Faster No-Regret Learning Dynamics for Extensive-Form Correlated Equilibrium

Ioannis Anagnostides,[1] **Gabriele Farina,**[1] **Christian Kroer,**[2] **Tuomas Sandholm**[1,3,4,5]

[1] Carnegie Mellon University
[2] Columbia University
[3] Strategy Robot, Inc.
[4] Optimized Markets, Inc.
[5] Strategic Machine, Inc.

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 establish 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, where the $O(\cdot)$ notation suppresses parameters polynomial in the description of the game. This significantly improves over the best prior rate of $O(T^{-1/2})$. To achieve this, we develop a framework for performing accelerated Φ-regret minimization via predictions. To employ our generic template, one of our key technical contributions is to characterize the stability of fixed points associated with trigger deviation functions through a refined perturbation analysis of a structured Markov chain. Finally, experiments on standard benchmarks corroborate our theoretical 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 and 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 and Sandholm 2017; Moravčík et al. 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, Kroer, and Sandholm 2021) 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, Goldberg, and Papadimitriou 2006; Etessami and Yannakakis 2007) and in practice (Berg and 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 and Mas-Colell 2000; Celli et al. 2019). More recently, it was shown that accelerated uncoupled no-regret learning dynamics can lead to CCE in normal-form games at a rate of $\widetilde{O}(T^{-1})$ (Daskalakis, Fishelson, and Golowich 2021), where the notation $\widetilde{O}(\cdot)$ suppresses $\mathrm{polylog}(T)$ factors.

In the context of EFGs, the idea of correlation is much more intricate, and there are several notions of correlated equilibria 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 and Forges 2008) for extensive-form games with *perfect recall*. 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). Moreover, in a follow-up work they also established an explicit rate of convergence of $O(T^{-1/2})$ (Farina et al. 2021). Our paper is concerned with the following fundamental question: *Can we develop faster uncoupled no-regret learning dynamics for* EFCE?

1.1 Contributions

Our primary contribution is to answer this question in the positive:

Theorem 1.1. *On any perfect-recall general-sum multiplayer extensive-form game, there exist uncoupled no-regret learning dynamics which lead to a correlated distribution of play that is an $O(T^{-3/4})$-approximate EFCE. Here 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 and Sridharan 2013b). One of our conceptual contributions is to connect this line of work with the framework of Φ-*regret* minimization of (Greenwald and Jafari 2003; Gordon, Greenwald, and Marks 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 (Gordon, Greenwald, and Marks 2008), including the powerful notion of *swap regret* (Blum and Mansour 2007). Moreover, one of the primary insights behind the result of Farina et al. (2021) is to cast convergence to an EFCE as a Φ-regret minimization problem. Given these prior connections, we believe that our stable-predictive 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 and 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 suitable distance-generating functions (Farina, Kroer, and Sandholm 2019b).

From a technical standpoint, in order to apply our generic template for accelerated Φ-regret minimization, we establish two separate ingredients. First, we develop a *stable-predictive* external regret minimizer for the set of transformations Φ associated with EFCE. This differs from the construction by (Farina et al. 2021) in that we have to additionally guarantee and preserve the stability—and subsequently the predictivity—throughout the construction. The second component consists of sharply characterizing the stability of fixed points of *trigger deviation functions*. This turns out to be particularly challenging, and direct extensions of prior techniques appear to only give 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 and Peng (2020). This leads to a rate of convergence that depends *polynomially* on the description of the game, which is crucial for the applicability of the accelerated dynamics. Finally, we support our theoretical findings with experiments on several general-sum benchmarks.

1.2 Further Related Work

The line of work on accelerated no-regret learning was pioneered by Daskalakis, Deckelbaum, and Kim (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 and 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 an $O(T^{-3/4})$ rate can be obtained for multiplayer general-sum normal-form games. In a recent result, Chen and Peng (2020) strengthened the regret bounds of (Syrgkanis et al. 2015) from external to swap regret using the celebrated construction of Blum and Mansour (2007). Even more recently, Daskalakis, Fishelson, and Golowich (2021) established a near-optimal rate of convergence of $\widetilde{O}(1/T)$ to a coarse correlated equilibrium when all players employ the Optimistic Multiplicative Weights Update (OMWU) algorithm in a normal-form game.

Correlated equilibrium in extensive-form games is much less understood than Nash equilibrium. A feasible EFCE can also be computed efficiently through a variant of the *Ellipsoid algorithm* (Papadimitriou and Roughgarden 2008;

Jiang and Leyton-Brown 2015), and an alternative sampling-based approach was given by Dudík and 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 converge to *normal-form coarse correlated equilibria*, a solution concept much less appealing than EFCE in extensive-form games (Gordon, Greenwald, and Marks 2008). Finally, (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

2.1 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 $z \in \mathcal{Z}$. In this context, the set of nodes in the game tree can be expressed as the (disjoint) union $\mathcal{H} := \bigcup_{i \in [n] \cup \{c\}} \mathcal{H}^{(i)} \cup \mathcal{Z}$.

Imperfect Information. To model imperfect information, the set of decision nodes $\mathcal{H}^{(i)}$ of player i are partitioned into a collection of sets $\mathcal{J}^{(i)}$, which are called *information sets*. Each information set $j \in \mathcal{J}^{(i)}$ groups nodes which cannot be distinguished by 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 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 in-

cluded) 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. 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 be equivalently represented in *sequence-form*:

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

$$\mathcal{Q}^{(i)} := \left\{ \boldsymbol{q} \in \mathbb{R}_{\geq 0}^{|\Sigma^{(i)}|} : \boldsymbol{q}[\varnothing] = 1, \right.$$

$$\left. \boldsymbol{q}[\sigma^{(i)}(j)] = \sum_{a \in \mathcal{A}(j)} \boldsymbol{q}[(j, a)], \quad \forall j \in \mathcal{J}^{(i)} \right\}.$$

This definition ensures the probability mass conservation for the sequence-form strategies along every possible decision point. Observe that the probability of playing action a at information set $j \in \mathcal{J}^{(i)}$ 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 $\Pi_j^{(i)}$. A well-known implication of Kuhn's theorem (Kuhn 1953) is that $\mathcal{Q}^{(i)} = \operatorname{co} \Pi^{(i)}$, and $\mathcal{Q}_j^{(i)} = \operatorname{co} \Pi_j^{(i)}$, for any $i \in [n]$ and $j \in \mathcal{J}^{(i)}$. 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, we overload notation by representing the utility of player $i \in [n]$ under a profile $\boldsymbol{\pi} \in \Pi$ as

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

2.2 Regret and Optimistic Regret Minimization

Consider a convex and compact set $\mathcal{X} \subseteq \mathbb{R}^d$ representing the set 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 linear utility function $\ell^t : \mathcal{X} \ni \boldsymbol{x} \mapsto \langle \ell^t, \boldsymbol{x} \rangle$, and may alter its internal state accordingly.

The error of a regret minimizer is typically measured in terms of *external regret*, defined, for a time horizon T, as follows:

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

Φ-Regret. A conceptual generalization of the concept of external regret is the so-called Φ-*regret*. 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:

$$R^T := \max_{\phi^* \in \Phi} \sum_{t=1}^T \langle \phi^*(\boldsymbol{x}^t), \boldsymbol{\ell}^t \rangle - \sum_{t=1}^T \langle \boldsymbol{x}^t, \boldsymbol{\ell}^t \rangle. \qquad (2)$$

When the set of transformations Φ coincides with the set of *constant* functions, one recovers the notion of external regret given in (1). However, Φ-regret is substantially more expressive and yields more appealing notions of hindsight rationality (Gordon, Greenwald, and Marks 2008), incorporating the notion of *swap regret* (Blum and Mansour 2007).
Optimistic Regret Minimization. We will employ the following definition, which is a slight modification of the RVU property introduced by (Syrgkanis et al. 2015).

Definition 2.2 (Stable-Predictive). *Let $\mathcal{R}$ be a regret 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*

$$\| \boldsymbol{x}^t - \boldsymbol{x}^{t-1} \| \leq \kappa.$$

Moreover, it is said to be (α, β)-predictive with respect to $\| \cdot \|$ if for all $t \geq 1$ its regret R^T satisfies

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

for any sequence of utility vectors $\boldsymbol{\ell}^1, \ldots, \boldsymbol{\ell}^T$, where $\| \cdot \|_*$ is the dual norm of $\| \cdot \|$.

We remark that, for additional flexibility, we have slightly abused the original definition in (Syrgkanis et al. 2015) by allowing $\alpha = \alpha(T)$.
Optimistic Follow the Regularized Leader. Let d be a 1-strongly convex function with respect to a norm $\| \cdot \|$ and $\eta > 0$ be the *learning rate*. OFTRL's update rule takes the following form for $t \geq 2$:

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

where $\boldsymbol{m}^t$ is the *prediction* at time t, and $\boldsymbol{x}^1 := \arg\min_{\boldsymbol{x} \in \mathcal{X}} d(\boldsymbol{x})$. Unless specified otherwise, it will be tacitly assumed that $\boldsymbol{m}^t := \boldsymbol{\ell}^{t-1}$, for $t \geq 1$, where by convention $\boldsymbol{\ell}^0 := \boldsymbol{0}$. The authors in (Syrgkanis et al. 2015) established the following property:

Lemma 2.3. *OFTRL is 3η-stable[1] and $(\Omega_d/\eta, \eta)$-predictive with respect to any norm $\| \cdot \|$ for which d is 1-strongly convex, where Ω_d is the range of d on $\mathcal{X}$, that is, $\Omega_d := \max_{\boldsymbol{x}, \boldsymbol{x}' \in \mathcal{X}} \{d(\boldsymbol{x}) - d(\boldsymbol{x}')\}$.*

---

[1] This assumes that $\|\boldsymbol{m}^t\|_* \leq 1$ and $\|\boldsymbol{u}^t\|_* \leq 1$, for any $t \geq 1$.

Here we consider the *entropic regularizer* with respect to the simplex $d(\boldsymbol{x}) := \sum_{i=1}^d \boldsymbol{x}_i \log \boldsymbol{x}_i$, which is well-known to be 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).[2]

2.3 Extensive-Form Correlated Equilibrium

We will work with the definition of EFCE used in (Farina et al. 2019b), which is equivalent to that of von Stengel and Forges (2008). First, let us introduce the concept of a *trigger deviation function*.

Definition 2.4. Consider some player $i \in [n]$, a sequence $\hat{\sigma} = (j, a) \in \Sigma_*^{(i)}$, and joint 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 and all its descendants is replaced by the behavior specified by the continuation strategy:

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

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

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

Definition 2.5 (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 $\phi_{\hat{\sigma} \to \hat{\boldsymbol{\pi}}}^{(i)} \in \Psi^{(i)}$,*

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

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

Theorem 2.6 ((Farina et al. 2021)). *For every player $i \in [n]$, let $\boldsymbol{\pi}^{(i),1}, \ldots, \boldsymbol{\pi}^{(i),T} \in \Pi^{(i)}$ 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*

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

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

---

[2] When $\boldsymbol{m}^t := \boldsymbol{0}$, for all $t \geq 1$, we recover *Multiplicative Weights Updates (MWU)*.

3 Accelerating Φ-Regret Minimization

In this section we first develop a general template for accelerating Φ-regret minimization. Then, we instantiate our template to obtain accelerated dynamics for EFCE. Our approach combines the framework of Gordon, Greenwald, and Marks (2008) with stable-predictive (aka. optimistic) regret minimization. As in (Gordon, Greenwald, and Marks 2008), our template combines 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, our construction further requires that 2) is stable (in the sense of Definition 2.2). To achieve this, we will focus on regret minimizers having 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 (ϕ^t). 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 $\boldsymbol{x}^{t+1} = \phi^{t+1}(\boldsymbol{x}^{t+1})$ 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 on the regret minimizer).

Given these two components our next theorem builds a stable-predictive Φ-regret minimizer. The ℓ_1 norm is used only for convenience; the theorem readily extends under any equivalent norm.

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 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*

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

where ϵ_t is the error of STABLEFPORACLE at time t, and D_ℓ 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, Greenwald, and Marks 2008), and it is included in Appendix B. As a result, this theorem 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.

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 $\Psi^{(i)}$. The authors in (Farina et al. 2021) 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 3.4. 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 $\hat{\sigma} \in \Sigma_*^{(i)}$. In (Farina et al. 2021) it was 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 $h_{\hat{\sigma}}^{(i)} : \boldsymbol{q} \mapsto \phi_{\hat{\sigma} \to \boldsymbol{q}}^{(i)}$. 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, 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, B)-predictive.*

This construction requires a stable-predictive regret minimizer for the set $\mathcal{Q}_j^{(i)}$, for each $j \in \mathcal{J}^{(i)}$. For this reason, in Appendix A we develop a stable-predictive variant of CFR using OMWU in each local regret minimizer; this follows 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 employ *regret circuits* (Farina, Kroer, and Sandholm 2019c), leading to the main result of this subsection:

Theorem 3.4. *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 utility vectors $\boldsymbol{L}^1, \ldots, \boldsymbol{L}^T$ the regret incurred can be bounded as

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

where $\|\boldsymbol{L}^t\|_{\infty} \leq D_{\boldsymbol{L}}$ for all $t \geq 1$.

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 $\phi \in \mathrm{co}\,\Psi^{(i)}$. All of the proofs of this section are included in Appendix B.2.

Multiplicative Stability. We will say that a sequence $(\boldsymbol{z}^t)$, with $\boldsymbol{z}^t \in \mathbb{R}_{\geq 0}^d$, is said to be κ-*multiplicative-stable*, with $\kappa \in (0,1)$, 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 $t \geq 2$. We will also say that $\boldsymbol{z}_i^t$ and $\boldsymbol{z}_i^{t-1}$ are κ-*multiplicative-close* when $(1-\kappa)\boldsymbol{z}_i^{t-1} \leq \boldsymbol{z}_i^t \leq (1+\kappa)\boldsymbol{z}_i^{t-1}$. Importantly, this strong notion of stability is guaranteed by OMWU (see Lemma B.2). Thus, if $\mathfrak{D}^{(i)}$ is the depth of i's subtree and $D_{\boldsymbol{x}}^{(i)}$ is an upper bound on the ℓ_1 norm of any vector in the sequence-form polytope $\mathcal{Q}^{(i)}$, we can show the following:

Lemma 3.5. *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. 2021), let us introduce the following definitions.

Definition 3.6. 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 3.7. Consider a player $i \in [n]$, a trunk $J \subseteq \mathcal{J}^{(i)}$, and $\phi \in \mathrm{co}\,\Psi^{(i)}$. A vector $\boldsymbol{x} \in \mathbb{R}_{\geq 0}^{|\Sigma^{(i)}|}$ 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 $a \in \mathcal{A}(j)$.

The main question that arises under the stable-predictive framework is whether this fixed point operation can 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 establish polynomial degradation in terms of the stability. More precisely, we first prove the following key lemma:

Lemma 3.8. *Let $\mathbf{M}$ be the transition matrix of an m-state Markov chain such that $\mathbf{M} := \boldsymbol{v}\mathbf{1}^{\top} + \mathbf{C}$, where $\mathbf{C}$ is a matrix with strictly positive entries and columns summing to $1-\lambda$, and $\boldsymbol{v}$ is a vector with strictly positive entries summing to λ. Then, if $\boldsymbol{\pi}$ is the stationary distribution of $\mathbf{M}$, there exists, for each $i \in [m]$, a (non-empty) finite set F_i and $F = \bigcup_i F_i$, and corresponding parameters $b_j \in \{0,1\}, 0 \leq p_j \leq m-2, |S_j| = m - p_j - b_j - 1$, for each $j \in F_j$, such that*

$$\boldsymbol{\pi}_i = \frac{\sum_{j \in F_i} \lambda^{p_j+1}(\boldsymbol{v}[q_j])^{b_j} \prod_{(s,w) \in S_j} \mathbf{C}[(s,w)]}{\sum_{j \in F} C_j \lambda^{p_j+b_j} \prod_{(s,w) \in S_j} \mathbf{C}[(s,w)]},$$

where $C_j = C_j(m)$ is a positive constant.

The main takeaway of this lemma is that the stationary distribution has an *affine* dependence on the vector $\boldsymbol{v}$; note that such a dependence is not at all apparent from the Markov chain tree theorem, and derives from the particular structure of the Markov chain. This will be crucial in order to obtain polynomial bounds in terms of stability. Specifically, using a slight variant of Lemma 3.8 (see Corollary B.8) leads to the following stability bound:

Corollary 3.9. *Let $\mathbf{M}, \mathbf{M}'$ be the 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 $\mathbf{C}$ and $\mathbf{C}'$ are matrices with strictly positive entries, and $\boldsymbol{v}, \boldsymbol{v}'$ are vectors with strictly positive entries such that $\boldsymbol{v} = \boldsymbol{r}/l$ and $\boldsymbol{v}' = \boldsymbol{r}'/l'$, for some $l > 0$ and $l' > 0$. If $\boldsymbol{\pi}$ and $\boldsymbol{\pi}'$ are the stationary distributions of $\mathbf{M}$ and $\mathbf{M}'$, let $\boldsymbol{w} := l\boldsymbol{\pi}$ and $\boldsymbol{w}' := l'\boldsymbol{\pi}'$. Finally, let λ and λ' be the sum of the entries of $\boldsymbol{v}$ and $\boldsymbol{v}'$ respectively. Then, if (i) the matrices $\mathbf{C}$ and $\mathbf{C}'$ are κ-multiplicative-close; (ii) the scalars λ and λ' are κ-multiplicative-close; (iii) the vectors $\boldsymbol{r}$ and $\boldsymbol{r}'$ are γ-multiplicative-close; and (iv) the numbers l and l' are also γ-multiplicative-close, then the vectors $\boldsymbol{w}$ and $\boldsymbol{w}'$ are $(\gamma + O(\kappa m))$-multiplicative-close, for a sufficiently small $\kappa = O(1/m)$.*

Importantly, observe that the "closeness" of $\boldsymbol{w}$ and $\boldsymbol{w}'$ does not scale with $\gamma \cdot m$. Using this bound we manage to show the following:

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

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

Theorem 3.11. *Consider a player $i \in [n]$, and let $\phi = \sum_{\hat{\sigma} \in \Sigma_*^{(i)}} \boldsymbol{\lambda}[\hat{\sigma}]\phi_{\hat{\sigma} \to \boldsymbol{q}_{\hat{\sigma}}}^{(i)}$ be a transformation in $\mathrm{co}\,\Psi^{(i)}$ such*
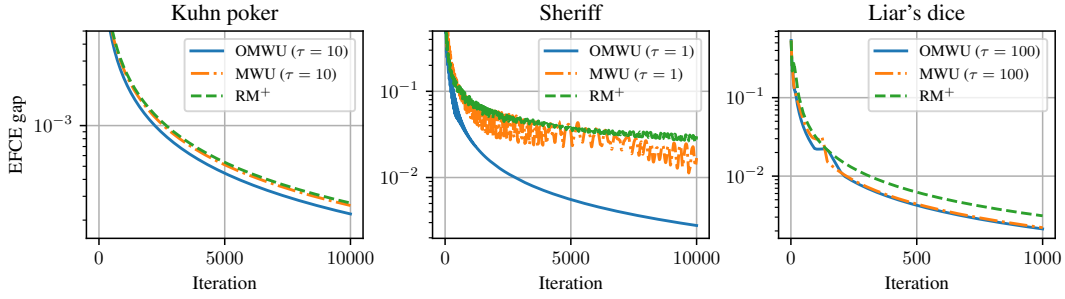
Figure 1: The performance of MWU, OMWU, and RM$^+$ on three general-sum EFGs.

that the sequence of $\boldsymbol{\lambda}^t$'s and $\boldsymbol{q}_{\hat{\sigma}}^t$'s is κ-multiplicative-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 $O(\kappa|\mathcal{A}^{(i)}|\mathfrak{D}^{(i)})$-multiplicative-stable, where $|\mathcal{A}^{(i)}| \coloneqq \max_{j \in \mathcal{J}^{(i)}} |\mathcal{A}(j)|$, for a sufficiently small $\kappa = O(1/(|\mathcal{A}^{(i)}|\mathfrak{D}^{(i)}))$.

Finally, using the stability bounds derived in Lemma 3.5, we arrive at the following conclusion:

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

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

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. (2021) 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.11), 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. 2019b)—a benchmark game introduced specifically for the study of correlated equilibria; and (iii) a three-player version of *Liar's dice* (Lisý, Lanctot, and Bowling 2015). A detailed description of each of the three game instances is available in Appendix D.

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, 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. It should be noted

that one iteration costs the same for every algorithm, up to constant factors. 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, Kroer, and Sandholm 2019a; Burch, Moravcik, and Schmid 2019), but it is unclear if they still guarantee convergence in the EFCE setting.

5 Conclusions

In this paper 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 main technical contributions was 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.

Following recent progress in normal-form games (Daskalakis, Fishelson, and Golowich 2021; Anagnostides et al. 2021), an important question for the future is to obtain a further acceleration of the order $\widetilde{O}(T^{-1})$. We believe that our characterization of the fixed points associated with trigger deviation functions is an important step towards achieving this goal.

References

Anagnostides, I.; Daskalakis, C.; Farina, G.; Fishelson, M.; Golowich, N.; and Sandholm, T. 2021. Near-Optimal No-Regret Learning for Correlated Equilibria in Multi-Player General-Sum Games. *CoRR*, abs/2111.06008.

Aumann, R. 1974. Subjectivity and Correlation in Randomized Strategies. *Journal of Mathematical Economics*, 1: 67–96.

Berg, K.; and Sandholm, T. 2017. Exclusion Method for Finding Nash Equilibrium in Multiplayer Games. In *AAAI Conference on Artificial Intelligence (AAAI)*.

Blum, A.; and Mansour, Y. 2007. From External to Internal Regret. *J. Mach. Learn. Res.*, 8: 1307–1324.

Bowling, M.; Burch, N.; Johanson, M.; and Tammelin, O. 2015. Heads-up Limit Hold'em Poker is Solved. *Science*, 347(6218).

Brown, N.; and Sandholm, T. 2017. Superhuman AI for heads-up no-limit poker: Libratus beats top professionals. *Science*, eaao1733.

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

Celli, A.; Marchesi, A.; Bianchi, T.; and Gatti, N. 2019. Learning to Correlate in Multi-Player General-Sum Sequential Games. In *Proceedings of the Annual Conference on Neural Information Processing Systems (NeurIPS)*, volume 32.

Celli, A.; Marchesi, A.; Farina, G.; and Gatti, N. 2020. No-Regret Learning Dynamics for Extensive-Form Correlated Equilibrium. In Larochelle, H.; Ranzato, M.; Hadsell, R.; Balcan, M.; and Lin, H., eds., *Advances in Neural Information Processing Systems 33: Annual Conference on Neural Information Processing Systems 2020*.

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

Chiang, C.-K.; Yang, T.; Lee, C.-J.; Mahdavi, M.; Lu, C.-J.; Jin, R.; and Zhu, S. 2012. Online optimization with gradual variations. In *Conference on Learning Theory*, 6–1.

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

Daskalakis, C.; Fishelson, M.; and Golowich, N. 2021. Near-Optimal No-Regret Learning in General Games. *CoRR*, abs/2108.06924.

Daskalakis, C.; Goldberg, P.; and Papadimitriou, C. 2006. The Complexity of Computing a Nash Equilibrium. In *Proceedings of the Annual Symposium on Theory of Computing (STOC)*.

Dudík, M.; and Gordon, G. J. 2009. A Sampling-Based Approach to Computing Equilibria in Succinct Extensive-Form Games. In Bilmes, J. A.; and Ng, A. Y., eds., *UAI 2009, Proceedings of the Twenty-Fifth Conference on Uncertainty in Artificial Intelligence, Montreal, QC, Canada, June 18-21, 2009*, 151–160. AUAI Press.

Etessami, K.; and Yannakakis, M. 2007. On the Complexity of Nash Equilibria and Other Fixed Points (Extended Abstract). In *Proceedings of the Annual Symposium on Foundations of Computer Science (FOCS)*, 113–123.

Farina, G.; Celli, A.; Marchesi, A.; and Gatti, N. 2021. Simple Uncoupled No-Regret Learning Dynamics for Extensive-Form Correlated Equilibrium.

Farina, G.; Celli, A.; and Sandholm, T. 2021. Efficient Decentralized Learning Dynamics for Extensive-Form Coarse Correlated Equilibrium: No Expensive Computation of Stationary Distributions Required. ArXiv preprint. arXiv:2109.08138.

Farina, G.; Kroer, C.; Brown, N.; and Sandholm, T. 2019a. Stable-Predictive Optimistic Counterfactual Regret Minimization. In *International Conference on Machine Learning (ICML)*.

Farina, G.; Kroer, C.; and Sandholm, T. 2019a. Online Convex Optimization for Sequential Decision Processes and Extensive-Form Games. In *AAAI Conference on Artificial Intelligence*.

Farina, G.; Kroer, C.; and Sandholm, T. 2019b. Optimistic Regret Minimization for Extensive-Form Games via Dilated Distance-Generating Functions. In *Advances in Neural Information Processing Systems, NeurIPS 2019,*, 5222–5232.

Farina, G.; Kroer, C.; and Sandholm, T. 2019c. Regret Circuits: Composability of Regret Minimizers. In *International Conference on Machine Learning*, 1863–1872.

Farina, G.; Kroer, C.; and Sandholm, T. 2021. Better Regularization for Sequential Decision Spaces: Fast Convergence Rates for Nash, Correlated, and Team Equilibria. In *ACM Conference on Economics and Computation*.

Farina, G.; Ling, C. K.; Fang, F.; and Sandholm, T. 2019b. Correlation in Extensive-Form Games: Saddle-Point Formulation and Benchmarks. In *Conference on Neural Information Processing Systems (NeurIPS)*.

Gilpin, A.; and Sandholm, T. 2007. Lossless Abstraction of Imperfect Information Games. *Journal of the ACM*, 54(5).

Gordon, G. J.; Greenwald, A.; and Marks, C. 2008. No-regret learning in convex games. In *Proceedings of the 25th international conference on Machine learning*, 360–367. ACM.

Greenwald, A.; and Jafari, A. 2003. A General Class of No-Regret Learning Algorithms and Game-Theoretic Equilibria. In *Conference on Learning Theory (COLT)*. Washington, D.C.

Hart, S.; and Mas-Colell, A. 2000. A Simple Adaptive Procedure Leading to Correlated Equilibrium. *Econometrica*, 68: 1127–1150.

Hoda, S.; Gilpin, A.; Peña, J.; and Sandholm, T. 2010. Smoothing Techniques for Computing Nash Equilibria of Sequential Games. *Mathematics of Operations Research*, 35(2).

Jiang, A. X.; and Leyton-Brown, K. 2015. Polynomial-time computation of exact correlated equilibrium in compact games. *Games Econ. Behav.*, 91: 347–359.

Kroer, C.; Waugh, K.; Kılınç-Karzan, F.; and Sandholm, T. 2020. Faster algorithms for extensive-form game solving via improved smoothing functions. *Mathematical Programming*.

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

Kuhn, H. W. 1950. A Simplified Two-Person Poker. In Kuhn, H. W.; and Tucker, A. W., eds., *Contributions to the Theory of Games*, volume 1 of *Annals of Mathematics Studies, 24*, 97–103. Princeton, New Jersey: Princeton University Press.

Kuhn, H. W. 1953. Extensive Games and the Problem of Information. In Kuhn, H. W.; and Tucker, A. W., eds., *Contributions to the Theory of Games*, volume 2 of *Annals of Mathematics Studies, 28*, 193–216. Princeton, NJ: Princeton University Press.

Lisý, V.; Lanctot, M.; and Bowling, M. 2015. Online Monte Carlo Counterfactual Regret Minimization for Search in Imperfect Information Games. In *Autonomous Agents and Multi-Agent Systems*, 27–36.

Moravčík, M.; Schmid, M.; Burch, N.; Lisý, V.; Morrill, D.; Bard, N.; Davis, T.; Waugh, K.; Johanson, M.; and Bowling, M. 2017. DeepStack: Expert-level artificial intelligence in heads-up no-limit poker. *Science*.

Morrill, D.; D'Orazio, R.; Lanctot, M.; Wright, J. R.; Bowling, M.; and Greenwald, A. R. 2021a. Efficient Deviation Types and Learning for Hindsight Rationality in Extensive-Form Games. In Meila, M.; and Zhang, T., eds., *Proceedings of the 38th International Conference on Machine Learning, ICML 2021*, volume 139 of *Proceedings of Machine Learning Research*, 7818–7828. PMLR.

Morrill, D.; D'Orazio, R.; Sarfati, R.; Lanctot, M.; Wright, J. R.; Greenwald, A. R.; and Bowling, M. 2021b. Hindsight and Sequential Rationality of Correlated Play. In *Thirty-Fifth AAAI Conference on Artificial Intelligence, AAAI 2021*, 5584–5594. AAAI Press.

Nesterov, Y. 2005. Excessive Gap Technique in Nonsmooth Convex Minimization. *SIAM Journal of Optimization*, 16(1).

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

Rakhlin, A.; and Sridharan, K. 2013a. Online Learning with Predictable Sequences. In *Conference on Learning Theory*, 993–1019.

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

Syrgkanis, V.; Agarwal, A.; Luo, H.; and Schapire, R. E. 2015. Fast convergence of regularized learning in games. In *Advances in Neural Information Processing Systems*, 2989–2997.

Tammelin, O. 2014. Solving Large Imperfect Information Games Using CFR+. arXiv preprint. arXiv:1407.5042.

Tammelin, O.; Burch, N.; Johanson, M.; and Bowling, M. 2015. Solving Heads-up Limit Texas Hold'em. In *Proceedings of the 24th International Joint Conference on Artificial Intelligence (IJCAI)*.

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

Zinkevich, M.; Bowling, M.; Johanson, M.; and Piccione, C. 2007. Regret Minimization in Games with Incomplete Information. In *Proceedings of the Annual Conference on Neural Information Processing Systems (NIPS)*.

A Sequential Decision Making and Stable-Predictive CFR

The main purpose of this section is to provide a stable-predictive variant of CFR following the construction of (Farina et al. 2019a). The main result is given in Theorem A.4. We begin by introducing the basic setting of *sequential decision making*.

A sequential decision process can be represented using a tree consisting of two types of nodes: *decision nodes* and *observation nodes*. The set of all decision nodes will be denoted by $\mathcal{J}$, while the set of observation nodes by $\mathcal{K}$. At every decision node $j \in \mathcal{J}$ the agent has to select a strategy $\boldsymbol{x}_j$ in the form of a probability distribution over all possible actions $\mathcal{A}(j)$. On the other hand, the agent receives some type of feedback at each observation point $k \in \mathcal{K}$. More precisely, it is assumed that the agent may receive a signal from a set $\mathcal{S}(k)$. At every decision point $j \in \mathcal{J}$ of the sequential decision process, the strategy $\boldsymbol{x}_j \in \Delta^{|\mathcal{A}(j)|}$ secures a (linear) utility of the form $\langle \boldsymbol{\ell}_j, \boldsymbol{x}_j \rangle$, for some utility vector $\boldsymbol{\ell}_j$. The expected utility throughout the entire decision process can be expressed as $\sum_{j \in \mathcal{J}} \pi_j \langle \boldsymbol{\ell}_j, \boldsymbol{x}_j \rangle$, where π_j is the probability that the agent reaches the decision point j. Before we proceed with the representation of the strategies, it is important to point out that in all extensive-form games of perfect recall the agents face a sequential decision process.

Decomposition of Sequence-Form Space. Our following construction will rely on a recursive decomposition of the sequence-form space $\mathcal{X}^\triangle$:

- Consider an observation node $k \in \mathcal{K}$, and let $\mathcal{C}_k$ be the children decision points of k. Then, $\mathcal{X}_k^\triangle$ can be decomposed as the following Cartesian product:

$$\mathcal{X}_k^\triangle := \bigtimes_{j \in \mathcal{C}_k} \mathcal{X}_j^\triangle; \tag{3}$$

- Consider a decision point $j \in \mathcal{J}$, and let $\mathcal{C}_j = \{k_1, \ldots, k_{n_j}\}$ be the children observation points of j, with $n_j = |\mathcal{A}(j)|$. Then, $\mathcal{X}_j^\triangle$ can be decomposed as follows:

$$\mathcal{X}_j^\triangle := \left\{ \begin{pmatrix} \boldsymbol{\lambda}_1 \\ \vdots \\ \boldsymbol{\lambda}_{n_j} \\ \boldsymbol{\lambda}_1 \boldsymbol{x}_1 \\ \vdots \\ \boldsymbol{\lambda}_{n_j} \boldsymbol{x}_{n_j} \end{pmatrix} : (\boldsymbol{\lambda}_1, \ldots, \boldsymbol{\lambda}_{n_j}) \in \Delta^{n_j}, \boldsymbol{x}_1 \in \mathcal{X}_{k_1}^\triangle, \ldots, \boldsymbol{x}_{n_j} \in \mathcal{X}_{k_{n_j}}^\triangle \right\}. \tag{4}$$

In view of this decomposition, our regret minimizer for the sequence-form strategy space $\mathcal{X}^\triangle$ will be established based on *local* regret minimizers. Specifically, the basic ingredients for the overall construction are given in Proposition A.1 and Proposition A.2. We should note that in the sequel the stability and the predictive bounds will be tacitly assumed with respect to the pair of norms $(\| \cdot \|_1, \| \cdot \|_\infty)$.

Proposition A.1. *Consider an observation node $k \in \mathcal{K}$, and assume access to a κ_j-multiplicatively-stable (α_j, β_j)-predictive regret minimizer $\mathcal{R}_j^\triangle$ over the sequence-form strategy space $\mathcal{X}_j^\triangle$, for each $j \in \mathcal{C}_k$. Then, we can construct a $\max_j\{\kappa_j\}$-multiplicative-stable (A, B)-predictive regret minimizer $\mathcal{R}_k^\triangle$ for the sequence-form strategy space $\mathcal{X}_k^\triangle$, where $A = \sum_{j \in \mathcal{C}_k} \alpha_j$ and $B = \sum_{j \in \mathcal{C}_k} \beta_j$.*

Proof. Given the decomposition of (3), the composite regret minimizer can be constructed based on the regret circuit for the Cartesian product given in (Farina, Kroer, and Sandholm 2019c). In particular, it is direct to verify that $R_k^{\triangle,T} = \sum_{j \in \mathcal{C}_k} R_j^{\triangle,T}$, where $R_k^{\triangle,T}$ is the regret accumulated by the composite regret minimizer, and $R_j^{\triangle,T}$ the cumulative regret of each individual regret minimizer. In particular, by assumption we know that

$$R_j^{\triangle,T} \leq \alpha_j + \beta_j \sum_{t=1}^T \|\boldsymbol{\ell}_j^{\triangle,t} - \boldsymbol{\ell}_j^{\triangle,t-1}\|_\infty^2.$$

As a result, we can conclude that

$$R_k^{\triangle,T} \leq \left(\sum_{j \in \mathcal{C}_k} \alpha_j \right) + \left(\sum_{j \in \mathcal{C}_k} \beta_j \right) \sum_{t=1}^T \|\boldsymbol{\ell}_k^{\triangle,t} - \boldsymbol{\ell}_k^{\triangle,t-1}\|_\infty^2,$$

where we used that $\|\boldsymbol{\ell}_j^{\triangle,t} - \boldsymbol{\ell}_j^{\triangle,t-1}\|_\infty \leq \|\boldsymbol{\ell}_k^{\triangle,t} - \boldsymbol{\ell}_k^{\triangle,t-1}\|_\infty$. Finally, the $\max_j\{\kappa_j\}$-multiplicative-stability of $\mathcal{R}_k^\triangle$ follows directly from the κ_j-multiplicative-stability of each $\mathcal{R}_j^\triangle$. $\square$

Proposition A.2. *Consider a decision node $j \in \mathcal{J}$, and assume access to a K-multiplicative-stable (α_k, β_k)-predictive regret minimizer $\mathcal{R}_k^{\triangle}$ over the sequence-form strategy space $\mathcal{X}_k^{\triangle}$, for each $k \in \mathcal{C}_j$. Moreover, assume access to a κ-multiplicatively-stable (α, β)-predictive regret minimizer $\mathcal{R}_{\Delta}$ over the simplex $\Delta^{|\mathcal{A}(j)|}$. Then, for a sufficiently small constant κ, we can construct a $(2\kappa + K)$-multiplicatively-stable (A, B)-predictive regret minimizer $\mathcal{R}_j^{\triangle}$ for the sequence-form strategy space $\mathcal{X}_j^{\triangle}$, where*

$$A(T) = \alpha + \max_{k \in \mathcal{C}_j}\{\alpha_k\} + 2\beta D_{\boldsymbol{\ell}}^2 D_{\boldsymbol{x}}^2 K^2 T;$$

$$B = \max_{k \in \mathcal{C}_j}\{\beta_k\} + 2\beta D_{\boldsymbol{x}}^2,$$

with $D_{\boldsymbol{\ell}}$ being an upper bound on the ℓ_{∞} norm of the sequence $(\boldsymbol{\ell}^{\triangle, t})$, and $D_{\boldsymbol{x}}$ an upper bound on the ℓ_1 norm of all $\boldsymbol{x} \in \mathcal{X}^{\triangle}$.

Proof. For this construction we will use the regret circuit for the convex hull, stated in Proposition B.1. In particular, first note that—by assumption—the regret $R_k^{\triangle, T}$ accumulated by each regret minimizer $\mathcal{R}_k^{\triangle}$ can be bounded as

$$R_k^{\triangle, T} \leq \alpha_k + \beta_k \sum_{t=1}^{T} \|\boldsymbol{\ell}_k^{\triangle, T} - \boldsymbol{\ell}_k^{\triangle, t-1}\|_{\infty}^2.$$

Moreover, by construction, each regret minimizer $\mathcal{R}_k^{\triangle}$ receives the same utility as $\mathcal{R}_j^{\triangle}$; this, along with the guarantee of Proposition B.1, imply that

$$R_j^{\triangle, T} \leq \alpha + \max_{k \in \mathcal{C}_j}\{\alpha_k\} + \max_{k \in \mathcal{C}_j}\{\beta_k\} \sum_{t=1}^{T} \|\boldsymbol{\ell}_j^{\triangle, t} - \boldsymbol{\ell}_j^{\triangle, t-1}\|_{\infty}^2 + \beta \sum_{t=1}^{T} \|\boldsymbol{\Lambda}^t - \boldsymbol{\Lambda}^{t-1}\|_{\infty}^2, \tag{5}$$

where $\boldsymbol{\Lambda}^t$ represents the utility function received as input by $\mathcal{R}_{\Delta}$. Next, similarly to the analysis of Theorem 3.4, we can deduce that for some $k \in \mathcal{C}_j$,

$$\|\boldsymbol{\Lambda}^t - \boldsymbol{\Lambda}^{t-1}\|_{\infty}^2 \leq 2\|\boldsymbol{\ell}_j^{\triangle, t} - \boldsymbol{\ell}_j^{\triangle, t-1}\|_{\infty}^2 \|\boldsymbol{x}_k^t\|_1^2 + 2\|\boldsymbol{\ell}_j^{\triangle, t-1}\|_{\infty}^2 \|\boldsymbol{x}_k^t - \boldsymbol{x}_k^{t-1}\|_1^2$$
$$\leq 2D_{\boldsymbol{x}}^2 \|\boldsymbol{\ell}_j^{\triangle, t} - \boldsymbol{\ell}_j^{\triangle, t-1}\|_{\infty}^2 + 2D_{\boldsymbol{\ell}}^2 D_{\boldsymbol{x}}^2 K^2,$$

where we used that $\|\boldsymbol{x}_k^t\|_1 \leq D_{\boldsymbol{x}}$, and the fact that $\|\boldsymbol{x}_k^t - \boldsymbol{x}_k^{t-1}\|_1 \leq K \|\boldsymbol{x}_k^{t-1}\|_1 \leq K D_{\boldsymbol{x}}$ by K-multiplicative-stability of the sequence $(\boldsymbol{x}_k^t)$. As a result, if we plug-in this bound to (5) we can conclude that

$$R_j^{\triangle, T} \leq \left(\alpha + \max_{k \in \mathcal{C}_j}\{\alpha_k\} + 2\beta D_{\boldsymbol{\ell}}^2 D_{\boldsymbol{x}}^2 K^2 T\right) + \left(\max_{k \in \mathcal{C}_j}\{\beta_k\} + 2\beta D_{\boldsymbol{x}}^2\right) \sum_{t=1}^{T} \|\boldsymbol{\ell}_j^{\triangle, t} - \boldsymbol{\ell}_j^{\triangle, t-1}\|_{\infty}^2.$$

Finally, the $(2\kappa + K)$-multiplicative-stability of $\mathcal{R}_j^{\triangle}$ can be directly verified from the decomposition given in (4). □

Remark A.3. Given the decomposition provided in Equation (4), the regret circuit for the convex hull should operate on the appropriate "lifted" subspace for each $\mathcal{X}_k^{\triangle}$, which does not essentially alter the analysis of the regret since the augmented entries remain invariant; see (Farina, Kroer, and Sandholm 2019c, Figure 7).

Finally, we can inductively combine Proposition A.1 and Proposition A.2 in order to establish a stable-predictive variant of CFR:

Theorem A.4 (Optimistic CFR). *If every local regret minimizer $\mathcal{R}_j^{\triangle}$ is updated using OMWU with a sufficiently small learning rate η, for each $j \in \mathcal{J}$, we can construct an $O(\eta \mathfrak{D} D_{\boldsymbol{x}}^2 D_{\boldsymbol{\ell}})$-stable (A, B)-predictive regret minimizer $\mathcal{R}^{\triangle}$ for the space of sequence-form strategies $\mathcal{X}^{\triangle}$, such that*

$$A(T) = O\left(\frac{\log |\mathcal{A}|}{\eta} |\mathcal{J}| + \eta^3 D_{\boldsymbol{\ell}}^4 D_{\boldsymbol{x}}^6 \mathfrak{D}^2 T |\mathcal{J}|\right);$$

$$B = O(\eta D_{\boldsymbol{x}}^2 |\mathcal{J}|), \tag{6}$$

where $|\mathcal{A}| \coloneqq \max_{j \in \mathcal{J}} |\mathcal{A}(j)|$, $D_{\boldsymbol{\ell}}$ is an upper bound on the ℓ_{∞} norm of the utility functions, $D_{\boldsymbol{x}}$ is an upper bound on the ℓ_1 norm of all $\boldsymbol{x} \in \mathcal{X}^{\triangle}$, and $\mathfrak{D}$ is the depth of the decision process. Moreover, $\mathcal{R}^{\triangle}$ is $O(\eta \mathfrak{D} D_{\boldsymbol{x}} D_{\boldsymbol{\ell}})$-multiplicatively-stable.

Proof. First of all, it is easy to see that all the (local) counterfactual losses (*e.g.*, see (Farina et al. 2019a, Section 4)) have ℓ_{∞} norm bounded by $O(D_{\boldsymbol{x}} D_{\boldsymbol{\ell}})$, where recall that $D_{\boldsymbol{x}}$ is an upper bound on the ℓ_1 norm of all $\boldsymbol{x} \in \mathcal{X}^{\triangle}$. As a result, we know from Lemma B.2 that the output of each local regret minimizer $\mathcal{R}_j^{\triangle}$ under OMWU with learning rate η is $O(\eta D_{\boldsymbol{x}} D_{\boldsymbol{\ell}})$-multiplicative-stable. Along with Proposition A.2, we can inductively infer that the output of $\mathcal{R}$ is $O(\eta \mathfrak{D} D_{\boldsymbol{x}} D_{\boldsymbol{\ell}})$-multiplicative-stable, where

$\mathfrak{D}$ is the depth of the decision process. Moreover, we can conclude from this property that $\mathcal{R}$ is $O(\eta \mathfrak{D} D_{\boldsymbol{x}}^2 D_{\boldsymbol{\ell}})$-stable (in the additive sense of Definition 2.2) given that the ℓ_1 norm of $\boldsymbol{x} \in \mathcal{X}^{\triangle}$ is bounded by $D_{\boldsymbol{x}}$.

For the predictive bound, first recall that the range of the entropic regularizer on the m-dimensional simplex is $\log m$. Thus, by Lemma 2.3 we know that each local regret minimizer at the information set $j \in \mathcal{J}$ instantiated with OMWU with learning rate η will be $(\log(|A_j|/\eta, \eta)$-predictive. Moreover, we previously argued that the stability parameter K appearing in the regret bound of Proposition A.2 will be bounded by $O(\eta \mathfrak{D} D_{\boldsymbol{x}}^2 D_{\boldsymbol{\ell}})$. As a result, our predictivity bound in (6) follows inductively from Proposition A.2. $\qquad \square$

Naturally, the same bounds apply for constructing a regret minimizer for the subspace $\mathcal{X}_j^{\triangle}$, for any $j \in \mathcal{J}$, as required in Proposition 3.3.

B Omitted Proofs

In this section we include all of the omitted proofs from the main body. First, let us introduce some additional useful notation. It will be convenient to instantiate a trigger deviation function (recall Definition 2.4) in the form of a linear mapping $\phi_{\hat{\sigma} \to \hat{\boldsymbol{\pi}}}^{(i)}$: $\mathbb{R}^{|\Sigma^{(i)}|} \ni \boldsymbol{x} \mapsto \mathbf{M}_{\hat{\sigma} \to \hat{\boldsymbol{\pi}}}^{(i)} \boldsymbol{x}$, where $\mathbf{M}_{\hat{\sigma} \to \hat{\boldsymbol{\pi}}}^{(i)}$ is such that for any $\sigma_r, \sigma_c \in \Sigma^{(i)}$,

$$\mathbf{M}_{\hat{\sigma} \to \hat{\boldsymbol{\pi}}}^{(i)}[\sigma_r, \sigma_c] = \begin{cases} 1 & \text{if } \sigma_c \not\succeq \hat{\sigma} \quad \& \quad \sigma_r = \sigma_c; \\ \hat{\boldsymbol{\pi}}[\sigma_r] & \text{if } \sigma_c = \hat{\sigma} \quad \& \quad \sigma_r \succeq j; \\ 0 & \text{otherwise}, \end{cases} \tag{7}$$

where $\hat{\sigma} = (j, a) \in \Sigma_*^{(i)}$. It is not hard to show that the linear mapping described in (7) is indeed a trigger deviation function in the sense of Definition 2.4. We will also sometimes use the notation $\boldsymbol{x} \otimes \boldsymbol{y} = \boldsymbol{x} \boldsymbol{y}^\top$ to denote the *outer product* of vectors $\boldsymbol{x}$ and $\boldsymbol{y}$. Moreover, we will write $(\mathbf{M})^\flat$ to represent the *vectorization* of matrix $\mathbf{M}$.

We commence with the proof of Theorem 3.2. For the convenience of the reader the theorem is restated below.

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 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*

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

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 $\boldsymbol{x} \in \mathcal{X}$.*

Proof. Fix some iteration $t \geq 2$. The first step is to obtain the next strategy of $\mathcal{R}_\Phi$: $\phi^t = \mathcal{R}_\Phi.\text{NEXTSTRATEGY}()$. Then, our regret minimizer $\mathcal{R}$ will simply output the strategy $\boldsymbol{x}^t$ such that $\boldsymbol{x}^t = \text{STABLEFPORACLE}(\phi^t; \boldsymbol{x}^{t-1}, G, \epsilon_t)$.[3] By assumption (recall Definition 3.1) we know that this is indeed well-defined, and $\boldsymbol{x}^t$ will be such that (i) $\|\phi^t(\boldsymbol{x}^t) - \boldsymbol{x}^t\| \leq \epsilon_t$, and (ii) $\|\boldsymbol{x}^t - \boldsymbol{x}^{t-1}\| \leq G$. This immediately implies that $\mathcal{R}$ will be G-stable.

Afterwards, we receive feedback from the environment in the form of a utility vector $\boldsymbol{\ell}^t$. This is used to construct the utility function $L^t : \phi \mapsto \langle \boldsymbol{\ell}^t, \phi(\boldsymbol{x}^t) \rangle$; we will let $\boldsymbol{L}^t = (\boldsymbol{\ell}^t \otimes \boldsymbol{x}^t)^\flat$ be the corresponding utility vector. Then, this function is given as feedback to $\mathcal{R}_\Phi$; that is, we invoke the subroutine $\mathcal{R}_\Phi.\text{OBSERVEUTILITY}(L^t)$. As a result, the (external) regret of $\mathcal{R}_\Phi$ can be expressed as

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

In particular, if R^T is the Φ-regret of $\mathcal{R}$, we have that

$$R^T - R_\Phi^T = \sum_{t=1}^{T} \langle \boldsymbol{\ell}^t, \phi^t(\boldsymbol{x}^t) \rangle - \sum_{t=1}^{T} \langle \boldsymbol{\ell}^t, \boldsymbol{x}^t \rangle = \sum_{t=1}^{T} \langle \boldsymbol{\ell}^t, \phi^t(\boldsymbol{x}^t) - \boldsymbol{x}^t \rangle$$

$$\leq \sum_{t=1}^{T} \|\boldsymbol{\ell}^t\|_* \|\phi^t(\boldsymbol{x}^t) - \boldsymbol{x}^t\| \leq D_{\boldsymbol{\ell}} \sum_{t=1}^{T} \epsilon_t, \tag{8}$$

where we used the Cauchy-Schwarz inequality, as well as the fact that $\|\phi^t(\boldsymbol{x}^t) - \boldsymbol{x}^t\| \leq \epsilon_t$. Next, we will bound the term $\|\boldsymbol{L}^t - \boldsymbol{L}^{t-1}\|_\infty$ in terms of $\|\boldsymbol{\ell}^t - \boldsymbol{\ell}^{t-1}\|_\infty$. To this end, it follows that

---

[3]For $t = 1$ it suffices to return any $\boldsymbol{x}^t$ such that $\boldsymbol{x}^1 = \phi^1(\boldsymbol{x}^1)$.

$$\|\boldsymbol{L}^t - \boldsymbol{L}^{t-1}\|_\infty^2 = \|(\boldsymbol{\ell}^t \otimes \boldsymbol{x}^t)^\flat - (\boldsymbol{\ell}^{t-1} \otimes \boldsymbol{x}^{t-1})^\flat\|_\infty^2$$

$$= \|(\boldsymbol{\ell}^t \otimes \boldsymbol{x}^t)^\flat - (\boldsymbol{\ell}^{t-1} \otimes \boldsymbol{x}^t)^\flat + (\boldsymbol{\ell}^{t-1} \otimes \boldsymbol{x}^t)^\flat - (\boldsymbol{\ell}^{t-1} \otimes \boldsymbol{x}^{t-1})^\flat\|_\infty^2$$

$$= \|((\boldsymbol{\ell}^t - \boldsymbol{\ell}^{t-1}) \otimes \boldsymbol{x}^t)^\flat + (\boldsymbol{\ell}^{t-1} \otimes (\boldsymbol{x}^t - \boldsymbol{x}^{t-1}))^\flat\|_\infty^2$$

$$\leq 2\|((\boldsymbol{\ell}^t - \boldsymbol{\ell}^{t-1}) \otimes \boldsymbol{x}^t)^\flat\|_\infty^2 + 2\|(\boldsymbol{\ell}^{t-1} \otimes (\boldsymbol{x}^t - \boldsymbol{x}^{t-1}))^\flat\|_\infty^2 \tag{9}$$

$$= 2\|\boldsymbol{\ell}^t - \boldsymbol{\ell}^{t-1}\|_\infty^2\|\boldsymbol{x}^t\|_\infty^2 + 2\|\boldsymbol{\ell}^{t-1}\|_\infty^2\|\boldsymbol{x}^t - \boldsymbol{x}^{t-1}\|_\infty^2 \tag{10}$$

$$\leq 2\|\boldsymbol{\ell}^t - \boldsymbol{\ell}^{t-1}\|_\infty^2 + 2D_{\boldsymbol{\ell}}^2\kappa^2, \tag{11}$$

where we used the triangle inequality together with Young's inequality in (9), the property that $\|(\boldsymbol{w} \otimes \boldsymbol{z})^\flat\|_\infty = \|\boldsymbol{w}\|_\infty\|\boldsymbol{z}\|_\infty$ in (10), and the stability property $\|\boldsymbol{x}^t - \boldsymbol{x}^{t-1}\|_\infty \leq \|\boldsymbol{x}^t - \boldsymbol{x}^{t-1}\|_1 \leq \kappa$ in (11). As a result, if we plug-in (11) to (8) and we use the (α, β)-predictivity of $\mathcal{R}_\Phi$ we can conclude that

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

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

concluding the proof. $\qquad\square$

B.1 Proofs for Section 3.1

In this section we present all the omitted proofs from Section 3.1. We commence with the proof of Proposition 3.3, established via Algorithm 2. We note that a similar construction appears in (Farina et al. 2021).

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, 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, B)-predictive.*

Proof. As suggested in Algorithm 1, let us consider the linear function $g_{\hat{\sigma}}^{(i),t} : \mathbb{R}^{|\Sigma_j^{(i)}|} \ni \boldsymbol{x} \mapsto L^t(h_{\hat{\sigma}}^{(i)}(\boldsymbol{x})) - L^t(h_{\hat{\sigma}}^{(i)}(\mathbf{0}))$, with $\boldsymbol{g}_{\hat{\sigma}}^t \in \mathbb{R}^{|\Sigma_j^{(i)}|}$ being the corresponding canonical vector of this function. In Algorithm 1 the observed utility function L^t at time t is used to construct $g_{\hat{\sigma}}^{(i),t}$, and the latter linear function is given as input to $\mathcal{R}_{\mathcal{Q}}^{(i)}$. As a result, it follows that

$$\sup_{\phi^* \in \Psi_{\hat{\sigma}}^{(i)}} \sum_{t=1}^T L^t(\phi^*) - \sum_{t=1}^T L^t\left(\phi_{\hat{\sigma} \to \boldsymbol{q}_{\hat{\sigma}}^t}^{(i)}\right) = \sup_{\boldsymbol{q}_{\hat{\sigma}}^* \in \mathcal{Q}_j^{(i)}} \sum_{t=1}^T g_{\hat{\sigma}}^{(i),t}(\boldsymbol{q}_{\hat{\sigma}}^*) - \sum_{t=1}^T g_{\hat{\sigma}}^{(i),t}(\boldsymbol{q}_{\hat{\sigma}}^t).$$

That is, the cumulative regret incurred by $\mathcal{R}_{\hat{\sigma}}^{(i)}$ under the sequence of utility functions $L^1, \ldots, L^T$ is equal to the regret incurred by $\mathcal{R}_{\mathcal{Q}}^{(i)}$ under the sequence of utility functions $g_{\hat{\sigma}}^{(i),t}$. As a result, if we invoke the (A, B)-predictive assumption for the regret minimizer $\mathcal{R}_{\mathcal{Q}}^{(i)}$, we can infer that the cumulative regret R^T of $\mathcal{R}_{\hat{\sigma}}^{(i)}$ can be bounded as

$$R^T \leq A(T) + B \sum_{t=1}^T \|\boldsymbol{g}_{\hat{\sigma}}^{(i),t} - \boldsymbol{g}_{\hat{\sigma}}^{(i),t-1}\|_\infty^2 \leq A(T) + B \sum_{t=1}^T \|\boldsymbol{L}^t - \boldsymbol{L}^{t-1}\|_\infty^2,$$

where the bound follows from the fact that $\boldsymbol{g}_{\hat{\sigma}}^{(i),t} = (\boldsymbol{L}^t[\sigma_r, \hat{\sigma}])_{\sigma_r \succeq j}$. It is also direct to verify that the stability is preserved since $\|(\mathbf{M}_{\hat{\sigma} \to \boldsymbol{q}_{\hat{\sigma}}^t}^{(i)})^\flat - (\mathbf{M}_{\hat{\sigma} \to \boldsymbol{q}_{\hat{\sigma}}^{t-1}}^{(i)})^\flat\|_1 = \|\boldsymbol{q}_{\hat{\sigma}}^t - \boldsymbol{q}_{\hat{\sigma}}^{t-1}\|_1 \leq K$, by K-stability of $\mathcal{R}_{\mathcal{Q}}^{(i)}$. Finally, the claim regarding the complexity of Algorithm 1 follows directly given that we can store the vector $\boldsymbol{g}_{\hat{\sigma}}^{(i),t}$ in $O(|\Sigma_j^{(i)}|)$ time. $\qquad\square$

Next, we conclude the construction by combining the individual regret minimizers using a regret circuit for the convex hull, leading to Theorem 3.4; the statement of the theorem is included below for convenience.

Theorem 3.4. *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*

Algorithm 1: Stable-Predictive Regret Minimizer $\mathcal{R}_{\hat{\sigma}}^{(i)}$ for the set $\Psi_{\hat{\sigma}}^{(i)}$

Input:
- Player $i \in [n]$
- A trigger sequence $\hat{\sigma} = (j, a) \in \Sigma_*^{(i)}$
- A (deterministic) K-stable (A_T, B)-predictive regret minimizer $\mathcal{R}_{\mathcal{Q}}^{(i)}$ for $\mathcal{Q}_j^{(i)}$

1   **function** NEXTSTRATEGY()**:**
2     $\boldsymbol{q}_{\hat{\sigma}}^t \leftarrow \mathcal{R}_{\mathcal{Q}}^{(i)}.\text{NEXTSTRATEGY}()$
3   **return** $\phi_{\hat{\sigma} \leftarrow \boldsymbol{q}_{\hat{\sigma}}^t}^{(i)}$

---

4   **function** OBSERVEUTILITY(L^t)**:**
5   Construct the linear function $g_{\hat{\sigma}}^{(i),t} : \mathbb{R}^{|\Sigma_j^{(i)}|} \ni \boldsymbol{x} \mapsto L^t(h_{\hat{\sigma}}^{(i)}(\boldsymbol{x})) - L^t(h_{\hat{\sigma}}^{(i)}(\boldsymbol{0}))$
6     $\mathcal{R}_{\mathcal{Q}}^{(i)}.\text{OBSERVEUTILITY}(g_{\hat{\sigma}}^{(i),t})$

---

an algorithm which constructs a regret minimizer $\mathcal{R}_{\Psi}^{(i)}$ for the set co $\Psi^{(i)}$ *such that (i) $\mathcal{R}_{\Psi}^{(i)}$ is $O(K + |\Sigma^{(i)}|\kappa)$-stable, and (ii) under any sequence of utility vectors $\boldsymbol{L}^1, \ldots, \boldsymbol{L}^T$ the regret incurred can be bounded as*

$$R_{\Psi}^T \leq O(\alpha(T) + A(T) + \beta D_{\boldsymbol{L}}^2 K^2 T)$$

$$+ O(B + \beta |\Sigma^{(i)}|^2) \sum_{t=1}^T \|\boldsymbol{L}^t - \boldsymbol{L}^{t-1}\|_{\infty}^2,$$

where $\|\boldsymbol{L}^t\|_{\infty} \leq D_{\boldsymbol{L}}$ for all $t \geq 1$.

For the proof we will employ the following construction.

Proposition B.1 ((Farina, Kroer, and Sandholm 2019c))**.** *Consider a collection of sets $\mathcal{X}_1, \ldots, \mathcal{X}_m$, 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* co$\{\mathcal{X}_1, \ldots, \mathcal{X}_m\}$ *can be constructed as follows:*

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

Then, if $R_1^T, \ldots, R_m^T$ is the regret accumulated by the regret minimizers $\mathcal{R}_1, \ldots, \mathcal{R}_m$, 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

$$R_{\text{co}}^T \leq R_{\Delta}^T + \max\{R_1^T, \ldots, R_m^T\}. \tag{12}$$

In our setting, this proposition can be cast in the form of Algorithm 2. We are now ready to show Theorem 3.4.

Proof of Theorem 3.4. First of all, Proposition B.1 implies that the accumulated regret of the regret circuit for the convex hull can be bounded as

$$R_{\Psi}^T \leq \alpha + A(T) + B \sum_{t=1}^T \|\boldsymbol{L}^t - \boldsymbol{L}^{t-1}\|_{\infty}^2 + \beta \sum_{t=1}^T \|\boldsymbol{\Lambda}^t - \boldsymbol{\Lambda}^{t-1}\|_{\infty}^2, \tag{13}$$

where we used the fact that each regret minimizer $\mathcal{R}_{\hat{\sigma}}^{(i)}$ obtains as input the same utility function as $\mathcal{R}_{\Psi}^{(i)}$, while we also used the notation $\boldsymbol{\Lambda}^t \in \mathbb{R}^{|\Sigma_*^{(i)}|}$ to represent the utility function given to $\mathcal{R}_{\Delta}^{(i)}$ in accordance to Proposition B.1. Next, let us focus on bounding the norm $\|\boldsymbol{\Lambda}^t - \boldsymbol{\Lambda}^{t-1}\|_{\infty}^2$; in particular, it follows that for some index $s \in \{1, \ldots, |\Sigma_*^{(i)}|\}$,

$$\|\boldsymbol{\Lambda}^t - \boldsymbol{\Lambda}^{t-1}\|_{\infty}^2 = \left(\langle \boldsymbol{L}^t, \boldsymbol{x}_s^t \rangle - \langle \boldsymbol{L}^{t-1}, \boldsymbol{x}_s^{t-1} \rangle\right)^2 \tag{14}$$

$$= \left(\langle \boldsymbol{L}^t, \boldsymbol{x}_s^t \rangle - \langle \boldsymbol{L}^{t-1}, \boldsymbol{x}_s^t \rangle + \langle \boldsymbol{L}^{t-1}, \boldsymbol{x}_s^t \rangle - \langle \boldsymbol{L}^{t-1}, \boldsymbol{x}_s^{t-1} \rangle\right)^2$$

$$= \left(\langle \boldsymbol{L}^t - \boldsymbol{L}^{t-1}, \boldsymbol{x}_s^t \rangle + \langle \boldsymbol{L}^{t-1}, \boldsymbol{x}_s^t - \boldsymbol{x}_s^{t-1} \rangle\right)^2$$

$$\leq 2 \left(\langle \boldsymbol{L}^t - \boldsymbol{L}^{t-1}, \boldsymbol{x}_s^t \rangle\right)^2 + 2 \left(\langle \boldsymbol{L}^{t-1}, \boldsymbol{x}_s^t - \boldsymbol{x}_s^{t-1} \rangle\right)^2 \tag{15}$$

$$\leq 2\|\boldsymbol{L}^t - \boldsymbol{L}^{t-1}\|_{\infty}^2 \|\boldsymbol{x}_s^t\|_1^2 + 2\|\boldsymbol{L}^{t-1}\|_{\infty}^2 \|\boldsymbol{x}_s^t - \boldsymbol{x}_s^{t-1}\|_1^2 \tag{16}$$

$$\leq 8\|\boldsymbol{L}^t - \boldsymbol{L}^{t-1}\|_{\infty}^2 |\Sigma^{(i)}|^2 + 2D_{\boldsymbol{L}}^2 K^2, \tag{17}$$

where (14) follows from the definition of the ℓ_∞ norm; (15) uses Young's inequality; (16) is an application of Cauchy-Schwarz; and (17) uses the fact that $\|\boldsymbol{x}_s^t\|_1 \le 2|\Sigma^{(i)}|$. As a result, if we plug the bound of Equation (17) to (13) we can conclude that

$$R_\Psi^T \le \alpha + A(T) + (B + 2\beta D_{\boldsymbol{x}}^2) \sum_{t=1}^T \|\boldsymbol{L}^t - \boldsymbol{L}^{t-1}\|_\infty^2 + 2\beta D_{\mathbf{L}}^2 K^2 T. \tag{18}$$

Moreover, regarding the stability of the composite regret minimizer we observe that

$$
\begin{aligned}
\left\| \sum_{k=1}^{|\Sigma_*^{(i)}|} \boldsymbol{\lambda}_k^t \boldsymbol{x}_k^t - \sum_{k=1}^{|\Sigma_*^{(i)}|} \boldsymbol{\lambda}_k^{t-1} \boldsymbol{x}_k^{t-1} \right\|_1 &\le \sum_{k=1}^{|\Sigma_*^{(i)}|} \|\boldsymbol{\lambda}_k^t \boldsymbol{x}_k^t - \boldsymbol{\lambda}_k^{t-1} \boldsymbol{x}_k^{t-1}\|_1 \\
&\le \sum_{k=1}^{|\Sigma_*^{(i)}|} \|\boldsymbol{\lambda}_k^t \boldsymbol{x}_k^t - \boldsymbol{\lambda}_k^t \boldsymbol{x}_k^{t-1} + \boldsymbol{\lambda}_k^t \boldsymbol{x}_k^{t-1} - \boldsymbol{\lambda}_k^{t-1} \boldsymbol{x}_k^{t-1}\|_1 \\
&\le \sum_{k=1}^{|\Sigma_*^{(i)}|} \left(\|\boldsymbol{\lambda}_k^t \boldsymbol{x}_k^t - \boldsymbol{\lambda}_k^t \boldsymbol{x}_k^{t-1}\|_1 + \|\boldsymbol{\lambda}_k^t \boldsymbol{x}_k^{t-1} - \boldsymbol{\lambda}_k^{t-1} \boldsymbol{x}_k^{t-1}\|_1 \right) \\
&\le \sum_{k=1}^{|\Sigma_*^{(i)}|} \boldsymbol{\lambda}_k^t \|\boldsymbol{x}_k^t - \boldsymbol{x}_k^{t-1}\|_1 + \sum_{k=1}^{|\Sigma_*^{(i)}|} \|\boldsymbol{x}_k^{t-1}\|_1 |\boldsymbol{\lambda}_k^t - \boldsymbol{\lambda}_k^{t-1}| \\
&\le K + 2|\Sigma^{(i)}| \|\boldsymbol{\lambda}^t - \boldsymbol{\lambda}^{t-1}\|_1 \\
&\le K + 2|\Sigma^{(i)}| \kappa,
\end{aligned}
$$

$$\tag{19}$$
$$\tag{20}$$
$$\tag{21}$$

where we used the triangle inequality in (19), the ℓ_1 stability of the sequence $(\boldsymbol{x}_k^t)$ in (20), and the ℓ_1 stability of $(\boldsymbol{\lambda}^t)$ in (21). This verifies our claim about the stability and the predictivity of $\mathcal{R}_\Psi^{(i)}$. Finally, the complexity analysis for the NEXTSTRATEGY function follows directly since the NEXTSTRATEGY operation of each individual regret minimizer runs in $O(|\Sigma^{(i)}|)$, while the analysis of the OBSERVEUTILITY routine follows similarly to (Farina et al. 2021, Theorem 4.6), and it is therefore omitted. $\quad\square$

Algorithm 2: Stable-Predictive Regret Minimizer $\mathcal{R}_\Psi^{(i)}$ for the set co $\Psi^{(i)}$

Input:
- Player $i \in [n]$
- A K-stable (A_T, B)-predictive regret minimizer $\mathcal{R}_{\hat{\sigma}}^{(i)}$ for $\Psi_{\hat{\sigma}}^{(i)}$, for each $\hat{\sigma} \in \Sigma_*^{(i)}$
- A deterministic κ-stable (α, β)-predictive regret minimizer $\mathcal{R}_\Delta^{(i)}$ for $\Delta^{|\Sigma_*^{(i)}|}$

1 **Function** NEXTSTRATEGY():
2 $\quad$ $\boldsymbol{\lambda}^t \leftarrow \mathcal{R}_\Delta^{(i)}.\text{NEXTSTRATEGY}()$
3 $\quad$ **for** $\hat{\sigma} \in \Sigma_*^{(i)}$ **do**
4 $\quad\quad$ $\phi_{\hat{\sigma} \to \boldsymbol{q}_{\hat{\sigma}}^t} \leftarrow \mathcal{R}_{\hat{\sigma}}^{(i)}.\text{NEXTSTRATEGY}()$
5 $\quad$ **return** $\sum_{\hat{\sigma} \in \Sigma_*^{(i)}} \boldsymbol{\lambda}^t[\hat{\sigma}] \phi_{\hat{\sigma} \to \boldsymbol{q}_{\hat{\sigma}}^t}$ represented implicitly as $\{\boldsymbol{\lambda}^t[\hat{\sigma}], \boldsymbol{q}_{\hat{\sigma}}^t\}_{\hat{\sigma} \in \Sigma_*^{(i)}}$
6 **Function** OBSERVEUTILITY(L^t):
7 $\quad$ **for** $\hat{\sigma} \in \Sigma_*^{(i)}$ **do**
8 $\quad\quad$ $\mathcal{R}_{\hat{\sigma}}^{(i)}.\text{OBSERVEUTILITY}(L^t)$
9 $\quad$ Construct the linear function $\ell_{\boldsymbol{\lambda}}^t : \boldsymbol{\lambda} \mapsto \sum_{\hat{\sigma} \in \Sigma_*^{(i)}} \boldsymbol{\lambda}[\hat{\sigma}] L^t \left(\phi_{\hat{\sigma} \to \boldsymbol{q}_{\hat{\sigma}}^t} \right)$
10 $\quad$ $\mathcal{R}_\Delta^{(i)}.\text{OBSERVEUTILITY}(\ell_{\boldsymbol{\lambda}}^t)$

B.2   Proofs for Section 3.2

We commence this section with the proof that OMWU guarantees multiplicative-stability.

Lemma B.2 (Multiplicative Stability of OMWU on the Simplex). *Consider the* OMWU *algorithm on the simplex Δ^m with learning rate $\eta > 0$. If all the utility functions $\boldsymbol{\ell}^t \in \mathbb{R}^m$ are such that $\|\boldsymbol{\ell}^t\|_\infty \le L$, and $\eta < 1/(12L)$, then the sequence $(\boldsymbol{x}^t)_{t \ge 2}$ produced by* OMWU *is $(12\eta L)$-multiplicatively-stable.*

Proof. It is well-known that the update rule of OMWU on the simplex has the following form:

$$x_i^t = \frac{e^{2\eta \ell_i^{t-1} - \eta \ell_i^{t-2}}}{\sum_{k=1}^m e^{2\eta \ell_k^{t-1} - \eta \ell_k^{t-2}} x_k^{t-1}} x_i^{t-1},$$

for $t \geq 2$. As a result, it follows that

$$x_i^t \leq \frac{e^{3\eta L}}{\sum_{k=1}^m e^{-3\eta L} x_k^{t-1}} x_i^{t-1} = e^{6\eta L} x_i^{t-1} \leq (1 + 12\eta L) x_i^{t-1},$$

where we used that $\ell_i^{t-1}, \ell^{t-2} \in [-L, L]$, the fact that $\boldsymbol{x}^{t-1} \in \Delta^m$, and that $e^x \leq 1 + 2x$, for all $x \in [0, 1/2]$. Similarly, we have that

$$x_i^t \geq \frac{e^{-3\eta L}}{\sum_{k=1}^m e^{3\eta L} x_k^{t-1}} x_i^{t-1} = e^{-6\eta L} x_i^{t-1} \geq (1 - 6\eta L) x_i^{t-1}.$$

$\square$

Proof of Lemma 3.5. Let us start with the regret minimizer $\mathcal{R}_\Delta^{(i)}$. By Lemma B.2 it suffices to bound the ℓ_∞ norm of the utility vectors. In particular, if $\boldsymbol{\Lambda}$ is the associated utility vector predicted from Proposition B.1, it follows that $\|\boldsymbol{\Lambda}\|_\infty = |\langle \mathbf{L}, \boldsymbol{x}_k \rangle| \leq \|\mathbf{L}\|_\infty \|\boldsymbol{x}_k\|_1 = O(D_{\mathbf{L}}|\Sigma^{(i)}|)$, for some $k \in \{1, \ldots, |\Sigma_*^{(i)}|\}$, where we used the (generalized) Cauchy–Schwarz inequality, and the fact that $\|\boldsymbol{x}_k\|_1 = O(|\Sigma^{(i)}|)$. Moreover, for $\mathbf{L}^t = (\boldsymbol{\ell}^t \otimes \boldsymbol{x}^t)^\flat$ and $\|\boldsymbol{x}^t\|_\infty \leq 1$ it follows that $D_{\mathbf{L}} \leq D_\ell$, and the claimed bound follows directly from Lemma B.2. Finally, for each regret minimizer $\mathcal{R}_{\hat\sigma}$ the multiplicative-stability bound can only be a factor of $\mathfrak{D}^{(i)}$ from the derived bound in Theorem A.4, concluding the proof. $\square$

Next, we focus on the proof of Theorem 3.11. To this end, we leverage the approach of Kruckman, Greenwald, and Wicks (2010), who provided an alternative proof of the classic Markov chain tree theorem using linear-algebraic techniques. We commence by stating some elementary properties of the determinant.

Fact B.3. *The following properties hold:*
- *The determinant is a multilinear function with respect to the rows and columns of the matrix. That is,*

$$\det(\boldsymbol{u}_1, \ldots, \alpha \boldsymbol{u}_k + \beta \boldsymbol{u}_k', \ldots, \boldsymbol{u}_m) = \alpha \det(\boldsymbol{u}_1, \ldots, \boldsymbol{u}_k, \ldots, \boldsymbol{u}_m) + \beta \det(\boldsymbol{u}_1, \ldots, \boldsymbol{u}_k', \ldots, \boldsymbol{u}_m),$$

  for any $\boldsymbol{u}_1, \ldots, \boldsymbol{u}_m \in \mathbb{R}^m$, $\boldsymbol{u}_k' \in \mathbb{R}^m$, and $\alpha, \beta \in \mathbb{R}$;
- *If any two rows or columns of a square matrix $\mathbf{A}$ are equal, then $\det(\mathbf{A}) = 0$;*
- *The determinant remains invariant under permutations.*

Given a matrix $\mathbf{A}$, the minor $\mathtt{mn}^{(i,j)}(\mathbf{A})$ is the matrix formed from $\mathbf{A}$ after deleting the i-th row and the j-th column. Then, the *cofactor* is defined as $\mathtt{co}^{(i,j)}(\mathbf{A}) = (-1)^{i+j} \det\left(\mathtt{mn}^{(i,j)}(\mathbf{A})\right)$, while the *adjugate* (or adjoint) matrix $\mathrm{adj}(\mathbf{A})^\top$ is the matrix with entries the corresponding cofactors of $\mathbf{A}$; that is, $\mathrm{adj}(\mathbf{A})[(i,j)] := \mathtt{co}^{(j,i)}(\mathbf{A})$. With this notation at hand, we are ready to state the following characterization due to (Kruckman, Greenwald, and Wicks 2010, Theorem 3.4):

Theorem B.4 ((Kruckman, Greenwald, and Wicks 2010)). *Consider an ergodic m-state Markov chain with transition matrix $\mathbf{M}$. If $\boldsymbol{x} \in \mathbb{R}^m$ is such that $\boldsymbol{x}_i := \mathrm{adj}(\mathcal{L})[(i,i)]$, where $\mathcal{L} := \mathbf{M} - \mathbf{I}_m$ is the Laplacian of the system, $\boldsymbol{x}$ is an eigenvector of $\mathbf{M}$ with a corresponding eigenvalue of 1. That is, $\mathbf{M}\boldsymbol{x} = \boldsymbol{x}$.*

An key step of our proof for Theorem 3.11 uses this theorem in order to characterize the stationary distribution of a certain (ergodic) Markov chain. Incidentally, an alternative characterization can be provided using the classic Markov chain tree theorem. In particular, a central component of the latter theorem is the notion of a *directed tree*:

Definition B.5 (Directed Tree). *A graph $G = (V, E)$ is said to be a *directed tree* rooted at $u \in V$ if (i) it does not contain any cycles, and (ii) u has *no* outgoing edges, while every other node has exactly *one* outgoing edge.*

We will represent with $\mathcal{D}_i$ the set of all graphs which have property (ii) with respect to a node $i \in [m]$. Moreover, we will use $\mathcal{T}_i$ to represent the subset of $\mathcal{D}_i$ which also has property (i) of Definition B.5. For a matrix $\mathbf{D} \in \mathcal{D}_i$, we define a matrix $\mathtt{mp}(\mathbf{D})$ so that $\mathtt{mp}(\mathbf{D})_{(j,k)} = 1$ if $(k, j) \in E(\mathbf{D})$, and 0 otherwise. The following lemma will be of particular use for our purposes:

Lemma B.6 ((Kruckman, Greenwald, and Wicks 2010)). *Consider some $m \times m$ matrix $\mathbf{D} \in \mathcal{D}_i$, and let R_i be the determinant of the Laplacian matrix $\mathcal{L} := \mathtt{mp}(\mathbf{D}) - \mathbf{I}$ after replacing the i-th column with the i-th standard unit vector $\boldsymbol{e}_i$. Then, $R_i = (-1)^{m-1}$ if $\mathbf{D} \in \mathcal{T}_i$, i.e. $\mathbf{D}$ contains no (directed) cycles. Otherwise, $R_i = 0$.*

Before we proceed with the technical proof of Lemma 3.8, we also state a useful elementary fact.

Fact B.7. *The adjugate matrix at (i, i) is equal to the determinant of $\mathbf{A}$ after we replace the i-th column with the vector $\boldsymbol{e}_i$.*

Lemma 3.8. *Let* $\mathbf{M}$ *be the transition matrix of an m-state Markov chain such that* $\mathbf{M} := \boldsymbol{v}\mathbf{1}^\top + \mathbf{C}$, *where* $\mathbf{C}$ *is a matrix with strictly positive entries and columns summing to $1 - \lambda$, and $\boldsymbol{v}$ is a vector with strictly positive entries summing to λ. Then, if $\boldsymbol{\pi}$ is the stationary distribution of* $\mathbf{M}$, *there exists, for each $i \in [m]$, a (non-empty) finite set F_i and $F = \bigcup_i F_i$, and corresponding parameters $b_j \in \{0,1\}, 0 \le p_j \le m - 2, |S_j| = m - p_j - b_j - 1$, for each $j \in F_i$, such that*

$$\boldsymbol{\pi}_i = \frac{\sum_{j \in F_i} \lambda^{p_j + 1} (\boldsymbol{v}[q_j])^{b_j} \prod_{(s,w) \in S_j} \mathbf{C}[(s,w)]}{\sum_{j \in F} C_j \lambda^{p_j + b_j} \prod_{(s,w) \in S_j} \mathbf{C}[(s,w)]},$$

where $C_j = C_j(m)$ is a positive constant.

Proof. Let us consider the Laplacian matrix $\mathcal{L} = \mathbf{M} - \mathbf{I}_m$, and the quantities $\Sigma_i := \mathrm{adj}(\mathcal{L})[(i,i)]$. We shall first characterize the structure of Σ_i's. By symmetry, we can focus without loss of generality on the term Σ_1. We know from Fact B.7 that Σ_1 can be expressed as

$$\Sigma_1 = \det(\boldsymbol{e}_1, \boldsymbol{v} + \boldsymbol{c}_2 - \boldsymbol{e}_2, \dots, \boldsymbol{v} + \boldsymbol{c}_m - \boldsymbol{e}_m), \tag{22}$$

where $\boldsymbol{c}_j$ represents the j-th column of $\mathbf{C}$. Now if $\boldsymbol{e}_{j,k} := \boldsymbol{e}_j - \boldsymbol{e}_k$, given that $\mathbf{M}$ is column-stochastic, we observe that

$$\boldsymbol{e}_j - \boldsymbol{v} - \boldsymbol{c}_j = \sum_{k=1}^m (\boldsymbol{e}_j - \boldsymbol{e}_k)\boldsymbol{v}[k] + \sum_{k=1}^m (\boldsymbol{e}_j - \boldsymbol{e}_k)\boldsymbol{c}_j[k] = \sum_{k=1}^m \boldsymbol{e}_{j,k}\boldsymbol{v}[k] + \sum_{k=1}^m \boldsymbol{e}_{j,k}\boldsymbol{c}_j[k].$$

Next, if we plug-in this expansion to (22) it follows that

$$\Sigma_1 = \det\left(\boldsymbol{e}_1, \sum_{k=1}^m \boldsymbol{e}_{k,2}\boldsymbol{v}[k] + \sum_{k=1}^m \boldsymbol{e}_{k,2}\boldsymbol{c}_2[k], \dots, \sum_{k=1}^m \boldsymbol{e}_{k,m}\boldsymbol{v}[k] + \sum_{k=1}^m \boldsymbol{e}_{k,m}\boldsymbol{c}_m[k]\right). \tag{23}$$

By multilinearity of the determinant (Fact B.3), Σ_1 can be expressed as the sum of terms, with a single term of the form

$$\det\left(\boldsymbol{e}_1, \sum_{k=1}^m \boldsymbol{e}_{k,2}\boldsymbol{c}_2[k], \dots, \sum_{k=1}^m \boldsymbol{e}_{k,m}\boldsymbol{c}_m[k]\right), \tag{24}$$

independent on $\boldsymbol{v}$, while any other term can be expressed in the form

$$\det\left(\boldsymbol{e}_1, \boldsymbol{z}_2, \dots, \sum_{k=1}^m \boldsymbol{e}_{k,j}\boldsymbol{v}[k], \dots, \boldsymbol{z}_m\right), \tag{25}$$

for some index j, where $\boldsymbol{z}_\ell$ is either $\sum_{k=1}^m \boldsymbol{e}_{k,\ell}\boldsymbol{v}[k]$ or $\sum_{k=1}^m \boldsymbol{e}_{k,\ell}\boldsymbol{c}_\ell[k]$. Now let us first analyze each term of (25). We will show that it can be equivalently expressed so that the vector $\boldsymbol{v}$ appears only in a single column. Indeed, consider any other column in the matrix involved in the determinant of (25), expressed in the form $\sum_{k=1}^m \boldsymbol{e}_{k,\ell}\boldsymbol{v}[k]$, for some index $\ell \ne j$, if such column exists. Then, if we subtract the j-th column from that column it would take the form

$$\sum_{k=1}^m \boldsymbol{e}_{k,\ell}\boldsymbol{v}[k] - \sum_{k=1}^m \boldsymbol{e}_{k,j}\boldsymbol{v}[k] = \sum_{k=1}^m (\boldsymbol{e}_j - \boldsymbol{e}_\ell)\boldsymbol{v}[k] = \lambda \boldsymbol{e}_{j,\ell},$$

where recall that λ is the sum of the entries of vector $\boldsymbol{v}$, while this operation does not modify the value of the determinant. Thus, by multinearity, the determinant (25) is equal to

$$\lambda^p \det\left(\boldsymbol{e}_1, \boldsymbol{z}_2', \dots, \sum_{k=1}^m \boldsymbol{e}_{k,j}\boldsymbol{v}[k], \dots, \boldsymbol{z}_m'\right), \tag{26}$$

where $\boldsymbol{z}_\ell'$ is either $\sum_{k=1}^m \boldsymbol{e}_{k,\ell}\boldsymbol{c}_\ell[k]$ or $\boldsymbol{e}_{j,\ell}$, and $0 \le p \le m - 2$. Next, if we use again the multilinearity property, the term in (26) can be expressed as a sum of terms each of which has the form

$$\left(\lambda^p \boldsymbol{v}[q] \prod_{(s,w) \in S} \mathbf{C}[(s,w)]\right) \det(\boldsymbol{e}_1, \boldsymbol{e}_{\cdot,2}, \dots, \boldsymbol{e}_{\cdot,m}),$$

where $|S| = m - p - 2$. (For notational simplicity we used the notation $\boldsymbol{e}_{\cdot,2}, \dots, \boldsymbol{e}_{\cdot,m}$ to hide the first index.) Thus, the induced determinant $\det(\boldsymbol{e}_1, \boldsymbol{e}_{\cdot,2}, \dots, \boldsymbol{e}_{\cdot,m})$ matches after a suitable permutation the form of Lemma B.6 associated with some matrix $\mathbf{D} \in \mathcal{D}_i$. As a result, it can either be 0 or $(-1)^{m-1}$, depending on whether the corresponding graph has a (directed) cycle. Similar reasoning applies for the determinant in (24), which can be expressed as a sum of terms

$$(-1)^{m-1} \prod_{(s,w) \in S} \mathbf{C}[(s,w)],$$

where $|S| = m - 1$. Overall, we have shown that each Σ_i (due to symmetry) can be expressed in the form

$$(-1)^{m-1} \sum_{j \in F_i} \lambda^{p_j} (\boldsymbol{v}[q_j])^{b_j} \prod_{(s,w) \in S_j} \mathbf{C}[(s,w)], \tag{27}$$

where for all j it holds that $b_j \in \{0, 1\}$, and $|S_j| = m - p_j - b_j - 1$. Next, we will focus on characterizing the term $\Sigma := \sum_{i=1}^m \Sigma_i$. In particular, the stationary distribution $\boldsymbol{\pi}$ of $\mathbf{M}$ is such that

$$\left(\mathbf{C} + \boldsymbol{v}\mathbf{1}^\top \right) \boldsymbol{\pi} = \boldsymbol{\pi} \iff \mathbf{C}\boldsymbol{\pi} + \boldsymbol{v} = \boldsymbol{\pi} \iff (\mathbf{I}_m - \mathbf{C})\boldsymbol{\pi} = \boldsymbol{v}, \tag{28}$$

where we used that $\mathbf{1}^\top \boldsymbol{\pi} = 1$ since $\boldsymbol{\pi} \in \Delta^m$. Moreover, we claim that the matrix $\mathbf{I}_m - \mathbf{C}$ is invertible. Indeed, the sum of the columns of $\mathbf{C}$ is $1 - \lambda$, and subsequently it follows that the maximum eigenvalue of $\mathbf{C}$ is $(1 - \lambda)$. In turn, this implies that all the eigenvalues of $\mathbf{I}_m - \mathbf{C}$ are at least $\lambda > 0$. As a result, we can use Cramer's rule to obtain an explicit formula for the solution of the linear system with respect to the first coordinate of $\boldsymbol{\pi}$:

$$\boldsymbol{\pi}_1 = \frac{\det(\boldsymbol{v}, \boldsymbol{e}_2 - \boldsymbol{c}_2, \dots, \boldsymbol{e}_m - \boldsymbol{c}_m)}{\det(\boldsymbol{e}_1 - \boldsymbol{c}_1, \boldsymbol{e}_2 - \boldsymbol{c}_2, \dots, \boldsymbol{e}_m - \boldsymbol{c}_m)}. \tag{29}$$

Moreover, it follows that

$$\begin{aligned}
\det(\boldsymbol{v}, \boldsymbol{e}_2 - \boldsymbol{c}_2, \dots, \boldsymbol{e}_m - \boldsymbol{c}_m) &= \det(\boldsymbol{v}, \boldsymbol{e}_2 - \boldsymbol{c}_2 - \boldsymbol{v}, \dots, \boldsymbol{e}_m - \boldsymbol{c}_m - \boldsymbol{v}) \\
&= \det(\boldsymbol{v} + (\lambda\boldsymbol{e}_1 - \boldsymbol{v}), \boldsymbol{e}_2 - \boldsymbol{c}_2 - \boldsymbol{v}, \dots, \boldsymbol{e}_m - \boldsymbol{c}_m - \boldsymbol{v}) \\
&= \lambda \det(\boldsymbol{e}_1, \boldsymbol{e}_2 - \boldsymbol{c}_2 - \boldsymbol{v}, \dots, \boldsymbol{e}_m - \boldsymbol{c}_m - \boldsymbol{v}),
\end{aligned} \tag{30}$$

where in (30) we used the fact that $\det(\lambda\boldsymbol{e}_1 - \boldsymbol{v}, \dots, \boldsymbol{e}_m - \boldsymbol{c}_m - \boldsymbol{v}) = 0$. Thus, if we use the definition of Σ_1, Fact B.7, and (29), we arrive at the following conclusion:

$$\boldsymbol{\pi}_1 = \lambda \frac{\Sigma_1}{\det(\mathbf{I}_m - \mathbf{C})}.$$

But we can also infer from Theorem B.4 that $\boldsymbol{\pi}_1 = \Sigma_1/\Sigma$, implying the following identity:

$$\det(\mathbf{I}_m - \mathbf{C}) = \lambda \sum_{i=1}^m \Sigma_i. \tag{31}$$

In fact, we have shown this formula for *any* vector $\lambda\boldsymbol{p}$, where $\boldsymbol{p}$ is a probability distribution, and $\lambda > 0$. Thus, it must also hold for $\mathbf{v} := \frac{\lambda}{m}\mathbf{1}$. That is,

$$\det(\mathbf{I}_m - \mathbf{C}) = \lambda(-1)^{m-1} \sum_{j \in F} C_j \lambda^{p_j + b_j} \prod_{(s,w) \in S_j} \mathbf{C}[(s,w)], \tag{32}$$

where $|S_j| \leq m - 1 - p_j$, $C_j = C_j(m)$ is a positive parameter independent on the entries of $\boldsymbol{v}$ and $\mathbf{C}$, and $F = \bigcup_i F_i$. Finally, given that the vector $\boldsymbol{\pi} \in \Delta^m$ with $\boldsymbol{\pi}_i = \Sigma_i/\Sigma$ is the (unique) stationary distribution of $\mathbf{M}$, the claim follows directly from (27), (31), and (32). $\qquad\square$

Corollary B.8. *Let $\mathbf{M}$ be the transition matrix of an m-state Markov chain such that $\mathbf{M} := \boldsymbol{v}\mathbf{1}^\top + \mathbf{C}$, where $\mathbf{C}$ is a matrix with strictly positive entries and columns summing to $1 - \lambda$, and $\boldsymbol{v}$ is a vector with strictly positive entries summing to λ. Moreover, let $\boldsymbol{v} = \boldsymbol{r}/l$, for some $l > 0$. Then, if $\boldsymbol{\pi}$ is the stationary distribution of $\mathbf{M}$, there exists, for each $i \in [m]$, a (non-empty) finite set F_i and $F = \bigcup_i F_i$, and corresponding parameters $b_j \in \{0, 1\}, 0 \leq p_j \leq m - 2, |S_j| = m - p_j - b_j - 1$, for each $j \in F_i$, such that the i-th coordinate of the vector $\boldsymbol{w} := l\boldsymbol{\pi}$ can be expressed as*

$$\boldsymbol{w}_i = \frac{\sum_{j \in F_i} \lambda^{p_j + 1} (\boldsymbol{r}[q_j])^{b_j} l^{1 - b_j} \prod_{(s,w) \in S_j} \mathbf{C}[(s,w)]}{\sum_{j \in F} C_j \lambda^{p_j + b_j} \prod_{(s,w) \in S_j} \mathbf{C}[(s,w)]}, \tag{33}$$

where $C_j = C_j(m)$ is a positive constant.

Proof. The proof follows directly from the formula derived in Lemma 3.8. $\qquad\square$

This expression for the stationary distribution was derived specifically to characterize the multiplicative stability of the fixed points associated with EFCE. In particular, this will be shown directly from Corollary 3.9, which is recalled next.

Corollary 3.9. *Let $\mathbf{M}, \mathbf{M}'$ be the 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 $\mathbf{C}$ and $\mathbf{C}'$ are matrices with strictly positive entries, and $\boldsymbol{v}, \boldsymbol{v}'$ are vectors with strictly positive entries such that $\boldsymbol{v} = \boldsymbol{r}/l$ and $\boldsymbol{v}' = \boldsymbol{r}'/l'$, for some $l > 0$ and $l' > 0$. If $\boldsymbol{\pi}$ and $\boldsymbol{\pi}'$ are the stationary distributions of $\mathbf{M}$ and $\mathbf{M}'$, let $\boldsymbol{w} := l\boldsymbol{\pi}$ and $\boldsymbol{w}' := l'\boldsymbol{\pi}'$. Finally, let λ and λ' be the sum of the entries of $\boldsymbol{v}$ and $\boldsymbol{v}'$ respectively. Then, if (i) the matrices $\mathbf{C}$ and $\mathbf{C}'$ are κ-multiplicative-close; (ii) the scalars λ and λ' are κ-multiplicative-close; (iii) the vectors $\boldsymbol{r}$ and $\boldsymbol{r}'$ are γ-multiplicative-close; and (iv) the numbers l and l' are also γ-multiplicative-close, then the vectors $\boldsymbol{w}$ and $\boldsymbol{w}'$ are $(\gamma + O(\kappa m))$-multiplicative-close, for a sufficiently small $\kappa = O(1/m)$.*

Proof. Consider some coordinate $i \in [m]$, and let

$$V_j := \lambda^{p_j+1}(\boldsymbol{r}[q_j])^{b_j} l^{1-b_j} \prod_{(s,w) \in S_j} \mathbf{C}[(s,w)],$$

for some $j \in F_i$. Also let V'_j be the corresponding quantity with respect to $\mathbf{M}'$. Then, by assumption we have that

$$V'_j \le (1+\kappa)^{p_j+1}(1+\gamma)(1+\kappa)^{|S_j|}V_j \le (1+\gamma)(1+\kappa)^m V_j,$$

where we used the fact that $|S_j| + p_j + 1 \le m$ by Corollary B.8. Moreover, for a sufficiently small $\kappa = O(1/m)$, we can infer that $V'_j \le (1+\gamma)(1+O(\kappa m))V'_j = (1+(\gamma+O(\kappa m)))V_j$. Thus, this implies that $\sum_{j \in F_i} V'_j \le (1+(\gamma+O(\kappa m))) \sum_{j \in F_i} V'_j$. Moreover, we can show that the denominator of (33) induces an extra additive factor of $O(\kappa m)$ in the multiplicative stability, implying that $\boldsymbol{w}'_i \le (1+(\gamma+O(\kappa m)))\boldsymbol{w}_i$, for any $i \in [m]$. Similarly, it follows that $\boldsymbol{w}'_i \ge (1-(\gamma+O(\kappa m)))\boldsymbol{w}_i$. $\qquad\square$

Next, we will use this statement to prove Proposition 3.10. We note that it is tacitly assumed that the vectors $\boldsymbol{\lambda}^t, \boldsymbol{q}^t_{\hat\sigma}$, and $\boldsymbol{x}_{(j \in J)}$, involved in Proposition 3.10 have strictly positive coordinates; this will be indeed the case when OMWU is used to construct the individual (local) regret minimizers.

Proof of Proposition 3.10. Let us focus on the stability analysis of Algorithm 3 as the rest of the claim follows from (Farina et al. 2021, Proposition 4.14). In particular, for consistency with the terminology of Corollary 3.9, let us define

$$\mathbf{C}[(a_r, a_c)] := \boldsymbol{\lambda}[(j^*, a_c)]\boldsymbol{q}_{(j^*,a_c)}[(j^*, a_r)] + \left(1 - \sum_{\hat\sigma \preceq (j^*,a_c)} \boldsymbol{\lambda}[\hat\sigma]\right)\mathbf{1}\{a_r = a_c\};$$

and $l := \boldsymbol{x}[\sigma_p]$. We will show that the requirements of Corollary 3.9 are satisfied:

(i) The entries of $\mathbf{C}$ are $O(\kappa)$-multiplicatively-stable. In particular, this follows from the fact that $1 - \sum_{\hat\sigma \preceq (j^*,a_c)} \boldsymbol{\lambda}[\hat\sigma] = \sum_{\tilde\sigma \in \tilde\Sigma} \boldsymbol{\lambda}[\tilde\sigma]$, and the latter term is clearly κ-multiplicatively-stable;

(ii) The sum of the entries of $\boldsymbol{v}^t := \boldsymbol{r}^t/l^t$ is κ-multiplicatively-stable. To see this, note that the sum of each column of $\mathbf{C}$ can be expressed as $\sum_{\tilde\sigma \in \tilde\Sigma} \boldsymbol{\lambda}[\tilde\sigma]$, and as a result, since the matrix $\mathbf{C} + \frac{1}{l}\boldsymbol{r}\mathbf{1}^\top$ is stochastic, we can infer that the sum of the entries of $\boldsymbol{v}$ can also be expressed as $\sum_{\tilde\sigma \in \tilde\Sigma} \boldsymbol{\lambda}[\tilde\sigma]$, since $\boldsymbol{\lambda}$ is a vector on the simplex. But the latter term is clearly κ-multiplicatively-stable, as desired;

(iii) The sequence of vectors $\boldsymbol{r}^t$ is $\gamma + O(\kappa)$-multiplicatively-stable. This assertion can be directly verified from the definition of $\boldsymbol{r}$ in Algorithm 3;

(iv) The sequence of numbers of l^t is γ-multiplicatively-stable. Indeed, this follows directly from the assumption that the sequence of $\boldsymbol{x}^t$'s is γ-multiplicatively-stable.

As a result, we can apply Corollary 3.9 to conclude the proof. $\qquad\square$

Proof of Theorem 3.11. Our argument proceeds inductively. For a root information set $j \in \mathcal{J}^{(i)}$, Proposition 3.10 implies $O(\kappa|\mathcal{A}|)$-multiplicative-stability for any induced partial fixed point; this follows given that the $\varnothing$-partial fixed point is trivially 0-multiplicatively-stable. Next, the theorem follows inductively given that by Proposition 3.10 each sequence can only incur an additive factor of $O(\kappa|\mathcal{A}|)$ in the multiplicative stability bound with respect to the preceding sequences. $\qquad\square$

More precisely, if $F^{(i)} := \max_{j_1 \prec j_2 \prec \cdots \prec j_d} \sum_{i=1}^d |\mathcal{A}(j_i)|$, with $j_1, \ldots, j_d \in \mathcal{J}^{(i)}$, we can show that the sequence of fixed points is $O(\kappa F^{(i)})$-multiplicatively-stable. Observe that $F^{(i)}$ can be trivially upper bounded by $|\mathcal{A}^{(i)}|\mathfrak{D}^{(i)}$, as well as the number of sequences $|\Sigma^{(i)}|$.

B.3 Putting Everything Together

First of all, if we combine Theorem A.4 with Theorem 3.4, instantiating the regret minimizer $\mathcal{R}^{(i)}_\Delta$ of Theorem 3.4 using OMWU with learning rate η, we can construct a regret minimizer for the set $\operatorname{co}\Psi^{(i)}$ with regret R^T_Ψ bounded as

$$R^T_\Psi \le \frac{\mathcal{P}^{(i)}}{\eta} + \eta^3 D^4_{\boldsymbol{\ell}} \mathcal{V}^{(i)} T + \eta \mathcal{B}^{(i)} \sum_{t=1}^T \|\boldsymbol{L}^t - \boldsymbol{L}^{t-1}\|^2_\infty, \tag{34}$$

where $\mathcal{P}^{(i)}, \mathcal{V}^{(i)}$, and $\mathcal{B}^{(i)}$ are game-specific parameters *polynomial* on the size of the game, and independent of T and η; and $D_{\boldsymbol{\ell}}$ is an upper bound on the ℓ_∞ norm of the utility sequences. In addition, this regret minimizer will be $(\eta\mathcal{K}^{(i)})$-stable, for a game-specific parameter $\mathcal{K}^{(i)}$ polynomial on the size of the game, and independent on η and T. Thus, this stable-predictive regret minimizer for the set of transformations $\Psi^{(i)}$ realizes the first requirement of Theorem 3.2.

Moreover, let $\eta\mathcal{G}^{(i)}$ be the parameter associated with the stability of the fixed points, as predicted by Corollary 3.12, for some game-specific parameter $\mathcal{G}^{(i)}$ polynomial on the size of the game, and independent on η and T. Now observe that this component realizes the STABLEFPORACLE; the second requirement of Theorem 3.2. As a result, we can apply Theorem 3.2 to conclude the following:

Algorithm 3: EXTEND$(\phi, J, j^*, \boldsymbol{x})$; (Farina et al. 2021)

    Input:
- $\phi = \sum_{\hat{\sigma} \in \Sigma_*^{(i)}} \boldsymbol{\lambda}[\hat{\sigma}] \phi_{\hat{\sigma} \to \boldsymbol{q}_{\hat{\sigma}}}^{(i)} \in \operatorname{co} \Psi^{(i)}$
- $J \subseteq \mathcal{J}^{(i)}$ trunk for player i
- $j^* \in \mathcal{J}^{(i)}$ information set not in J with an immediate predecessor in J
- $\boldsymbol{x} \in \mathbb{R}_{\geq 0}^{|\Sigma^{(i)}|}$ J-partial fixed point of ϕ

    Output: $\boldsymbol{x}' \in \mathbb{R}_{\geq 0}^{|\Sigma^{(i)}|}$ $(J \cup \{j^*\})$-partial fixed point of ϕ

1   $\sigma_p \leftarrow \sigma^{(i)}(j^*)$

2   Let $\boldsymbol{r} \in \mathbb{R}_{\geq 0}^{|\mathcal{A}(j^*)|}$ be defined as $\boldsymbol{r}[a] := \sum_{j' \preceq \sigma_p} \sum_{a' \in \mathcal{A}(j')} \boldsymbol{\lambda}[(j', a')] \boldsymbol{q}_{(j', a')}[(j^*, a)] \boldsymbol{x}[(j', a')]$

3   Let $\mathbf{W} \in \boldsymbol{x}[\sigma_p] \mathbb{S}^{|\mathcal{A}(j^*)|}$ be the matrix with entries $\mathbf{W}[a_r, a_c]$ defined, for $a_r, a_c \in \mathcal{A}(j^*)$, as

    $\boldsymbol{r}[a_r] + \left(\boldsymbol{\lambda}[(j^*, a_c)] \boldsymbol{q}_{(j^*, a_c)}[(j^*, a_r)] + \left(1 - \sum_{\hat{\sigma} \preceq (j^*, a_c)} \boldsymbol{\lambda}[\hat{\sigma}] \right) \mathbf{1}\{a_r = a_c\} \right) \boldsymbol{x}[\sigma_p]$

4   **if** $\boldsymbol{x}[\sigma_p] = 0$ **then**

5      $\boldsymbol{w} \leftarrow \boldsymbol{0} \in \mathbb{R}_{\geq 0}^{|\mathcal{A}(j^*)|}$

6   **else**

7      $\boldsymbol{b} \in \Delta^{|\mathcal{A}(j^*)|} \leftarrow$ stationary distribution of $\frac{1}{\boldsymbol{x}[\sigma_p]} \mathbf{W}$

8      $\boldsymbol{w} \to \boldsymbol{x}[\sigma_p] \boldsymbol{b}$

9   $\boldsymbol{x}' \leftarrow \boldsymbol{x}$

10   **for** $a \in \mathcal{A}(j^*)$ **do**

11      $\boldsymbol{x}'[(j^*, a)] \leftarrow \boldsymbol{w}[(j^*, a)]$

Algorithm 4: FIXEDPOINT(ϕ); (Farina et al. 2021)

    Input: $\phi = \sum_{\hat{\sigma} \in \Sigma_*^{(i)}} \boldsymbol{\lambda}[\hat{\sigma}] \phi_{\hat{\sigma} \to \boldsymbol{q}_{\hat{\sigma}}}^{(i)} \in \operatorname{co} \Psi^{(i)}$

    Output: $\boldsymbol{q}^* \in \mathcal{Q}^{(i)}$ such that $\boldsymbol{q} = \phi(\boldsymbol{q}^*)$

1   $\boldsymbol{q} \leftarrow \boldsymbol{0} \in \mathbb{R}^{|\Sigma^{(i)}|}, \boldsymbol{q}[\varnothing] \leftarrow \varnothing$

2   $J \leftarrow \varnothing$

3   **for** $j \in \mathcal{J}^{(i)}$ in top-down order **do**

4      $\boldsymbol{q}^* \leftarrow$ EXTEND$(\phi, J, j, \boldsymbol{q}^*)$

5      $J = J \cup \{j\}$

6   **return** $\boldsymbol{q}$

Corollary B.9. *Consider any player $i \in [n]$. There exists a regret minimizing algorithm such that under any sequence of utility vectors $\boldsymbol{\ell}^1, \dots, \boldsymbol{\ell}^T$, with ℓ_∞ norm bounded by $D_{\boldsymbol{\ell}}$, the accumulated $\Psi^{(i)}$-regret R^T can be bounded as*

$$R^T \leq \frac{\mathcal{P}^{(i)}}{\eta} + \eta^3 D_{\boldsymbol{\ell}}^4 \mathcal{V}^{(i)} T + \eta \mathcal{B}^{(i)} \sum_{t=1}^{T} \|\boldsymbol{\ell}^t - \boldsymbol{\ell}^{t-1}\|_\infty^2. \tag{35}$$

In addition, the regret minimizer is $\eta \mathcal{G}^{(i)}$-stable, for a sufficiently small learning rate η. All of the parameters $\mathcal{P}^{(i)}, \mathcal{V}^{(i)}, \mathcal{B}^{(i)}$, and $\mathcal{G}^{(i)}$ are game-specific parameters, polynomial on the size of the game, and independent on η and T.

    We note that the parameters appearing in this corollary are in general different from the ones in (34), and we use the same symbols with an abuse of notation. So far we have focused on bounding the regret of a player without any assumptions on the observed utility functions. A crucial observation is that if all players employ a regularized (or smooth) learning algorithm, then the observed utility functions change slowly over time. To formalize this observation, we start with the following auxiliary claim:

Claim B.10. *For any player $i \in [n]$ the observed utilities satisfy*

$$\|\boldsymbol{\ell}^{(i),t} - \boldsymbol{\ell}^{(i),t-1}\|_\infty \leq |\mathcal{Z}| \sum_{k \neq i}^{n} \|\boldsymbol{q}^{(k),t} - \boldsymbol{q}^{(k),t-1}\|_1. \tag{36}$$

Proof. For a profile of mixed sequence-form strategies $(\boldsymbol{q}^{(1)}, \ldots, \boldsymbol{q}^{(n)})$, the utility of player i can be expressed as

$$u^{(i)}(\boldsymbol{q}^{(1)}, \ldots, \boldsymbol{q}^{(n)}) = \sum_{z \in \mathcal{Z}} p^{(c)}(z) u^{(i)}(z) \prod_{k=1}^{n} \boldsymbol{q}^{(k)}(\sigma^{(k)}(z)).$$

As a result, given that (by assumption) $|u^{(i)}(z)| \leq 1$ for all $z \in \mathcal{Z}$, it follows that

$$\|\boldsymbol{\ell}^{(i),t} - \boldsymbol{\ell}^{(i),t-1}\|_\infty \leq \sum_{z \in \mathcal{Z}} \left| \prod_{k \neq i}^{n} \boldsymbol{q}^{(k),t}(\sigma^{(k)}(z)) - \prod_{k \neq i}^{n} \boldsymbol{q}^{(k),t-1}(\sigma^{(k)}(z)) \right|$$

$$\leq \sum_{z \in \mathcal{Z}} \sum_{k \neq i}^{n} \left| \boldsymbol{q}^{(k),t}(\sigma^{(k)}(z)) - \boldsymbol{q}^{(k),t-1}(\sigma^{(k)}(z)) \right|, \tag{37}$$

where in the last bound we used the inequality

$$|(a_1 a_2 \ldots a_m) - (b_1 b_2 \ldots b_m)| \leq \sum_{i=1}^{m} |a_i - b_i|(a_1 \ldots a_{i-1})(b_{i+1} \ldots b_m) \leq \sum_{i=1}^{m} |a_i - b_i|,$$

for any $a_1, \ldots, a_m, b_1, \ldots, b_m \in [0, 1]$. Finally, from (37) we can conclude that

$$\|\boldsymbol{\ell}^{(i),t} - \boldsymbol{\ell}^{(i),t-1}\|_\infty \leq \sum_{k \neq i}^{n} \sum_{z \in \mathcal{Z}} \left| \boldsymbol{q}^{(k),t}(\sigma^{(k)}(z)) - \boldsymbol{q}^{(k),t-1}(\sigma^{(k)}(z)) \right| \leq |\mathcal{Z}| \sum_{k \neq i}^{n} \|\boldsymbol{q}^{(k),t} - \boldsymbol{q}^{(k),t-1}\|_1,$$

as desired. □

As a result, if *all* players employ a stable algorithm to update their strategies, the observed utilities will also be stable. This leads to the following conclusion:

Corollary B.11. *Assume that all agents play according to the dynamics associated with Corollary B.9. Then, for learning rate $\eta = O(T^{-1/4})$, the $\Psi^{(i)}$-regret of every player i is bounded by $O(T^{1/4})$, where the $O(\cdot)$ notation suppresses game-specific parameters polynomial in the size of the game.*

Proof. Let us use the notation $\mathcal{P} := \max_i \mathcal{P}^{(i)}, \mathcal{V} := \max_i \mathcal{V}^{(i)}, \mathcal{B} := \max_i \mathcal{B}^{(i)}$, and $\mathcal{G} := \max_i \mathcal{G}^{(i)}$. For any player $i \in [n]$, Claim B.10 implies that $\|\boldsymbol{\ell}^{(i),t} - \boldsymbol{\ell}^{(i),t-1}\|_\infty \leq |\mathcal{Z}|(n-1)\eta\mathcal{G}$, since every player updates her strategy through an $\eta\mathcal{G}$-stable learning algorithm. Thus, by the guarantee of Corollary B.9 we can conclude that for any player i,

$$R^{(i),T} \leq \frac{\mathcal{P}}{\eta} + \eta^3 \mathcal{V} T + \eta^3 (n-1)^2 \mathcal{B} |\mathcal{Z}|^2 \mathcal{G}^2 T,$$

where we used the fact that $D_\ell \leq 1$, which follows from the normalization assumption on the utilities. Finally, taking $\eta = O(T^{-1/4})$ yields that $R^{(i),T} = O(T^{1/4})$, as desired. □

Finally, we conclude with the proof of Theorem 1.1.

Proof of Theorem 1.1. The guarantee of Corollary B.11 along with Theorem 2.6 suffice in order to establish Theorem 1.1. In particular, for a player i let $\boldsymbol{\mu}^{(i),t}$ be any probability distribution on the set $\Pi^{(i)}$ such that $\mathbb{E}_{\boldsymbol{\pi} \sim \mu^{(i),t}}[\boldsymbol{\pi}] = \boldsymbol{q}^{(i),t}$, where $\boldsymbol{q}^{(i),t}$ is the output of the regret minimizer, on the *mixed* sequence-form strategy polytope $\mathcal{Q}^{(i)}$, realized with the dynamics of Corollary B.11. Moreover, let $\boldsymbol{\mu}^t := \boldsymbol{\mu}^{(1),t} \otimes \cdots \otimes \boldsymbol{\mu}^{(n),t}$ be the joint probability distribution, and $\bar{\boldsymbol{\mu}} := \frac{1}{T} \sum_{t=1}^{T} \boldsymbol{\mu}^t$. Then, it follows by the linearity of expectation, as well as the linearity of the expression in Definition 2.5 (recall that the set of transformations $\Psi^{(i)}$ is linear) that $\bar{\boldsymbol{\mu}}$ is an ϵ-EFCE, where $\epsilon := \frac{1}{T} \max_i R^{(i),T}$, and $R^{(i),T}$ is the cumulative $\Psi^{(i)}$-regret with respect to the set $Q^{(i)}$. Finally, the proof follows given that by virtue of Corollary B.11, $R^{(i),T} = O(T^{1/4})$ for *all* $i \in [n]$. □

We also point out that the complexity of each iteration in the dynamics is analogous to that in (Farina et al. 2021). Finally, we remark that it is easy to make the overall regret minimizer robust against adversarial losses using an adaptive learning rate.

C Acceleration for Extensive-Form Coarse Correlated Equilibrium

In this section we develop accelerated uncoupled no-regret dynamics for extensive-form *coarse* correlated equilibria (EFCCE). In particular, we will provide a slightly improved bound compared to that for EFCE. At the heart of this improvement is a recent algorithm due to (Farina, Celli, and Sandholm 2021) which circumvents the need of computing the stationary distribution of a Markov chain; this will allow us to show more favorable stability properties. First, let us introduce some basic concepts, commencing with that of *coarse* trigger deviation functions:

Definition C.1 (Coarse Trigger Deviation Functions). Consider some player $i \in [n]$, some information set $j \in \mathcal{J}^{(i)}$, and a continuation strategy $\hat{\boldsymbol{\pi}} \in \Pi_j^{(i)}$. A *coarse trigger deviation function* with respect to the information set j and the continuation strategy $\hat{\boldsymbol{\pi}}$ is any linear function $f : \mathbb{R}^{|\Sigma^{(i)}|} \to \mathbb{R}^{|\Sigma^{(i)}|}$ with the following properties:

- $f(\boldsymbol{\pi}) = \boldsymbol{\pi}$ for any $\boldsymbol{\pi} \in \Pi^{(i)}$ such that $\boldsymbol{\pi}[\sigma^{(i)}(j)] = 0$;
- For any $\sigma \in \Sigma^{(i)}$ and $\boldsymbol{\pi} \in \Pi^{(i)}$ such that $\boldsymbol{\pi}[\sigma^{(i)}(j)] = 1$,

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

It will be convenient to instantiate a coarse trigger deviation function in the form of a linear mapping $\phi_{j \to \hat{\boldsymbol{\pi}}}^{(i)} : \mathbb{R}^{|\Sigma^{(i)}|} \ni \boldsymbol{x} \mapsto \mathbf{M}_{j \to \hat{\boldsymbol{\pi}}}^{(i)} \boldsymbol{x}$, where $\mathbf{M}_{j \to \hat{\boldsymbol{\pi}}}^{(i)}$ is such that for any $\sigma_r, \sigma_c \in \Sigma^{(i)}$,

$$\mathbf{M}_{j \to \hat{\boldsymbol{\pi}}}^{(i)}[\sigma_r, \sigma_c] = \begin{cases} 1 & \text{if } \sigma_c \not\succeq j \quad \& \quad \sigma_r = \sigma_c; \\ \hat{\boldsymbol{\pi}}[\sigma_r] & \text{if } \sigma_c = \sigma^{(i)}(j) \quad \& \quad \sigma_r \succeq j; \\ 0 & \text{otherwise.} \end{cases}$$

We will also let $\tilde{\Psi}^{(i)} := \{\phi_{j \to \hat{\boldsymbol{\pi}}}^{(i)} : j \in \mathcal{J}^{(i)}, \hat{\boldsymbol{\pi}} \in \Pi_j^{(i)}\}$ be the set of all possible linear mappings defining coarse trigger deviations functions for player i.

Similarly to EFCE, we define the notion of *extensive-form coarse correlated equilibrium (EFCCE)* based on *coarse* trigger deviation functions:

Definition C.2 (EFCCE). For $\epsilon \geq 0$, a probability distribution $\boldsymbol{\mu} \in \Delta^{|\Pi|}$ is an ϵ-approximate EFCCE if for every player $i \in [n]$ and every coarse trigger deviation function $\phi_{j \to \hat{\boldsymbol{\pi}}}^{(i)} \in \tilde{\Psi}^{(i)}$, it holds that

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

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

This equilibrium concept requires that the expected utility of any *trigger agent* $(j, \hat{\boldsymbol{\pi}})$ is never larger (by more than an amount of ϵ) than the expected utility of the player when the mediator's recommendations are followed.

Theorem C.3 ((Farina, Celli, and Sandholm 2021)). *For every player $i \in [n]$, let $\boldsymbol{\pi}^{(i),1}, \ldots, \boldsymbol{\pi}^{(i),T} \in \Pi^{(i)}$ be a sequence of deterministic sequence-form strategies whose cumulative $\tilde{\Psi}^{(i)}$-regret is $R^{(i),T}$ with respect to the sequence of linear utility functions*

$$\ell^{(i),t} : \Pi^{(i)} \ni \boldsymbol{\pi}^{(i)} \mapsto u^{(i)} \left(\boldsymbol{\pi}^{(i)}, \boldsymbol{\pi}^{(-i),t} \right). \tag{38}$$

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

In this context, we will employ the general template we presented in Section 3, and in particular Theorem 3.2, in order to obtain accelerated no-regret dynamics for EFCCE. First, observe that a stable-predictive (external) regret minimizer for the set $\tilde{\Psi}^{(i)}$ can be directly obtained from Theorem 3.4 since $\tilde{\Psi}^{(i)} \subseteq \Psi^{(i)}$. Thus, for the rest of the section we will focus on the second ingredient required in Theorem 3.2: establishing the stability of the associated fixed points.

C.1 Stability of the Fixed Points

Unlike all known methods for computing fixed points related to EFCE, it was recently shown by (Farina, Celli, and Sandholm 2021) (see also (Morrill et al. 2021a)) that one can bypass the computation of a stationary distribution of a Markov chain using a succinct closed-form solution. This algorithm is presented in Algorithm 5.

Theorem C.4 (Theorem 3, (Farina, Celli, and Sandholm 2021)). *Algorithm 5 returns a fixed point of $\phi \in \operatorname{co} \tilde{\Psi}^{(i)}$, and it runs in time $O(|\Sigma^{(i)}| \mathfrak{D}^{(i)})$.*

Proposition C.5. *Under the conditions of Lemma 3.5, Algorithm 5 yields a sequence of $(12\kappa \mathfrak{D}^{(i)})$-multiplicatively-stable strategies, where $\kappa = O(\eta D_{\boldsymbol{\ell}}(|\Sigma^{(i)}| + D_{\boldsymbol{x}}^{(i)}(\mathfrak{D}^{(i)})^2))$ for a sufficiently small $\kappa = O(1/\mathfrak{D}^{(i)})$.*

Algorithm 5: FIXEDPOINT(ϕ) for $\phi \in \text{co } \tilde{\Psi}^{(i)}$

Input: $\phi = \sum_{j \in \mathcal{J}^{(i)}} \boldsymbol{\lambda}[j] \phi_{j \to \boldsymbol{q}_j} \in \text{co } \tilde{\Psi}^{(i)}$

Output: $\boldsymbol{q}^* \in \mathcal{Q}^{(i)}$ such that $\phi(\boldsymbol{q}^*) = \boldsymbol{q}^*$

1   $\boldsymbol{q}^* \leftarrow \mathbf{0} \in \mathbb{R}_{\geq 0}^{|\Sigma^{(i)}|}, \boldsymbol{q}^*[\varnothing] \leftarrow 1$

2   **for** $\sigma = (j, a) \in \Sigma_*^{(i)}$ *in top-down ($\prec$) order* **do**

3      $d_\sigma \leftarrow \sum_{j' \preceq j} \boldsymbol{\lambda}[j']$

4   **if** $d_\sigma = 0$ **then**

5         $\boldsymbol{q}^*[\sigma] \leftarrow \frac{\boldsymbol{q}^*[\sigma^{(i)}(j)]}{|\mathcal{A}(j)|}$

6   **else**

7         $\boldsymbol{q}^*[\sigma] \leftarrow \frac{1}{d_\sigma} \sum_{j' \preceq j} \boldsymbol{\lambda}[j'] \boldsymbol{q}_{j'}[\sigma] \boldsymbol{q}^*[\sigma^{(i)}(j')]$

8   **return** $\boldsymbol{q}^*$

Proof. By Lemma 3.5 the sequences $(\boldsymbol{\lambda}^t)$ and $(\boldsymbol{q}_\sigma^t)$ are κ-multiplicatively-stable. Moreover, given that the sequence $(\boldsymbol{\lambda}^t)$ is updated using OMWU, it will always be the case that $d_\sigma > 0$, for any $\sigma \in \Sigma^{(i)}$; in other words, Algorithm 5 will never visit the first "if" branch. Now fix any $t \geq 2$. We will show by induction that $\boldsymbol{q}^{*,t}[\sigma]$ is such that $\boldsymbol{q}^{*,t}[\sigma] \leq (1+2\kappa)^{3\mathfrak{D}[\sigma]-2} \boldsymbol{q}^{*,t-1}[\sigma]$ and $\boldsymbol{q}^{*,t}[\sigma] \geq (1-\kappa)^{3\mathfrak{D}^{(i)}[\sigma]-2} \boldsymbol{q}^{*,t-1}[\sigma]$, where $\mathfrak{D}^{(i)}[\sigma] \geq 1$ is the depth of sequence $\sigma \in \Sigma_*^{(i)}$ with respect to i's subtree. For the base case, let $\sigma = (j, a)$ such that $j \in \mathcal{J}^{(i)}$ corresponds to a root information set for player i. Then, it follows that $d_\sigma = \boldsymbol{\lambda}[j]$, in turn implying that $\boldsymbol{q}^{*,t} = \boldsymbol{q}_j^t[\sigma]$. Thus, $\boldsymbol{q}^{*,t}$ is indeed κ-multiplicatively stable. Next, consider some sequence $\sigma = (j, a)$ at depth $\mathfrak{D}^{(i)}[\sigma] \geq 2$ such that all ancestor sequences—*i.e.* all $\sigma^{(i)}(j')$ for $j' \preceq j$—satisfy the inductive hypothesis. Then, it follows that

$$\boldsymbol{q}^{*,t}[\sigma] = \frac{\sum_{j' \preceq j} \boldsymbol{\lambda}^t[j'] \boldsymbol{q}_{j'}^t[\sigma] \boldsymbol{q}^{*,t}[\sigma^{(i)}(j')]}{\sum_{j' \preceq j} \boldsymbol{\lambda}^t[j']} \tag{39}$$

$$\leq \frac{(1+\kappa)^2}{1-\kappa} \frac{\sum_{j' \preceq j} \boldsymbol{\lambda}^{t-1}[j'] \boldsymbol{q}_{j'}^{t-1}[\sigma] \boldsymbol{q}^{*,t}[\sigma^{(i)}(j')]}{\sum_{j' \preceq j} \boldsymbol{\lambda}^{t-1}[j']} \tag{40}$$

$$\leq (1+2\kappa)^3 (1+2\kappa)^{3\mathfrak{D}^{(i)}[\sigma]-5} \boldsymbol{q}^{*,t-1}[\sigma] \tag{41}$$

$$\leq (1+2\kappa)^{3\mathfrak{D}^{(i)}[\sigma]-2} \boldsymbol{q}^{*,t-1}[\sigma],$$

where (39) derives from Algorithm 5, (40) uses the κ-multiplicative stability of $\boldsymbol{\lambda}^t$'s and $\boldsymbol{q}_{\cdot j}^t$'s, and (41) leverages the inductive hypothesis, as well as the fact that $1/(1-\kappa)/ \leq (1+2\kappa)$ for $\kappa \leq 1/2$. Similar reasoning yields:

$$\boldsymbol{q}^{*,t}[\sigma] = \frac{\sum_{j' \preceq j} \boldsymbol{\lambda}^t[j'] \boldsymbol{q}_{j'}^t[\sigma] \boldsymbol{q}^{*,t}[\sigma^{(i)}(j')]}{\sum_{j' \preceq j} \boldsymbol{\lambda}^t[j']}$$

$$\geq \frac{(1-\kappa)^2}{1+\kappa} \frac{\sum_{j' \preceq j} \boldsymbol{\lambda}^{t-1}[j'] \boldsymbol{q}_{j'}^{t-1}[\sigma] \boldsymbol{q}^{*,t}[\sigma^{(i)}(j')]}{\sum_{j' \preceq j} \boldsymbol{\lambda}^{t-1}[j']}$$

$$\geq (1-\kappa)^3 (1-\kappa)^{3\mathfrak{D}^{(i)}[\sigma]-5} \boldsymbol{q}^{*,t-1}[\sigma]$$

$$\geq (1-\kappa)^{3\mathfrak{D}^{(i)}[\sigma]-2} \boldsymbol{q}^{*,t-1}[\sigma].$$

As a result, if $\mathfrak{D}^{(i)}$ is the depth of i's subtree, we conclude that $\boldsymbol{q}^{*,t}[\sigma] \leq (1+2\kappa)^{3\mathfrak{D}^{(i)}-2} \boldsymbol{q}^{*,t-1}[\sigma] \leq e^{6\mathfrak{D}^{(i)}\kappa - 4\kappa} \boldsymbol{q}^{*,t-1}[\sigma] \leq (1 + 12\mathfrak{D}^{(i)}\kappa) \boldsymbol{q}^{*,t-1}[\sigma]$, where we used the inequality $e^x \leq 1 + 2x$ for $x \in [0, 1/2]$. Moreover, we have that $\boldsymbol{q}^{*,t}[\sigma] \geq (1-\kappa)^{3\mathfrak{D}^{(i)}-2} \boldsymbol{q}^{*,t-1}[\sigma] \geq (1 - 3\mathfrak{D}^{(i)}\kappa) \boldsymbol{q}^{*,t-1}[\sigma]$, by Bernoulli's inequality. This concludes the proof. $\square$

Observe that the derived bound on stability is slightly better compared to that for EFCE (Corollary 3.12). Consequently, having established the stability of the fixed points, we can apply Theorem 3.2 to derive a $\tilde{\Psi}^{(i)}$ stable-predictive regret minimizer for each player $i \in [n]$. Then, using the same steps as in Appendix B.3 we arrive at the following conclusion:

Corollary C.6 (Accelerated Convergence to EFCCE). *There exist uncoupled no-regret learning dynamics so that after T iterations the correlated distribution of play converges to an extensive-form coarse correlated equilibrium at a rate of $O(T^{-3/4})$.*

D Description of Game Instances used in the Experiments

In this section we provide a description of the game instances used in our experiments. The parameters associated with each game are summarized in Table 1.

Kuhn poker. First, we experimented on a *three-player* variant of the popular benchmark game known as *Kuhn poker* (Kuhn 1950). In our version, a deck of three cards—a Jack, a Queen, and a King—is employed. Players initially commit a single chip to the pot, and privately receive a single card. The first player can either *check* or *bet* (*i.e.* place an extra chip). Then, the second player can in turn check or bet if the first player checked, or *folded/called* in response to the first player's bet. If no betting occurred in the previous rounds, the third player can either check or bet. In the contrary case, the player can either fold or call. Following a bet of the second player (or respectively the third player), the first player (or respectively the first and the second players) has to decide whether to fold or to call. At the *showdown*, the player with the *highest* card—who has not folded in a previous round—gets to win all the chips committed in the pot.

Sheriff. Our second benchmark is a bargaining game inspired by the board game *Sheriff of Nottingham*, introduced by (Farina et al. 2019b). In particular, we used the *baseline* version of the game. This game consists of two players: the *Smuggler* and the *Sheriff*. The smuggler must originally come up with a number $n \in \{0, 1, 2, 3\}$ which corresponds to the number of illegal items to be loaded in the cargo. It is assumed that each illegal item has a fixed value of 1. Subsequently, 2 rounds of bargaining between the two players follow. At each round, the Smuggler decides on a bribe ranging from 0 to 3, and the Sheriff must decide whether or not the cargo will be inspected given the bribe amount. The Sheriff's decision is binding only in the last round of bargaining. In particular, if during the last round the Sheriff accepts the bribe, the game stops with the Smuggler obtaining a utility of n minus the bribe amount b that was proposed in the last bargaining round, while the Sheriff receives a utility equal to b. On the other hand, if the Sheriff does not accept the bribe in last bargaining round and decides to inspect the cargo, there are two possible alternatives. If the cargo has no illegal items (*i.e.* $n = 0$), the Smuggler receives the fixed amount of 3, while the utility of the Sheriff is set to be -3. In the contrary case, the utility of the smuggler is assumed to be $-2n$, while the utility of the Sheriff is $2n$.

Liar's dice. The final benchmark we experimented on is the game of *Liar's dice*, introduced by Lisý, Lanctot, and Bowling (2015). In the three-player variant, the beginning of the game sees each of the three players privately roll an unbiased 3-face die. Then, the players have to sequentially make claims about their private information. In particular, the first player may announce any face value up to 3, as well as the minimum number of dice that the player claims are showing that value among the dice of *all* players. Then, each player can either make a higher bid, or challenge the previous claim by declaring the previous agent a "liar". More precisely, it is assumed that a bid is higher than the previous one if either the face value is higher, or if the claimed number of dice is greater. If the current claim is challenged, all the dices must be revealed. If the claim was valid, the last bidder wins and receives a reward of $+1$, while the challenger suffers a negative payoff of -1. Otherwise, the utilities obtained are reversed. Any other player will receive 0 utility.

Game	Players	Decision points	Sequences	Leaves
Kuhn poker	3	36	75	78
Sheriff	2	73	222	256
Liar's dice	3	1536	3069	13797

Table 1: The parameters associated with each game.