}}}(\boldsymbol{z}^*, \widehat{\boldsymbol{z}}_t) \le \frac{C_2}{t}$, *where* $C_2 > 0$ *is some constant depending on the game,* $\widehat{\boldsymbol{z}}_1$, *and* η.

We note that the uniqueness assumption is often required in the analysis of OMWU even for normal-form games [13, 46] (although [46, Appendix A.5] provides empirical evidence to show that this may be an artifact of the analysis). Also note that for normal-form games, [46, Theorem 3] show a linear convergence rate, whereas here we only show a slower sub-linear rate, due to additional complications introduced by treeplexes (see more discussions in the next section). Whether this can be improved is left as a future direction.

On the other hand, thanks to the closed-form updates of DOMWU, we are able to show the following linear convergence rate for this algorithm.

Theorem 7. *If the EFG has a unique Nash equilibrium* $\boldsymbol{z}^*$, *then* DOMWU *with step size* $\eta \le \frac{1}{8P}$ *guarantees* $\frac{1}{2}\|\boldsymbol{z}_t - \boldsymbol{z}^*\|^2 \le D_{\Psi^{\text{dil}}_{\boldsymbol{\alpha}}}(\boldsymbol{z}^*, \boldsymbol{z}_t) \le C_3(1 + C_4)^{-t}$, *where* $C_3, C_4 > 0$ *are constants that depend on the game,* $\widehat{\boldsymbol{z}}_1$, *and* η.

To the best of our knowledge, this is the first last-iterate convergence result for algorithms with dilated regularizers. Unfortunately, due to technical difficulties, we were unable to prove similar results for DOGDA (see Appendix D for more discussion). We leave that as an important future direction.

6 Analysis Overview

In this section, we provide an overview of our analysis. It starts from the following standard one-step regret analysis of OOMD (see, for example, [46, Lemma 1]):

Lemma 8. *Consider the update rules in Eq. (2). Suppose that* ψ *is 1-strongly convex with respect to the 2-norm,* $\|F(\boldsymbol{z}_1) - F(\boldsymbol{z}_2)\| \le L\|\boldsymbol{z}_1 - \boldsymbol{z}_2\|$ *for all* $\boldsymbol{z}_1, \boldsymbol{z}_2 \in \mathcal{Z}$ *and some* $L > 0$, *and* $\eta \le \frac{1}{8L}$. *Then for any* $\boldsymbol{z} \in \mathcal{Z}$ *and any* $t \ge 1$, *we have*

$$\eta F(\boldsymbol{z}_t)^\top(\boldsymbol{z}_t - \boldsymbol{z}) \le D_\psi(\boldsymbol{z}, \widehat{\boldsymbol{z}}_t) - D_\psi(\boldsymbol{z}, \widehat{\boldsymbol{z}}_{t+1}) - D_\psi(\widehat{\boldsymbol{z}}_{t+1}, \boldsymbol{z}_t) - \tfrac{15}{16}D_\psi(\boldsymbol{z}_t, \widehat{\boldsymbol{z}}_t) + \tfrac{1}{16}D_\psi(\widehat{\boldsymbol{z}}_t, \boldsymbol{z}_{t-1}).$$

Note that the Lipschitz condition on F holds in our case with $L = P$ since

$$\|F(\boldsymbol{z}_1) - F(\boldsymbol{z}_2)\| = \sqrt{\|\boldsymbol{G}(\boldsymbol{y}_1 - \boldsymbol{y}_2)\|^2 + \|\boldsymbol{G}^\top(\boldsymbol{x}_1 - \boldsymbol{x}_2)\|^2} \le \sqrt{P\|\boldsymbol{z}_1 - \boldsymbol{z}_2\|_1^2} \le P\|\boldsymbol{z}_1 - \boldsymbol{z}_2\|,$$

which is also why the step size is chosen to be $\eta \le \frac{1}{8P}$ in all our results. In the following, we first prove Theorem 4. Then, we review the convergence analysis of [46] for OMWU in normal-form games, and finally demonstrate how to prove Theorem 6 and Theorem 7 by building upon this previous work and addressing the additional complications from EFGs.

6.1 Proof of Theorem 4

For any $\boldsymbol{z}^* \in \mathcal{Z}^*$, by optimality of $\boldsymbol{z}^*$ we have:

$$F(\boldsymbol{z}_t)^\top(\boldsymbol{z}_t - \boldsymbol{z}^*) = \boldsymbol{x}_t^\top \boldsymbol{G}\boldsymbol{y}_t - \boldsymbol{x}_t^\top \boldsymbol{G}\boldsymbol{y}_t + \boldsymbol{x}_t^\top \boldsymbol{G}\boldsymbol{y}^* - \boldsymbol{x}^{*\top} \boldsymbol{G}\boldsymbol{y}_t \geq \boldsymbol{x}^{*\top} \boldsymbol{G}\boldsymbol{y}^* - \boldsymbol{x}^{*\top} \boldsymbol{G}\boldsymbol{y}^* = 0.$$

Thus, taking $\boldsymbol{z} = \boldsymbol{z}^*$ in Lemma 8 and rearranging, we arrive at

$$D_\psi(\boldsymbol{z}^*, \widehat{\boldsymbol{z}}_{t+1}) \leq D_\psi(\boldsymbol{z}^*, \widehat{\boldsymbol{z}}_t) - D_\psi(\widehat{\boldsymbol{z}}_{t+1}, \boldsymbol{z}_t) - \tfrac{15}{16} D_\psi(\boldsymbol{z}_t, \widehat{\boldsymbol{z}}_t) + \tfrac{1}{16} D_\psi(\widehat{\boldsymbol{z}}_t, \boldsymbol{z}_{t-1}).$$

Defining $\Theta_t = D_\psi(\boldsymbol{z}^*, \widehat{\boldsymbol{z}}_t) + \frac{1}{16} D_\psi(\widehat{\boldsymbol{z}}_t, \boldsymbol{z}_{t-1})$ and $\zeta_t = D_\psi(\widehat{\boldsymbol{z}}_{t+1}, \boldsymbol{z}_t) + D_\psi(\boldsymbol{z}_t, \widehat{\boldsymbol{z}}_t)$, we rewrite the inequality above as

$$\Theta_{t+1} \leq \Theta_t - \tfrac{15}{16}\zeta_t. \tag{5}$$

We remark that similar inequalities appear in [46, Eq. (3) and Eq. (4)], but here we pick $\boldsymbol{z}^* \in \mathcal{Z}^*$ arbitrarily while they have to pick a particular $\boldsymbol{z}^* \in \mathcal{Z}^*$ (such as the projection of $\widehat{\boldsymbol{z}}_t$ onto $\mathcal{Z}^*$). Summing Eq. (5) over t, telescoping, and applying the strong convexity of ψ, we have

$$\Theta_1 \geq \Theta_1 - \Theta_T \geq \frac{15}{16}\sum_{t=1}^{T-1} \zeta_t \geq \frac{15}{32}\sum_{t=1}^{T-1} \|\widehat{\boldsymbol{z}}_{t+1} - \boldsymbol{z}_t\|^2 + \|\boldsymbol{z}_t - \widehat{\boldsymbol{z}}_t\|^2 \geq \frac{15}{64}\sum_{t=2}^{T-1} \|\boldsymbol{z}_t - \boldsymbol{z}_{t-1}\|^2.$$

Since the rightmost term is a summation of nonnegative terms and is upper bounded by a constant, we conclude that $\|\boldsymbol{z}_{t+1} - \boldsymbol{z}_t\| \to 0$ as $T \to \infty$. Therefore, $\boldsymbol{z}_t$ converges as $T \to \infty$. On the other hand, since OOMD is a regret-minimization algorithm, it is well known that the average iterate converges to a Nash equilibrium [18]. Consequently, combining the two facts above implies that $\boldsymbol{z}_t$ has to converge to a Nash equilibrium, which proves Theorem 4.

We remark that Lemma 8 holds for general closed convex domains as shown in [46]. Consequently, with the same argument, Theorem 4 holds more generally as long as $\mathcal{X}$ and $\mathcal{Y}$ are closed convex sets. While the argument is straightforward, we are not aware of similar results in prior works. Also note that unlike Theorem 6 and Theorem 7, Theorem 4 holds without the uniqueness assumption for VOMWU and DOMWU.

6.2 Review for normal-form games

To better explain our analysis and highlight its novelty, we first review the two-stage analysis of [46] for OMWU in normal-form games, a special case of our setting when $\mathcal{X}$ and $\mathcal{Y}$ are simplexes. Note that both VOMWU and DOMWU reduce to OMWU in this case. As with Theorem 6 and Theorem 7, the normal-form OMWU results assume a unique Nash equilibrium $\boldsymbol{z}^*$. With this uniqueness assumption and [35, Lemma C.4], Wei et al. [46] show the following inequality

$$\zeta_t = D_\psi(\widehat{\boldsymbol{z}}_{t+1}, \boldsymbol{z}_t) + D_\psi(\boldsymbol{z}_t, \widehat{\boldsymbol{z}}_t) \geq C_5 \|\boldsymbol{z}^* - \widehat{\boldsymbol{z}}_{t+1}\|^2 \tag{6}$$

for some problem-dependent constant $C_5 > 0$, which, when combined with Eq. (5), implies that if the algorithm's current iterate is far from $\boldsymbol{z}^*$, then the decrease in Θ_t is more substantial, that is, the algorithm makes more progress on approaching $\boldsymbol{z}^*$. To establish a recursion, however, we need to connect the 2-norm back to the Bregman divergence (a reverse direction of strong convexity). To do so, Wei et al. [46] argue that $\widehat{\boldsymbol{z}}_{t+1,i}$ can be lower bounded by another problem-dependent constant for $i \in \mathrm{supp}(\boldsymbol{z}^*)$ [46, Lemma 19], where $\mathrm{supp}(\boldsymbol{z}^*)$ denotes the support of $\boldsymbol{z}^*$. This further allows them to lower bound $\|\boldsymbol{z}^* - \widehat{\boldsymbol{z}}_{t+1}\|$ in terms of $D_\psi(\boldsymbol{z}^*, \widehat{\boldsymbol{z}}_{t+1})$ (which is just $\mathrm{KL}(\boldsymbol{z}^*, \widehat{\boldsymbol{z}}_{t+1})$), leading to

$$\zeta_t = D_\psi(\widehat{\boldsymbol{z}}_{t+1}, \boldsymbol{z}_t) + D_\psi(\boldsymbol{z}_t, \widehat{\boldsymbol{z}}_t) \geq C_6 D_\psi(\boldsymbol{z}^*, \widehat{\boldsymbol{z}}_{t+1})^2, \tag{7}$$

for some $C_6 > 0$. On the other hand, ignoring the nonnegative term $D_\psi(\boldsymbol{z}_t, \widehat{\boldsymbol{z}}_t)$, we also have:

$$\zeta_t = D_\psi(\widehat{\boldsymbol{z}}_{t+1}, \boldsymbol{z}_t) + D_\psi(\boldsymbol{z}_t, \widehat{\boldsymbol{z}}_t) \geq D_\psi(\widehat{\boldsymbol{z}}_{t+1}, \boldsymbol{z}_t) \geq \frac{1}{4} D_\psi(\widehat{\boldsymbol{z}}_{t+1}, \boldsymbol{z}_t)^2, \tag{8}$$

where the last step uses the fact that $\widehat{\boldsymbol{z}}_{t+1}$ and $\boldsymbol{z}_t$ are close [46, Lemma 17 and Lemma 18]. Now, Eq. (7) and Eq. (8) together imply $6\zeta_t \geq 2C_6 D_\psi(\boldsymbol{z}^*, \widehat{\boldsymbol{z}}_{t+1})^2 + D_\psi(\widehat{\boldsymbol{z}}_{t+1}, \boldsymbol{z}_t)^2 \geq \min\left\{C_6, \frac{1}{2}\right\} \Theta_{t+1}^2$. Plugging this back into Eq. (5), we obtain a recursion

$$\Theta_{t+1} \leq \Theta_t - C_7 \Theta_{t+1}^2 \tag{9}$$

for some $C_7 > 0$, which then implies $\Theta_t = O(1/t)$ [46, Lemma 12]. This can be seen as the first and slower stage of the convergence behavior of the algorithm.

To further show a linear convergence rate, they argue that there exists a constant $C_8 > 0$ such that when the algorithm's iterate is reasonably close to $\boldsymbol{z}^*$ in the following sense:

$$\max\{\|\boldsymbol{z}^* - \widehat{\boldsymbol{z}}_t\|_1, \|\boldsymbol{z}^* - \boldsymbol{z}_t\|_1\} \le C_8, \tag{10}$$

the following improved version of Eq. (7) holds (note the lack of square on the right-hand side):

$$\zeta_t = D_\psi(\widehat{\boldsymbol{z}}_{t+1}, \boldsymbol{z}_t) + D_\psi(\boldsymbol{z}_t, \widehat{\boldsymbol{z}}_t) \ge C_9 D_\psi(\boldsymbol{z}^*, \widehat{\boldsymbol{z}}_{t+1}) \tag{11}$$

for some constant $0 < C_9 < 1$. Therefore, using the $1/t$ convergence rate derived in the first stage, there exists a T_0 such that when $t \ge T_0$, Eq. (10) holds and the algorithm enters the second stage. In this stage, combining Eq. (11) and the fact $\zeta_t \ge D_\psi(\widehat{\boldsymbol{z}}_{t+1}, \boldsymbol{z}_t)$ gives $\zeta_t \ge \frac{C_9}{2}\Theta_{t+1}$, which, together with Eq. (5) again, implies an improved recursion $\Theta_{t+1} \le \Theta_t - \frac{15}{32}C_9\Theta_{t+1}$. This finally shows a linear convergence rate $\Theta_t = O((1+\rho)^{-t})$ for some problem-dependent constant $\rho > 0$.

6.3 Analysis of Theorem 6 and Theorem 7

While we mainly follow the steps of the analysis of [46] discussed above to prove Theorem 6 and Theorem 7, we remark that the generalization is highly non-trivial. First of all, we have to prove Eq. (6) for $\mathcal{Z}$ being a general treeplex, which does not follow [35, Lemma C.4] since its proof is very specific to simplexes. Instead, we prove it by writing down the primal-dual linear program of Eq. (1) and applying the strict complementary slackness; see Appendix D.1 for details.

Next, to connect the 2-norm back to the Bregman divergence (which is not the simple KL divergence anymore, especially for DOMWU), we prove the following for VOMWU:

$$D_\psi(\boldsymbol{z}^*, \widehat{\boldsymbol{z}}_{t+1}) \le \sum_{i \in \text{supp}(\boldsymbol{z}^*)} \frac{(z_i^* - \widehat{z}_{t+1,i})^2}{\widehat{z}_{t+1,i}} + \sum_{i \notin \text{supp}(\boldsymbol{z}^*)} \widehat{z}_{t+1,i} \le \frac{3P\|\boldsymbol{z}^* - \widehat{\boldsymbol{z}}_{t+1}\|}{\min_{i \in \text{supp}(\boldsymbol{z}^*)} \widehat{z}_{t+1,i}}, \tag{12}$$

and the following for DOMWU:

$$\frac{D_\psi(\boldsymbol{z}^*, \widehat{\boldsymbol{z}}_{t+1})}{C'} \le \sum_{i \in \text{supp}(\boldsymbol{z}^*)} \frac{(z_i^* - \widehat{z}_{t+1,i})^2}{z_i^* \widehat{q}_{t+1,i}} + \sum_{i \notin \text{supp}(\boldsymbol{z}^*)} z_{p_i}^* \widehat{q}_{t+1,i} \le \frac{\|\boldsymbol{z}^* - \widehat{\boldsymbol{z}}_{t+1}\|_1}{\min_{i \in \text{supp}(\boldsymbol{z}^*)} z_i^* \widehat{z}_{t+1,i}}, \tag{13}$$

where $C' = 4P\|\boldsymbol{\alpha}\|_\infty$ (see Appendix D.2). We then show a lower bound on $z_{t+1,i}$ and $\widehat{z}_{t+1,i}$ for all $i \in \text{supp}(z^*)$, using similar arguments of [46] (see Appendix D.3). Combining Eq. (12) and Eq. (13) with Eq. (6), we have the counterpart of Eq. (7) for both VOMWU and DOMWU.

Showing Eq. (8) also involves extra complication if we follow their analysis, especially for VOMWU which does not admit a closed-form update. Instead, we find a simple workaround: by applying Eq. (5) repeatedly, we get $D_\psi(\boldsymbol{z}^*, \widehat{\boldsymbol{z}}_1) = \Theta_1 \ge \cdots \ge \Theta_{t+1} \ge \frac{1}{16} D_\psi(\widehat{\boldsymbol{z}}_{t+1}, \boldsymbol{z}_t)$, thus, $\zeta_t \ge D_\psi(\widehat{\boldsymbol{z}}_{t+1}, \boldsymbol{z}_t) \ge C_{10} D_\psi(\widehat{\boldsymbol{z}}_{t+1}, \boldsymbol{z}_t)^2$ for some $C_{10} > 0$ depending on $D_\psi(\boldsymbol{z}^*, \widehat{\boldsymbol{z}}_1)$. Combining this with Eq. (7), and applying them to Eq. (5), we obtain the recursion $\Theta_{t+1} \le \Theta_t - C_{11}\Theta_{t+1}^2$ for some $C_{11} > 0$ similar to Eq. (9), which implies $\Theta_t = O(1/t)$ for both VOMWU and DOMWU and proves Theorem 6.

Finally, to show a linear convergence rate, we need to show the counterpart of Eq. (11), which is again more involved compared to the normal-form game case. Indeed, we are only able to do so for DOMWU by making use of its closed-form update described in Lemma 2. Specifically, observe that in Eq. (13), the term $\sum_{i \notin \text{supp}(\boldsymbol{z}^*)} z_{p_i}^* \widehat{q}_{t+1,i}$ is the one that prevents us from bounding $D_\psi(\boldsymbol{z}^*, \widehat{\boldsymbol{z}}_{t+1})$ by $\mathcal{O}(\|\boldsymbol{z}^* - \widehat{\boldsymbol{z}}_{t+1}\|^2)$. Thus, our high-level idea is to argue that $\sum_{i \notin \text{supp}(\boldsymbol{z}^*)} z_{p_i}^* \widehat{q}_{t+1,i}$ decreases significantly as $\widehat{\boldsymbol{z}}_t$ gets close enough to $\boldsymbol{z}^*$. To do so, we use a bottom-up induction to prove that, for any information set $h \in \mathcal{H}^{\mathcal{Z}}$, indices $i, j \in \Omega_h$ such that $i \notin \text{supp}(\boldsymbol{z}^*)$ and $j \in \text{supp}(\boldsymbol{z}^*)$, $\widehat{L}_{t,i}$ is significantly larger than $\widehat{L}_{t,j}$ when $\widehat{\boldsymbol{z}}_t$ is close to $\boldsymbol{z}^*$, where $\widehat{\boldsymbol{L}}_t$ is the counterpart of $\boldsymbol{L}$ in Lemma 2 when computing of $\widehat{\boldsymbol{q}}_{t+1}$. This makes sure that the term $\sum_{i \notin \text{supp}(\boldsymbol{z}^*)} z_{p_i}^* \widehat{q}_{t+1,i}$ is dominated by the other term involving $i \in \text{supp}(\boldsymbol{z}^*)$ in Eq. (13), which eventually helps us show Eq. (11) and the final linear convergence rate in Theorem 7. See Appendix D.5 for details.

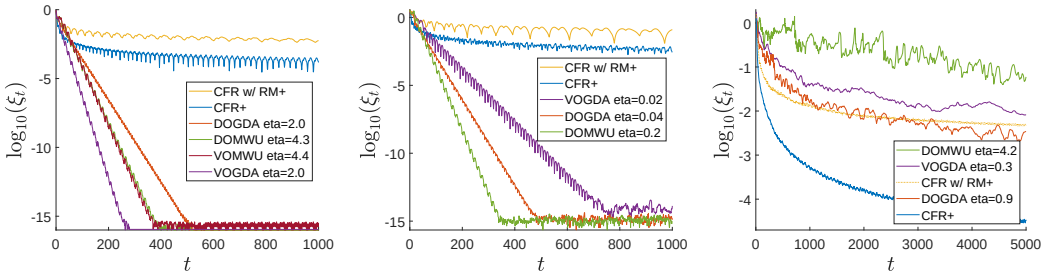

Figure 1: Experiments on Kuhn poker (left), Pursuit-evasion (middle), and Leduc poker (right). A description of each game is given in Appendix A. $\xi_t = \max_{\boldsymbol{y}} \overline{\boldsymbol{x}}_t^\top \boldsymbol{G}\boldsymbol{y} - \min_{\boldsymbol{x}} \boldsymbol{x}^\top \boldsymbol{G}\overline{\boldsymbol{y}}_t$ is the duality gap at time step t, where $(\overline{\boldsymbol{x}}_t, \overline{\boldsymbol{y}}_t)$ is the approximate Nash equilibrium computed by the algorithm at time t (for the optimistic algorithms, $(\overline{\boldsymbol{x}}_t, \overline{\boldsymbol{y}}_t)$ is $(\boldsymbol{x}_t, \boldsymbol{y}_t)$ while for the CFR-based algorithms, $(\overline{\boldsymbol{x}}_t, \overline{\boldsymbol{y}}_t)$ is the linear average). The legend order reflects the curve order at the right-most point. Due to much higher computation overhead than all the other algorithms, we only run VOMWU on Kuhn poker, the game with the smallest size among the three games. For each optimistic algorithm, we fine-tune step size η to get better convergence results and show its value in the legends. There is no hyperparameter for the CFR-based algorithms. All the experiments are run on CPU in a personal computer and the total computation time is less than an hour. There is no random seed and the results are all deterministic.

7 Experiments

In this section, we experimentally evaluate the algorithms on three standard EFG benchmarks: Kuhn poker [31], Pursuit-evasion [29], and Leduc poker [40]. The results are shown in Figure 1. Besides the optimistic algorithms, we also show two CFR-based algorithms as reference points. "CFR+" refers to CFR with alternating updates, linear averaging [42], and regret matching+ as the regret minimizer. "CFR w/ RM+" is CFR with regret matching+ and linear averaging. We provide the formal descriptions of these two algorithms in Appendix C for completeness. For the optimistic algorithms, we plot the last iterate performance. For the CFR-based algorithms, we plot the linear average iterate performance.

For Kuhn poker and Pursuit-evasion (on the left and in the middle of Figure 1), all of the optimistic algorithms perform much better than CFR+, and their curves are nearly straight, showing their linear last-iterate convergence on these games.

For Leduc poker, although CFR+ performs the best, we can still observe the last-iterate convergence trends of the optimistic algorithms. We remark that although VOGDA and DOMWU have linear convergence rate in theory, the experiment on Leduc uses a step size η which is much larger than Corollary 5 and Theorem 7 suggest, which may void the linear convergence guarantee. This is done because the theoretical step size takes too many iterations before it starts to show improvement. It is worth noting that CFR+ improves significantly when changing simultaneous updates (that is, CFR w/ RM+) to alternating updates. Analyzing alternation and combining it with optimistic algorithms is a promising direction. We provide a description of each game, more discussions, and details of the experiments in Appendix A.

8 Conclusions

In this work, we developed the first general last-iterate convergence results for solving EFGs. The paper opens up many potential future directions. On the theoretical side, natural questions include whether it is possible to obtain better convergence rates for VOMWU and DOGDA, whether one can remove the uniqueness assumption for VOMWU and DOMWU, and finally whether it is possible to obtain last-iterate convergence rates for CFR-like optimistic algorithms such as those in [17]. On the practical side, optimistic algorithms with last-iterate convergence guarantees may allow more efficient computation and better incorporation with deep learning-based game-solving approaches.

References

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

[2] M. Bowling, N. Burch, M. Johanson, and O. Tammelin. Heads-up limit hold'em poker is solved. *Science*, 347(6218):145–149, 2015.

[3] G. W. Brown. Iterative solution of games by fictitious play. *Activity analysis of production and allocation*, 13(1):374–376, 1951.

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

[5] N. Brown and T. Sandholm. Solving imperfect-information games via discounted regret minimization. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 33, pages 1829–1836, 2019.

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

[7] N. Brown, S. Ganzfried, and T. Sandholm. Hierarchical abstraction, distributed equilibrium computation, and post-processing, with application to a champion no-limit texas hold'em agent. In *AAAI Workshop: Computer Poker and Imperfect Information*, 2015.

[8] N. Brown, A. Lerer, S. Gross, and T. Sandholm. Deep counterfactual regret minimization. In *International conference on machine learning*, pages 793–802. PMLR, 2019.

[9] N. Burch. *Time and space: Why imperfect information games are hard.* PhD thesis, University of Alberta, 2018.

[10] N. Burch, M. Johanson, and M. Bowling. Solving imperfect information games using decomposition. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 28, 2014.

[11] N. Burch, M. Moravcik, and M. Schmid. Revisiting CFR+ and alternating updates. *Journal of Artificial Intelligence Research*, 64:429–443, 2019.

[12] L. Condat. Fast projection onto the simplex and the ℓ_1 ball. *Mathematical Programming*, 158 (1):575–585, 2016.

[13] C. Daskalakis and I. Panageas. Last-iterate convergence: Zero-sum games and constrained min-max optimization. *Innovations in Theoretical Computer Science*, 2019.

[14] C. Daskalakis, A. Ilyas, V. Syrgkanis, and H. Zeng. Training gans with optimism. In *International Conference on Learning Representations*, 2018.

[15] G. Farina, C. Kroer, N. Brown, and T. Sandholm. Stable-predictive optimistic counterfactual regret minimization. In *International conference on machine learning*, pages 1853–1862. PMLR, 2019.

[16] G. Farina, C. Kroer, and T. Sandholm. Optimistic regret minimization for extensive-form games via dilated distance-generating functions. In *Advances in Neural Information Processing Systems*, pages 5222–5232, 2019.

[17] G. Farina, C. Kroer, and T. Sandholm. Faster game solving via predictive Blackwell approachability: Connecting regret matching and mirror descent. In *Proceedings of the AAAI Conference on Artificial Intelligence*. AAAI, 2021.

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

[19] S. Ganzfried and T. Sandholm. Potential-aware imperfect-recall abstraction with earth mover's distance in imperfect-information games. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 28, 2014.

[20] S. Ganzfried and T. Sandholm. Endgame solving in large imperfect-information games. In *Proceedings of the 2015 International Conference on Autonomous Agents and Multiagent Systems*, pages 37–45. Citeseer, 2015.

[21] Y. Gao, C. Kroer, and D. Goldfarb. Increasing iterate averaging for solving saddle-point problems. In *Proceedings of the AAAI Conference on Artificial Intelligence*, 2021.

[22] A. Gilpin, J. Peña, and T. Sandholm. First-order algorithm with $\mathcal{O}(\ln(1/\epsilon))$ convergence for ϵ-equilibrium in two-person zero-sum games. In *AAAI*, 2008.

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

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

[25] J. Heinrich and D. Silver. Deep reinforcement learning from self-play in imperfect-information games. *arXiv preprint arXiv:1603.01121*, 2016.

[26] S. Hoda, A. Gilpin, J. Pena, and T. Sandholm. Smoothing techniques for computing nash equilibria of sequential games. *Mathematics of Operations Research*, 35(2):494–512, 2010.

[27] C. Kroer and T. Sandholm. Extensive-form game abstraction with bounds. In *Proceedings of the fifteenth ACM conference on Economics and computation*, pages 621–638, 2014.

[28] C. Kroer, G. Farina, and T. Sandholm. Smoothing method for approximate extensive-form perfect equilibrium. In *Proceedings of the 26th International Joint Conference on Artificial Intelligence*, pages 295–301, 2017.

[29] C. Kroer, G. Farina, and T. Sandholm. Robust stackelberg equilibria in extensive-form games and extension to limited lookahead. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 32, 2018.

[30] C. Kroer, K. Waugh, F. Kılınç-Karzan, and T. Sandholm. Faster algorithms for extensive-form game solving via improved smoothing functions. *Mathematical Programming*, 179(1):385–417, 2020.

[31] H. Kuhn. Simplified two-person poker. *Contributions to the Theory of Games*, pages 97–103, 1950.

[32] M. Lanctot, K. Waugh, M. Zinkevich, and M. H. Bowling. Monte carlo sampling for regret minimization in extensive games. In *NIPS*, pages 1078–1086, 2009.

[33] Q. Lei, S. G. Nagarajan, I. Panageas, and X. Wang. Last iterate convergence in no-regret learning: constrained min-max optimization for convex-concave landscapes. *The 24nd International Conference on Artificial Intelligence and Statistics*, 2021.

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

[35] P. Mertikopoulos, C. Papadimitriou, and G. Piliouras. Cycles in adversarial regularized learning. In *Proceedings of the Twenty-Ninth Annual ACM-SIAM Symposium on Discrete Algorithms*, 2018.

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

[37] M. Moravcik, M. Schmid, K. Ha, M. Hladik, and S. Gaukrodger. Refining subgames in large imperfect information games. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 30, 2016.

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

[39] S. Rakhlin and K. Sridharan. Optimization, learning, and games with predictable sequences. In *Advances in Neural Information Processing Systems*, pages 3066–3074, 2013.

[40] F. Southey, M. Bowling, B. Larson, C. Piccione, N. Burch, D. Billings, and C. Rayner. Bayes' bluff: opponent modelling in poker. In *Proceedings of the Twenty-First Conference on Uncertainty in Artificial Intelligence*, pages 550–558, 2005.

[41] V. Syrgkanis, A. Agarwal, H. Luo, and R. E. Schapire. Fast convergence of regularized learning in games. In *Advances in Neural Information Processing Systems*, 2015.

[42] O. Tammelin. Solving large imperfect information games using CFR+. *arXiv preprint arXiv:1407.5042*, 2014.

[43] P. Tseng. On linear convergence of iterative methods for the variational inequality problem. In *Journal of Computational and Applied Mathematics*, 1995.

[44] R. J. Vanderbei et al. *Linear programming*, volume 3. Springer, 2015.

[45] B. von Stengel. Efficient computation of behavior strategies. *Games and Economic Behavior*, 14(2):220–246, 1996.

[46] C.-Y. Wei, C.-W. Lee, M. Zhang, and H. Luo. Linear last-iterate convergence in constrained saddle-point optimization. In *International Conference on Learning Representations*, 2021.

[47] A. Zimin and G. Neu. Online learning in episodic Markovian decision processes by relative entropy policy search. In *Neural Information Processing Systems 26*, 2013.

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

[49] M. Zinkevich, M. Johanson, M. Bowling, and C. Piccione. Regret minimization in games with incomplete information. *Advances in neural information processing systems*, 20:1729–1736, 2007.

Checklist

1. For all authors...
   (a) Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope? [Yes]
   (b) Did you describe the limitations of your work? [Yes]
   (c) Did you discuss any potential negative societal impacts of your work? [N/A] This work is mostly theoretical, and we do not foresee any negative ethical or societal outcomes.
   (d) Have you read the ethics review guidelines and ensured that your paper conforms to them? [Yes]
2. If you are including theoretical results...
   (a) Did you state the full set of assumptions of all theoretical results? [Yes]
   (b) Did you include complete proofs of all theoretical results? [Yes]
3. If you ran experiments...
   (a) Did you include the code, data, and instructions needed to reproduce the main experimental results (either in the supplemental material or as a URL)? [Yes] included in the supplemental material
   (b) Did you specify all the training details (e.g., data splits, hyperparameters, how they were chosen)? [Yes] See Figure 1.
   (c) Did you report error bars (e.g., with respect to the random seed after running experiments multiple times)? [N/A] All the experiments are deterministic.
   (d) Did you include the total amount of compute and the type of resources used (e.g., type of GPUs, internal cluster, or cloud provider)? [Yes] See Figure 1.
4. If you are using existing assets (e.g., code, data, models) or curating/releasing new assets...
   (a) If your work uses existing assets, did you cite the creators? [N/A]
   (b) Did you mention the license of the assets? [N/A]
   (c) Did you include any new assets either in the supplemental material or as a URL? [N/A]
   (d) Did you discuss whether and how consent was obtained from people whose data you're using/curating? [N/A]
   (e) Did you discuss whether the data you are using/curating contains personally identifiable information or offensive content? [N/A]
5. If you used crowdsourcing or conducted research with human subjects...
   (a) Did you include the full text of instructions given to participants and screenshots, if applicable? [N/A]

(b) Did you describe any potential participant risks, with links to Institutional Review Board (IRB) approvals, if applicable? [N/A]

(c) Did you include the estimated hourly wage paid to participants and the total amount spent on participant compensation? [N/A]