FASTER NO-REGRET LEARNING DYNAMICS FOR EXTENSIVE-FORM CORRELATED EQUILIBRIUM

Anonymous authors
Paper under double-blind review

ABSTRACT

A recent emerging trend in the literature on learning in games has been concerned with providing accelerated learning dynamics for correlated and coarse correlated equilibria in normal-form games. Much less is known about the significantly more challenging setting of extensive-form games, which can capture sequential and simultaneous moves, as well as imperfect information. In this paper, we develop faster no-regret learning dynamics for *extensive-form correlated equilibrium (EFCE)* in multiplayer general-sum imperfect-information extensive-form games. When all agents play T repetitions of the game according to the accelerated dynamics, the correlated distribution of play is an $O(T^{-3/4})$-approximate EFCE. This significantly improves over the best prior rate of $O(T^{-1/2})$. One of our conceptual contributions is to connect predictive (that is, optimistic) regret minimization with the framework of Φ-regret. One of our main technical contributions is to characterize the stability of certain fixed point strategies through a refined perturbation analysis of a structured Markov chain, which may be of independent interest. Finally, experiments on standard benchmarks corroborate our findings.

1 INTRODUCTION

Game-theoretic solution concepts describe how agents should rationally act in games. Over the last two decades there has been tremendous progress in imperfect-information game solving and algorithms based on game-theoretic solution concepts have become the state of the art. Prominent milestones of this were an optimal strategy for Rhode Island hold'em poker (Gilpin & Sandholm, 2007), a near-optimal strategy for limit Texas hold'em (Bowling et al., 2015), and a superhuman strategy for no-limit Texas hold'em (Brown & Sandholm, 2017). In particular, these advances rely on algorithms that approximate *Nash equilibria* (*NE*) of two-player zero-sum *extensive-form games* (*EFGs*). EFGs are a broad class of games that capture sequential and simultaneous interaction, and imperfect information. For two-player zero-sum EFGs, it is by now well-understood how to compute a Nash equilibrium at scale: in theory this can be achieved using accelerated uncoupled no-regret learning dynamics, for example by having each player use an *optimistic* regret minimizer and leveraging suitable *distance-generating functions* (Hoda et al., 2010; Kroer et al., 2020; Farina et al., 2021c) for the EFG decision space. Such a setup converges to an equilibrium at a rate of $O(T^{-1})$. In practice, modern variants of the *counterfactual regret minimization* (*CFR*) framework typically lead to better performance, although the worst-case convergence rate is $O(T^{-1/2})$ (Zinkevich et al., 2007). CFR is also an uncoupled no-regret learning dynamic.

However, many real-world applications are not two-player zero-sum games, but instead have *general-sum* utilities and often more than two players. In such settings, Nash equilibrium suffers from several drawbacks when used as a prescriptive tool. First, there can be multiple equilibria, and an equilibrium strategy may perform very poorly when played against the "wrong" equilibrium strategies of the other player(s). Thus, the players effectively would need to communicate in order to find an equilibrium, or hope to converge to it via some sort of learning dynamics. Second, finding a Nash equilibrium is computationally hard both in theory (Daskalakis et al., 2006; Etessami & Yannakakis, 2007) and in practice (Berg & Sandholm, 2017). This effectively squashes any hope of developing efficient learning dynamics that converge to general-sum Nash equilibria.

A competing notion of rationality proposed by Aumann (1974) is that of *correlated equilibrium* (*CE*), typically modeled via a trusted mediator who privately recommends actions to the players.

Unlike NE, it is known that the latter can be computed in polynomial time and, perhaps even more importantly, it can be attained through *uncoupled* learning dynamics, where the players only need to reason about their own observed utilities. This overcomes the often unreasonable presumption that players have knowledge about the other players' utilities. At the same time, uncoupled learning algorithms have proven to be a remarkably *scalable* approach for computing equilibria in large-scale games, as described above. The basic CE notion is defined for normal-form games, and there it has long been known that uncoupled no-regret learning dynamics can converge to CE or the *coarse correlated equilibrium* (*CCE*) variant at a rate of $O(T^{-1/2})$ (Hart & Mas-Colell, 2000; Celli et al., 2019). More recently, it was shown that accelerated uncoupled no-regret learning dynamics can compute CCE and CE at a rate of $O(T^{-3/4})$ (Syrgkanis et al., 2015; Chen & Peng, 2020).

In the context of EFGs, the idea of correlation is much more intricate, and there are several notions of correlated equilibrium, based on when the mediator gives recommendations and how the mediator reacts to players who disregard the advice. One of the most compelling notions for EFGs is the *extensive-form correlated equilibrium* (henceforth EFCE) (von Stengel & Forges, 2008) for extensive-form games with *perfect recall*. In an EFCE, the mediator forms recommendations for each of the possible decision points an agent may encounter in the game, and recommended actions are gradually revealed to players as they reach new information sets; thus, the mediator must take into account the *evolution* of the players' beliefs throughout the game. Because of the sequential nature, the presence of private information in the game, and the gradual revelation of recommendations, the constraints associated with EFCE are significantly more complex than for normal-form games. For these reasons, the question of whether uncoupled learning dynamics can converge to an EFCE was only very recently resolved by Celli et al. (2020). Very recent follow-up work established an explicit rate of convergence of $O(T^{-1/2})$ (Farina et al., 2021a). Our paper is concerned with the following fundamental question: *Can one develop faster uncoupled no-regret learning dynamics for* EFCE*?* We answer this in the positive:

Theorem 1.1. *On any finite perfect-recall general-sum multiplayer extensive-form game, the uncoupled no-regret learning dynamics described in this paper lead to a correlated distribution of play that is an $O(T^{-3/4})$-approximate* EFCE*, where the $O(\cdot)$ notation suppresses game-specific parameters polynomial in the size of the game.*

We achieve this result using the framework of *predictive* (also known as *optimistic*) regret minimization (Chiang et al., 2012; Rakhlin & Sridharan, 2013b). One of our conceptual contributions is to connect this line of work with the framework of Φ-*regret* minimization of Greenwald & Jafari (2003); Gordon et al. (2008), by providing a general template for stable-predictive Φ-regret minimization. The importance of Φ-regret is that it leads to substantially more powerful notions of hindsight rationality, beyond the usual *external* regret. These include the powerful notion of *swap regret* (Blum & Mansour, 2007). One of the primary insights behind the result of Farina et al. (2021a) is to cast convergence to an EFCE as a Φ-regret minimization problem. Given these prior connections, we believe that our stable-predictive Φ template is of independent interest, and could lead to further applications in the future.

Theorem 1.1 extends and strengthens several prior papers in the literature, including the seminal work of Syrgkanis et al. (2015) that provides accelerated dynamics for *coarse* correlated equilibrium in normal-form games, as well as the more recent result of Chen & Peng (2020) which showed $O(T^{-3/4})$ convergence to a correlated equilibrium in normal-form games. For the more challenging class of extensive-form games, accelerated rates were previously known only for finding a *Nash* equilibrium in the special case of *two-player zero-sum* games, where an $O(T^{-3/4})$ rate was achieved via a stable-predictive CFR setup (Farina et al., 2019a) and an $O(T^{-1})$ rate was achieved via optimistic regret minimizers coupled with good distance-generating functions (Farina et al., 2019c).

The rate of convergence of our dynamics depends *polynomially* on the size of the game. To achieve this, one of the primary challenges consists of sharply characterizing the stability of fixed points related to *trigger deviation functions*. This turns out to be particularly challenging, and direct extensions of prior techniques appear to give only a bound that is *exponential* in the size of the game. In this context, one of our key technical contributions is to provide a refined perturbation analysis for a Markov chain consisting of a rank-one stochastic matrix, employing tools that have not been used before in this line of work, and substantially extending the techniques of Chen & Peng (2020). Finally, we support our theoretical findings with experiments on several general-sum benchmarks.

Further Related Work. The line of work on accelerated no-regret learning for *Nash* equilibrium was pioneered by Daskalakis et al. (2015), showing that one can bypass the adversarial $\Omega(T^{-1/2})$ barrier for the incurred average regret if *both* players in a zero-sum game employ an uncoupled variant of the excessive gap technique (Nesterov, 2005), leading to a near-optimal rate of $O(\log T/T)$. Subsequently, Rakhlin & Sridharan (2013a) showed that the optimal rate of $O(1/T)$ can be obtained with a remarkably simple variant of Online Mirror Descent which incorporates a *prediction* term in the update step. While these results only hold for zero-sum games, Syrgkanis et al. (2015) showed that $O(T^{-3/4})$ rate can be obtained for multiplayer general-sum normal-form games. In a recent result, Chen & Peng (2020) strengthened the regret bounds of Syrgkanis et al. (2015) from external to swap regret using the celebrated construction of Blum & Mansour (2007). While the latter involves strategies arising as stationary distributions of a Markov chain, which in general may behave highly nonsmoothly, one of the key observations of Chen & Peng (2020) is that updating each individual regret minimizer via Optimistic Multiplicative Weights (OMWU) suffices to guarantee stability.

Correlated equilibrium in extensive-form games is much less understood than Nash equilibrium, although there has been substantial recent progress. For EFCE we already discussed Celli et al. (2020) and Farina et al. (2021a). A feasible EFCE can also be computed efficiently through a variant of the *Ellipsoid algorithm* (Papadimitriou & Roughgarden, 2008; Jiang & Leyton-Brown, 2015), and an alternative sampling-based approach was given by Dudík & Gordon (2009). However, those approaches perform poorly in large-scale problems, and do not allow the players to arrive at EFCE via distributed learning. Celli et al. (2019) devised variants of the CFR algorithm that provably convergence to *normal-form coarse correlated equilibria*, a solution concept much less appealing than EFCE in extensive-form games. Morrill et al. (2021a;b) characterize hindsight rationality notions and associate a set of solution concepts with suitable $O(T^{-1/2})$ no-regret learning dynamics.

2 Preliminaries

Extensive-form Games. An extensive-form game is abstracted on a directed and rooted *game tree* $\mathcal{T}$. The set of nodes of $\mathcal{T}$ is denoted with $\mathcal{H}$; non-terminal nodes are referred as *decision nodes*, and are associated with a player who acts by selecting an action from a set of possible actions $\mathcal{A}(h)$, where $h \in \mathcal{H}$ represents the decision node. By convention, the set of players $[n] \cup \{c\}$ includes a *fictitious* agent c who "selects" actions according to fixed probability distributions dictated by the nature of the game (e.g., the roll of a dice); this intends to model external stochastic phenomena occurring during the game. For a player $i \in [n] \cup \{c\}$, we let $\mathcal{H}^{(i)} \subseteq \mathcal{H}$ be the subset of decision nodes wherein a player i makes a decision. The set of *leaves* $\mathcal{Z} \subseteq \mathcal{H}$, or equivalently the *terminal nodes*, correspond to different outcomes; once the game transitions to a terminal node $z \in \mathcal{Z}$, payoffs are assigned to each player based on a set of *normalized* utility functions $\{u^{(i)} : \mathcal{Z} \to [-1, 1]\}_{i \in [n]}$. It will also be convenient to represent with $p^{(c)}(z)$ the product of probabilities of "chance" moves encountered in the path from the root until the terminal node $z \in \mathcal{Z}$.

Imperfect Information. To capture imperfect information, the set of decision nodes $\mathcal{H}^{(i)}$ of player i are partitioned into a collection of sets of nodes $\mathcal{J}^{(i)}$, which are called *information sets*. Each information set $j \in \mathcal{J}^{(i)}$ groups nodes which cannot be distinguished by player i. Thus, for any nodes $h, h' \in j$ we have $\mathcal{A}(h) = \mathcal{A}(h')$. As usual, we assume that the game satisfies *perfect recall*: players never forget information once acquired. This implies, in particular, that for any nodes $h, h' \in j$ the sequence of i's actions from the root until h must coincide with the sequence from the root to node h'; otherwise, i would be able to distinguish between the nodes h and h' by virtue of perfect recall. We will also define a partial order $\prec$ on $\mathcal{J}^{(i)}$, so that $j \prec j'$, for $j, j' \in \mathcal{J}^{(i)}$, if there exist nodes $h \in j$ and $h' \in j'$ such that the path from the root to h' passes through node h. If $j \prec j'$, we will say that j is an *ancestor* of j', or equivalently, j is a descendant of j'.

Sequence-form Strategies. For a player $i \in [n]$, an information set $j \in \mathcal{J}^{(i)}$, and an action $a \in \mathcal{A}(j)$, we will denote with $\sigma = (j, a)$ the *sequence* of i's actions encountered on the path from the root of the game until (and included) action a. For notational convenience, we will use the special symbol $\varnothing$ to denote the *empty sequence*. Then, i's set of sequences is defined as $\Sigma^{(i)} := \{(j, a) : j \in \mathcal{J}^{(i)}, a \in \mathcal{A}(j)\} \cup \{\varnothing\}$; we will also use the notation $\Sigma^{(i)}_* := \Sigma^{(i)} \setminus \{\varnothing\}$. For a given information set $j \in \mathcal{J}^{(i)}$ we will use $\sigma^{(i)}(j) \in \Sigma^{(i)}$ to represent the *parent sequence*; i.e. the last sequence encountered by player i before reaching any node in the information set j, assuming

that it exists. Otherwise, we let $\sigma^{(i)}(j) = \varnothing$, and we say that j is the *root information set* of player i. At a high-level, a *strategy* for a player specifies a probability distribution for every possible information set encountered in the game tree. For perfect-recall EFGs, strategies can equivalently be represented in *sequence-form*, where each sequence is associated with a value specifying the realization probability of the sequence of information set-action pairs taken by that player in order to reach a given information set-action pair (Koller et al., 1996; Von Stengel, 2002):

Definition 2.1 (Sequence-form Polytope)**.** The *sequence-form strategy polytope* for player $i \in [n]$ is defined as the following (convex) polytope:

$$\mathcal{Q}^{(i)} := \left\{ \boldsymbol{q} \in \mathbb{R}_{\geq 0}^{|\Sigma^{(i)}|} : \boldsymbol{q}[\varnothing] = 1, \quad \boldsymbol{q}[\sigma^{(i)}(j)] = \sum_{a \in \mathcal{A}(j)} \boldsymbol{q}[(j,a)], \quad \forall j \in \mathcal{J}^{(i)} \right\}. \tag{1}$$

The probability of playing action a at information set j can be obtained by dividing $\boldsymbol{q}[(j,a)]$ by $\boldsymbol{q}[\sigma^{(i)}(j)]$. Analogously, one can define the sequence-form strategy polytope for the *subtree* of the partially ordered set $(\mathcal{J}^{(i)}, \prec)$ *rooted* at $j \in \mathcal{J}^{((i)}$, which will be denoted as $\mathcal{Q}_j^{(i)}$. Moreover, the set of *deterministic* sequence-form strategies for player $i \in [n]$ is the set $\Pi^{(i)} = \mathcal{Q}^{(i)} \cap \{0,1\}^{|\Sigma^{(i)}|}$, and similarly for $\Pi_j^{(i)}$. The *joint* set of deterministic sequence-form strategies of the players will be represented with $\Pi := \bigtimes_{i \in [n]} \Pi^{(i)}$. As such, an element $\boldsymbol{\pi} \in \Pi$ is an n-tuple $(\boldsymbol{\pi}^{(1)}, \dots, \boldsymbol{\pi}^{(n)})$ specifying a deterministic sequence-form strategy for every player $i \in [n]$. Finally, the utility of player $i \in [n]$ under a profile $\boldsymbol{\pi} \in \Pi$ can be expressed as

$$u^{(i)}(\boldsymbol{\pi}) := \sum_{z \in \mathcal{Z}} p^{(c)}(z) u^{(i)}(z) \mathbf{1}\{\boldsymbol{\pi}^{(k)}[\sigma^{(k)}(z)] = 1, \forall k \in [n]\}. \tag{2}$$

Extensive-Form Correlated Equilibrium. We will work with the definition of EFCE due to Farina et al. (2019e) (which is equivalent to that of von Stengel & Forges (2008)), which in turn is based on the notion of *trigger agents* (Gordon et al. (2008)):

Definition 2.2 (Trigger Agent)**.** Consider a player $i \in [n]$, some sequence $\hat{\sigma} = (j,a) \in \Sigma_*^{(i)}$, and $\hat{\boldsymbol{\pi}} \in \Pi_j^{(i)}$. A $(\hat{\sigma}, \hat{\boldsymbol{\pi}})$-*trigger agent* engages in the game according to the following rules:

- If the action a at information set j has not been recommended, the trigger agent faithfully adheres to the recommendations put forward by the mediator;
- Otherwise, if at the information set j the mediator recommends to play action a, the trigger agent gets "trigerred" by the sequence $\hat{\sigma} = (j,a)$. That is, the agent will henceforth disregard any recommendations, and will instead commit to the *continuation strategy* $\hat{\boldsymbol{\pi}}$.

With this definition, an EFCE is a probability distribution $\boldsymbol{\mu} \in \Delta^{|\Pi|}$ over Π such that for any player $i \in [n]$, trigger sequence $\hat{\sigma} = (j,a) \in \Sigma_*^{(i)}$, and continuation strategy $\hat{\boldsymbol{\pi}} \in \Pi_j^{(i)}$, the expected utility of the $(\hat{\sigma}, \hat{\boldsymbol{\pi}})$-trigger agent is not strictly greater than the expected utility of i when the mediator's recommendations are followed. Formally, let us first introduce *trigger deviation functions*:

Definition 2.3 (Trigger Deviation Functions)**.** Consider some player $i \in [n]$, some sequence $\hat{\sigma} = (j,a) \in \Sigma_*^{(i)}$, and joint deterministic sequence-form strategies $\boldsymbol{\pi} \in \Pi_j^{(i)}$. A *trigger deviation function* with respect to a trigger sequence $\hat{\sigma}$ and continuation strategy $\hat{\boldsymbol{\pi}}$ is any linear function $f : \mathbb{R}^{|\Sigma^{(i)}|} \to \mathbb{R}^{|\Sigma^{(i)}|}$ with the following properties:

- Any strategy $\boldsymbol{\pi} \in \Pi^{(i)}$ which does not prescribe the sequence $\hat{\sigma}$ remains invariant. That is, $f(\boldsymbol{\pi}) = \boldsymbol{\pi}$ for any $\boldsymbol{\pi} \in \Pi^{(i)}$ such that $\boldsymbol{\pi}[\hat{\sigma}] = 0$;
- Otherwise, the prescribed sequence $\hat{\sigma} = (j,a)$ is modified so that the behavior at j, as well as all its descendants is replaced by the behavior specified by the continuation strategy:

$$f(\boldsymbol{\pi})[\sigma] = \begin{cases} \boldsymbol{\pi}[\sigma] & \text{if } \sigma \not\succeq j; \\ \hat{\boldsymbol{\pi}}[\sigma] & \text{if } \sigma \succeq j, \end{cases} \tag{3}$$

  for all $\sigma \in \Sigma^{(i)}$ and $\boldsymbol{\pi} \in \Pi^{(i)}$ such that $\boldsymbol{\pi}[\hat{\sigma}] = 1$.

We will let $\Psi^{(i)} := \{\phi^{(i)}_{\hat{\sigma} \to \hat{\boldsymbol{\pi}}} : \hat{\sigma} = (j,a) \in \Sigma_*^{(i)}, \hat{\boldsymbol{\pi}} \in \Pi_j^{(i)}\}$ be the set of all possible linear mappings defining trigger deviation functions for player i.

Definition 2.4 (EFCE)**.** For $\epsilon \geq 0$, a probability distribution $\boldsymbol{\mu} \in \Delta^{|\Pi|}$ is an ϵ-approximate EFCE if for every player $i \in [n]$ and every trigger deviation function $\phi^{(i)}_{\hat{\sigma} \to \hat{\boldsymbol{\pi}}} \in \Psi^{(i)}$, it holds that

$$\mathbb{E}_{\boldsymbol{\pi} \sim \boldsymbol{\mu}} \left[u^{(i)} \left(\phi^{(i)}_{\hat{\sigma} \to \hat{\boldsymbol{\pi}}}(\boldsymbol{\pi}^{(i)}), \boldsymbol{\pi}^{(-i)} \right) - u^{(i)}(\boldsymbol{\pi}) \right] \leq \epsilon, \tag{4}$$

where $\boldsymbol{\pi} = (\boldsymbol{\pi}_1, \dots, \boldsymbol{\pi}_n) \in \Pi$. A probability distribution $\boldsymbol{\mu} \in \Delta^{|\Pi|}$ is an EFCE if it is a 0-EFCE.

Theorem 2.5 (Farina et al. (2021a))**.** *For every player $i \in [n]$, let $\boldsymbol{\pi}^{(i),1}, \dots, \boldsymbol{\pi}^{(i),T} \in \Pi^{(i)}$ be a sequence of deterministic sequence-form strategies whose cumulative $\Psi^{(i)}$-regret is $R^{(i),T}$ with respect to the sequence of linear utility functions*

$$\ell^{(i),t} : \Pi^{(i)} \ni \boldsymbol{\pi}^{(i)} \mapsto u^{(i)} \left(\boldsymbol{\pi}^{(i)}, \boldsymbol{\pi}^{(-i),t} \right). \tag{5}$$

Then, the empirical frequency of play $\boldsymbol{\mu} \in \Delta^{|\Pi|}$ is an ϵ-EFCE, where $\epsilon := \frac{1}{T} \max_{i \in [n]} R^{(i),T}$.

Regret, Φ-Regret and Optimistic Regret Minimization

Consider a convex and compact set $\mathcal{X} \subseteq \mathbb{R}^d$ representing the space of strategies of some agent. In the online decision making framework, a *regret minimizer* $\mathcal{R}$ can be thought of as a black-box device which interacts with the external environment via the following two basic subroutines:

- $\mathcal{R}.\textsc{NextStrategy}()$: The regret minimizer returns the strategy $\boldsymbol{x}^t \in \mathcal{X}$ at time t;
- $\mathcal{R}.\textsc{ObserveUtility}(\ell^t)$: The regret minimizer receives as feedback a utility function $\ell^t : \mathcal{X} \to \mathbb{R}$, and may alter its internal state accordingly. It is assumed that the utility function is linear, so that $\ell^t(\boldsymbol{x}) = \langle \boldsymbol{\ell}^t, \boldsymbol{x} \rangle$, for a (uniquely-defined) vector $\boldsymbol{\ell}^t \in \mathbb{R}^d$.

The decision making is *online* in the sense that at each time t the regret minimizer can adapt to previously received information, but no information about future utilities is available. The error of a regret minimizer is typically measured in terms of *external regret*, defined as follows:

$$R^T := \max_{\boldsymbol{x}^* \in \mathcal{X}} \sum_{t=1}^{T} \langle \boldsymbol{x}^*, \boldsymbol{\ell}^t \rangle - \sum_{t=1}^{T} \langle \boldsymbol{x}^t, \boldsymbol{\ell}^t \rangle, \tag{6}$$

where T represents the time horizon. That is, the performance of the online algorithm is compared with the best *fixed* strategy in *hindsight*. A regret minimizer is called *Hannan consistent* if, under any sequence of utility functions, the regret grows sublinearly with T; i.e. $\mathcal{R} = o(T)$.

Φ-*Regret.* A conceptual generalization of the concept of external regret is the so-called Φ-*regret*. Specifically, in this framework the performance of the learning algorithm is measured based on a *set of transformations* $\Phi : \mathcal{X} \to \mathcal{X}$, leading to the notion of cumulative Φ-regret:

$$R^T := \max_{\phi^* \in \Phi} \sum_{t=1}^{T} \langle \phi^*(\boldsymbol{x}^t), \boldsymbol{\ell}^t \rangle - \sum_{t=1}^{T} \langle \boldsymbol{x}^t, \boldsymbol{\ell}^t \rangle. \tag{7}$$

Observe that when the set of transformations Φ coincides with the set of *constant* functions, one recovers the ordinal notion of (external) regret given in Equation (6). However, Φ-regret is substantially more expressive and yields a more appealing notion of hindsight rationality, incorporating the notion of *swap regret* (Blum & Mansour, 2007).

An emerging subfield of online learning (Chiang et al. (2012); Rakhlin & Sridharan (2013a)) studies the improved performance guarantees one can obtain if the utility functions observed by the no-regret algorithm possess additional *structure*, usually in the form of *small variation*. For our purposes, we will use the following definition, which is a slight modification of the RVU property introduced by (Syrgkanis et al., 2015, Definition 3).

Definition 2.6 (Stable-predictivity)**.** Let $\mathcal{R}$ be a regert minimizer and let $\| \cdot \|$ be a norm. $\mathcal{R}$ is said to be κ-*stable* with respect to $\| \cdot \|$ if for all $t \geq 2$, the strategies output by $\mathcal{R}$ satisfy

$$\|\boldsymbol{x}^t - \boldsymbol{x}^{t-1}\| \leq \kappa, \tag{8}$$

Moreover, it is said to be (α, β)-*predictive* with respect to $\| \cdot \|$, where $\alpha : \mathbb{N} \to \mathbb{R}_{>0}$ and $\beta \in \mathbb{R}_{>0}$, if at all $t \geq 1$ its regret R^T satisfies

$$R^T \leq \alpha(T) + \beta \sum_{t=2}^{T} \|\boldsymbol{\ell}^t - \boldsymbol{\ell}^{t-1}\|_*^2, \tag{9}$$

no matter the sequence of utility vectors $\boldsymbol{\ell}^1, \dots, \boldsymbol{\ell}^T$, where $\|\cdot\|_*$ is the dual norm of $\|\cdot\|$.

Optimistic Follow the Regularized Leader. The class of algorithms termed Optimistic Follow the Regularized Leader (OFTRL) constitutes a simple variant of FTRL, incorporating a *prediction* into the optimization step. To be more precise, let d be a 1-strongly convex function with respect to some norm $\|\cdot\|$, and $\eta > 0$ the *learning rate*. OFTRL's update rule takes the following form:

$$\boldsymbol{x}^t \coloneqq \arg\max_{\boldsymbol{x} \in \mathcal{X}} \left\{ \left\langle \boldsymbol{x}, 2\boldsymbol{\ell}^{t-1} + \sum_{\tau=1}^{t-2} \boldsymbol{\ell}^\tau \right\rangle - \frac{d(\boldsymbol{x})}{\eta} \right\}, \tag{OFTRL}$$

where $\boldsymbol{x}^1 \coloneqq \arg\min_{\boldsymbol{x} \in \mathcal{X}} d(\boldsymbol{x})$. Syrgkanis et al. (2015) established the following property:

Lemma 2.7 (Syrgkanis et al. (2015, Proposition 7 and Lemma 20)[1]). (OFTRL) *with learning rate η is 2η-stable and $(\Omega_d/\eta, \eta)$-predictive with respect to any norm $\|\cdot\|$ with respect to which d is 1-strongly convex, where Ω_d is the range of d on $\mathcal{X}$, that is, $\Omega_d \coloneqq \max_{\boldsymbol{x}, \boldsymbol{x}' \in \mathcal{X}} \{d(\boldsymbol{x}) - d(\boldsymbol{x}')\}$.*

In this paper, we consider the entropic regularizer, which is 1-strongly convex with respect to the ℓ_1 norm. The pair of dual norms in the predictivity bound will therefore be $(\|\cdot\|_1, \|\cdot\|_\infty)$. We call this OFTRL setup *Optimistic Multiplicative Weights Updates* (OMWU).

3 Accelerating Φ-Regret Minimization via Optimism

In this section we develop a general template for accelerated Φ-regret minimization for general sets, and then we instantiate the template for dynamics for EFCE. Our approach combines a framework of Gordon et al. (2008) with the framework of stable-predictive (aka. optimistic) regret minimization. As in Gordon et al. (2008), in our template we combine 1) a regret minimizer that outputs a linear transformation $\phi^t \in \Phi$ at every time t, and 2) a fixed-point oracle for each $\phi^t \in \Phi$. However, in our framework, we further require that (2) is stable (in the sense of Definition 2.6). To make (2) stable, we will focus on regret minimizers that have the following property.

Definition 3.1. Consider a set of functions Φ such that $\phi(\mathcal{X}) \subseteq \mathcal{X}$ for all $\phi \in \Phi$, and a no-regret algorithm $\mathcal{R}_\Phi$ for the set of transformations Φ which returns a sequence $\phi^t \in \Phi$. We say that $\mathcal{R}_\Phi$ is *fixed point G-stable*, for $G \geq 0$, if the following conditions hold:

- Every ϕ^t admits a fixed point. That is, there exists $\boldsymbol{x}^t \in \mathcal{X}$ such that $\phi^t(\boldsymbol{x}^t) = \boldsymbol{x}^t$.
- For any $\boldsymbol{x}^t$ such that $\boldsymbol{x}^t = \phi^t(\boldsymbol{x}^t)$, there exists $\boldsymbol{x}^{t+1}$ with $\boldsymbol{x}^{t+1} = \phi^{t+1}(\boldsymbol{x}^{t+1})$ such that $\|\boldsymbol{x}^{t+1} - \boldsymbol{x}^t\| \leq G$.

We will show how to construct an accelerated Φ-regret minimizer starting from the following:

1. $\mathcal{R}_\Phi$: A κ-stable (α, β)-predictive fixed point G-stable regret minimizer for Φ;
2. StableFPOracle$(\phi; \widetilde{\boldsymbol{x}}, G, \epsilon)$: A *stable fixed point oracle* which returns a point $\boldsymbol{x} \in \mathcal{X}$ such that (i) $\|\phi(\boldsymbol{x}) - \boldsymbol{x}\| \leq \epsilon$, and (ii) $\|\boldsymbol{x} - \widetilde{\boldsymbol{x}}\| \leq G$ (the existence of such a fixed point is guaranteed by the fixed point G-stability assumption for the regret minimizer).

Given these two components, our next theorem builds a stable-predictive Φ-regret minimizer.

Theorem 3.2 (Accelerated Φ-Regret Minimization). *Consider a κ-stable (α, β)-predictive regret minimizer $\mathcal{R}_\Phi$ for a set of linear transformations Φ, with respect to the ℓ_1 norm $\|\cdot\|_1$. Moreover, assume that $\mathcal{R}_\Phi$ is fixed point G-stable with respect to Φ. Then, if we have access to a* StableFPOracle*, we can construct a G-stable algorithm with Φ-regret R^T bounded as*

$$R^T \leq \alpha(T) + 2\beta D_{\boldsymbol{\ell}}^2 \kappa^2 T + 2\beta \sum_{t=2}^{T} \|\boldsymbol{\ell}^t - \boldsymbol{\ell}^{t-1}\|_\infty^2 + D_{\boldsymbol{\ell}} \sum_{t=1}^{T} \epsilon_t, \tag{10}$$

where ϵ_t is the error of StableFPOracle *at time t, and $D_{\boldsymbol{\ell}}$ is an upper bound on the ℓ_∞ norm of $\boldsymbol{\ell}^t$'s. It is also assumed that $\|\boldsymbol{x}\|_\infty \leq 1$ for all $\boldsymbol{x} \in \mathcal{X}$.*

The proof is similar to that of Gordon et al. (2008), and is included in Appendix B. This framework reduces performing accelerated Φ-regret minimization to (i) developing a stable-predictive regret minimizer for the set $\Psi^{(i)}$, for any player $i \in [n]$, and (ii) establishing a StableFPOracle. In this context, Section 3.1 is concerned with the former task, while Section 3.2 with the latter.

[1] Technically, the argument in Syrgkanis et al. (2015) was developed with respect to the simplex, but it is easy to verify that the claim directly extends for any convex and compact set $\mathcal{X}$.

3.1 CONSTRUCTING A STABLE-PREDICTIVE REGRET MINIMIZER FOR $\Psi^{(i)}$

Here we develop a regret minimizer for the set $\operatorname{co}\Psi^{(i)}$, the convex hull of the set of trigger deviation functions. Given that $\operatorname{co}\Psi^{(i)} \supseteq \Psi^{(i)}$, this will immediately imply a regret minimizer for the set $\Psi^{(i)}$. Farina et al. (2021a) observed that the set $\operatorname{co}\Psi^{(i)}$ can be evaluated in two stages. First, for a fixed sequence $\hat{\sigma} = (j, a) \in \Sigma_*^{(i)}$ we define the set $\Psi_{\hat{\sigma}}^{(i)} := \operatorname{co}\left\{\phi_{\hat{\sigma}\to\hat{\boldsymbol{\pi}}} : \hat{\boldsymbol{\pi}} \in \Pi_j^{(i)}\right\}$; then, we take the convex hull of all $\Psi_{\hat{\sigma}}^{(i)}$, that is, $\operatorname{co}\Psi^{(i)} = \operatorname{co}\{\Psi_{\hat{\sigma}}^{(i)} : \hat{\sigma} \in \Sigma_*^{(i)}\}$. Correspondingly, we first develop a stable-predictive regret minimizer for the set $\Psi_{\hat{\sigma}}^{(i)}$, for any $\hat{\sigma} \in \Sigma_*^{(i)}$, and these individual regret minimizers are then combined using a *regret circuit* to conclude the construction in Theorem 3.5. All the omitted poofs and pseudocode for this section are included in Appendix B.1.

Stable-Predictive Regret Minimizer for the set $\Psi_{\hat{\sigma}}^{(i)}$. Consider a sequence $\hat{\sigma} \in \Sigma_*^{(i)}$. Farina et al. (2021a) observed that the set of transformations $\Psi_{\hat{\sigma}}^{(i)} := \operatorname{co}\left\{\phi_{\hat{\sigma}\to\hat{\boldsymbol{\pi}}} : \hat{\boldsymbol{\pi}} \in \Pi_j^{(i)}\right\}$ is the image of $\mathcal{Q}_j^{(i)}$ under the affine mapping $h_{\hat{\sigma}}^{(i)} : \phi_{\hat{\sigma}\to\boldsymbol{\pi}}^{(i)}$. Hence, it is well-known that a regret minimizer for $\Psi_{\hat{\sigma}}^{(i)}$ can be constructed starting from a regret minimizer for $\mathcal{Q}_j^{(i)}$. We now show that the same can be said if one restricts to *stable-predictive* regret minimizers. In particular, we have the following.

Proposition 3.3. *Consider a player $i \in [n]$ and any trigger sequence $\hat{\sigma} = (j, a) \in \Sigma_*^{(i)}$. There exists an algorithm which constructs a deterministic regret minimizer $\mathcal{R}_{\hat{\sigma}}^{(i)}$ with access to a K-stable (A_T, B)-predictive deterministic regret minimizer $\mathcal{R}_{\mathcal{Q}}^{(i)}$ for the set $\mathcal{Q}_j^{(i)}$, such that $\mathcal{R}_{\hat{\sigma}}^{(i)}$ is K-stable and (A_T, B)-predictive. Moreover, if the routines* NEXTSTRATEGY *and* OBSERVEUTILITY *of $\mathcal{R}_{\mathcal{Q}}^{(i)}$ run in linear time on $|\Sigma_j^{(i)}|$, the complexity of $\mathcal{R}_{\hat{\sigma}}^{(i)}$ is $O(|\Sigma_j^i|)$.*

In Appendix A we describe a stable-predictive variant of CFR for the set $\mathcal{Q}_j^{(i)}$, for each $j \in \mathcal{J}^{(i)}$, following the construction of Farina et al. (2019a).

Stable-Predictive Regret Minimizer for $\operatorname{co}\Psi^{(i)}$. The next step consists of combining the regret minimizers $\Psi_{\hat{\sigma}}^{(i)}$, for all $\hat{\sigma} \in \Sigma_*^{(i)}$, to a composite regret minimizer for the set $\operatorname{co}\Psi^{(i)}$. To this end, we will employ *regret circuits* (Farina et al., 2019d), and in particular the following construction:

Proposition 3.4 (Farina et al. (2019d))**.** *Consider a collection of sets $\mathcal{X}_1, \dots, \mathcal{X}_m$, and let $\mathcal{R}_i$ be a regret minimizer for the set $\mathcal{X}_i$, for each $i \in [m]$. Moreover, let $\mathcal{R}_\Delta$ be a regret minimizer for the m-simplex Δ^m. A regret minimizer $\mathcal{R}_{\text{co}}$ for the set $\operatorname{co}\{\mathcal{X}_1, \dots, \mathcal{X}_m\}$ can be constructed as follows:*

- *$\mathcal{R}_{\text{co}}$.* NEXTSTRATEGY *obtains the next strategy $\boldsymbol{x}_i^t$ of each regret minimizer $\mathcal{R}_i$, as well as the next strategy $\boldsymbol{\lambda}^t = (\lambda_1^t, \dots, \lambda_m^t) \in \Delta^m$ of $\mathcal{R}_\Delta$, and returns the strategy $\lambda_1^t \boldsymbol{x}_1^t + \dots + \lambda_m^t \boldsymbol{x}_m^t$.*
- *$\mathcal{R}_{\text{co}}$.* OBSERVEUTILITY(L^t) *forwards the function L^t to each of the regret minimizers $\mathcal{R}_1, \dots, \mathcal{R}_m$, while it also forwards the utility $(\lambda_1, \dots, \lambda_m) \mapsto \lambda_1 L^t(\boldsymbol{x}_1^t) + \dots + \lambda_m L^t(\boldsymbol{x}_m^t)$.*

Then, if $R_1^T, \dots, R_m^T$ is the regret accumulated by the regret minimizers $\mathcal{R}_1, \dots, \mathcal{R}_m$, and R_Δ^T is the regret of $\mathcal{R}_\Delta$, then the cumulative regret of the composite regret minmizers $\mathcal{R}_{\text{co}}$ can be bounded as

$$R_{\text{co}}^T \le R_\Delta^T + \max\{R_1^T, \dots, R_m^T\}. \tag{11}$$

We are now ready to establish the main result of this section.

Theorem 3.5. *Consider a κ-stable (α, β)-predictive regret minimizer $\mathcal{R}_\Delta^{(i)}$ for the the simplex $\Delta^{|\Sigma_*^{(i)}|}$, and K-stable (A, B)-predictive regret minimizers $\mathcal{R}_{\hat{\sigma}}^{(i)}$ for each $\hat{\sigma} \in \Sigma_*^{(i)}$, all with respect to the pair of norms $(\|\cdot\|_1, \|\cdot\|_\infty)$. Then, there exists an algorithm which constructs a regret minimizer $\mathcal{R}_\Psi^{(i)}$ for the set $\operatorname{co}\Psi^{(i)}$ such that (i) $\mathcal{R}_\Psi^{(i)}$ is $O(K + |\Sigma^{(i)}|\kappa)$-stable, and (ii) under any sequence of linear utility functions $L^1, \dots, L^T$ the regret incurred can be bounded as*

$$R_\Psi^T \le O(\alpha(T) + A(T) + \beta D_{\mathbf{L}}^2 K^2 T) + O(B + \beta|\Sigma^{(i)}|^2) \sum_{t=2}^{T} \|\mathbf{L}^t - \mathbf{L}^{t-1}\|_\infty^2, \tag{12}$$

where $\|\mathbf{L}^t\|_\infty \le D_{\mathbf{L}}$. Moreover, if the routines NEXTSTRATEGY *and* OBSERVEUTILITY *of $\mathcal{R}_\Delta^{(i)}$ and $\mathcal{R}_{\hat{\sigma}}^{(i)}$, for each $\hat{\sigma} \in \Sigma_*^{(i)}$, run in linear time on $|\Sigma^{(i)}|$, then the complexity of $\mathcal{R}_\Psi^{(i)}$ is $O(|\Sigma^{(i)})|^2)$.*

3.2 STABILITY OF THE FIXED POINTS

In this subsection we complete the construction of the $\Psi^{(i)}$-regret minimizer by establishing a *stable* fixed point oracle for any $\phi \in \operatorname{co}\Psi^{(i)}$. All of the proofs of this section are included in Appendix B.2.

Multiplicative Stability. A sequence $\{\boldsymbol{z}^t\}$, with $\boldsymbol{z}^t \in \mathbb{R}^d_{\geq 0}$, is said to be *$\kappa$-multiplicative-stable* if $(1-\kappa)\boldsymbol{z}_i^{t-1} \leq \boldsymbol{z}_i^t \leq (1+\kappa)\boldsymbol{z}_i^{t-1}$, for any $i \in [d]$, and for all $t \geq 2$. Chen & Peng (2020) observed that a Markov chain whose associated transition matrix is multiplicatively stable will also have a stable stationary distribution in the sense of Definition 2.6. Importantly, this notion of multiplicative stability is guaranteed by OMWU (see Lemma B.1). Thus, if $\mathfrak{D}^{(i)}$ is the depth of i's actions and $D_{\boldsymbol{x}}^{(i)}$ is an upper bound on the ℓ_1 norm in the treeplex, we can show the following:

Lemma 3.6. *When each regret minimizer $\mathcal{R}_{\hat{\sigma}}^{(i)}$ is constructed using predictive* CFR *instantiated with* OMWU *with learning rate η (Theorem A.4) such that for all $\hat{\sigma} \in \Sigma_*^{(i)}$, the output sequence is $O(\eta(\mathfrak{D}^{(i)})^2 D_{\boldsymbol{x}}^{(i)} D_{\boldsymbol{\ell}})$-multiplicatively-stable. Moreover, if the regret minimizer $\mathcal{R}_{\Delta}^{(i)}$ is realized using* OMWU *with learning rate η, it will output an $O(\eta|\Sigma^{(i)}|D_{\boldsymbol{\ell}})$-multiplicatively-stable sequence.*

This characterization will be crucial for establishing the stability of the fixed points. In particular, following the approach of Farina et al. (2021a), let us introduce the following definitions:

Definition 3.7. Consider a player $i \in [n]$ and let $J \subseteq \mathcal{J}^{(i)}$ be a subset of i's information sets. We say than J is a *trunk* of $\mathcal{J}^{(i)}$ if, for every $j \in J$, all predecessors of j are also in J.

Definition 3.8. Consider a player $i \in [n]$, a trunk $J \subseteq \mathcal{J}^{(i)}$, and $\phi \in \operatorname{co}\Psi^{(i)}$. A vector $\boldsymbol{x} \in \mathbb{R}^{|\Sigma^{(i)}|}_{\geq 0}$ is a J-partial fixed point of ϕ if the following conditions hold:

- $\boldsymbol{x}[\varnothing] = 1$ and $\boldsymbol{x}[\sigma^{(i)}(j)] = \sum_{a \in \mathcal{A}(j)} \boldsymbol{x}[(j,a)]$, for all $j \in J$;
- $\phi(\boldsymbol{x})[\varnothing] = \boldsymbol{x}[\varnothing] = 1$, and $\phi(\boldsymbol{x})[(j,a)] = \boldsymbol{x}[(j,a)]$, for all $j \in J$, and $a \in \mathcal{A}(j)$.

An important property is that a J-partial fixed point can be efficiently "promoted" to a $J \cup \{j^*\}$-partial fixed point by computing the stationary distribution of a certain Markov chain. However, a significant concern is whether this fixed point operation can potentially cause a substantial degradation in terms of stability. One of our key results is that the associated Markov chain has a particular structure, which enables us to substantially improve the stability bound and thereby obtain a polynomial degradation in stability. More precisely, we show the following technical lemma.

Lemma 3.9. *Let $\mathbf{M}$ and $\mathbf{M}'$ be transition matrices of m-state Markov chains such that $\mathbf{M} = \boldsymbol{v}\mathbf{1}^\top + \mathbf{C}$ and $\mathbf{M}' = \boldsymbol{v}'\mathbf{1}^\top + \mathbf{C}'$, where $\mathbf{C}, \mathbf{C}', \boldsymbol{v}, \boldsymbol{v}'$ have strictly positive entries. Moreover, let $\boldsymbol{\pi}$ and $\boldsymbol{\pi}'$ be the (unique) stationary distributions of $\mathbf{M}$ and $\mathbf{M}'$ respectively. Then, if the entries of the matrices $\mathbf{C}$ and $\mathbf{C}'$ are κ-multiplicatively-close, (ii) the entries of the vectors $\boldsymbol{v}$ and $\boldsymbol{v}'$ are γ-multiplicatively-close, and (iii) the sum of the entries of $\boldsymbol{v}$ and $\boldsymbol{v}'$ are κ-multiplicatively-close, then $\boldsymbol{\pi}$ and $\boldsymbol{\pi}'$ are $(\gamma + O(\kappa m))$-multiplicatively-close, for a sufficiently small $\kappa = O(1/m)$.*

Using a slight variant of this result (Corollary B.9), we manage to obtain the following:

Proposition 3.10. *Consider a player $i \in [n]$, and let $\phi = \sum_{\hat{\sigma} \in \Sigma_*^{(i)}} \boldsymbol{\lambda}[\hat{\sigma}]\phi^{(i)}_{\hat{\sigma} \to \boldsymbol{q}_{\hat{\sigma}}}$ be a transformation in $\operatorname{co}\Psi^{(i)}$ such that the sequence of $\boldsymbol{\lambda}^t$'s and $\boldsymbol{q}_{\hat{\sigma}}^t$'s is κ-multiplicatively-stable, for all $\hat{\sigma} \in \Sigma_*^{(i)}$. If $\boldsymbol{x}^t$ is a γ-multiplicatively-stable J-partial fixed point sequence, there is an algorithm which computes a $(J \cup \{j^*\})$-partial fixed point $(\boldsymbol{x}^t)'$ of ϕ such that the sequence of $(\boldsymbol{x}')^t$'s is $(\gamma + O(\kappa|\mathcal{A}(j^*)|))$-multiplicatively-stable, for any sufficiently small $\kappa = O(1/|\mathcal{A}(j^*)|)$.*

Thus, using our technical lemma, we manage to bypass the substantial overhead of the term $\gamma|\mathcal{A}(j^*)|$, which would follow using techniques similar to Chen & Peng (2020). This turns out to be crucial for obtaining a polynomial dependence on the size of the game. Finally, we can inductively employ this proposition to show the overall stability of the fixed points:

Theorem 3.11. *Consider a player $i \in [n]$, and let $\phi = \sum_{\hat{\sigma} \in \Sigma_*^{(i)}} \boldsymbol{\lambda}[\hat{\sigma}]\phi^{(i)}_{\hat{\sigma} \to \boldsymbol{q}_{\hat{\sigma}}}$ be a transformation in $\operatorname{co}\Psi^{(i)}$ such that the sequence of $\boldsymbol{\lambda}^t$'s and $\boldsymbol{q}_{\hat{\sigma}}^t$'s is κ-multiplicatively-stable, for all $\hat{\sigma} \in \Sigma_*^{(i)}$. Then, there exists an algorithm which computes a fixed point $\boldsymbol{q}^t \in \mathcal{Q}^{(i)}$ of ϕ such that the sequence of $\boldsymbol{q}^t$'s is $O(\kappa|\mathcal{A}^{(i)}|\mathfrak{D}^{(i)})$-multiplicatively-stable, where $|\mathcal{A}^{(i)}| := \max_{j \in \mathcal{J}^{(i)}} |\mathcal{A}(j)|$, and for a sufficiently small $\kappa = O(1/(|\mathcal{A}^{(i)}|\mathfrak{D}^{(i)}))$.*

Finally, if we use the stability values derived in Lemma 3.6, we arrive at the following conclusion:

Corollary 3.12. *For $\kappa = O((D_{\boldsymbol{x}}^{(i)}(\mathfrak{D}^{(i)})^2 + |\Sigma^{(i)}|)|\mathcal{A}^{(i)}|\mathfrak{D}^{(i)} D_{\boldsymbol{\ell}})$, the sequence of fixed points will be $(\eta\kappa)$-multiplicatively-stable, for any sufficiently small $\eta = O(1/\kappa)$.*

Putting It All Together. Finally, having established these ingredients, we can use the template of Theorem 3.2 to obtain Theorem 1.1, as we formally show in Appendix B.3.

The pseudocode for the final algorithm is given in Appendix B.

4 Experiments

In this section we experimentally investigate the performance of our stable-predictive algorithm compared to two other popular approaches based on a CFR-style decomposition of regrets into local regret-minimization problems: the existing algorithm by Farina et al. (2021b) instantiated with (i) *regret matching*$^+$ (RM$^+$) (Tammelin, 2014) for each simplex (in place of regret matching), and (ii) using the vanilla MWU algorithm for each simplex. In accordance to the theoretical predictions, the stepsize for OMWU is set as $\eta_t = \tau \cdot t^{-1/4}$ (cf. Corollary B.12), and for MWU it is set as $\eta_t = \tau \cdot t^{-1/2}$, where the parameter τ is chosen by picking the best-performing value among $\{0.01, 0.1, 1, 10, 100\}$. In particular, we evaluate their performance based on the following popular benchmark games: (i) a three-player variant of *Kuhn poker* (Kuhn, 1950); (ii) a two-player bargaining game known as *Sheriff* (Farina et al., 2019e)—a benchmark game introduced specifically for the study of correlated equilibria; and (iii) a three-player version of *Liar's dice* (Lisý et al., 2015). A detailed description of each of the three game instances is available in Appendix D.

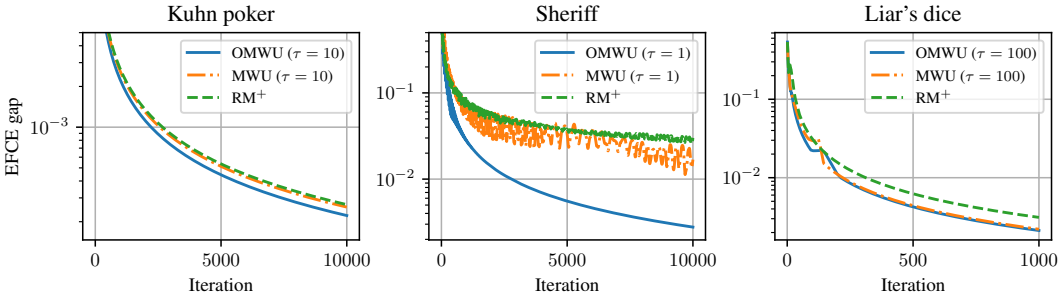

Figure 1: The performance of MWU, OMWU, and RM$^+$ on three general-sum EFGs.

Figure 1 shows the performance of each of the three learning dynamics for computing EFCE. On the x-axis we plot the number of iterations performed by each algorithm, (one iteration costs the same for every algorithm, up to constant factors), and on the y-axis we plot the EFCE gap, defined as the maximum advantage that any player can gain by defecting optimally from the mediator's recommendations. We see that on every game, OMWU performs better than or on par with RM$^+$ and MWU. On Sheriff, OMWU performs significantly better than both RM$^+$ and MWU, by about an order of magnitude. One caveat to these results is that we did not use two tricks that help CFR$^+$ in two-player zero-sum EFG solving: alternation and linear averaging. These tricks are known to retain convergence guarantees in that context (Tammelin et al., 2015; Farina et al., 2019b; Burch et al., 2019), but it is unclear if they still guarantee convergence in the EFCE setting.

5 Conclusions

We described uncoupled no-regret learning dynamics so that if all agents play T repetitions of the game according to the dynamics, the correlated distribution of play is an $O(T^{-3/4})$-approximate EFCE. This substantially improves over the prior best rate of $O(T^{-1/2})$. One of our conceptual contributions is to connect the line of work on optimistic regret minimization with the framework of Φ-regret. One of our main technical contributions is to characterize the stability of the fixed points associated with trigger deviation functions through a refined perturbation analysis of a certain structured Markov chain, which may be of independent interest. Finally, experiments conducted on standard benchmarks corroborated our theoretical findings.

REFERENCES

Robert Aumann. Subjectivity and correlation in randomized strategies. *Journal of Mathematical Economics*, 1:67–96, 1974.

Kimmo Berg and Tuomas Sandholm. Exclusion method for finding nash equilibrium in multiplayer games. In *AAAI Conference on Artificial Intelligence (AAAI)*, 2017.

Avrim Blum and Yishay Mansour. From external to internal regret. *J. Mach. Learn. Res.*, 8:1307–1324, 2007.

Michael Bowling, Neil Burch, Michael Johanson, and Oskari Tammelin. Heads-up limit hold'em poker is solved. *Science*, 347(6218), January 2015.

Noam Brown and Tuomas Sandholm. Superhuman AI for heads-up no-limit poker: Libratus beats top professionals. *Science*, pp. eaao1733, Dec. 2017.

Neil Burch, Matej Moravcik, and Martin Schmid. Revisiting CFR+ and alternating updates. *Journal of Artificial Intelligence Research*, 64:429–443, 2019.

Andrea Celli, Alberto Marchesi, Tommaso Bianchi, and Nicola Gatti. Learning to correlate in multi-player general-sum sequential games. In *Proceedings of the Annual Conference on Neural Information Processing Systems (NeurIPS)*, volume 32, 2019.

Andrea Celli, Alberto Marchesi, Gabriele Farina, and Nicola Gatti. No-regret learning dynamics for extensive-form correlated equilibrium. In Hugo Larochelle, Marc'Aurelio Ranzato, Raia Hadsell, Maria-Florina Balcan, and Hsuan-Tien Lin (eds.), *Advances in Neural Information Processing Systems 33: Annual Conference on Neural Information Processing Systems 2020*, 2020.

Xi Chen and Binghui Peng. Hedging in games: Faster convergence of external and swap regrets. In *Proceedings of the Annual Conference on Neural Information Processing Systems (NeurIPS)*, 2020.

Chao-Kai Chiang, Tianbao Yang, Chia-Jung Lee, Mehrdad Mahdavi, Chi-Jen Lu, Rong Jin, and Shenghuo Zhu. Online optimization with gradual variations. In *Conference on Learning Theory*, pp. 6–1, 2012.

Constantinos Daskalakis, Paul Goldberg, and Christos Papadimitriou. The complexity of computing a Nash equilibrium. In *Proceedings of the Annual Symposium on Theory of Computing (STOC)*, 2006.

Constantinos Daskalakis, Alan Deckelbaum, and Anthony Kim. Near-optimal no-regret algorithms for zero-sum games. *Games and Economic Behavior*, 92:327–348, 2015.

Miroslav Dudík and Geoffrey J. Gordon. A sampling-based approach to computing equilibria in succinct extensive-form games. In Jeff A. Bilmes and Andrew Y. Ng (eds.), *UAI 2009, Proceedings of the Twenty-Fifth Conference on Uncertainty in Artificial Intelligence, Montreal, QC, Canada, June 18-21, 2009*, pp. 151–160. AUAI Press, 2009.

Kousha Etessami and Mihalis Yannakakis. On the complexity of Nash equilibria and other fixed points (extended abstract). In *Proceedings of the Annual Symposium on Foundations of Computer Science (FOCS)*, pp. 113–123, 2007.

Gabriele Farina, Christian Kroer, Noam Brown, and Tuomas Sandholm. Stable-predictive optimistic counterfactual regret minimization. In *International Conference on Machine Learning (ICML)*, 2019a.

Gabriele Farina, Christian Kroer, and Tuomas Sandholm. Online convex optimization for sequential decision processes and extensive-form games. In *AAAI Conference on Artificial Intelligence*, 2019b.

Gabriele Farina, Christian Kroer, and Tuomas Sandholm. Optimistic regret minimization for extensive-form games via dilated distance-generating functions. In *Advances in Neural Information Processing Systems, NeurIPS 2019,*, pp. 5222–5232, 2019c.

Gabriele Farina, Christian Kroer, and Tuomas Sandholm. Regret circuits: Composability of regret minimizers. In *International Conference on Machine Learning*, pp. 1863–1872, 2019d.

Gabriele Farina, Chun Kai Ling, Fei Fang, and Tuomas Sandholm. Correlation in extensive-form games: Saddle-point formulation and benchmarks. In *Conference on Neural Information Processing Systems (NeurIPS)*, 2019e.

Gabriele Farina, Andrea Celli, Alberto Marchesi, and Nicola Gatti. Simple uncoupled no-regret learning dynamics for extensive-form correlated equilibrium, 2021a.

Gabriele Farina, Andrea Celli, and Tuomas Sandholm. Efficient decentralized learning dynamics for extensive-form coarse correlated equilibrium: No expensive computation of stationary distributions required. ArXiv preprint, 2021b.

Gabriele Farina, Christian Kroer, and Tuomas Sandholm. Better regularization for sequential decision spaces: Fast convergence rates for Nash, correlated, and team equilibria. In *ACM Conference on Economics and Computation*, 2021c.

Andrew Gilpin and Tuomas Sandholm. Lossless abstraction of imperfect information games. *Journal of the ACM*, 54(5), 2007.

Geoffrey J Gordon, Amy Greenwald, and Casey Marks. No-regret learning in convex games. In *Proceedings of the 25th international conference on Machine learning*, pp. 360–367. ACM, 2008.

Amy Greenwald and Amir Jafari. A general class of no-regret learning algorithms and game-theoretic equilibria. In *Conference on Learning Theory (COLT)*, Washington, D.C., 2003.

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

Samid Hoda, Andrew Gilpin, Javier Peña, and Tuomas Sandholm. Smoothing techniques for computing Nash equilibria of sequential games. *Mathematics of Operations Research*, 35(2), 2010.

Albert Xin Jiang and Kevin Leyton-Brown. Polynomial-time computation of exact correlated equilibrium in compact games. *Games Econ. Behav.*, 91:347–359, 2015.

Daphne Koller, Nimrod Megiddo, and Bernhard von Stengel. Efficient computation of equilibria for extensive two-person games. *Games and Economic Behavior*, 14(2), 1996.

Christian Kroer, Kevin Waugh, Fatma Kılınç-Karzan, and Tuomas Sandholm. Faster algorithms for extensive-form game solving via improved smoothing functions. *Mathematical Programming*, 2020.

Alex Kruckman, Amy Greenwald, and John R. Wicks. An elementary proof of the Markov chain tree theorem. Technical Report 10-04, Brown University, 2010.

H. W. Kuhn. A simplified two-person poker. In H. W. Kuhn and A. W. Tucker (eds.), *Contributions to the Theory of Games*, volume 1 of *Annals of Mathematics Studies, 24*, pp. 97–103. Princeton University Press, Princeton, New Jersey, 1950.

Viliam Lisý, Marc Lanctot, and Michael Bowling. Online Monte Carlo counterfactual regret minimization for search in imperfect information games. In *Autonomous Agents and Multi-Agent Systems*, pp. 27–36, 2015.

Dustin Morrill, Ryan D'Orazio, Marc Lanctot, James R. Wright, Michael Bowling, and Amy R. Greenwald. Efficient deviation types and learning for hindsight rationality in extensive-form games. In Marina Meila and Tong Zhang (eds.), *Proceedings of the 38th International Conference on Machine Learning, ICML 2021*, volume 139 of *Proceedings of Machine Learning Research*, pp. 7818–7828. PMLR, 2021a.

Dustin Morrill, Ryan D'Orazio, Reca Sarfati, Marc Lanctot, James R. Wright, Amy R. Greenwald, and Michael Bowling. Hindsight and sequential rationality of correlated play. In *Thirty-Fifth AAAI Conference on Artificial Intelligence, AAAI 2021*, pp. 5584–5594. AAAI Press, 2021b.

Yurii Nesterov. Excessive gap technique in nonsmooth convex minimization. *SIAM Journal of Optimization*, 16(1), 2005.

Christos H. Papadimitriou and Tim Roughgarden. Computing correlated equilibria in multi-player games. *J. ACM*, 55(3):14:1–14:29, 2008.

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

Alexander Rakhlin and Karthik Sridharan. Optimization, learning, and games with predictable sequences. In *Advances in Neural Information Processing Systems*, pp. 3066–3074, 2013b.

Vasilis Syrgkanis, Alekh Agarwal, Haipeng Luo, and Robert E Schapire. Fast convergence of regularized learning in games. In *Advances in Neural Information Processing Systems*, pp. 2989–2997, 2015.

Oskari Tammelin. Solving large imperfect information games using CFR+. arXiv preprint, 2014.

Oskari Tammelin, Neil Burch, Michael Johanson, and Michael Bowling. Solving heads-up limit Texas hold'em. In *Proceedings of the 24th International Joint Conference on Artificial Intelligence (IJCAI)*, 2015.

Bernhard Von Stengel. Chapter 45 computing equilibria for two-person games. volume 3 of *Handbook of Game Theory with Economic Applications*, pp. 1723–1759. Elsevier, 2002.

Bernhard von Stengel and Françoise Forges. Extensive-form correlated equilibrium: Definition and computational complexity. *Mathematics of Operations Research*, 33(4):1002–1022, 2008.

Martin Zinkevich, Michael Bowling, Michael Johanson, and Carmelo Piccione. Regret minimization in games with incomplete information. In *Proceedings of the Annual Conference on Neural Information Processing Systems (NIPS)*, 2007.