Faster No-Regret Learning Dynamics for Extensive-Form Correlated and Coarse Correlated Equilibria

IOANNIS ANAGNOSTIDES*, Carnegie Mellon University, USA
GABRIELE FARINA*, Carnegie Mellon University, USA
CHRISTIAN KROER, Columbia University, USA
ANDREA CELLI, Bocconi University, Italy
TUOMAS SANDHOLM, Carnegie Mellon University & Strategy Robot, Inc. & Optimized Markets, Inc. & Strategic Machine, Inc., USA

A recent emerging trend in the literature on learning in games has been concerned with providing faster 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 both sequential and simultaneous moves, as well as imperfect information. In this paper we establish faster no-regret learning dynamics for *extensive-form correlated equilibria (EFCE)* in multiplayer general-sum imperfect-information extensive-form games. When all players follow our 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 *Phi-regret minimization* via predictions. One of our key technical contributions—that enables us to employ our generic template—is to characterize the stability of fixed points associated with *trigger deviation functions* through a refined perturbation analysis of a structured Markov chain. Furthermore, for the simpler solution concept of extensive-form *coarse* correlated equilibrium (EFCCE) we give a new succinct closed-form characterization of the associated fixed points, bypassing the expensive computation of stationary distributions required for EFCE. Our results place EFCCE closer to *normal-form coarse correlated equilibria* in terms of the per-iteration complexity, although the former prescribes a much more compelling notion of correlation. Finally, experiments conducted on standard benchmarks corroborate our theoretical findings.

*Both authors contributed equally to the paper.

Authors' addresses: Ioannis Anagnostides, ianagnos@cs.cmu.edu, Carnegie Mellon University, 5000 Forbes Avenue, Pittsburgh, Pennsylvania, USA, 15232; Gabriele Farina, gfarina@cs.cmu.edu, Carnegie Mellon University, 5000 Forbes Avenue, Pittsburgh, Pennsylvania, USA, 15232; Christian Kroer, christian.kroer@columbia.edu, Columbia University, New York City, New York, USA, 10027; Andrea Celli, andrea.celli2@unibocconi.it, Bocconi University, Via Roberto Sarfatti, 25, Milan, Italy, 20100; Tuomas Sandholm, sandholm@cs.cmu.edu, Carnegie Mellon University & Strategy Robot, Inc. & Optimized Markets, Inc. & Strategic Machine, Inc., Pittsburgh, Pennsylvania, USA, 15232.

ArXiv preprint

Contents

1 INTRODUCTION

Game-theoretic solution concepts describe how rational agents should 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 include 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* [Farina et al., 2021b, Hoda et al., 2010, Kroer et al., 2020] 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 [Zinkevich et al., 2007] typically lead to better practical performance, although the worst-case convergence rate known in theory remains inferior. 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 decentralized learning dynamics. Second, finding a Nash equilibrium is computationally hard both in theory [Daskalakis et al., 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 Nash equilibria in general games.

A competing notion of rationality proposed by Aumann [1974] is that of *correlated equilibrium* (*CE*). Unlike NE, it is known that the former can be computed in polynomial time and, perhaps even more importantly, it can be attained through *uncoupled* learning dynamics where 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. In normal-form games (NFGs), a *correlated strategy* is defined as a probability distribution over joint action profiles, customarily modeled via a trusted external mediator that draws an action profile from this distribution and then privately recommends to each player their component. A correlated strategy is a CE if, for each player, the mediator's recommendation is the best action in expectation, assuming that all the other players follow their recommended actions [Aumann, 1974]. In NFGs it has long been known that uncoupled no-regret learning dynamics can converge to CE and *coarse correlated equilibria* (*CCE*) at a rate of $O(T^{-1/2})$ [Foster and Vohra, 1997, Hart and Mas-Colell, 2000]. More recently, it has been established that accelerated dynamics can converge at a rate of $\widetilde{O}(T^{-1})$ [Anagnostides et al., 2021, Daskalakis et al., 2021] in NFGs, where the notation $\widetilde{O}(\cdot)$ suppresses $\mathrm{polylog}(T)$ factors.

However, 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. Three natural extensions of CE to

extensive-form games are the *extensive-form correlated equilibrium (EFCE)* by von Stengel and Forges [2008], the *extensive-form coarse correlated equilibrium (EFCCE)* by Farina et al. [2020], and the *normal-form coarse correlated equilibrium (NFCCE)* by Celli et al. [2019a]. The set of those equilibria are such that, for any extensive-form game, EFCE $\subseteq$ EFCCE $\subseteq$ NFCCE. In an EFCE, the stronger of those notions of correlation, the mediator forms recommendations for each of the possible decision points an agent may encounter in the game, and recommended actions are gradually revealed to players as they reach new information sets; thus, the mediator must take into account the *evolution* of the players' beliefs throughout the game. Because of the sequential nature, the presence of private information in the game, and the gradual revelation of recommendations, the constraints associated with EFCE are significantly more complex than for normal-form games. For these reasons, the question of whether uncoupled learning dynamics can converge to an EFCE was only recently resolved by Celli et al. [2020]. Moreover, in a follow-up work the authors also established an explicit rate of convergence of $O(T^{-1/2})$ [Farina et al., 2021a]. Our paper is primarily concerned with the following fundamental question:

Can we develop faster uncoupled no-regret learning dynamics for EFCE?

We affirmatively answer this question by developing dynamics converging at a rate of $O(T^{-3/4})$ to an EFCE. Furthermore, we also study learning dynamics for the simpler solution concept of EFCCE. More precisely, although accelerated learning dynamics for EFCE can be automatically employed for EFCCE (since the set of EFCEs forms a subset of the set of EFCCEs), all the known learning dynamics for EFCE have large per-iteration complexity. Indeed, they require as an intermediate step the expensive computation of the stationary distributions of multiple Markov chains. Thus, the following natural question arises: *Are there learning dynamics for EFCCE that avoid the expensive computation of stationary distributions?* We answer this question in the positive. Our results reveal that EFCCE is more akin to NFCCE than to EFCE from a learning perspective, although EFCE prescribes a much more compelling notion of correlation than NFCCE.

1.1 Contributions

Our first primary contribution is to develop faster no-regret learning dynamics for EFCE:

Theorem 1.1. *On any 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.*

This substantially improves over the prior best known rate of $O(T^{-1/2})$ recently established by Farina et al. [2021a]. To achieve this result we employ 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 *Phi-regret* minimization [Gordon et al., 2008, Greenwald and Jafari, 2003] by providing a general template for stable-predictive Phi-regret minimization (Theorem 3.2). The importance of Phi-regret is that it leads to substantially more compelling notions of hindsight rationality, well-beyond the usual *external* regret [Gordon et al., 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. [2021a] is to cast convergence to an EFCE as a Phi-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.

From a technical standpoint, in order to apply our generic template for accelerated Phi-regret minimization (Theorem 3.2), we establish two separate ingredients. First, we develop a *predictive*

external regret minimizer for the set of transformations associated with EFCE. This deviates from the construction of Farina et al. [2021a] in that we have to additionally guarantee and preserve the predictive bounds throughout the construction. Further, our algorithm combines optimistic regret minimization—under suitable DGFs—for the sequence-form polytope, with *regret decomposition* in the style of CFR. While these have been the two main paradigms employed in EFGs, they were used separately in the past. We refer to Figure 2 for a detailed description of our algorithm.

The second central 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 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 (Lemma 4.11). To do this, we deviate from prior techniques (*e.g.*, [Candogan et al., 2013, Chen and Peng, 2020]) that used the Markov chain tree theorem, and instead use an alternative linear-algebraic characterization for the eigenvectors of the underlying Laplacian system. This leads to a rate of convergence that depends *polynomially* on the description of the game, which is crucial for the practical applicability of the dynamics.

Next, we shift our attention to learning dynamics for EFCCE. We first introduce the notion of *coarse trigger deviation functions*, and we show that if each player employs a no-coarse-trigger-regret algorithm, the correlated distribution of play converges to an EFCCE (Theorem 2.11). This allows for a unifying treatment of EFCE and EFCCE. Moreover, we show that, unlike all existing methods for computing fixed points of trigger deviation functions, the fixed points of *coarse* trigger deviation functions admit a succinct closed-form characterization (Theorem 5.1); in turn, this enables us to obtain a much more efficient algorithm for computing them (Algorithm 1). From a practical standpoint, this is crucial as it substantially reduces the per-iteration complexity of the dynamics, placing EFCCE closer to NFCCE in terms of the underlying complexity, even though EFCCE prescribes a stronger notion of correlation. Another implication of our closed-form characterization is an improved stability analysis for the fixed points, which is much less technical than the one we give for EFCE (Proposition 5.2). 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 et al. [2015], showing that one can bypass the adversarial $\Omega(T^{-1/2})$ barrier for the incurred average regret if *both* players in a zero-sum game employ an uncoupled variant of Nesterov's 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 $O(T^{-1})$ can be obtained with a remarkably simple variant of 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 in [Syrgkanis et al., 2015] from external to swap regret using the celebrated construction of Blum and Mansour [2007], thereby establishing a rate of convergence of $O(T^{-3/4})$ to CE. Even more recent work [Anagnostides et al., 2021, Daskalakis et al., 2021] has established a near-optimal rate of convergence of $\widetilde{O}(T^{-1})$ to correlated equilibria in normal-form games when all players leverage *optimistic multiplicative weights update (OMWU)*, where $\widetilde{O}(\cdot)$ hides $\mathrm{polylog}(T)$ factors. Extending these results to EFCE presents a considerable challenge since their techniques crucially rely on the softmax-type strucure of OMWU on the simplex, as well as the particular structure of the associated fixed points.

Correlated equilibria in extensive-form games are much less understood than Nash equilibria. It is known that a feasible EFCE can also be computed efficiently through a variant of the *Ellipsoid algorithm* [Jiang and Leyton-Brown, 2015, Papadimitriou and Roughgarden, 2008], while 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. [2019b] devised variants of the CFR algorithm that provably converge to an NFCCE, a solution concept much less appealing than EFCE in extensive-form games [Gordon et al., 2008]. Finally, Morrill et al. [2021a,b] characterize different hindsight rationality notions in EFGs, associating each solution concept with suitable $O(T^{-1/2})$ no-regret learning dynamics.

2 PRELIMINARIES

In this section we introduce the necessary background related to extensive-form games (EFGs), correlated equilibria in EFGs, and regret minimization. A comprehensive treatment on EFGs can be found in [Shoham and Leyton-Brown, 2009], while for an introduction to the theory of learning in games the reader is referred to the excellent book of Cesa-Bianchi and Lugosi [2006].

Conventions. In the sequel we use the $O(\cdot)$ notation to suppress (universal) constants. We typically use subscripts to indicate the player or some element in the game tree uniquely associated with a given player, such as a decision point; to lighten our notation, the associated player is not made explicit in the latter case. Superscripts are reserved almost exclusively for time indexes. Finally, the k-th coordinate of a vector $\boldsymbol{x} \in \mathbb{R}^d$ will be denoted by $\boldsymbol{x}[k]$.

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 to 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 some 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 of the game. 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 player i. Thus, for any nodes $h, h' \in j$ we have $\mathcal{A}_h = \mathcal{A}_{h'}$. As usual, we assume that the game satisfies *perfect recall*: players never forget information once acquired. This implies, in particular, that for any nodes $h, h' \in j$ the sequence of i's actions from the root until h must coincide with the sequence from the root to node h'; otherwise, i would be able to distinguish between 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.

	Description
$\mathcal{J}_i$	Information sets of player i
$\mathcal{A}_j$	Set of actions at information set j
Σ_i	Set of sequences of player i
Σ_i^*	Set of sequences of player i excluding $\varnothing$
Σ_j	Set of sequences at $j \in \mathcal{J}_i$ and all of its descendants
$\mathfrak{D}_i$	Maximum depth of any $j \in \mathcal{J}_i$
$\mathcal{Z}$	Number of leaves
$\|\mathcal{Q}_i\|_1$	Upper bound on the ℓ_1-norm of any $\boldsymbol{x} \in \mathcal{Q}_i$
Π_i	Deterministic sequence-form strategies of player i
Π_j	Deterministic sequence-form strategies rooted at $j \in \mathcal{J}_i$
$\mathcal{Q}_i$	Sequence-form strategies of player i
$\mathcal{Q}_j$	Sequence-form strategies rooted at $j \in \mathcal{J}^{(i)}$
Π	Set of joint deterministic sequence-form strategies

Table 1. Summary of EFG notation.

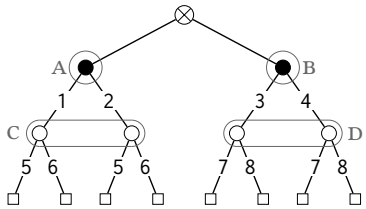

Fig. 1. Example of a two-player EFG.

Sequence-form Strategies. For a player $i \in [n]$, an information set $j \in \mathcal{J}_i$, and an action $a \in \mathcal{A}_j$, we will denote with $\sigma = (j, a)$ the *sequence* of i's actions encountered on the path from the root of the game until (and included) action a. For notational convenience, we will use the special symbol $\varnothing$ to denote the *empty sequence*. Then, i's set of sequences is defined as $\Sigma_i := \{(j, a) : j \in \mathcal{J}_i, a \in \mathcal{A}_j\} \cup \{\varnothing\}$; we will also use the notation $\Sigma_i^* := \Sigma_i \setminus \{\varnothing\}$. For a given information set $j \in \mathcal{J}_i$ we will use $\sigma_j \in \Sigma_i$ 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 $\sigma_j = \varnothing$, and we say that j is a *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:

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

This definition ensures the probability mass conservation for the sequence-form strategies along every possible decision point. The probability of playing action a at information set $j \in \mathcal{J}_i$ can be obtained by dividing $\boldsymbol{q}[(j, a)]$ by $\boldsymbol{q}[\sigma_j]$. 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 by $\mathcal{Q}_j$. 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 Π_j. A well-known implication of Kuhn's theorem [Kuhn, 1953] is that $\mathcal{Q}_i = \operatorname{co} \Pi_i$, and $\mathcal{Q}_j = \operatorname{co} \Pi_j$, 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

$$u_i(\boldsymbol{\pi}) := \sum_{z \in \mathcal{Z}} p_c(z) u_i(z) \mathbb{1}\{\boldsymbol{\pi}_k[\sigma_{k,z}] = 1, \forall k \in [n]\},$$

where $\sigma_{i,z}$ denotes the last sequence of player i before reaching the terminal node $z \in \mathcal{Z}$. For the convenience of the reader, in Table 1 we have summarized the main notation related to EFGs used throughout this paper.

An Illustrative Example. To further clarify some of the concepts we have introduced so far, we illustrate a simple two-player EFG in Figure 1. Black nodes belong to player 1, white round nodes to player 2, square nodes are terminal nodes (aka leaves), and the crossed node is a chance node. Player 2 has two information sets, $\mathcal{J}_2 \coloneqq \{\mathrm{C}, \mathrm{D}\}$, each containing two nodes. This captures the lack of knowledge regarding the action played by player 1. In contrast, the outcome of the chance move is observed by both players. At the information set C, player 2 has two possible actions, $\mathcal{A}_\mathrm{C} \coloneqq \{5, 6\}$. Thus, one possible sequence for player 2 is the pair $\sigma = (\mathrm{C}, 5) \in \Sigma_2$.

2.2 Online Learning 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 a 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* $\ell^{(t)} : \mathcal{X} \ni \boldsymbol{x} \mapsto \langle \boldsymbol{\ell}^{(t)}, \boldsymbol{x} \rangle$, and may alter its internal state accordingly.

The utility function $\ell^{(t)}$ could depend adversarially on the previous outputs $\boldsymbol{x}^{(1)}, \dots, \boldsymbol{x}^{(t-1)}$, but not on $\boldsymbol{x}^{(t)}$. The decision making is *online* in the sense that the regret minimizer can adapt to previously received information, but no information about future utilities is available. The performance of a regret minimizer is typically measured in terms of its *cumulative external regret* (or simply regret), defined, for a time horizon $T \in \mathbb{N}$, as follows.

$$\mathrm{Reg}^T \coloneqq \max_{\boldsymbol{x}^* \in \mathcal{X}} \sum_{t=1}^{T} \langle \boldsymbol{x}^*, \boldsymbol{\ell}^{(t)} \rangle - \sum_{t=1}^{T} \langle \boldsymbol{x}^{(t)}, \boldsymbol{\ell}^{(t)} \rangle. \tag{1}$$

That is, the performance of the online algorithm is compared to the best *fixed* strategy in *hindsight*. A regret minimizer is called *Hannan consistent* if, under any sequence of (bounded) utility functions, its regret grows sublinearly with T; that is, $\mathrm{Reg}^T = o(T)$. It is well-known that broad families of learning algorithms incur $O(\sqrt{T})$ regret under *any* sequence of utility functions, which also matches the lower bound in the adversarial regime (see [Cesa-Bianchi and Lugosi, 2006]).

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

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

When the set of transformations Φ coincides with the set of *constant functions* we recover the notion of external regret given in (1). However, Phi-regret is substantially stronger and it yields more appealing notions of hindsight rationality [Gordon et al., 2008], incorporating the notion of *swap regret* [Blum and Mansour, 2007].

Optimistic Regret Minimization. An emerging subfield of online learning ([Chiang et al., 2012, Rakhlin and Sridharan, 2013a]) studies the improved performance guarantees one can obtain when the utilities observed by the regret minimization algorithm possess additional structure, typically in the form of *small variation*. Such considerations diverge from the adversarial regime we previously described, and are motivated—among others—by the fact that in many settings the utility functions are themselves selected by *regularized learning algorithms*. For our purposes we shall employ the following definition, which is a modification of the RVU property [Syrgkanis et al., 2015].

Definition 2.2 (Stable-Predictive)**.** Let $\mathcal{R}$ be a regert minimizer and $\|\cdot\|$ be any norm. $\mathcal{R}$ is said to be κ-*stable* with respect to $\|\cdot\|$ if for all $t \geq 2$ the strategies output by $\mathcal{R}$ are such that

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

Moreover, $\mathcal{R}$ is said to be (α, β)-*predictive* with respect to $\|\cdot\|$ if its regret Reg^T can be bounded as

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

for any sequence of utilities $\boldsymbol{\ell}^{(1)}, \dots, \boldsymbol{\ell}^{(T)}$, where $\|\cdot\|_*$ is the dual norm of $\|\cdot\|$.

In the above definition $\boldsymbol{m}^{(t)}$ serves as the *prediction* of the regret minimizer $\mathcal{R}$ at time $t \geq 1$. While traditional online algorithms are not known to satisfy (2), we will next present natural variants which are indeed stable-predictive in the sense of Definition 2.2.

Optimistic Follow the Regularized Leader. Let d be a 1-strongly convex *distance generating function (DGF)* with respect to a norm $\|\cdot\|$, and $\eta > 0$ be the *learning rate*. The update rule of *optimistic follow the regularized leader (OFTRL)* takes the following form for $t \geq 2$:

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

where $\boldsymbol{m}^{(t)}$ is the prediction at time t, and $\boldsymbol{x}^{(1)} \coloneqq \arg\min_{\boldsymbol{x} \in \mathcal{X}} d(\boldsymbol{x})$. Unless specified otherwise, it will be tacitly assumed that $\boldsymbol{m}^{(t)} \coloneqq \boldsymbol{\ell}^{(t-1)}$, for $t \geq 1$, where we conventionally let $\boldsymbol{\ell}^{(0)} \coloneqq \mathbf{0}$. Syrgkanis et al. [2015] established the following property:

Lemma 2.3. (OFTRL) *is* $(\Omega_d/\eta, \eta)$*-predictive*[1] *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,* $\Omega_d \coloneqq \max_{\boldsymbol{x}, \boldsymbol{x}' \in \mathcal{X}} \{d(\boldsymbol{x}) - d(\boldsymbol{x}')\}$.

The *entropic regularizer* on the simplex is defined as $d(\boldsymbol{x}) \coloneqq \sum_{k=1}^{d} \boldsymbol{x}[k] \log \boldsymbol{x}[k]$, and it is well-known to be 1-strongly convex with respect to the ℓ_1-norm. This OFTRL setup will be referred to as *optimistic multiplicative weights updates (OMWU).*[2]

We will also require a suitable DGF for the sequence-form polytope. To this end, we will employ the *dilatable global entropy* DGF, recently introduced by Farina et al. [2021b].

Definition 2.4 ([Farina et al., 2021b])**.** The *dilatable global entropy distance generating function* $d : \mathcal{Q} \to \mathbb{R}_{\geq 0}$ is defined as

$$d(\boldsymbol{x}) \coloneqq \sum_{\sigma \in \Sigma} \boldsymbol{w}[\sigma] \boldsymbol{x}[\sigma] \log(\boldsymbol{x}[\sigma]).$$

The vector $\boldsymbol{w} \in \mathbb{R}^{|\Sigma|}$ is defined recursively as

$$\begin{aligned} \boldsymbol{w}[\varnothing] &= 1; \\ \boldsymbol{w}[(j,a)] &= \boldsymbol{y}[j] - \sum_{j' : \sigma_{j'} = (j,a)} \boldsymbol{y}[j'], \quad \forall (j,a) \in \Sigma, \end{aligned}$$

where

$$\boldsymbol{y}[j] = 1 + \max_{a \in \mathcal{A}_j} \left\{ \sum_{j' : \sigma_{j'} = (j,a)} \boldsymbol{y}[j'] \right\}, \quad \forall j \in \mathcal{J}. \tag{3}$$

[1] Syrgkanis et al. [2015] only stated this for the simplex, but their proof readily extends to arbitrary convex and compact sets.
[2] When $\boldsymbol{m}^{(t)} \coloneqq \mathbf{0}$, for all $t \geq 1$, we recover *multiplicative weights updates (MWU).*

This DGF is "nice" (in the parlance of Hoda et al. [2010]) since its gradient, as well as the gradient of its convex conjugate, can be computed *exactly* in linear time in $|\Sigma|$—the dimension of the domain. Our analysis will require the following characterization.

Lemma 2.5 ([Farina et al., 2021b]). *The dilatable global entropy d of Definition 2.4 is a DGF for the sequence-form polytope $\mathcal{Q}$. Moreover, it is $1/\|\mathcal{Q}\|_1$-strongly convex on* $\operatorname{relint}\mathcal{Q}$ *with respect to the $\|\cdot\|_1$ norm, where $\|\mathcal{Q}\|_1 = \max_{\boldsymbol{q}\in\mathcal{Q}} \|\boldsymbol{q}\|_1$. Finally, the d-diameter Ω_d of $\mathcal{Q}$ is at most $\|\mathcal{Q}\|_1^2 \max_{j\in\mathcal{J}} \log|\mathcal{A}_j|$.*

In the sequel we will instantiate (OFTRL) with dilatable global entropy as DGF to construct a stable-predictive regret minimizer for the sequence-form strategy polytope.

2.3 Extensive-Form Correlated and Coarse Correlated Equilibrium

In this subsection we introduce the notion of an *extensive-form correlated and coarse correlated equilibrium* (henceforth EFCE and EFCCE respectively). First, for EFCE we will work with the definition used in [Farina et al., 2019d], which is equivalent to the original one due to von Stengel and Forges [2008]. To this end, let us introduce the concept of a *trigger deviation function*.

Definition 2.6. Consider some player $i \in [n]$. A *trigger deviation function* with respect to a *trigger sequence* $\hat{\sigma} = (j, a) \in \Sigma_i^*$ and a *continuation strategy* $\hat{\boldsymbol{\pi}}_i \in \Pi_j$ is any linear function $f : \mathbb{R}^{|\Sigma_i|} \to \mathbb{R}^{|\Sigma_i|}$ with the following properties.

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

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

  for all $\sigma \in \Sigma_i$.

We will let $\Psi_i \coloneqq \{\phi_{\hat{\sigma}\to\hat{\boldsymbol{\pi}}_i} : \hat{\sigma} = (j, a) \in \Sigma_i^*, \hat{\boldsymbol{\pi}}_i \in \Pi_j\}$ 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 2.7 (EFCE)**.** A probability distribution $\boldsymbol{\mu} \in \Delta(\Pi)$ is an ϵ-EFCE, for $\epsilon \geq 0$, 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 $\boldsymbol{\pi} = (\boldsymbol{\pi}_1, \ldots, \boldsymbol{\pi}_n) \in \Pi$. We say that $\boldsymbol{\mu} \in \Delta(\Pi)$ is an EFCE if it is a 0-EFCE.

Theorem 2.8 ([Farina et al., 2021a])**.** *Suppose that for every player $i \in [n]$ the sequence of deterministic sequence-form strategies $\boldsymbol{\pi}_i^{(1)}, \ldots, \boldsymbol{\pi}_i^{(T)} \in \Pi_i$ incurs Ψ_i-regret at most Reg_i^T under 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 $\epsilon \coloneqq \frac{1}{T}\max_{i\in[n]} \operatorname{Reg}_i^T$.

Similarly, we introduce the closely related notion of a *coarse* trigger deviation function.

Definition 2.9 (Coarse Trigger Deviation Functions)**.** Consider some player $i \in [n]$. A *coarse trigger deviation function* with respect to an information set $j \in \mathcal{J}_i$ and a continuation strategy $\hat{\boldsymbol{\pi}}_i \in \Pi_j$ is any linear function $f : \mathbb{R}^{|\Sigma_i|} \to \mathbb{R}^{|\Sigma_i|}$ with the following properties:

- For any $\boldsymbol{\pi}_i \in \Pi_i$ such that $\boldsymbol{\pi}_i[\sigma_j] = 0$ it holds that $f(\boldsymbol{\pi}_i) = \boldsymbol{\pi}_i$;

- Otherwise,

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

for all $\sigma \in \Sigma_i$.

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

Definition 2.10 (EFCCE)**.** A probability distribution $\boldsymbol{\mu} \in \Delta(\Pi)$ is an ϵ-EFCCE, for $\epsilon \geq 0$, if for every player $i \in [n]$ and every coarse trigger deviation function $\phi_{j \to \hat{\boldsymbol{\pi}}_i} \in \widetilde{\Psi}_i$,

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

where $\boldsymbol{\pi} = (\boldsymbol{\pi}_1, \dots, \boldsymbol{\pi}_n) \in \Pi$. We say that $\boldsymbol{\mu} \in \Delta(\Pi)$ is an EFCCE if it is a 0-EFCCE.

Analogously to Theorem 2.8, we show (in Appendix A.2) that if all players employ a $\widetilde{\Psi}_i$-regret minimizer, the correlated distribution of play converges to an EFCCE.

Theorem 2.11. *Suppose that for every player $i \in [n]$ the sequence of deterministic sequence-form strategies $\boldsymbol{\pi}_i^{(1)}, \dots, \boldsymbol{\pi}_i^{(T)} \in \Pi_i$ incurs $\widetilde{\Psi}_i$-regret at most Reg_i^T under 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-EFCCE, where $\epsilon \coloneqq \frac{1}{T} \max_{i \in [n]} \mathrm{Reg}_i^T$.

3 ACCELERATING PHI-REGRET MINIMIZATION WITH OPTIMISM

In this section we present a general construction for obtaining improved Phi-regret guarantees. Our template is then instantiated in Sections 4 and 5 to obtain faster dynamics for EFCE and EFCCE.

Our approach combines the framework of Gordon et al. [2008] with stable-predictive (aka. optimistic) regret minimization. As in [Gordon et al., 2008], we combine 1) a regret minimizer that outputs a linear transformation $\phi^{(t)} \in \Phi$ at every time t, and 2) a fixed-point oracle for each $\phi^{(t)} \in \Phi$. However, 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 $(\phi^{(t)})$. We say that $\mathcal{R}_\Phi$ is *fixed point κ-stable* with respect to a norm $\|\cdot\|$ if the following conditions hold.

- Every $\phi^{(t)}$ 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 $\boldsymbol{x}^{(t)}$ with $\boldsymbol{x}^{(t)} = \phi^{(t)}(\boldsymbol{x}^{(t)})$, there is $\boldsymbol{x}^{(t+1)} = \phi^{(t+1)}(\boldsymbol{x}^{(t+1)})$ such that $\|\boldsymbol{x}^{(t+1)} - \boldsymbol{x}^{(t)}\| \leq \kappa$.

In this context, we will show how to construct a stable-predictive Φ-regret minimizer starting from the following two components.

(1) $\mathcal{R}_\Phi$: An (A, B)-predictive fixed point κ-stable regret minimizer for the set Φ;
(2) StableFPOracle$(\phi; \widetilde{\boldsymbol{x}}, \kappa, \epsilon)$: 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) $\|\boldsymbol{x} - \widetilde{\boldsymbol{x}}\| \leq \kappa$ (the existence of such a fixed point is guaranteed by the fixed point κ-stability assumption on the regret minimizer).

Theorem 3.2 (Stable-Predictive Phi-Regret Minimization)**.** *Consider an (A, B)-predictive regret minimizer $\mathcal{R}_\Phi$ with respect to $\|\cdot\|_1$ for a set of linear transformations Φ on $\mathcal{X}$. Moreover, suppose that*

$\mathcal{R}_\Phi$ is fixed point κ-stable. Then, if we have access to a StableFPOracle*, we can construct a κ-stable algorithm with Φ-regret Reg^T bounded as*

$$\mathrm{Reg}^T \le A + 2B\sum_{t=1}^{T}\|\boldsymbol{\ell}^{(t)} - \boldsymbol{\ell}^{(t-1)}\|_\infty^2 + 2B\|\boldsymbol{\ell}\|_\infty^2\sum_{t=1}^{T}\|\boldsymbol{x}^{(t)} - \boldsymbol{x}^{(t-1)}\|_\infty^2 + \|\boldsymbol{\ell}\|_\infty\sum_{t=1}^{T}\epsilon^{(t)},$$

where $\epsilon^{(t)}$ is the error of StableFPOracle *at time t, and $\|\boldsymbol{\ell}^{(t)}\|_\infty \le \|\boldsymbol{\ell}\|_\infty$ for any $t \ge 1$. It is also assumed that $\|\boldsymbol{x}\|_\infty \le 1$ for all $\boldsymbol{x} \in \mathcal{X}$.*

The ℓ_1 norm is used only for convenience; the theorem readily extends under any equivalent norm. The proof of Theorem 3.2 builds on the construction of Gordon et al. [2008], and it is included in Appendix A.3.

4 FASTER CONVERGENCE TO EFCE

Our framework (Theorem 3.2) reduces accelerating Φ-regret minimization to (i) developing a predictive regret minimizer for the set Φ, and (ii) establishing the stability of the fixed points (StableFPOracle). In this section we establish these components for the set of all possible trigger deviations functions (Definition 2.6), leading to faster convergence to EFCE. In particular, Section 4.1 is concerned with the former task while Section 4.2 is concerned with the latter.

4.1 Constructing a Predictive Regret Minimizer for Ψ_i

Here we develop a regret minimizer for the set $\mathrm{co}\,\Psi_i$, the convex hull of all trigger deviation functions (Definition 2.6) of player $i \in [n]$. Given that $\mathrm{co}\,\Psi_i \supseteq \Psi_i$, this will immediately imply a Ψ_i-regret minimizer—after applying Theorem 3.2. To this end, the set $\mathrm{co}\,\Psi_i$ can be evaluated in two stages. First, for a fixed sequence $\hat{\sigma} = (j, a) \in \Sigma_i^*$ we define the set $\Psi_{\hat{\sigma}} \coloneqq \mathrm{co}\left\{\phi_{\hat{\sigma}\to\hat{\boldsymbol{\pi}}_i} : \hat{\boldsymbol{\pi}}_i \in \Pi_j\right\}$. Then, we take the convex hull of all $\Psi_{\hat{\sigma}}$; that is, $\mathrm{co}\,\Psi_i = \mathrm{co}\{\Psi_{\hat{\sigma}} : \hat{\sigma} \in \Sigma_i^*\}$. In light of this, we first develop a predictive regret minimizer for the set $\Psi_{\hat{\sigma}}$, for any $\hat{\sigma} \in \Sigma_i^*$. These individual regret minimizers are then combined using a *regret circuit* to conclude the construction in Theorem 4.5. The overall algorithm is illustrated in Figure 2. All of the omitted poofs and pseudocode for this section are included in Appendix A.4.

4.1.1 Predictive Regret Minimizer for the set $\Psi_{\hat{\sigma}}$. Consider a sequence $\hat{\sigma} \in \Sigma_i^*$. We claim that the set of transformations $\Psi_{\hat{\sigma}} \coloneqq \mathrm{co}\left\{\phi_{\hat{\sigma}\to\hat{\boldsymbol{\pi}}_i} : \hat{\boldsymbol{\pi}}_i \in \Pi_j\right\}$ is the image of $\boldsymbol{Q}_j$ under the affine mapping $h_{\hat{\sigma}} : \boldsymbol{q} \mapsto \phi_{\hat{\sigma}\to\boldsymbol{q}}$. Hence, it is not hard to see that a regret minimizer for $\Psi_{\hat{\sigma}}$ can be constructed starting from a regret minimizer for $\boldsymbol{Q}_j$. We now show that the predictive bound is preserved through this construction.

Proposition 4.1. *Consider a player $i \in [n]$ and any trigger sequence $\hat{\sigma} = (j, a) \in \Sigma_i^*$. There exists an algorithm which constructs a regret minimizer $\mathcal{R}_{\hat{\sigma}}$ with access to an (A, B)-predictive regret minimizer $\mathcal{R}_{\boldsymbol{Q}_j}$ for the set $\boldsymbol{Q}_j$ such that $\mathcal{R}_{\hat{\sigma}}$ is (A, B)-predictive.*

This proposition requires a predictive regret minimizer for the set $\boldsymbol{Q}_j$, for each $j \in \mathcal{J}_i$. To this end, we instantiate (OFTRL) with dilatable global entropy as DGF (Definition 2.4). Then, combining Lemma 2.3 with Lemma 2.5 leads to the following predictive bound.

Lemma 4.2. *Suppose that the regret minimizer $\mathcal{R}_{\boldsymbol{Q}_j}$ is instantiated with dilatable global entropy. Then, $\mathcal{R}_{\boldsymbol{Q}_j}$ is (A, B)-predictive with respect to $\|\cdot\|_1$, where $A = \frac{\|\boldsymbol{Q}_i\|_1^2 \max_{j\in\mathcal{J}_i}\log|\mathcal{A}_j|}{\eta}$ and $B = \eta\|\boldsymbol{Q}_i\|_1$.*

The discrepancy between this bound and the one in Lemma 2.3 derives from the fact that the modulus of convexity with respect to $\|\cdot\|_1$ for the dilatable global entropy is $1/\|\boldsymbol{Q}_i\|_1$ instead of 1. Alternatively, we also establish a predictive variant of CFR which can be used in place of OFTRL for performing regret minimization over the set $\boldsymbol{Q}_j$.

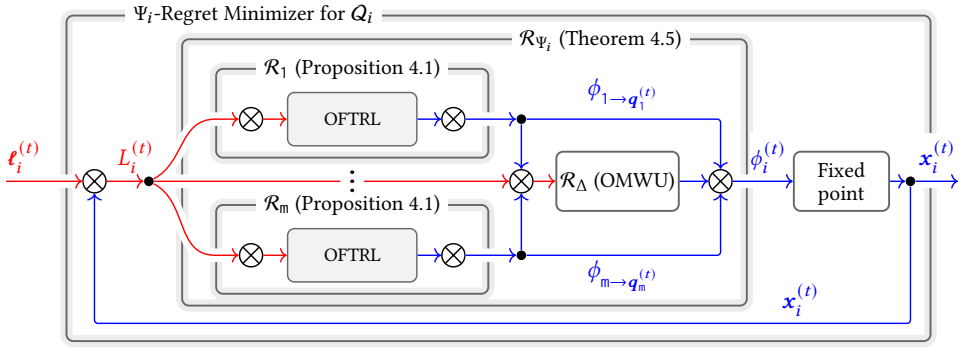

Fig. 2. An overview of the overall construction. For notational convenience we have let $\Sigma_i^* := \{1, 2, \dots, \mathtt{m}\}$. The symbol $\otimes$ in the figure denotes a multilinear transformation. We have used blue color for the iterates and red for the utilities. The algorithm first constructs a regret minimizer $\mathcal{R}_{\Psi_i}$ for the set Ψ_i (Theorem 4.5). This internally uses a regret minimizer $\mathcal{R}_\Delta$ which "mixes" the strategies of $\mathcal{R}_1, \dots, \mathcal{R}_{\mathtt{m}}$. In turn, the latter regret minimizers internally employ (OFTRL) with dilatable global entropy as DGF (Proposition 4.1). The last step can also be implemented using stable-predictive CFR (Theorem B.4), as we leverage for our experiments. Finally, $\mathcal{R}_{\Psi_i}$ is used to construct a stable-predictive Ψ_i-regret minimizer using the construction of Theorem 3.2.

Proposition 4.3 (Predictive CFR; Full Version in Theorem B.4)**.** *There exists a variant of CFR using OMWU which is (A, B)-predictive, where $A = O(\frac{\max_{j\in\mathcal{J}} \log|\mathcal{A}_j|}{\eta}\|\boldsymbol{Q}\|_1)$ and $B = O(\eta\|\boldsymbol{Q}\|_1^3)$.*

This construction follows the approach of Farina et al. [2019c], but here we make the dependencies on the size of the game explicit. The predictive bound we obtain for CFR is inferior to the one in Lemma 4.2, so the rest of our theoretical analysis will follow the "global" approach.

4.1.2 Predictive Regret Minimizer for co Ψ_i. The next step consists of appropriately combining the regret minimizers $\Psi_{\hat{\sigma}}$, for all $\hat{\sigma} \in \Sigma_i^*$, to a composite regret minimizer for the set co Ψ_i. To this end, we will use a *regret circuit* for the convex hull, formally introduced below.

Proposition 4.4 ([Farina et al., 2019b])**.** *Consider a collection of sets $\mathcal{X}_1, \dots, \mathcal{X}_m$, and let $\mathcal{R}_i$ be a regret minimizer for the set $\mathcal{X}_i$, for each $i \in [m]$. Moreover, let $\mathcal{R}_\Delta$ be a regret minimizer for the m-simplex Δ^m. A regret minimizer $\mathcal{R}_{\text{co}}$ for the set* co$\{\mathcal{X}_1, \dots, \mathcal{X}_m\}$ *can be constructed as follows.*

- *$\mathcal{R}_{\text{co}}$.* NextStrategy *obtains the next strategy $\boldsymbol{x}_i^{(t)}$ of each regret minimizer $\mathcal{R}_i$, as well as the next strategy $\boldsymbol{\lambda}^{(t)} = (\boldsymbol{\lambda}^{(t)}[1], \dots, \boldsymbol{\lambda}^{(t)}[m]) \in \Delta^m$ of $\mathcal{R}_\Delta$, and returns the corresponding convex combination: $\boldsymbol{\lambda}^{(t)}[1]\boldsymbol{x}_1^{(t)} + \cdots + \boldsymbol{\lambda}^{(t)}[m]\boldsymbol{x}_m^{(t)}$.*
- *$\mathcal{R}_{\text{co}}$.* ObserveUtility$(L^{(t)})$ *forwards $L^{(t)}$ to each of the regret minimizers $\mathcal{R}_1, \dots, \mathcal{R}_m$, while it forwards the utility function $(\boldsymbol{\lambda}[1], \dots, \boldsymbol{\lambda}[m]) \mapsto \boldsymbol{\lambda}[1]L^{(t)}(\boldsymbol{x}_1^{(t)}) + \cdots + \boldsymbol{\lambda}[m]L^{(t)}(\boldsymbol{x}_m^{(t)})$ to $\mathcal{R}_\Delta$.*

Then, if $\text{Reg}_1^T, \dots, \text{Reg}_m^T$ are the regrets accumulated by the regret minimizers $\mathcal{R}_1, \dots, \mathcal{R}_m$, and Reg_Δ^T is the regret of $\mathcal{R}_\Delta$, the regret Reg_{co}^T of the composite regret minmizers $\mathcal{R}_{\text{co}}$ can be bounded as

$$\text{Reg}_{\text{co}}^T \le \text{Reg}_\Delta^T + \max\{\text{Reg}_1^T, \dots, \text{Reg}_m^T\}.$$

Next, we leverage this construction to obtain the main result of this subsection: a predictive regret minimizer for the set of transformations co Ψ_i.

Theorem 4.5. *There exists a regret minimization algorithm $\mathcal{R}_{\Psi_i}$ for the set* co Ψ_i *(Figure 2) such that under any sequence of utility vectors $\boldsymbol{L}_i^{(1)}, \dots, \boldsymbol{L}_i^{(T)}$ its regret* $\mathrm{Reg}_{\Psi_i}^T$ *can be bounded as*

$$\mathrm{Reg}_{\Psi_i}^T \le \frac{\log|\Sigma_i| + \|Q_i\|_1^2 \max_{j \in \mathcal{J}_i} \log|\mathcal{A}_j|}{\eta} + \eta(\|Q_i\|_1 + 4|\Sigma_i|^2) \sum_{t=1}^{T} \|\boldsymbol{L}_i^{(t)} - \boldsymbol{L}_i^{(t-1)}\|_\infty^2.$$

As illustrated in Figure 2, the "mixer" $\mathcal{R}_\Delta$ is instantiated with OMWU, while each regret minimizer $\mathcal{R}_{\hat{\sigma}}$, for $\hat{\sigma} \in \hat{\sigma} \in \Sigma_i^*$, internally employs the dilatable global entropy as DGF to construct a regert minimizer over Q_j. A notable ingredient of our predictive regret circuit (Proposition A.1) is that we employ an advanced prediction mechanism in place of the usual "one-recency bias" wherein the prediction is simply the previously observed utility. This leads to an improved regret bound as we further explain in Remark A.2.

4.2 Stability of the Fixed Points

As suggested by Theorem 3.2, employing a predictive regret minimizer is of little gain if we cannot guarantee that the observed utilities will be stable. For this reason, in this subsection we focus on characterizing the stability of the fixed points, eventually leading to our stable-predictive co Ψ_i-regret minimizer. In the context of Theorem 3.2, this establishes the *stable* fixed point oracle. All of the omitted proofs of this section are included in Appendix A.5.

Multiplicative Stability. Our analysis will reveal a particularly strong notion of stability we refer to as *multiplicative stability*. More precisely, we say that a sequence $(z^{(t)})$, with $z^{(t)} \in \mathbb{R}^d_{>0}$, is κ*-multiplicative-stable*, with $\kappa \in (0,1)$, if $(1+\kappa)^{-1} z^{(t-1)}[k] \le z^{(t)}[k] \le (1+\kappa) z^{(t-1)}[k]$, for any $k \in [d]$ and for all $t \ge 2$. When $z^{(t)}[k]$ and $z^{(t-1)}[k]$ are such that $(1+\kappa)^{-1} z^{(t-1)}[k] \le z^{(t)}[k] \le (1+\kappa) z^{(t-1)}[k]$, we say that they are κ*-multiplicative-close*. We begin by showing that OMWU on the simplex and OFTRL with dilatable global entropy as DGF guarantee multiplicative stability.

Lemma 4.6. *Consider the OMWU algorithm on the simplex Δ^m with $\eta > 0$. If all the observed utilities and the predictions are such that $\|\boldsymbol{\ell}^{(t)}\|_\infty, \|\boldsymbol{m}^{(t)}\|_\infty \le \|\boldsymbol{\ell}\|_\infty$, and $\eta < 1/(12\|\boldsymbol{\ell}\|_\infty)$, then the sequence $(\boldsymbol{x}^{(t)})$ produced by OMWU is $(12\eta\|\boldsymbol{\ell}\|_\infty)$-multiplicative-stable.*

Lemma 4.7. *Consider the* (OFTRL) *algorithm on the sequence-form strategy polytope Q with dilatable global entropy as DGF and $\eta > 0$. If all the utility functions are such that $\|\boldsymbol{\ell}^{(t)}\|_\infty \le 1$, and $\eta = O(1/\mathfrak{D})$ is sufficiently small, then the sequence $(\boldsymbol{x}^{(t)})$ produced is $O(\eta\mathfrak{D})$-multiplicative-stable.*

To establish multiplicative stability of (OFTRL) under the dilatable global entropy DGF we first derive a closed-form solution which reveals the multiplicative structure of the update rule for the behavioral strategies at every "local" decision point. Then, the conversion to the sequence-form representation leads to a slight degradation of an $O(\mathfrak{D})$ (depth) factor in the multiplicative stability. Next, we use Lemmas 4.6 and 4.7 to arrive at the following conclusion.

Corollary 4.8. *Consider the regret minimization algorithm of Figure 2, and suppose that $\mathcal{R}_\Delta$ is instantiated using OMWU with $\eta > 0$, while each $\mathcal{R}_{\hat{\sigma}}$ is instantiated using* (OFTRL) *with dilatable global entropy as DGF and $\eta > 0$, for all $\hat{\sigma} \in \Sigma_i^*$. Then, for a sufficiently small $\eta = O(1/\|Q_i\|_1)$,*

(i) The output sequence of each $\mathcal{R}_{\hat{\sigma}}$ is $O(\eta\mathfrak{D}_i)$-multiplicative-stable;
(ii) The output sequence of $\mathcal{R}_\Delta$ is $O(\eta\|Q_i\|_1)$-multiplicative-stable.

Armed with this characterization, we will next establish the multiplicative stability of the fixed points associated with trigger deviation functions. To this end, building on the approach of Farina et al. [2021a], let us introduce the following definitions.

Definition 4.9. 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 4.10. Consider a player $i \in [n]$, a trunk $J \subseteq \mathcal{J}_i$, and $\phi_i \in \operatorname{co} \Psi_i$. A vector $\boldsymbol{x}_i \in \mathbb{R}_{\geq 0}^{|\Sigma_i|}$ is a *J-partial fixed point* of ϕ_i if the following conditions hold:

- $\boldsymbol{x}_i[\varnothing] = 1$ and $\boldsymbol{x}_i[\sigma_j] = \sum_{a \in \mathcal{A}_j} \boldsymbol{x}_i[(j,a)]$, for all $j \in J$;
- $\phi_i(\boldsymbol{x}_i)[\varnothing] = \boldsymbol{x}_i[\varnothing] = 1$, and $\phi_i(\boldsymbol{x}_i)[(j,a)] = \boldsymbol{x}_i[(j,a)]$, for all $j \in J$ and $a \in \mathcal{A}_j$.

An important property is that a J-partial fixed point can be efficiently "promoted" to a $J \cup \{j^*\}$-partial fixed point by computing the stationary distribution of a certain Markov chain (see Algorithm 4). However, it is a priori unclear how this fixed point operation would affect the stability of the produced strategies. In fact, even for a 2-state Markov chain, the stationary distribution could behave very unsmoothly under slight perturbations in the transition probabilities; *e.g.*, see [Chen and Peng, 2020, Haviv and Heyden, 1984, Meyer, 1980]. This is where the stronger notion of multiplicative stability comes into play. Indeed, it turns out that as long as the transition probabilities are multiplicative-stable, the stationary distribution will also be stable [Candogan et al., 2013]. This observation was also leveraged by Chen and Peng [2020] to obtain an $O(T^{-3/4})$ rate of convergence to correlated equilibria in normal-form games.

However, our setting is substantially more complex, and direct extensions of those prior techniques appears to only give a bound *exponential* in the size of the game. In light of this, one of our key observations is that the associated Markov chains has a particular structure which enables us to establish a polynomial degradation in terms of stability. At a high level, we observe that the underlying Markov chain can be expressed as the convex combination of a stable chain with a much less stable rank-one component. The main concern is that the unstable rank-one chain could cause a substantial degradation in terms of the stability of the fixed points. We address this by proving the following key lemma.

Lemma 4.11. *Let $\mathbf{M}$ be the transition matrix of an m-state Markov chain such that $\mathbf{M} \coloneqq \boldsymbol{v}\mathbf{1}^\top + \mathbf{C}$, 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 $j \in F_i$, 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 parameter.

The main takeaway of this lemma is that the stationary distribution has only an *affine dependence* on the vector $\boldsymbol{v}$. This will be crucial as $\boldsymbol{v}$ will be much less stable than the entries of $\mathbf{C}$, as we make precise in the sequel. Naturally, Lemma 4.11 is not at all apparent from the Markov chain tree theorem, and derives from the particular structure of the Markov chain. Indeed, to establish Lemma 4.11 we deviate from the existing techniques which are relying on the Markov chain tree theorem, and we instead leverage linear-algebraic techniques to characterize the corresponding eigenvector of the underlying Laplacian system. As a result, using a slight variant of Lemma 4.11 (see Corollary A.8) leads to the following stability bound.

Corollary 4.12. *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 $\boldsymbol{w} \coloneqq l\boldsymbol{\pi}$ and $\boldsymbol{w}' \coloneqq l'\boldsymbol{\pi}'$. 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 scalars 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 $\kappa = O(1/m)$.

Under the assertion that $\gamma \gg \kappa$, the key takeaway is that the "closeness" of $\boldsymbol{w}$ and $\boldsymbol{w}'$ does not scale with $O((\gamma+\kappa)m)$, but only as $\gamma + O(\kappa m)$. Using this bound we are ready to characterize the degradation in stability after a "promotion" (Algorithm 4) of a partial fixed point (in the formal sense of Definition 4.10).

Proposition 4.13. *Consider a player $i \in [n]$, and let $\phi_i^{(t)} = \sum_{\hat{\sigma} \in \Sigma_i^*} \boldsymbol{\lambda}_i^{(t)}[\hat{\sigma}] \phi_{\hat{\sigma} \to \boldsymbol{q}_{\hat{\sigma}}^{(t)}}$ be a transformation in $\operatorname{co}\Psi_i$ such that the sequences $(\boldsymbol{\lambda}_i^{(t)})$ and $(\boldsymbol{q}_{\hat{\sigma}}^{(t)})$ are κ-multiplicative-stable, for all $\hat{\sigma} \in \Sigma_i^*$. If $(\boldsymbol{x}_i^{(t)})$ is a γ-multiplicative-stable J-partial fixed point sequence, the sequence of $(J \cup \{j^*\})$-partial fixed points of ϕ_i is $(\gamma + O(\kappa|\mathcal{A}_{j^*}|))$-multiplicative-stable, for any sufficiently small $\kappa = O(1/|\mathcal{A}_{j^*}|)$.*

Moreover, we employ this proposition as the inductive step to derive sharp multiplicative-stability bounds for the sequence of fixed points. The underlying algorithm gradually invokes the "promotion" subroutine (Algorithm 4) in a top-down traversal of the tree, and it is given in Algorithm 5.

Theorem 4.14. *Consider a player $i \in [n]$, and let $\phi_i^{(t)} = \sum_{\hat{\sigma} \in \Sigma_i^*} \boldsymbol{\lambda}_i^{(t)}[\hat{\sigma}] \phi_{\hat{\sigma} \to \boldsymbol{q}_{\hat{\sigma}}^{(t)}}$ be a transformation in $\operatorname{co}\Psi_i$ such that the sequences $(\boldsymbol{\lambda}_i^{(t)})$ and $(\boldsymbol{q}_{\hat{\sigma}}^{(t)})$ are κ-multiplicative-stable, for all $\hat{\sigma} \in \Sigma_i^*$. Then, the sequence of fixed points $\boldsymbol{q}_i^{(t)} \in \mathcal{Q}_i$ of $\phi_i^{(t)}$ is $O(\kappa|\mathcal{A}_i|\mathfrak{D}_i)$-multiplicative-stable, for a sufficiently small $\kappa = O(1/(|\mathcal{A}_i|\mathfrak{D}_i))$, where $|\mathcal{A}_i| := \max_{j \in \mathcal{J}_i} |\mathcal{A}_j|$.*

A more refined bound is discussed in Remark A.9. The important insight of Theorem 4.14 is that the stability degrades according to the *sum* of the actions at the decision points encountered along each path, and not as the *product* of the actions. This is crucial as the latter bound—which would follow from prior techniques—need not be polynomial in the description of the game. At the heart of this improvement lies our refined characterization obtained in Lemma 4.11. Using the stability bounds derived in Corollary 4.8, we are ready to establish the multiplicative-stability of the sequence of fixed points.

Corollary 4.15 (Stability of Fixed Points)**.** *For any sufficiently small $\eta = O(1/(\mathfrak{D}_i|\mathcal{A}_i|\|\mathcal{Q}_i\|_1))$, the sequence of fixed points $(\boldsymbol{q}_i^{(t)})$ of player $i \in [n]$ is $O(\eta\mathfrak{D}_i|\mathcal{A}_i|\|\mathcal{Q}_i\|_1)$-multiplicative-stable.*

4.3 Completing the Proof

Finally, we combine all of the previous pieces to complete the construction. First, we apply Theorem 3.2 using the predictive bound obtained in Theorem 4.5 to conclude that the Ψ_i-regret of each player $i \in [n]$ can be bounded as

$$\mathrm{Reg}_i^T \le \frac{\log|\Sigma_i| + \|\mathcal{Q}_i\|_1^2 \log|\mathcal{A}_i|}{\eta} + 10\eta|\Sigma_i|^2 \sum_{t=1}^{T} \|\boldsymbol{\ell}_i^{(t)} - \boldsymbol{\ell}_i^{(t-1)}\|_\infty^2 + 10\eta|\Sigma_i|^2 \sum_{t=1}^{T} \|\boldsymbol{q}_i^{(t)} - \boldsymbol{q}_i^{(t-1)}\|_\infty^2, \quad (4)$$

where we assumed—for simplicity—exact computation of each fixed point, *i.e.*, $\epsilon^{(t)} = 0$ for any $t \ge 1$, while we also used the fact that $\|\boldsymbol{\ell}_i^{(t)}\|_\infty \le 1$ which follows from the normalization assumption. So far we have focused on bounding the regret of each player without making any assumptions about the stability of the observed utility functions. A crucial observation is that if all players employ a regularized (or smooth) learning algorithm, then the observed utility functions will also change slowly over time. This is formalized in the following auxiliary claim.

Claim 4.16. *For any player $i \in [n]$ the observed utilities satisfy*

$$\|\boldsymbol{\ell}_i^{(t)} - \boldsymbol{\ell}_i^{(t-1)}\|_\infty^2 \leq (n-1)|\mathcal{Z}|^2 \sum_{k \neq i} \|\boldsymbol{q}_k^{(t)} - \boldsymbol{q}_k^{(t-1)}\|_1^2.$$

Thus, plugging this bound to (4) yields that the Ψ_i-regret Reg_i^T of each player i can be bounded as

$$\frac{\log|\Sigma_i| + \|Q_i\|_1^2 \log|\mathcal{A}_i|}{\eta} + 10\eta(n-1)|\Sigma_i|^2|\mathcal{Z}|^2 \sum_{t=1}^{T} \sum_{k \neq i} \|\boldsymbol{q}_k^{(t)} - \boldsymbol{q}_k^{(t-1)}\|_1^2 + 10\eta|\Sigma_i|^2 \sum_{t=1}^{T} \|\boldsymbol{q}_i^{(t)} - \boldsymbol{q}_i^{(t-1)}\|_\infty^2.$$

As a result, using Corollary 4.15 we arrive at the following conclusion.

Corollary 4.17. *Suppose that each player follows the dynamics of Figure 2 with a sufficiently small learning rate $\eta = O(1/(T^{1/4}\mathfrak{D}_i|\mathcal{A}_i|\|Q_i\|_1))$. Then, the Ψ_i-regret of each player will be bounded as $\mathrm{Reg}_i^T \leq \mathcal{P}T^{1/4}$, where $\mathcal{P}$ is independent on T and polynomial on the description of the game.*

Finally, Theorem 1.1 follows from Theorem 2.8 after performing sampling in order to transition to deterministic strategies, as we explain in Appendix A.6. We also point out that the complexity of every iteration in the proposed dynamics is analogous to that in [Farina et al., 2021a], although the dynamics developed in the latter paper only attain a rate of convergence of $O(T^{-1/2})$. Finally, we remark that it is easy to make the overall regret minimizer robust against adversarial losses using an adaptive choice of learning rate.

5 FASTER CONVERGENCE TO EFCCE

In this section we turn our attention to learning dynamics for extensive-form *coarse* correlated equilibrium (EFCCE). While the dynamics we previously developed for EFCE would also trivially converge to EFCCE, as the former is a subset of the latter [Farina et al., 2020], our main contribution is to show that each iteration of EFCCE dynamics can be substantially more efficient compared to EFCE. Indeed, unlike all known methods for EFCE, we obtain in Section 5.1 a succinct closed-form solution for the fixed points associated with EFCCE which does not require the expensive computation of the stationary distribution of a Markov chain. This places EFCCE closer to normal-form coarse correlated equilibria (NFCCE) in terms of the per-iteration complexity, even thought EFCCE prescribes a much more compelling notion of correlation. Furthermore, we use this closed-form characterization in Section 5.2 to obtain improved stability bounds for the fixed points associated with EFCCE, and with a much simpler analysis compared to the one for EFCE.

5.1 Closed-Form Fixed Point Computation

As suggested by our general template introduced in Theorem 3.2, we first have to construct a predictive regret minimizer for the set of *coarse* trigger deviation functions $\widetilde{\Psi}_i$ (Definition 2.9). This construction is very similar to the one for Ψ_i we previously described in detail in Section 4.1. For this reason, here we focus on the computation and the stability properties of the fixed points associated with any $\phi_i \in \operatorname{co}\widetilde{\Psi}_i$. Specifically, we will first show that it it possible to compute a sequence-form strategy $\boldsymbol{q}_i$ such that $\phi_i(\boldsymbol{q}_i) = \boldsymbol{q}_i$ in linear time on $O(|\Sigma_i|\mathfrak{D}_i)$.

Indeed, let $\phi_i = \sum_{j \in \mathcal{J}_i} \boldsymbol{\lambda}_i[j]\phi_{j \to \boldsymbol{q}_j}$ be any coarse trigger deviation function, where $\boldsymbol{\lambda}_i \in \Delta(\mathcal{J}_i)$, and $\boldsymbol{q}_j \in Q_j$ for each $j \in \mathcal{J}_i$. Algorithm 1 describes an efficient procedure to compute a fixed point of a given transformation $\phi_i \in \operatorname{co}\widetilde{\Psi}_i$. In particular, the algorithm iterates over the sequences of player i according to their partial ordering $\prec$. That is, it is never the case that a sequence $\sigma = (j, a)$ is considered before σ_j. For every sequence $\sigma = (j, a) \in \Sigma_i^*$ the algorithm computes $d_\sigma \in \mathbb{R}_{\geq 0}$ as the sum of the weights corresponding to information sets preceding j (Line 3). If $d_\sigma = 0$, the choice we make at σ is indifferent as long as the resulting vector $\boldsymbol{q}_i$ is a well-formed sequence-form strategy.

For this reason, we simply set $\boldsymbol{q}_i[\sigma]$ so that the probability-mass flow is evenly divided among sequences originating in j (Line 5). Otherwise, when $d_\sigma \neq 0$, Line 7 assigns to $\boldsymbol{q}_i[\sigma]$ a value equal to the weighted sum of $\boldsymbol{q}_{j'}[\sigma]\boldsymbol{q}_i[\sigma']$ for sequences $\sigma' = (j', a')$ preceding information set $j \in \mathcal{J}_i$. In the next theorem we show that Algorithm 1 is indeed correct, and runs in time $O(|\Sigma_i|\mathfrak{D}_i)$.

Theorem 5.1. *For any player $i \in [n]$ and any transformation $\phi_i = \sum_{j \in \mathcal{J}_i} \boldsymbol{\lambda}_i[j]\phi_{j \to \boldsymbol{q}_j} \in \operatorname{co} \widetilde{\Psi}_i$, the output $\boldsymbol{q}_i \in \mathbb{R}^{|\Sigma_i|}$ of Algorithm 1 is such that $\boldsymbol{q}_i \in \mathcal{Q}_i$ and $\phi_i(\boldsymbol{q}_i) = \boldsymbol{q}_i$. Furthermore, Algorithm 1 runs in $O(|\Sigma_i|\mathfrak{D}_i)$.*

ALGORITHM 1: FixedPoint(ϕ_i) for $\phi_i \in \operatorname{co} \widetilde{\Psi}_i$

Input: $\phi_i = \sum_{j \in \mathcal{J}_i} \boldsymbol{\lambda}_i[j]\phi_{j \to \boldsymbol{q}_j} \in \operatorname{co} \widetilde{\Psi}_i$
Output: $\boldsymbol{q}_i \in \mathcal{Q}_i$ such that $\phi_i(\boldsymbol{q}_i) = \boldsymbol{q}_i$
1 $\boldsymbol{q}_i \leftarrow \mathbf{0} \in \mathbb{R}^{|\Sigma_i|}$, $\boldsymbol{q}_i[\varnothing] \leftarrow 1$
2 **for** $\sigma = (j, a) \in \Sigma_i^*$ *in top-down ($\prec$) order* **do**
3 $d_\sigma \leftarrow \sum_{j' \preceq j} \boldsymbol{\lambda}_i[j']$
4 **if** $d_\sigma = 0$ **then**
5 $\boldsymbol{q}_i[\sigma] \leftarrow \frac{\boldsymbol{q}_i[\sigma_j]}{|\mathcal{A}_j|}$
6 **else**
7 $\boldsymbol{q}_i[\sigma] \leftarrow \frac{1}{d_\sigma} \sum_{j' \preceq j} \boldsymbol{\lambda}_i[j']\boldsymbol{q}_{j'}[\sigma]\boldsymbol{q}_i[\sigma_{j'}]$
8 **return** $\boldsymbol{q}_i$

5.2 Stability of the Fixed Points

Another important application of our closed-form solution in Algorithm 1 is that it allows us to obtain through a simple analysis sharp bounds on the stability of the fixed points. Indeed, we show that the fixed point operation only leads to (multiplicative) degradation linear in the depth of each player's subtree.

Proposition 5.2. *Suppose that the sequences $(\boldsymbol{\lambda}_i^{(t)})$ and $(\boldsymbol{q}_j^{(t)})$, for all $j \in \mathcal{J}_i$, are κ-multiplicative-stable. Then, Algorithm 1 yields a sequence of $(12\kappa\mathfrak{D}_i)$-multiplicative-stable strategies, assuming that $\kappa < 1/(12\mathfrak{D}_i)$.*

Observe that the derived bound on stability is slightly better compared to that for EFCE (Theorem 4.14). Consequently, having established the stability of the fixed points, we can apply Theorem 3.2 to derive a stable-predictive $\widetilde{\Psi}_i$-regret minimizer for each player $i \in [n]$. This leads to a result analogous to Corollary 4.17 we showed for EFCE, but our dynamics for EFCCE have a substantially improved per-iteration complexity.

6 EXPERIMENTS

In this section we support our theoretical findings through experiments conducted on benchmark general-sum games. Namely, we experiment with the following popular games: (i) a three-player variant of *Kuhn poker* [Kuhn, 1950]; (ii) a two-player bargaining game known as *Sheriff* [Farina et al., 2019d]; (iii) a three-player version of *Liar's dice* [Lisý et al., 2015]; and (iv) three-player Goofspiel [Ross, 1971]. A detailed description of each of these games and the precise parameters we use is given in Appendix C. The rest of this section is organized as follows. Section 6.1 shows the convergence of our accelerated dynamics for EFCE (as presented in Section 4) compared to the prior state of the art. Next, Section 6.2 illustrates the convergence of our dynamics for EFCCE.

6.1 Convergence to EFCE

Here we investigate the performance of our accelerated dynamics for EFCE (Figure 2) compared to the existing algorithm by Farina et al. [2021a]. Both of these dynamics will be based on a CFR-style decomposition into "local" regret-minimization problems. In particular, our stable-predictive dynamics will use OMWU at every local decision point (as in Proposition 4.3), while the algorithm of Farina et al. [2021a] will be 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 our theoretical predictions (Corollary 4.17), the stepsize for OMWU is set as $\eta^{(t)} = \tau \cdot t^{-1/4}$, while for MWU it is set as $\eta^{(t)} = \tau \cdot t^{-1/2}$. Here τ is treated as a hyperparameter, chosen by picking the best-performing value among $\{0.01, 0.1, 1, 10, 100\}$.

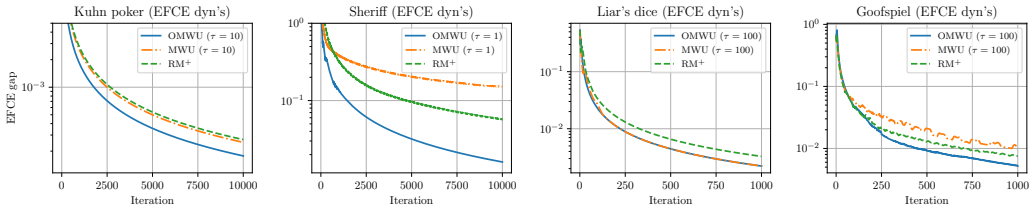

Fig. 3. The performance of EFCE dynamics based on MWU, OMWU, and RM$^+$ on four general-sum EFGs.

Figure 3 shows the rate of convergence for each of the three learning dynamics we described. On the x-axis we indicate 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, a benchmark introduced specifically for the study of correlated equilibria, OMWU outperforms both RM$^+$ and MWU by about an order of magnitude.

In the context of two-player zero-sum games, CFR with RM$^+$ is a formidable algorithm, typically outperforming theoretically superior dynamics. With that in mind, it is quite interesting that for EFCE computation we are able to achieve better performance using OMWU with only a modest amount of stepsize tuning. We hypothesize that this is due to the inherent differences between solving a zero-sum game via Nash equilibrium versus the problem of computing correlated equilibria. 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 [Burch et al., 2019, Farina et al., 2019a, Tammelin et al., 2015], but it is unclear if they still guarantee convergence in the EFCE setting.

6.2 EFCCE

Next, we investigate the convergence of our learning dynamics for EFCCE, obtained within the same framework we developed for EFCE. We first measure the rate of convergence in an analogous to the previous subsection setup. The results are illustrated in Figure 4.

Interestingly, we observe a noticeable qualitative difference for convergence to EFCCE. Indeed, unlike EFCE (Figure 3), RM$^+$ outperforms OMWU in both Liar's dice and Goofspiel. It is also surprising that MWU converges faster than its optimistic counterpart in Kuhn poker. These results

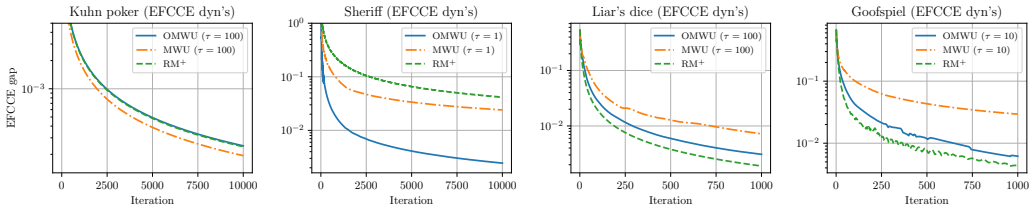

Fig. 4. The performance of EFCCE dynamics based on MWU, OMWU, and RM$^+$ on four general-sum EFGs.

suggest a substantial difference in the convergence properties between EFCE and EFCCE. Furthermore, we illustrate in Figure 5 the running time complexity of EFCE versus EFCCE dynamics (both instantiated with RM$^+$), measured in terms of the EFCCE gap.

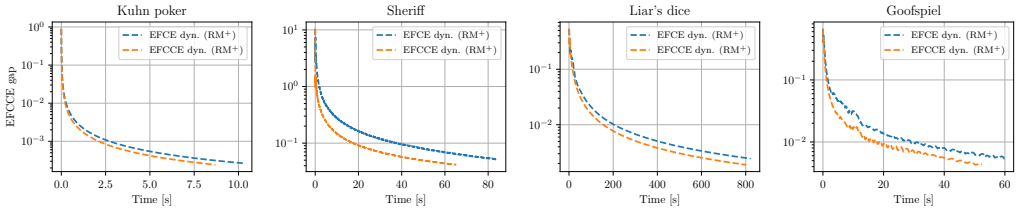

Fig. 5. The convergence of EFCE and EFCCE dynamics to EFCCE, measured through the EFCCE gap.

In each game, the fixed point computation for the EFCE dynamics was performed through an optimized implementation of the power iteration method, interrupted when the Euclidean norm of the residual was below the value of 10^{-6}. On the other hand, the fixed points for EFCCE were computed using our closed-form solution (Algorithm 1). In all four games, we see that our EFCCE dynamics outperform the EFCE dynamics in terms of the running time complexity, often by a significant margin. This is consistent with our intuition since EFCE dynamics are solving a strictly harder problem—minimizing the EFCE gap, instead of the EFCCE gap.

7 CONCLUSIONS

In this paper we developed uncoupled no-regret learning dynamics so that if all agents play T repetitions of the game according to our dynamics, the correlated distribution of play is an $O(T^{-3/4})$-approximate extensive-form correlated equilibrium. 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 structured Markov chain, which may be of independent interest. On the other hand, for fixed points associated with extensive-form *coarse* correlated equilibria we established a closed-form solution, circumventing the computation of the stationary distribution of any Markov chain. Finally, experiments conducted on standard benchmarks corroborated our theoretical findings.

Following recent progress in normal-form games [Anagnostides et al., 2021, Daskalakis et al., 2021], an important question for the future is to obtain a further acceleration of the order $\widetilde{O}(T^{-1})$. As we pointed out in Section 1.2, this would inevitably require new techniques since the known methods do not apply for the substantially more complex problem of extensive-form correlated equilibria. We believe that our characterization of the fixed points associated with trigger deviation functions could be an important step towards achieving this goal.

ACKNOWLEDGMENTS

Tuomas Sandholm is supported by the National Science Foundation under grants IIS-1901403 and CCF-1733556.

REFERENCES

Ioannis Anagnostides, Constantinos Daskalakis, Gabriele Farina, Maxwell Fishelson, Noah Golowich, and Tuomas Sandholm. 2021. Near-Optimal No-Regret Learning for Correlated Equilibria in Multi-Player General-Sum Games. *CoRR* abs/2111.06008 (2021). https://arxiv.org/abs/2111.06008

Robert Aumann. 1974. Subjectivity and Correlation in Randomized Strategies. *Journal of Mathematical Economics* 1 (1974), 67–96.

Kimmo Berg and Tuomas Sandholm. 2017. Exclusion Method for Finding Nash Equilibrium in Multiplayer Games. In *AAAI Conference on Artificial Intelligence (AAAI)*.

Avrim Blum and Yishay Mansour. 2007. From External to Internal Regret. *J. Mach. Learn. Res.* 8 (2007), 1307–1324.

Michael Bowling, Neil Burch, Michael Johanson, and Oskari Tammelin. 2015. Heads-up Limit Hold'em Poker is Solved. *Science* 347, 6218 (January 2015).

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

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

Ozan Candogan, Asuman E. Ozdaglar, and Pablo A. Parrilo. 2013. Dynamics in near-potential games. *Games Econ. Behav.* 82 (2013), 66–90. https://doi.org/10.1016/j.geb.2013.07.001

Andrea Celli, Stefano Coniglio, and Nicola Gatti. 2019a. Computing Optimal Ex Ante Correlated Equilibria in Two-Player Sequential Games. In *Proceedings of the 18th International Conference on Autonomous Agents and MultiAgent Systems (AAMAS)*. 909–917.

Andrea Celli, Alberto Marchesi, Tommaso Bianchi, and Nicola Gatti. 2019b. Learning to Correlate in Multi-Player General-Sum Sequential Games. In *Proceedings of the Annual Conference on Neural Information Processing Systems (NeurIPS)*, Vol. 32.

Andrea Celli, Alberto Marchesi, Gabriele Farina, and Nicola Gatti. 2020. No-Regret Learning Dynamics for Extensive-Form Correlated Equilibrium. In *Advances in Neural Information Processing Systems 33: Annual Conference on Neural Information Processing Systems 2020*, Hugo Larochelle, Marc'Aurelio Ranzato, Raia Hadsell, Maria-Florina Balcan, and Hsuan-Tien Lin (Eds.).

Nicolo Cesa-Bianchi and Gábor Lugosi. 2006. *Prediction, learning, and games*. Cambridge university press.

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

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

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

Constantinos Daskalakis, Maxwell Fishelson, and Noah Golowich. 2021. Near-Optimal No-Regret Learning in General Games. *CoRR* abs/2108.06924 (2021).

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

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

Kousha Etessami and Mihalis Yannakakis. 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.

Gabriele Farina, Tommaso Bianchi, and Tuomas Sandholm. 2020. Coarse correlation in extensive-form games. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 34. 1934–1941.

Gabriele Farina, Andrea Celli, Alberto Marchesi, and Nicola Gatti. 2021a. Simple Uncoupled No-Regret Learning Dynamics for Extensive-Form Correlated Equilibrium.

Gabriele Farina, Christian Kroer, Noam Brown, and Tuomas Sandholm. 2019c. Stable-Predictive Optimistic Counterfactual Regret Minimization. In *International Conference on Machine Learning (ICML)*.

Gabriele Farina, Christian Kroer, and Tuomas Sandholm. 2019a. Online Convex Optimization for Sequential Decision Processes and Extensive-Form Games. In *AAAI Conference on Artificial Intelligence*.

Gabriele Farina, Christian Kroer, and Tuomas Sandholm. 2019b. Regret Circuits: Composability of Regret Minimizers. In *International Conference on Machine Learning*. 1863–1872.

Gabriele Farina, Christian Kroer, and Tuomas Sandholm. 2021b. Better Regularization for Sequential Decision Spaces: Fast Convergence Rates for Nash, Correlated, and Team Equilibria. In *ACM Conference on Economics and Computation*.

Gabriele Farina, Chun Kai Ling, Fei Fang, and Tuomas Sandholm. 2019d. Correlation in Extensive-Form Games: Saddle-Point Formulation and Benchmarks. In *Conference on Neural Information Processing Systems (NeurIPS)*.

Dean Foster and Rakesh Vohra. 1997. Calibrated Learning and Correlated Equilibrium. *Games and Economic Behavior* 21 (1997), 40–55.

Andrew Gilpin and Tuomas Sandholm. 2007. Lossless Abstraction of Imperfect Information Games. *J. ACM* 54, 5 (2007).

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

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

Sergiu Hart and Andreu Mas-Colell. 2000. A Simple Adaptive Procedure Leading to Correlated Equilibrium. *Econometrica* 68 (2000), 1127–1150.

Moshe Haviv and Ludo Van Der Heyden. 1984. Perturbation Bounds for the Stationary Probabilities of a Finite Markov Chain. *Advances in Applied Probability* 16, 4 (1984), 804–818.

Samid Hoda, Andrew Gilpin, Javier Peña, and Tuomas Sandholm. 2010. Smoothing Techniques for Computing Nash Equilibria of Sequential Games. *Mathematics of Operations Research* 35, 2 (2010).

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

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

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

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

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

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

Carl D. Meyer, Jr. 1980. The Condition of a Finite Markov Chain and Perturbation Bounds for the Limiting Probabilities. *SIAM Journal on Algebraic Discrete Methods* 1, 3 (1980), 273–283.

Matej Moravčík, Martin Schmid, Neil Burch, Viliam Lisý, Dustin Morrill, Nolan Bard, Trevor Davis, Kevin Waugh, Michael Johanson, and Michael Bowling. 2017. DeepStack: Expert-level artificial intelligence in heads-up no-limit poker. *Science* (May 2017).

Dustin Morrill, Ryan D'Orazio, Marc Lanctot, James R. Wright, Michael Bowling, and Amy R. Greenwald. 2021a. Efficient Deviation Types and Learning for Hindsight Rationality in Extensive-Form Games. In *Proceedings of the 38th International Conference on Machine Learning, ICML 2021 (Proceedings of Machine Learning Research, Vol. 139)*, Marina Meila and Tong Zhang (Eds.). PMLR, 7818–7828.

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

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

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

Alexander Rakhlin and Karthik Sridharan. 2013a. Online Learning with Predictable Sequences. In *Conference on Learning Theory*. 993–1019.

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

Sheldon M Ross. 1971. Goofspiel—the game of pure strategy. *Journal of Applied Probability* 8, 3 (1971), 621–625.

Yoav Shoham and Kevin Leyton-Brown. 2009. *Multiagent systems: Algorithmic, game-theoretic, and logical foundations*. Cambridge University Press.

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

Oskari Tammelin. 2014. Solving Large Imperfect Information Games Using CFR+. arXiv preprint. arXiv:1407.5042 [cs.GT]

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

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

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

A OMITTED PROOFS

This section includes all of the proofs we omitted from the main body. Let us first introduce some additional useful notation.

A.1 Further Notation

It will be convenient to instantiate a trigger deviation function (recall Definition 2.6) in the form of a linear mapping $\phi_{\hat{\sigma}\to\hat{\pi}_i} : \mathbb{R}^{|\Sigma_i|} \ni \boldsymbol{x} \mapsto \mathbf{M}_{\hat{\sigma}\to\hat{\pi}_i}\boldsymbol{x}$, where $\mathbf{M}_{\hat{\sigma}\to\hat{\pi}_i}$ is such that for any $\sigma_r, \sigma_c \in \Sigma_i$,

$$\mathbf{M}_{\hat{\sigma}\to\hat{\pi}_i}[\sigma_r, \sigma_c] = \begin{cases} 1 & \text{if } \sigma_c \not\succeq \hat{\sigma} \quad \& \quad \sigma_r = \sigma_c; \\ \hat{\pi}_i[\sigma_r] & \text{if } \sigma_c = \hat{\sigma} \quad \& \quad \sigma_r \succeq j; \\ 0 & \text{otherwise,} \end{cases} \tag{5}$$

where $\hat{\sigma} = (j, a) \in \Sigma_i^*$. It is not hard to show that the linear mapping described in (5) is indeed a trigger deviation function in the sense of Definition 2.6. Similarly, we express a coarse trigger deviation function in the form of a linear mapping $\phi_{j\to\hat{\pi}_i} : \mathbb{R}^{|\Sigma_i|} \ni \boldsymbol{x} \mapsto \mathbf{M}_{j\to\hat{\pi}_i}\boldsymbol{x}$, where $\mathbf{M}_{j\to\hat{\pi}_i}$ is such that for any $\sigma_r, \sigma_c \in \Sigma_i$,

$$\mathbf{M}_{j\to\hat{\pi}_i}[\sigma_r, \sigma_c] = \begin{cases} 1 & \text{if } \sigma_c \not\succeq j \quad \& \quad \sigma_r = \sigma_c; \\ \hat{\pi}_i[\sigma_r] & \text{if } \sigma_c = \sigma_j \quad \& \quad \sigma_r \succeq j; \\ 0 & \text{otherwise.} \end{cases}$$

Furthermore, we will use the notation $\boldsymbol{x} \otimes \boldsymbol{y} = \boldsymbol{x}\boldsymbol{y}^\top$ to denote the *outer product* of (compatible) vectors $\boldsymbol{x}$ and $\boldsymbol{y}$, while we will also write $(\mathbf{M})^\flat$ to represent the standard *vectorization* of matrix $\mathbf{M}$.

A.2 Proofs from Section 2

Theorem 2.11. *Suppose that for every player $i \in [n]$ the sequence of deterministic sequence-form strategies $\boldsymbol{\pi}_i^{(1)}, \dots, \boldsymbol{\pi}_i^{(T)} \in \Pi_i$ incurs $\widetilde{\Psi}_i$-regret at most Reg_i^T under 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-EFCCE, where $\epsilon \coloneqq \frac{1}{T}\max_{i\in[n]} \mathrm{Reg}_i^T$.

Proof. By assumption, we know that for any $i \in [n]$ it holds that $\mathrm{Reg}_i^T \le \epsilon T$. Thus, by definition of Reg_i^T, it follows that for any $i \in [n]$ and any coarse trigger deviation function $\phi_i \in \widetilde{\Psi}_i$,

$$\begin{aligned} T\epsilon &\ge \sum_{t=1}^{T}\left(\ell_i^{(t)}\left(\phi_i(\boldsymbol{\pi}_i^{(t)})\right) - \ell_i^{(t)}\left(\boldsymbol{\pi}_i^{(t)}\right)\right) = \sum_{t=1}^{T}\left(u_i\left(\phi_i(\boldsymbol{\pi}_i^{(t)}), \boldsymbol{\pi}_{-i}^{(t)}\right) - u_i\left(\boldsymbol{\pi}^{(t)}\right)\right) \\ &= \sum_{t=1}^{T}\sum_{\boldsymbol{\pi}\in\Pi} \mathbb{1}\left\{\boldsymbol{\pi} = \boldsymbol{\pi}^{(t)}\right\}\left(u_i(\phi_i(\boldsymbol{\pi}_i), \boldsymbol{\pi}_{-i}) - u_i(\boldsymbol{\pi})\right) \\ &= \sum_{\boldsymbol{\pi}\in\Pi}\sum_{t=1}^{T}\left(\mathbb{1}\left\{\boldsymbol{\pi} = \boldsymbol{\pi}^{(t)}\right\}\right)\left(u_i(\phi_i(\boldsymbol{\pi}_i), \boldsymbol{\pi}_{-i}) - u_i(\boldsymbol{\pi})\right) \\ &= T\sum_{\boldsymbol{\pi}\in\Pi}\boldsymbol{\mu}[\boldsymbol{\pi}]\left(u_i(\phi_i(\boldsymbol{\pi}_i), \boldsymbol{\pi}_{-i}) - u_i(\boldsymbol{\pi})\right). \end{aligned}$$

This is precisely the definition of an ϵ-EFCCE (Definition 2.10), as we wanted to show. □

A.3 Proofs from Section 3

Here we prove Theorem 3.2. For the convenience of the reader the theorem is restated below.

Theorem 3.2 (Stable-Predictive Phi-Regret Minimization)**.** *Consider an (A, B)-predictive regret minimizer $\mathcal{R}_\Phi$ with respect to $\|\cdot\|_1$ for a set of linear transformations Φ on $\mathcal{X}$. Moreover, suppose that $\mathcal{R}_\Phi$ is fixed point κ-stable. Then, if we have access to a* StableFPOracle*, we can construct a κ-stable algorithm with Φ-regret Reg^T bounded as*

$$\mathrm{Reg}^T \le A + 2B \sum_{t=1}^{T} \|\boldsymbol{\ell}^{(t)} - \boldsymbol{\ell}^{(t-1)}\|_\infty^2 + 2B\|\boldsymbol{\ell}\|_\infty^2 \sum_{t=1}^{T} \|\boldsymbol{x}^{(t)} - \boldsymbol{x}^{(t-1)}\|_\infty^2 + \|\boldsymbol{\ell}\|_\infty \sum_{t=1}^{T} \epsilon^{(t)},$$

where $\epsilon^{(t)}$ is the error of StableFPOracle *at time t, and $\|\boldsymbol{\ell}^{(t)}\|_\infty \le \|\boldsymbol{\ell}\|_\infty$ for any $t \ge 1$. It is also assumed that $\|\boldsymbol{x}\|_\infty \le 1$ for all $\boldsymbol{x} \in \mathcal{X}$.*

Proof. Fix any iteration $t \ge 2$. The first step is to obtain the next strategy of $\mathcal{R}_\Phi$: $\phi^{(t)} = \mathcal{R}_\Phi.\textsc{NextStrategy}()$. Then, our regret minimizer $\mathcal{R}$ will simply output the strategy $\boldsymbol{x}^{(t)}$ such that $\boldsymbol{x}^{(t)} = \textsc{StableFPOracle}(\phi^{(t)}; \boldsymbol{x}^{(t-1)}, \kappa, \epsilon^{(t)})$.[3] By assumption (recall Definition 3.1), we know that this is indeed well-defined and $\boldsymbol{x}^{(t)}$ will be such that (i) $\|\phi^{(t)}(\boldsymbol{x}^{(t)}) - \boldsymbol{x}^{(t)}\|_1 \le \epsilon^{(t)}$, and (ii) $\|\boldsymbol{x}^{(t)} - \boldsymbol{x}^{(t-1)}\|_1 \le \kappa$. This immediately implies that $\mathcal{R}$ will be κ-stable.

Afterwards, we receive feedback from the environment in the form of a utility vector $\boldsymbol{\ell}^{(t)}$, which in turn is used to construct the utility function $L^{(t)} : \phi \mapsto \langle \boldsymbol{\ell}^{(t)}, \phi(\boldsymbol{x}^{(t)}) \rangle$. Since Φ is a set of linear transformations, we can represent the corresponding utility vector as $\boldsymbol{L}^{(t)} = (\boldsymbol{\ell}^{(t)} \otimes \boldsymbol{x}^{(t)})^\flat$. This function is then given as feedback to $\mathcal{R}_\Phi$; that is, we invoke the subroutine $\mathcal{R}_\Phi.\textsc{ObserveUtility}(L^{(t)})$. As a result, the (external) regret of $\mathcal{R}_\Phi$ can be expressed as

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

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

$$\begin{aligned}
\mathrm{Reg}^T - \mathrm{Reg}_\Phi^T &= \sum_{t=1}^{T} \langle \boldsymbol{\ell}^{(t)}, \phi^{(t)}(\boldsymbol{x}^{(t)}) \rangle - \sum_{t=1}^{T} \langle \boldsymbol{\ell}^{(t)}, \boldsymbol{x}^{(t)} \rangle = \sum_{t=1}^{T} \langle \boldsymbol{\ell}^{(t)}, \phi^{(t)}(\boldsymbol{x}^{(t)}) - \boldsymbol{x}^{(t)} \rangle \\
&\le \sum_{t=1}^{T} \|\boldsymbol{\ell}^{(t)}\|_* \|\phi^{(t)}(\boldsymbol{x}^{(t)}) - \boldsymbol{x}^{(t)}\| \le \|\boldsymbol{\ell}\|_\infty \sum_{t=1}^{T} \epsilon^{(t)}, \qquad (6)
\end{aligned}$$

where we used the Cauchy-Schwarz inequality, as well as the assumption that $\|\phi^{(t)}(\boldsymbol{x}^{(t)}) - \boldsymbol{x}^{(t)}\| \le \epsilon^{(t)}$. 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

$$\begin{aligned}
\|\boldsymbol{L}^{(t)} - \boldsymbol{L}^{(t-1)}\|_\infty^2 &= \|(\boldsymbol{\ell}^{(t)} \otimes \boldsymbol{x}^{(t)})^\flat - (\boldsymbol{\ell}^{(t-1)} \otimes \boldsymbol{x}^{(t-1)})^\flat\|_\infty^2 \\
&= \|(\boldsymbol{\ell}^{(t)} \otimes \boldsymbol{x}^{(t)})^\flat - (\boldsymbol{\ell}^{(t-1)} \otimes \boldsymbol{x}^{(t)})^\flat + (\boldsymbol{\ell}^{(t-1)} \otimes \boldsymbol{x}^{(t)})^\flat - (\boldsymbol{\ell}^{(t-1)} \otimes \boldsymbol{x}^{(t-1)})^\flat\|_\infty^2 \\
&= \|((\boldsymbol{\ell}^{(t)} - \boldsymbol{\ell}^{(t-1)}) \otimes \boldsymbol{x}^{(t)})^\flat + (\boldsymbol{\ell}^{(t-1)} \otimes (\boldsymbol{x}^{(t)} - \boldsymbol{x}^{(t-1)}))^\flat\|_\infty^2 \\
&\le 2\|((\boldsymbol{\ell}^{(t)} - \boldsymbol{\ell}^{(t-1)}) \otimes \boldsymbol{x}^{(t)})^\flat\|_\infty^2 + 2\|(\boldsymbol{\ell}^{(t-1)} \otimes (\boldsymbol{x}^{(t)} - \boldsymbol{x}^{(t-1)}))^\flat\|_\infty^2 \qquad (7) \\
&= 2\|\boldsymbol{\ell}^{(t)} - \boldsymbol{\ell}^{(t-1)}\|_\infty^2 \|\boldsymbol{x}^{(t)}\|_\infty^2 + 2\|\boldsymbol{\ell}^{(t-1)}\|_\infty^2 \|\boldsymbol{x}^{(t)} - \boldsymbol{x}^{(t-1)}\|_\infty^2 \qquad (8) \\
&\le 2\|\boldsymbol{\ell}^{(t)} - \boldsymbol{\ell}^{(t-1)}\|_\infty^2 + 2\|\boldsymbol{\ell}\|_\infty^2 \|\boldsymbol{x}^{(t)} - \boldsymbol{x}^{(t-1)}\|_\infty^2, \qquad (9)
\end{aligned}$$

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

where we used the triangle inequality together with Young's inequality in (7); the property that $\|(\boldsymbol{w} \otimes \boldsymbol{z})^\flat\|_\infty = \|\boldsymbol{w}\|_\infty \|\boldsymbol{z}\|_\infty$ in (8); and the fact that $\|\boldsymbol{x}^{(t)}\|_\infty \leq 1$ in (9). As a result, if we plug-in (9) to (6) and we use the (A, B)-predictive bound of $\mathcal{R}_\Phi$ we can conclude that

$$\begin{aligned}\mathrm{Reg}^T &\leq A + \|\boldsymbol{\ell}\|_\infty \sum_{t=1}^T \epsilon^{(t)} + B \sum_{t=1}^T \left(2\|\boldsymbol{\ell}^{(t)} - \boldsymbol{\ell}^{(t-1)}\|_\infty^2 + 2\|\boldsymbol{\ell}\|_\infty^2 \|\boldsymbol{x}^{(t)} - \boldsymbol{x}^{(t-1)}\|_\infty^2\right) \\ &= A + 2B \sum_{t=1}^T \|\boldsymbol{\ell}^{(t)} - \boldsymbol{\ell}^{(t-1)}\|_\infty^2 + 2B\|\boldsymbol{\ell}\|_\infty^2 \sum_{t=1}^T \|\boldsymbol{x}^{(t)} - \boldsymbol{x}^{(t-1)}\|_\infty^2 + \|\boldsymbol{\ell}\|_\infty \sum_{t=1}^T \epsilon^{(t)},\end{aligned}$$

concluding the proof. □

A.4 Proofs for Section 4.1

In this subsection we include the omitted proofs from Section 4.1. We commence with the proof of Proposition 4.1. The corresponding construction follows that due to Farina et al. [2021a], and it is highlighted in Algorithm 2.

Proposition 4.1. *Consider a player $i \in [n]$ and any trigger sequence $\hat{\sigma} = (j, a) \in \Sigma_i^*$. There exists an algorithm which constructs a regret minimizer $\mathcal{R}_{\hat{\sigma}}$ with access to an (A, B)-predictive regret minimizer $\mathcal{R}_{Q_j}$ for the set Q_j such that $\mathcal{R}_{\hat{\sigma}}$ is (A, B)-predictive.*

Proof. Consider the (linear) function $g_{\hat{\sigma}}^{(t)} : \mathbb{R}^{|\Sigma_j|} \ni \boldsymbol{x} \mapsto L_i^{(t)}(h_{\hat{\sigma}}(\boldsymbol{x})) - L_i^{(t)}(h_{\hat{\sigma}}(\boldsymbol{0}))$, and let $\boldsymbol{g}_{\hat{\sigma}}^{(t)} = (\boldsymbol{L}_i^{(t)}[\sigma_r, \hat{\sigma}])_{\sigma_r \succeq j}$ be the associated utility vector. As suggested in Algorithm 2, the observed utility function $L_i^{(t)}$ at time t is first used to construct $g_{\hat{\sigma}}^{(t)}$. Then, the latter function is given as input to $\mathcal{R}_{Q_j}$. Thus, we may conclude that

$$\max_{\phi^* \in \Psi_{\hat{\sigma}}} \sum_{t=1}^T L_i^{(t)}(\phi^*) - \sum_{t=1}^T L_i^{(t)}\left(\phi_{\hat{\sigma} \to \boldsymbol{q}_{\hat{\sigma}}^{(t)}}\right) = \max_{\boldsymbol{q}_{\hat{\sigma}}^* \in Q_j} \sum_{t=1}^T g_{\hat{\sigma}}^{(t)}(\boldsymbol{q}_{\hat{\sigma}}^*) - \sum_{t=1}^T g_{\hat{\sigma}}^{(t)}(\boldsymbol{q}_{\hat{\sigma}}^{(t)}).$$

In words, the cumulative regret incurred by $\mathcal{R}_{\hat{\sigma}}$ under the sequence of utility functions $L_i^{(1)}, \dots, L_i^{(T)}$ is equal to the regret incurred by $\mathcal{R}_{Q_j}$ under the sequence of utility functions $g_{\hat{\sigma}}^{(t)}$. As a result, if we use the (A, B)-predictive bound assumed for the regret minimizer $\mathcal{R}_{Q_j}$, it follows that the cumulative regret Reg^T of $\mathcal{R}_{\hat{\sigma}}$ can be bounded as

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

where we used the fact that $\boldsymbol{g}_{\hat{\sigma}}^{(t)} = (\boldsymbol{L}_i^{(t)}[\sigma_r, \hat{\sigma}])_{\sigma_r \succeq j}$. Finally, the claim regarding the complexity of Algorithm 2 is direct since we can store the vector $\boldsymbol{g}_{\hat{\sigma}}^{(t)}$ in $O(|\Sigma_j|)$ time.

□

Next, we conclude the construction by combining the individual regret minimizers for all possible trigger sequences. In particular, we leverage the regret circuit of Proposition 4.4 to obtain the following result.

Proposition A.1. *Consider an (α, β)-predictive regret minimizer $\mathcal{R}_\Delta$ for the the simplex $\Delta(\Sigma_i^*)$, and (A, B)-predictive regret minimizers $\mathcal{R}_{\hat{\sigma}}$ for each $\hat{\sigma} \in \Sigma_i^*$, all with respect to the pair of dual norms $(\|\cdot\|_1, \|\cdot\|_\infty)$. Then, there exists an algorithm which constructs a regret minimizer $\mathcal{R}_{\Psi_i}$ for the set*

ALGORITHM 2: Predictive Regret Minimizer $\mathcal{R}_{\hat{\sigma}}$ for the set $\Psi_{\hat{\sigma}}$

Input:
- Player $i \in [n]$
- A trigger sequence $\hat{\sigma} = (j, a) \in \Sigma_i^*$
- An (A, B)-predictive regret minimizer $\mathcal{R}_{Q_j}$ for the set Q_j

function *NextStrategy*():
  $\boldsymbol{q}_{\hat{\sigma}}^{(t)} \leftarrow \mathcal{R}_{Q_j}.\textsc{NextStrategy}()$
  return $\phi_{\hat{\sigma} \leftarrow \boldsymbol{q}_{\hat{\sigma}}^{(t)}}$

function *ObserveUtility*($L_i^{(t)}$):
  Construct the linear function $g_{\hat{\sigma}}^{(t)} : \mathbb{R}^{|\Sigma_j|} \ni \boldsymbol{x} \mapsto L_i^{(t)}(h_{\hat{\sigma}}(\boldsymbol{x})) - L_i^{(t)}(h_{\hat{\sigma}}(\boldsymbol{0}))$
  $\mathcal{R}_{Q_j}.\textsc{ObserveUtility}(g_{\hat{\sigma}}^{(t)})$

co Ψ_i *such that under any sequence of utility vectors* $\boldsymbol{L}_i^{(1)}, \dots, \boldsymbol{L}_i^{(T)}$ *its regret* $\mathrm{Reg}_{\Psi_i}^T$ *can be bounded as*

$$\mathrm{Reg}_{\Psi_i}^T \le \alpha + A + (B + 4\beta|\Sigma_i|^2) \sum_{t=1}^{T} \|\boldsymbol{L}_i^{(t)} - \boldsymbol{L}_i^{(t-1)}\|_\infty^2.$$

Moreover, if the routines NextStrategy *and* ObserveUtility *of* $\mathcal{R}_\Delta$ *and* $\mathcal{R}_{\hat{\sigma}}$*, for each* $\hat{\sigma} \in \Sigma_i^*$*, run in linear time on* $|\Sigma_i|$*, then the complexity of* $\mathcal{R}_\Psi$ *is* $O(|\Sigma_i|^2)$*.*

The overall algorithm associated with this construction has been summarized in Algorithm 3.

Remark A.2. To obtain better predictive bounds, the regret minimizer $\mathcal{R}_\Delta$ acting over the simplex in Proposition A.1 will leverage the "future" iterates of all the individual regret minimizers. In particular, instead of using the typical one-recency bias mechanism $\boldsymbol{m}_\lambda^{(t)}[k] \coloneqq \langle \boldsymbol{L}_i^{(t-1)}, \boldsymbol{x}_k^{(t-1)} \rangle$, we will let $\boldsymbol{m}_\lambda^{(t)}[k] \coloneqq \langle \boldsymbol{L}_i^{(t-1)}, \boldsymbol{x}_k^{(t)} \rangle$. To this end, $\mathcal{R}_\Delta$ has to obtain the next iterate from each regret minimizer $\mathcal{R}_{\hat{\sigma}}$. This does not create complications given that the output of each $\mathcal{R}_{\hat{\sigma}}$ in the construction only depends on the observed utilities up to that time. On the other hand, it seems that there is no straightforward extension of this trick for Theorem 3.2, at the cost of a mismatch term of the form $\sum_{t=1}^T \|\boldsymbol{x}^{(t)} - \boldsymbol{x}^{(t-1)}\|_\infty$.

Proof of Proposition A.1. First of all, Proposition 4.4 implies that the accumulated regret can be bounded as

$$\mathrm{Reg}_{\Psi_i}^T \le \alpha + A + B \sum_{t=1}^{T} \|\boldsymbol{L}_i^{(t)} - \boldsymbol{L}_i^{(t-1)}\|_\infty^2 + \beta \sum_{t=1}^{T} \|\boldsymbol{\ell}_\lambda^{(t)} - \boldsymbol{m}_\lambda^{(t)}\|_\infty^2, \tag{10}$$

where we used the fact that each regret minimizer $\mathcal{R}_{\hat{\sigma}}$ obtains as input the same utility function as $\mathcal{R}_{\Psi_i}$. We also used the notation $\boldsymbol{\ell}_{\boldsymbol{\lambda}}^t \in \mathbb{R}^{|\Sigma_i^*|}$ to represent the utility function given to $\mathcal{R}_\Delta$ as predicted by Proposition 4.4. Next, let us focus on bounding the norm $\|\boldsymbol{\ell}_\lambda^{(t)} - \boldsymbol{m}_\lambda^{(t)}\|_\infty^2$. In particular, it follows that for some index $s \in \{1, \dots, |\Sigma_i^*|\}$,

$$\begin{aligned}
\|\boldsymbol{\ell}_\lambda^{(t)} - \boldsymbol{m}_\lambda^{(t)}\|_\infty^2 &= \left(\langle \boldsymbol{L}_i^{(t)}, \boldsymbol{x}_s^{(t)} \rangle - \langle \boldsymbol{L}_i^{(t-1)}, \boldsymbol{x}_s^{(t)} \rangle \right)^2 \\
&\le \|\boldsymbol{L}_i^{(t)} - \boldsymbol{L}_i^{(t-1)}\|_\infty^2 \|\boldsymbol{x}_s^{(t)}\|_1^2 \\
&\le 4|\Sigma_i|^2 \|\boldsymbol{L}_i^{(t)} - \boldsymbol{L}_i^{(t-1)}\|_\infty^2,
\end{aligned}$$

where we used the fact that $\|\boldsymbol{x}_s\|_1 \le 2|\Sigma_i|$. Thus, plugging this bound to (10) gives the desired predictive bound. 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., 2021a, Theorem 4.6], and it is therefore omitted. □

ALGORITHM 3: Predictive Regret Minimizer $\mathcal{R}_{\Psi_i}$ for the set $\operatorname{co}\Psi_i$

Input:
- Player $i \in [n]$
- An (A, B)-predictive regret minimizer $\mathcal{R}_{\hat{\sigma}}$ for $\Psi_{\hat{\sigma}}$, for each $\hat{\sigma} \in \Sigma_i^*$
- An (α, β)-predictive regret minimizer $\mathcal{R}_\Delta$ for $\Delta(\Sigma_i^*)$

Function *NextStrategy*():
- $\boldsymbol{\lambda}_i^{(t)} \leftarrow \mathcal{R}_\Delta.\,\textsc{NextStrategy}()$
- **for** $\hat{\sigma} \in \Sigma_i^*$ **do**
  - $\phi_{\hat{\sigma} \to \boldsymbol{q}_{\hat{\sigma}}^{(t)}} \leftarrow \mathcal{R}_{\hat{\sigma}}.\,\textsc{NextStrategy}()$
- **return** $\sum_{\hat{\sigma} \in \Sigma_i^*} \boldsymbol{\lambda}_i^{(t)}[\hat{\sigma}]\phi_{\hat{\sigma} \to \boldsymbol{q}_{\hat{\sigma}}^{(t)}}$ represented implicitly as $\{\boldsymbol{\lambda}_i^{(t)}[\hat{\sigma}], \boldsymbol{q}_{\hat{\sigma}}^{(t)}\}_{\hat{\sigma} \in \Sigma_i^*}$

Function *ObserveUtility*($L_i^{(t)}$):
- **for** $\hat{\sigma} \in \Sigma_i^*$ **do**
  - $\mathcal{R}_{\hat{\sigma}}.\,\textsc{ObserveUtility}(L_i^{(t)})$
- Construct the linear function $\ell_{\boldsymbol{\lambda}}^{(t)} : \boldsymbol{\lambda} \mapsto \sum_{\hat{\sigma} \in \Sigma_i^*} \boldsymbol{\lambda}[\hat{\sigma}] L_i^{(t)} \left(\phi_{\hat{\sigma} \to \boldsymbol{q}_{\hat{\sigma}}^{(t)}}\right)$
- $\mathcal{R}_\Delta.\,\textsc{ObserveUtility}(\ell_\lambda^{(t)})$

Finally, we combine the previous pieces to prove Theorem 4.5, which is recalled below.

Theorem 4.5. *There exists a regret minimization algorithm $\mathcal{R}_{\Psi_i}$ for the set $\operatorname{co}\Psi_i$ (Figure 2) such that under any sequence of utility vectors $\boldsymbol{L}_i^{(1)}, \dots, \boldsymbol{L}_i^{(T)}$ its regret $\operatorname{Reg}_{\Psi_i}^T$ can be bounded as*

$$\operatorname{Reg}_{\Psi_i}^T \le \frac{\log|\Sigma_i| + \|\boldsymbol{Q}_i\|_1^2 \max_{j \in \mathcal{J}_i} \log|\mathcal{A}_j|}{\eta} + \eta(\|\boldsymbol{Q}_i\|_1 + 4|\Sigma_i|^2) \sum_{t=1}^{T} \|\boldsymbol{L}_i^{(t)} - \boldsymbol{L}_i^{(t-1)}\|_\infty^2.$$

Proof. The claim follows directly from Lemma 2.3 using the fact that the range of the negative entropy DGF on the simplex $\Delta(\Sigma_i^*)$ is at most $\log|\Sigma_i|$; the predictive bound of Lemma 4.2; Proposition 4.1 with the regret minimizer $\mathcal{R}_{\boldsymbol{Q}_j}$ instantiated using the dilatable global entropy DGF (Lemma 4.2); and the predictive bound of the regret circuit for the convex hull derived in Proposition A.1. □

A.5 Proofs for Section 4.2

We start this subsection with the proof that OMWU guarantees multiplicative stability.

Lemma 4.6. *Consider the OMWU algorithm on the simplex Δ^m with $\eta > 0$. If all the observed utilities and the predictions are such that $\|\boldsymbol{\ell}^{(t)}\|_\infty, \|\boldsymbol{m}^{(t)}\|_\infty \le \|\boldsymbol{\ell}\|_\infty$, and $\eta < 1/(12\|\boldsymbol{\ell}\|_\infty)$, then the sequence $(\boldsymbol{x}^{(t)})$ produced by OMWU is $(12\eta\|\boldsymbol{\ell}\|_\infty)$-multiplicative-stable.*

Proof. It is well-known that the update rule of OMWU on the simplex can be expressed in the following form:

$$\boldsymbol{x}^{(t)}[k] = \frac{e^{\eta\boldsymbol{\ell}^{(t-1)}[k] + \eta\boldsymbol{m}^{(t)}[k] - \eta\boldsymbol{m}^{(t-1)}[k]}}{\sum_{k'=1}^m e^{\eta\boldsymbol{\ell}^{(t-1)}[k'] + \eta\boldsymbol{m}^{(t)}[k'] - \eta\boldsymbol{m}^{(t-1)}[k']}\boldsymbol{x}^{(t-1)}[k']} \boldsymbol{x}^{(t-1)}[k],$$

for all $k \in [m]$ and $t \geq 2$. As a result, we have that

$$\boldsymbol{x}^{(t)}[k] \leq \frac{e^{3\eta\|\boldsymbol{\ell}\|_\infty}}{\sum_{k'=1}^m e^{-3\eta\|\boldsymbol{\ell}\|_\infty}\boldsymbol{x}^{(t-1)}[k']}\boldsymbol{x}^{(t-1)}[k] = e^{6\eta\|\boldsymbol{\ell}\|_\infty}\boldsymbol{x}^{(t-1)}[k] \leq (1+12\eta\|\boldsymbol{\ell}\|_\infty)\boldsymbol{x}^{(t-1)}[k],$$

where we used that $\boldsymbol{\ell}^{(t-1)}[k'], \boldsymbol{m}^{(t)}[k'], \boldsymbol{m}^{(t-1)}[k'] \in [-\|\boldsymbol{\ell}\|_\infty, \|\boldsymbol{\ell}\|_\infty]$, for all $k' \in [m]$, the fact that $\sum_{k'} \boldsymbol{x}^{(t-1)}[k'] = 1$ since $\boldsymbol{x}^{(t-1)} \in \Delta^m$, and that $e^x \leq 1 + 2x$, for all $x \in [0, 1/2]$. Similarly, we have that

$$\begin{aligned}\boldsymbol{x}^{(t)}[k] \geq \frac{e^{-3\eta\|\boldsymbol{\ell}\|_\infty}}{\sum_{k'=1}^m e^{3\eta\|\boldsymbol{\ell}\|_\infty}\boldsymbol{x}^{(t-1)}[k']}\boldsymbol{x}^{(t-1)}[k] = e^{-6\eta\|\boldsymbol{\ell}\|_\infty}\boldsymbol{x}^{(t-1)}[k] &\geq (1-6\eta\|\boldsymbol{\ell}\|_\infty)\boldsymbol{x}^{(t-1)}[k] \\ &\geq (1+12\eta\|\boldsymbol{\ell}\|_\infty)^{-1}\boldsymbol{x}^{(t-1)}[k],\end{aligned}$$

for $\eta \leq 1/(12\|\boldsymbol{\ell}\|_\infty)$. □

Lemma 4.7. *Consider the* (OFTRL) *algorithm on the sequence-form strategy polytope $\mathcal{Q}$ with dilatable global entropy as DGF and $\eta > 0$. If all the utility functions are such that $\|\boldsymbol{\ell}^{(t)}\|_\infty \leq 1$, and $\eta = O(1/\mathfrak{D})$ is sufficiently small, then the sequence $(\boldsymbol{x}^{(t)})$ produced is $O(\eta\mathfrak{D})$-multiplicative-stable.*

Proof. Let $\boldsymbol{S}^{(t-1)} \coloneqq \sum_{\tau=1}^{t-1} \boldsymbol{\ell}^{(\tau)}$. We claim that the next iterate of (OFTRL) with dilatable global entropy as DGF can be computed as follows. First, we compute recursively the quantities

$$\boldsymbol{r}^{(t)}[j] \coloneqq \boldsymbol{\gamma}[j]\log\left(\sum_{a\in\mathcal{A}_j}\exp\left\{\frac{\eta\boldsymbol{S}^{(t-1)}[(j,a)] + \eta\boldsymbol{m}^{(t)}[(j,a)] - \sum_{j':\sigma_{j'}=(j,a)}\boldsymbol{r}^{(t)}[j']}{\boldsymbol{\gamma}[j]}\right\}\right) \tag{11}$$

through a bottom-up tree traversal. Then, we determine the (local) behavioral strategies $\boldsymbol{b}_j \in \Delta(\mathcal{A}_j)$ at every decision point $j \in \mathcal{J}$ based on the following update rule:

$$\boldsymbol{b}_j[a] \propto \exp\left\{\frac{\eta\boldsymbol{S}^{(t-1)}[(j,a)] + \eta\boldsymbol{m}^{(t)}[(j,a)] - \sum_{j':\sigma_{j'}=(j,a)}\boldsymbol{r}^{(t)}[j']}{\boldsymbol{\gamma}[j]}\right\}. \tag{12}$$

Finally, the computed behavioral strategies are converted to the sequence-form representation. To argue about the multiplicative stability of the induced sequence, let us use the notation

$$\boldsymbol{s}^{(t)}[(j,a)] \coloneqq \frac{1}{\boldsymbol{\gamma}[j]}\left(2\eta\boldsymbol{\ell}^{(t-1)}[(j,a)] - \eta\boldsymbol{\ell}^{(t-2)}[(j,a)] - \sum_{j':\sigma_{j'}=(j,a)}\left(\boldsymbol{r}^{(t)}[j'] - \boldsymbol{r}^{(t-1)}[j']\right)\right). \tag{13}$$

Assuming that $\boldsymbol{m}^{(t)} \coloneqq \boldsymbol{\ell}^{(t-1)}$, it follows from (11) that

$$\begin{aligned}\boldsymbol{r}^{(t)}[j] &= \boldsymbol{\gamma}[j]\log\left(\sum_{a\in\mathcal{A}_j}\exp\left\{\frac{\eta\boldsymbol{S}^{(t-2)}[(j,a)] + \eta\boldsymbol{\ell}^{(t-2)}[(j,a)] - \sum_{j':\sigma_{j'}=(j,a)}\boldsymbol{r}^{(t-1)}[j']}{\boldsymbol{\gamma}[j]}\right\}e^{\boldsymbol{s}^{(t)}[(j,a)]}\right) \\ &\leq \boldsymbol{r}^{(t-1)}[j] + \boldsymbol{\gamma}[j]\max_{a\in\mathcal{A}_j}\boldsymbol{s}^{(t)}[(j,a)].\end{aligned}$$

Similarly, we have that

$$\boldsymbol{r}^{(t)}[j] \geq \boldsymbol{r}^{(t-1)}[j] + \boldsymbol{\gamma}[j]\min_{a\in\mathcal{A}_j}\boldsymbol{s}^{(t)}[(j,a)] = \boldsymbol{r}^{(t-1)}[j] - \boldsymbol{\gamma}[j]\max_{a\in\mathcal{A}_j}(-\boldsymbol{s}^{(t)}[(j,a)]).$$

Thus, we have shown that

$$\left|\boldsymbol{r}^{(t)}[j] - \boldsymbol{r}^{(t-1)}[j]\right| \leq \boldsymbol{\gamma}[j]\max_{a\in\mathcal{A}_j}|\boldsymbol{s}^{(t)}[(j,a)]|.$$

Recalling the definition of $\boldsymbol{s}^{(t)}[(j,a)]$ given in (13) we find that

$$\left|\boldsymbol{r}^{(t)}[j]-\boldsymbol{r}^{(t-1)}[j]\right| \le \max_{a\in\mathcal{A}_j}\left|2\eta\boldsymbol{\ell}^{(t-1)}[(j,a)]-\eta\boldsymbol{\ell}^{(t-2)}[(j,a)]-\sum_{\sigma_{j'}=(j,a)}\left(\boldsymbol{r}^{(t)}[j']-\boldsymbol{r}^{(t-1)}[j']\right)\right|$$
$$\le 3\eta+\max_{a\in\mathcal{A}_j}\sum_{j':\sigma_{j'}=(j,a)}\left|\boldsymbol{r}^{(t)}[j']-\boldsymbol{r}^{(t-1)}[j']\right|, \tag{14}$$

where we used the assumption that $\|\boldsymbol{\ell}^{(t-1)}\|_\infty, \|\boldsymbol{\ell}^{(t-2)}\|_\infty \le 1$. Now (12) can be equivalently expressed as

$$\boldsymbol{b}_j^{(t)}[a] \propto \boldsymbol{b}_j^{(t-1)}[a]\exp\left\{\frac{2\eta\boldsymbol{\ell}^{(t-1)}[(j,a)]-\eta\boldsymbol{\ell}^{(t-2)}[(j,a)]-\sum_{j':\sigma_{j'}=(j,a)}(\boldsymbol{r}^{(t)}[j']-\boldsymbol{r}^{(t-1)}[j'])}{\boldsymbol{\gamma}[j]}\right\}.$$

Using (14) and the assumption that $\|\boldsymbol{\ell}^{(t-1)}\|_\infty, \|\boldsymbol{\ell}^{(t-2)}\|_\infty \le 1$, it follows that

$$\left|\frac{2\eta\boldsymbol{\ell}^{(t-1)}[(j,a)]-\eta\boldsymbol{\ell}^{(t-2)}[(j,a)]-\sum_{j':\sigma_{j'}=(j,a)}(\boldsymbol{r}^{(t)}[j']-\boldsymbol{r}^{(t-1)}[j'])}{\boldsymbol{\gamma}[j]}\right| = O(\eta),$$

where we used the definition of $\boldsymbol{\gamma}$ given in (3). As a result, similarly to the argument in the proof of Lemma 4.6 we conclude that the sequence $(\boldsymbol{b}_j^{(t)})$ is $O(\eta)$-multiplicative-stable. Finally, the sequence-form strategy $\boldsymbol{x}^{(t)}[(j,a)]$ is computed by taking the product of all $\boldsymbol{b}_{j'}^{(t)}[a']$ for all sequences (j',a') on the path from the root to (j,a). Given that there are at most $\mathfrak{D}$ sequences on every path, we may conclude that for any $\sigma\in\Sigma$,

$$\boldsymbol{x}^{(t)}[\sigma] \le (1+O(\eta))^{\mathfrak{D}}\boldsymbol{x}^{(t-1)}[\sigma] \le (1+O(\eta\mathfrak{D}))\boldsymbol{x}^{(t-1)}[\sigma],$$

for a sufficiently small $\eta = O(1/\mathfrak{D})$. Similar reasoning yields that $\boldsymbol{x}^{(t)}[\sigma] \ge (1+O(\eta\mathfrak{D}))^{-1}\boldsymbol{x}^{(t-1)}[\sigma]$, concluding the proof. □

Next, we combine Lemmas 4.6 and 4.7 to show Corollary 4.8.

Proof of Corollary 4.8. Let us first focus on the regret minimizer $\mathcal{R}_{\hat{\sigma}}$, for some arbitrary $\hat{\sigma} = (j,a) \in \Sigma_i^*$. First, as predicted by Theorem 3.2, the utility function $\boldsymbol{L}_i^{(t)}$ is constructed as $\boldsymbol{L}_i^{(t)} := (\boldsymbol{\ell}_i^{(t)} \otimes \boldsymbol{x}_i^{(t)})^\flat$. Proposition 4.4 implies that this is the same utility observed by $\mathcal{R}_{\hat{\sigma}}$. Moreover, from the construction of Algorithm 2 we can conclude that the utility $\boldsymbol{g}_{\hat{\sigma}}^{(t)}$ observed by $\mathcal{R}_{\mathcal{Q}_j}$ will be such that $\|\boldsymbol{g}_{\hat{\sigma}}^{(t)}\| \le 1$ given that $\|\boldsymbol{x}_i^{(t)}\|_\infty \le 1$ (since $\boldsymbol{x}_i^{(t)} \in \mathcal{Q}_i$) and $\|\boldsymbol{\ell}_i^{(t)}\|_\infty \le 1$ by the normalization assumption. Thus, we conclude from Lemma 4.7 that the output sequence of $\mathcal{R}_{\mathcal{Q}_j}$ will be $O(\eta\mathfrak{D}_i)$-multiplicative-stable. Furthermore, the construction of Algorithm 2 immediately implies that the output sequence of $\mathcal{R}_{\hat{\sigma}}$ will also be $O(\eta\mathfrak{D}_i)$.

Next, we establish the claim regarding the stability of $\mathcal{R}_\Delta$. Indeed, it is easy to see that the utility $\boldsymbol{\ell}_\lambda^{(t)}$ observed by $\mathcal{R}_\Delta$ is such that $\|\boldsymbol{\ell}_\lambda\|_\infty = O(\|\boldsymbol{Q}_i\|_1)$, and the same holds for the prediction $\boldsymbol{m}_\lambda^{(t)}$. Thus, Lemma 4.6 completes the proof. □

Next, we focus on the proof of Theorem 4.14. To this end, we leverage the approach of Kruckman et al. [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 A.3. *The following properties hold:*

- *The determinant is a multilinear function with respect to the rows and columns of the matrix:*

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

  for any $\boldsymbol{u}_1, \dots, \boldsymbol{u}_m \in \mathbb{R}^m$, $\boldsymbol{u}'_k \in \mathbb{R}^m$, *and* $\alpha, \beta \in \mathbb{R}$*;*
- *If any two rows or columns of* $\mathbf{A}$ *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 its i-th row and its j-th column. Then, the *cofactor* is defined as $\mathrm{co}^{(i,j)}(\mathbf{A}) = (-1)^{i+j} \det\left(\mathtt{mn}^{(i,j)}(\mathbf{A})\right)$, while the *adjugate* (or adjoint) matrix $\mathrm{adj}(\mathbf{A})^\top$ is the matrix with entries the corresponding cofactors of $\mathbf{A}$; that is, $\mathrm{adj}(\mathbf{A})[(i,j)] \coloneqq \mathrm{co}^{(j,i)}(\mathbf{A})$. With this notation at hand, we are ready to state the following characterization due to [Kruckman et al., 2010, Theorem 3.4]:

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

A key step of our proof for Theorem 4.14 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 A.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 A.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 A.6 ([Kruckman et al., 2010])**.** *Consider some* $m \times m$ *matrix* $\mathbf{D} \in \mathcal{D}_i$*, and let* R_i *be the determinant of the Laplacian matrix* $\mathcal{L} \coloneqq \mathtt{mp}(\mathbf{D}) - \mathbf{I}$ *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 4.11, we also state a useful elementary fact.

Fact A.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 4.11. *Let* $\mathbf{M}$ *be the transition matrix of an* m*-state Markov chain such that* $\mathbf{M} \coloneqq \boldsymbol{v}\mathbf{1}^\top + \mathbf{C}$*, 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 parameter.*

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 A.7 that Σ_1 can be expressed as

$$\Sigma_1 = \det(\boldsymbol{e}[1], \boldsymbol{v} + \boldsymbol{c}_2 - \boldsymbol{e}[2], \dots, \boldsymbol{v} + \boldsymbol{c}_m - \boldsymbol{e}[m]), \tag{15}$$

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 have 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 (15) it follows that

$$\Sigma_1 = \det\left(\boldsymbol{e}[1], \sum_{k=1}^{m} \boldsymbol{e}_{k,2}\boldsymbol{v}[k] + \sum_{k=1}^{m} \boldsymbol{e}_{k,2}\boldsymbol{c}_2[k], \dots, \sum_{k=1}^{m} \boldsymbol{e}_{k,m}\boldsymbol{v}[k] + \sum_{k=1}^{m} \boldsymbol{e}_{k,m}\boldsymbol{c}_m[k]\right). \tag{16}$$

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

$$\det\left(\boldsymbol{e}[1], \sum_{k=1}^{m} \boldsymbol{e}_{k,2}\boldsymbol{c}_2[k], \dots, \sum_{k=1}^{m} \boldsymbol{e}_{k,m}\boldsymbol{c}_m[k]\right), \tag{17}$$

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

$$\det\left(\boldsymbol{e}[1], z_2, \dots, \sum_{k=1}^{m} \boldsymbol{e}_{k,j}\boldsymbol{v}[k], \dots, z_m\right), \tag{18}$$

for some index j, where z_ℓ 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 (18). 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 (18), expressed in the form $\sum_{k=1}^{m} \boldsymbol{e}_{k,\ell}\boldsymbol{v}[k]$, for some index $\ell \neq j$, 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 subtraction operation does not modify the value of the determinant. Thus, by multinearity, the determinant (18) is equal to

$$\lambda^p \det\left(\boldsymbol{e}[1], z'_2, \dots, \sum_{k=1}^{m} \boldsymbol{e}_{k,j}\boldsymbol{v}[k], \dots, z'_m\right), \tag{19}$$

where z'_ℓ is either $\sum_{k=1}^{m} \boldsymbol{e}_{k,\ell}\boldsymbol{c}_\ell[k]$ or $\boldsymbol{e}_{j,\ell}$, and $0 \leq p \leq m-2$. Next, if we use again the multilinearity property, the term in (19) 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 suppress 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 A.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 (17), 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

$$(-1)^{m-1} \sum_{j\in F_i} \lambda^{p_j} (\boldsymbol{v}[q_j])^{b_j} \prod_{(s,w)\in S_j} \mathbf{C}[(s,w)], \tag{20}$$

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 $\Sigma \coloneqq \sum_{i=1}^m \Sigma_i$. In particular, the stationary distribution $\boldsymbol{\pi}$ of $\mathbf{M}$ is such that

$$\left(\mathbf{C} + \boldsymbol{v}\mathbf{1}^\top\right)\boldsymbol{\pi} = \boldsymbol{\pi} \iff \mathbf{C}\boldsymbol{\pi} + \boldsymbol{v} = \boldsymbol{\pi} \iff (\mathbf{I}_m - \mathbf{C})\boldsymbol{\pi} = \boldsymbol{v}, \tag{21}$$

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 $\boldsymbol{\pi}$:

$$\boldsymbol{\pi}[1] = \frac{\det(\boldsymbol{v}, \boldsymbol{e}[2] - \boldsymbol{c}_2, \dots, \boldsymbol{e}[m] - \boldsymbol{c}_m)}{\det(\boldsymbol{e}[1] - \boldsymbol{c}_1, \boldsymbol{e}[2] - \boldsymbol{c}_2, \dots, \boldsymbol{e}[m] - \boldsymbol{c}_m)}. \tag{22}$$

Moreover, it follows that

$$\begin{aligned}
\det(\boldsymbol{v}, \boldsymbol{e}[2] - \boldsymbol{c}_2, \dots, \boldsymbol{e}[m] - \boldsymbol{c}_m) &= \det(\boldsymbol{v}, \boldsymbol{e}[2] - \boldsymbol{c}_2 - \boldsymbol{v}, \dots, \boldsymbol{e}[m] - \boldsymbol{c}_m - \boldsymbol{v}) \\
&= \det(\boldsymbol{v} + (\lambda\boldsymbol{e}[1] - \boldsymbol{v}), \boldsymbol{e}[2] - \boldsymbol{c}_2 - \boldsymbol{v}, \dots, \boldsymbol{e}[m] - \boldsymbol{c}_m - \boldsymbol{v}) \qquad (23) \\
&= \lambda \det(\boldsymbol{e}[1], \boldsymbol{e}[2] - \boldsymbol{c}_2 - \boldsymbol{v}, \dots, \boldsymbol{e}[m] - \boldsymbol{c}_m - \boldsymbol{v}),
\end{aligned}$$

where in (23) 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 A.7, and (22), 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 A.4 that $\boldsymbol{\pi}_1 = \Sigma_1/\Sigma$, implying the following identity:

$$\det(\mathbf{I}_m - \mathbf{C}) = \lambda \sum_{i=1}^m \Sigma_i. \tag{24}$$

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 $\mathbf{v} \coloneqq \frac{\lambda}{m}\mathbf{1}$. That is,

$$\det(\mathbf{I}_m - \mathbf{C}) = \lambda(-1)^{m-1} \sum_{j\in F} C_j \lambda^{p_j + b_j} \prod_{(s,w)\in S_j} \mathbf{C}[(s,w)], \tag{25}$$

where $|S_j| \le m - 1 - p_j$, $C_j = C_j(m)$ 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 (20), (24), and (25). □

Corollary A.8. *Let $\mathbf{M}$ be the transition matrix of an m-state Markov chain such that $\mathbf{M} \coloneqq \boldsymbol{v}\mathbf{1}^\top + \mathbf{C}$, 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 $b_j \in \{0,1\}, 0 \le p_j \le m-2, |S_j| = m - p_j - b_j - 1$, for each $j \in F_i$, such that the i-th coordinate of the vector $\boldsymbol{w} \coloneqq l\boldsymbol{\pi}$ can be expressed as

$$\boldsymbol{w}[i] = \frac{\sum_{j \in F_i} \lambda^{p_j+1} (\boldsymbol{r}[q_j])^{b_j} l^{1-b_j} \prod_{(s,w)\in S_j} \mathbf{C}[(s,w)]}{\sum_{j \in F} C_j \lambda^{p_j+b_j} \prod_{(s,w)\in S_j} \mathbf{C}[(s,w)]}, \tag{26}$$

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

Proof. The proof follows directly from the formula derived in Lemma 4.11. □

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 4.12, which is recalled next.

Corollary 4.12. *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 $\boldsymbol{w} \coloneqq l\boldsymbol{\pi}$ and $\boldsymbol{w}' \coloneqq l'\boldsymbol{\pi}'$. 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 scalars 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

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

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 A.8. Moreover, for a sufficiently small $\kappa = O(1/m)$, we can infer that $V_j' \le (1+\gamma)(1+O(\kappa m))V_j' = (1+(\gamma+O(\kappa m)))V_j$. In turn, 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 (26) 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 $\boldsymbol{w}'[i] \ge (1+(\gamma+O(\kappa m)))^{-1}\boldsymbol{w}[i]$. □

Next, we will use this statement to prove Proposition 4.13, which is recalled below.

Proposition 4.13. *Consider a player $i \in [n]$, and let $\phi_i^{(t)} = \sum_{\hat{\sigma}\in\Sigma_i^*} \boldsymbol{\lambda}_i^{(t)}[\hat{\sigma}]\phi_{\hat{\sigma}\to \boldsymbol{q}_{\hat{\sigma}}^{(t)}}$ be a transformation in $\operatorname{co}\Psi_i$ such that the sequences $(\boldsymbol{\lambda}_i^{(t)})$ and $(\boldsymbol{q}_{\hat{\sigma}}^{(t)})$ are κ-multiplicative-stable, for all $\hat{\sigma} \in \Sigma_i^*$. If $(\boldsymbol{x}_i^{(t)})$ is a γ-multiplicative-stable J-partial fixed point sequence, the sequence of $(J \cup \{j^*\})$-partial fixed points of ϕ_i is $(\gamma + O(\kappa|\mathcal{A}_{j^*}|))$-multiplicative-stable, for any sufficiently small $\kappa = O(1/|\mathcal{A}_{j^*}|)$.*

We note that it is tacitly assumed that the vectors $\boldsymbol{\lambda}_i^{(t)}, \boldsymbol{q}_{\hat{\sigma}}^{(t)}$ and $\boldsymbol{x}_{(j\in J)}$, involved in Proposition 4.13, have strictly positive coordinates; this is indeed the case under our dynamics (Figure 2).

Proof of Proposition 4.13. Let us focus on the stability analysis of Algorithm 4 as the rest of the claim follows from [Farina et al., 2021a, Proposition 4.14]. In particular, for consistency with the terminology of Corollary 4.12, let us define

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

and $l := \boldsymbol{x}_i[\sigma_p]$. We will show that the conditions of Corollary 4.12 are satisfied:

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

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

(iii) The sequence $(\boldsymbol{r}^{(t)})$ is $\gamma + O(\kappa)$-multiplicative-stable. This assertion can be directly verified from the definition of $\boldsymbol{r}$ in Algorithm 4;

(iv) The sequence of scalars $(l^{(t)})$ is γ-multiplicative-stable. Indeed, this follows directly from the assumption that the sequence $(\boldsymbol{x}_i^{(t)})$ is γ-multiplicative-stable.

As a result, we can apply Corollary 4.12 to conclude the proof. □

Theorem 4.14. *Consider a player $i \in [n]$, and let $\phi_i^{(t)} = \sum_{\hat{\sigma} \in \Sigma_i^*} \boldsymbol{\lambda}_i^{(t)}[\hat{\sigma}] \phi_{\hat{\sigma} \to \boldsymbol{q}_{\hat{\sigma}}^{(t)}}$ be a transformation in $\operatorname{co} \Psi_i$ such that the sequences $(\boldsymbol{\lambda}_i^{(t)})$ and $(\boldsymbol{q}_{\hat{\sigma}}^{(t)})$ are κ-multiplicative-stable, for all $\hat{\sigma} \in \Sigma_i^*$. Then, the sequence of fixed points $\boldsymbol{q}_i^{(t)} \in \mathcal{Q}_i$ of $\phi_i^{(t)}$ is $O(\kappa |\mathcal{A}_i| |\mathfrak{D}_i|)$-multiplicative-stable, for a sufficiently small $\kappa = O(1/(|\mathcal{A}_i| |\mathfrak{D}_i|))$, where $|\mathcal{A}_i| := \max_{j \in \mathcal{J}_i} |\mathcal{A}_j|$.*

Proof. Our argument proceeds inductively. For a root information set $j \in \mathcal{J}_i$, Proposition 4.13 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-multiplicative-stable. Next, the theorem follows inductively given that by Proposition 4.13 each sequence can only incur an additive factor of $O(\kappa |\mathcal{A}|)$ in the multiplicative stability bound with respect to the preceding sequences. □

Remark A.9. More precisely, if $F_i := \max_{j_1 < j_2 < \cdots < j_d} \sum_{i=1}^d |\mathcal{A}_{j_i}|$, with $j_1, \dots, j_d \in \mathcal{J}_i$, we can show that the sequence of fixed points is $O(\kappa F_i)$-multiplicative-stable. Observe that F_i can be trivially upper bounded by $|\mathcal{A}_i| |\mathfrak{D}_i|$, as well as the number of sequences $|\Sigma_i|$.

A.6 Proofs from Section 4.3

We begin this subsection with the proof of Claim 4.16, which is recalled below.

Claim 4.16. *For any player $i \in [n]$ the observed utilities satisfy*

$$\|\boldsymbol{\ell}_i^{(t)} - \boldsymbol{\ell}_i^{(t-1)}\|_\infty^2 \le (n-1) |\mathcal{Z}|^2 \sum_{k \neq i} \|\boldsymbol{q}_k^{(t)} - \boldsymbol{q}_k^{(t-1)}\|_1^2.$$

Proof. For a profile of mixed sequence-form strategies $(\boldsymbol{q}_1, \dots, \boldsymbol{q}_n)$, the utility of player i can be expressed as

$$u_i(\boldsymbol{q}_1, \dots, \boldsymbol{q}_n) = \sum_{z \in \mathcal{Z}} p_c(z) u_i(z) \prod_{k=1}^n \boldsymbol{q}_k(\sigma_{k,z}).$$

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

$$\begin{aligned} \|\boldsymbol{\ell}_i^{(t)} - \boldsymbol{\ell}_i^{(t-1)}\|_\infty &\le \sum_{z \in \mathcal{Z}} \left| \prod_{k \neq i}^n \boldsymbol{q}_k^{(t)}(\sigma_{k,z}) - \prod_{k \neq i}^n \boldsymbol{q}_k^{(t-1)}(\sigma_{k,z}) \right| \\ &\le \sum_{z \in \mathcal{Z}} \sum_{k \neq i}^n \left| \boldsymbol{q}_k^{(t)}(\sigma_{k,z}) - \boldsymbol{q}_k^{(t-1)}(\sigma_{k,z}) \right|, \end{aligned} \tag{27}$$

ALGORITHM 4: EXTEND($\phi_i, J, j^*, \boldsymbol{x}$); [Farina et al., 2021a]

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

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

Let $\boldsymbol{r} \in \mathbb{R}_{\geq 0}^{|\mathcal{A}_{j^*}|}$ be defined as $\boldsymbol{r}[a] \coloneqq \sum_{j' \leq \sigma_{j^*}} \sum_{a' \in \mathcal{A}_{j'}} \boldsymbol{\lambda}_i[(j', a')] \boldsymbol{q}_{(j',a')}[(j^*, a)] \boldsymbol{x}_i[(j', a')]$

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

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

if $\boldsymbol{x}_i[\sigma_{j^*}] = 0$ **then**
  $\boldsymbol{w} \leftarrow \mathbf{0} \in \mathbb{R}_{\geq 0}^{|\mathcal{A}_{j^*}|}$
else
  $\boldsymbol{b} \in \Delta(\mathcal{A}_{j^*}) \leftarrow$ stationary distribution of $\frac{1}{\boldsymbol{x}_i[\sigma_{j^*}]} \mathbf{W}$
  $\boldsymbol{w} \to \boldsymbol{x}_i[\sigma_{j^*}] \boldsymbol{b}$
$\boldsymbol{x}_i' \leftarrow \boldsymbol{x}_i$
for $a \in \mathcal{A}_{j^*}$ **do**
  $\boldsymbol{x}_i'[(j^*, a)] \leftarrow \boldsymbol{w}[(j^*, a)]$

ALGORITHM 5: FIXEDPOINT(ϕ_i); [Farina et al., 2021a]

Input: $\phi_i = \sum_{\hat{\sigma} \in \Sigma_i^*} \boldsymbol{\lambda}_i[\hat{\sigma}] \phi_{\hat{\sigma} \to \boldsymbol{q}_{\hat{\sigma}}} \in \operatorname{co} \Psi_i$

Output: $\boldsymbol{q}_i \in \mathcal{Q}_i$ such that $\boldsymbol{q}_i = \phi_i(\boldsymbol{q}_i)$

$\boldsymbol{q}_i \leftarrow \mathbf{0} \in \mathbb{R}^{|\Sigma_i|}, \boldsymbol{q}_i[\varnothing] \leftarrow \varnothing$
$J \leftarrow \varnothing$
for $j \in \mathcal{J}_i$ *in top-down order* **do**
  $\boldsymbol{q}_i \leftarrow$ EXTEND$(\phi_i, J, j, \boldsymbol{q}_i^*)$
  $J = J \cup \{j\}$
return $\boldsymbol{q}_i^*$

where in the last bound we used the well-known inequality

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

for any $a_1, \dots, a_m, b_1, \dots, b_m \in [0, 1]$. Finally, from (27) we can conclude that

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

Finally, the claim follows from a standard application of Young's inequality. □

Next, we include the proof of Theorem 1.1.

PROOF OF THEOREM 1.1. For a player $i \in [n]$ we let $\boldsymbol{\mu}_i^{(t)}$ be any probability distribution on the set Π_i such that $\mathbb{E}_{\boldsymbol{\pi}_i \sim \mu_i^{(t)}}[\boldsymbol{\pi}_i] = \boldsymbol{q}_i^{(t)}$, where $\boldsymbol{q}_i^{(t)}$ is the output of the regret minimizer operating over the *mixed* sequence-form strategy polytope $\mathcal{Q}_i$, realized with the dynamics associated with Corollary 4.17. Moreover, let $\boldsymbol{\mu}^{(t)} \coloneqq \boldsymbol{\mu}_1^{(t)} \otimes \dots \otimes \boldsymbol{\mu}_n^{(t)}$ be the associated joint probability distribution,

and $\bar{\boldsymbol{\mu}}^{(t)} \coloneqq \frac{1}{T}\sum_{t=1}^{T} \boldsymbol{\mu}^{(t)}$ be their average over time. Then, since the expression in Definition 2.7 is linear (recall that the set of transformations Ψ_i is linear), it follows from the linearity of expectation that $\bar{\boldsymbol{\mu}}^{(t)}$ is an ϵ-EFCE, where, if Reg_i^T is the cumulative Ψ_i-regret of player i with respect to $\mathcal{Q}_i$, it holds that $\epsilon \coloneqq \frac{1}{T}\max_i \mathrm{Reg}_i^T$. Finally, the proof follows given that $\mathrm{Reg}_i^T = \mathcal{P}T^{1/4}$, for every player $i \in [n]$, where $\mathcal{P}$ is a parameter polynomial in the game (Corollary 4.17). □

A.7 Proofs from Section 5

Before we proceed with the proof of Theorem 5.1, we first show the following useful claim.

Lemma A.10. *For any $\phi_i = \sum_{j' \in \mathcal{J}_i} \boldsymbol{\lambda}_i[j']\phi_{j' \to \boldsymbol{q}_{j'}} \in \operatorname{co}\Psi_i$, $\boldsymbol{q}_i \in \mathcal{Q}_i$, and $\sigma = (j, a) \in \Sigma_i$,*

$$\phi_i(\boldsymbol{q}_i)[\sigma] - \boldsymbol{q}_i[\sigma] = \left(\sum_{j' \preceq j} \boldsymbol{\lambda}_i[j']\boldsymbol{q}_{j'}[\sigma]\boldsymbol{q}_i[\sigma_{j'}]\right) - d_\sigma \boldsymbol{q}_i[\sigma].$$

Proof. By definition of the linear mapping $\phi_{j' \to \boldsymbol{q}_{j'}}$, we have that

$$\begin{aligned}
\phi_i(\boldsymbol{q}_i)[\sigma] &= \sum_{j' \in \mathcal{J}_i} \boldsymbol{\lambda}_i[j']\phi_{j' \to \boldsymbol{q}_{j'}}(\boldsymbol{q}_i)[\sigma] \\
&= \sum_{j' \in \mathcal{J}_i} \boldsymbol{\lambda}_i[j'] \begin{cases} \boldsymbol{q}_{j'}[\sigma]\boldsymbol{q}_i[\sigma_{j'}] & \text{if } \sigma \succeq j' \\ \boldsymbol{q}_i[\sigma] & \text{otherwise} \end{cases} \\
&= \left(1 - \sum_{j' \preceq \sigma} \boldsymbol{\lambda}_i[j']\right)\boldsymbol{q}_i[\sigma] + \sum_{j' \preceq \sigma} \boldsymbol{\lambda}_i[j']\boldsymbol{q}_{j'}[\sigma]\boldsymbol{q}_i[\sigma_{j'}].
\end{aligned}$$

A rearrangement of the last equation completes the proof. □

Theorem 5.1. *For any player $i \in [n]$ and any transformation $\phi_i = \sum_{j \in \mathcal{J}_i} \boldsymbol{\lambda}_i[j]\phi_{j \to \boldsymbol{q}_j} \in \operatorname{co}\widetilde{\Psi}_i$, the output $\boldsymbol{q}_i \in \mathbb{R}^{|\Sigma_i|}$ of Algorithm 1 is such that $\boldsymbol{q}_i \in \mathcal{Q}_i$ and $\phi_i(\boldsymbol{q}_i) = \boldsymbol{q}_i$. Furthermore, Algorithm 1 runs in $O(|\Sigma_i|\mathfrak{D}_i)$.*

Proof. Consider some arbitrary $\phi_i \in \operatorname{co}\Psi_i$. The proof is divided into three claims: (i) the vector $\boldsymbol{q}_i \in \mathbb{R}^{|\Sigma_i|}$ obtained through Algorithm 1 is such that $\boldsymbol{q}_i \in \mathcal{Q}_i$ (*i.e.*, it is a proper sequence-form strategy); (ii) the sequence-form strategy $\boldsymbol{q}_i$ obtained through Algorithm 1 is such that $\phi_i(\boldsymbol{q}_i) = \boldsymbol{q}_i$; and (iii) Algorithm 1 runs in time $O(|\Sigma_i|\mathfrak{D}_i)$.

Part 1: $\boldsymbol{q}_i$ is a sequence-form strategy. First, by construction Line 1) we have that $\boldsymbol{q}_i[\varnothing] = 1$. Thus, we need to show that, for each $j \in \mathcal{J}_i$, it holds that $\sum_{a \in \mathcal{A}_j} \boldsymbol{q}_i[(j, a)] = \boldsymbol{q}_i[\sigma_j]$ (recall Definition 2.1). Indeed, for any $j \in \mathcal{J}_i$ such that $d_\sigma = 0$, it is immediate to see that the above constraint holds by construction (Line 5). On the other hand, for each $j \in \mathcal{J}_i$ such that $d_\sigma \neq 0$, we have that

$$\begin{aligned}
\sum_{a \in \mathcal{A}_j} \boldsymbol{q}_i[(j, a)] &= \frac{1}{d_\sigma}\left(\sum_{a \in \mathcal{A}_j}\sum_{j' \preceq j} \boldsymbol{\lambda}_i[j']\boldsymbol{q}_{j'}[(j, a)]\boldsymbol{q}_i[\sigma_{j'}]\right) \\
&= \frac{1}{d_\sigma}\left(\sum_{j' \preceq j} \boldsymbol{\lambda}_i[j']\boldsymbol{q}_i[\sigma_{j'}]\left(\sum_{a \in \mathcal{A}_j} \boldsymbol{q}_{j'}[(j, a)]\right)\right) \\
&= \frac{1}{d_\sigma}\left(\sum_{j' \preceq j} \boldsymbol{\lambda}_i[j']\boldsymbol{q}_i[\sigma_{j'}] \cdot \begin{cases} \boldsymbol{q}_{j'}[\sigma_j] & \text{if } j' \prec j \\ 1 & \text{otherwise} \end{cases}\right),
\end{aligned}$$

where the first equality holds by Line 7, and the last equality holds since $\boldsymbol{q}_{j'} \in \mathcal{Q}_{j'}$. Next, we distinguish between two cases: if $d_{\sigma_j} = 0$, then $\boldsymbol{\lambda}_i[j'] = 0$ for each $j' \prec j$. Therefore, since we are

assuming $d_\sigma \neq 0$, it must be the case that $d_\sigma = \boldsymbol{\lambda}_i[j] \neq 0$. This yields that

$$\sum_{a \in \mathcal{A}_j} \boldsymbol{q}_i[(j,a)] = \frac{1}{d_\sigma}\left(\sum_{j' \preceq j} \boldsymbol{\lambda}_i[j']\boldsymbol{q}_i[\sigma_{j'}] \cdot \begin{cases} \boldsymbol{q}_{j'}[\sigma_j] & \text{if } I' \prec I \\ 1 & \text{otherwise} \end{cases}\right)$$
$$= \frac{1}{\boldsymbol{\lambda}_i[j]}\left(\boldsymbol{\lambda}_i[j]\boldsymbol{q}_i[\sigma_j]\right) = \boldsymbol{q}_i[\sigma_j].$$

On the other hand, if $d_{\sigma_j} \neq 0$, then $\boldsymbol{q}_i[\sigma_j]$ was set according to Line 7, and thus,

$$\boldsymbol{q}_i[\sigma_j] = \frac{1}{d_{\sigma_j}}\left(\sum_{j' \prec j} \boldsymbol{\lambda}_i[j']\boldsymbol{q}_i[\sigma_{j'}]\boldsymbol{q}_{j'}[\sigma_j]\right). \tag{28}$$

By definition of d_σ (Line 3), it holds that $d_\sigma = d_{\sigma_j} + \boldsymbol{\lambda}_i[j]$. Thus,

$$\sum_{a \in \mathcal{A}_j} \boldsymbol{q}_i[(j,a)] = \frac{1}{d_\sigma}\left(\sum_{j' \preceq j} \boldsymbol{\lambda}_i[j']\boldsymbol{q}_i[\sigma_{j'}] \cdot \begin{cases} \boldsymbol{q}_{j'}[\sigma_j] & \text{if } j' \prec j \\ 1 & \text{otherwise} \end{cases}\right)$$
$$= \frac{1}{d_{\sigma_j} + \boldsymbol{\lambda}_i[j]}\left(\boldsymbol{\lambda}_i[j]\boldsymbol{q}_i[\sigma_j] + \sum_{j' \prec j} \boldsymbol{\lambda}_i[j']\boldsymbol{q}_i[\sigma_{j'}]\boldsymbol{q}_{j'}[\sigma_j]\right)$$
$$= \frac{1}{d_{\sigma_j} + \boldsymbol{\lambda}_i[j]}\left(\boldsymbol{\lambda}_i[j]\boldsymbol{q}_i[\sigma_j] + d_{\sigma_j}\boldsymbol{q}_i[\sigma_j]\right) = \boldsymbol{q}_i[\sigma_j],$$

where the second to last equality is obtained from (28). This concludes the first part of the proof.

Part 2: $\boldsymbol{q}_i$ is a fixed point of ϕ_i. Fix a sequence $\sigma = (j,a) \in \Sigma_i$. We want to show that $\phi(\boldsymbol{q}_i)[\sigma] - \boldsymbol{q}_i[\sigma] = 0$. If $\sum_{j' \preceq j} \boldsymbol{\lambda}_i[j'] = 0$, then it immediately follows that $\phi_i(\boldsymbol{q}_i)[\sigma] = \boldsymbol{q}_i[\sigma]$. Otherwise, applying Lemma A.10 and substituting $\boldsymbol{q}_i[\sigma]$ according to Line 7 yields that

$$\phi_i(\boldsymbol{q}_i)[\sigma] - \boldsymbol{q}_i[\sigma] = \left(\sum_{j' \preceq j} \boldsymbol{\lambda}_i[j']\boldsymbol{q}_{j'}[\sigma]\boldsymbol{q}_i[\sigma_{j'}]\right) - d_\sigma \boldsymbol{q}_i[\sigma]$$
$$= \left(\sum_{j' \preceq j} \boldsymbol{\lambda}_i[j']\boldsymbol{q}_{j'}[\sigma]\boldsymbol{q}_i[\sigma_{j'}]\right) - \frac{d_\sigma}{d_\sigma}\left(\sum_{j' \preceq j} \boldsymbol{\lambda}_i[j']\boldsymbol{q}_{j'}[\sigma]\boldsymbol{q}_i[\sigma_{j'}]\right) = 0.$$

This concludes the second part of the proof.

Part 3: time complexity. For each sequence in Σ_i^* Algorithm 1 has to visit at most $\mathfrak{D}_i$ information sets as part of Lines 3 and 7. This completes the proof. □

Proposition 5.2. *Suppose that the sequences $(\boldsymbol{\lambda}_i^{(t)})$ and $(\boldsymbol{q}_j^{(t)})$, for all $j \in \mathcal{J}_i$, are κ-multiplicative-stable. Then, Algorithm 1 yields a sequence of $(12\kappa\mathfrak{D}_i)$-multiplicative-stable strategies, assuming that $\kappa < 1/(12\mathfrak{D}_i)$.*

Proof. By assumption, we know that $\boldsymbol{\lambda}_i[j] > 0$ for all $j \in \mathcal{J}_i$. Thus, it will always be the case that $d_\sigma > 0$, for any $\sigma \in \Sigma_i$. Hence, Algorithm 1 will never visit the first "if" branch.

Now fix any $t \geq 2$. We will show by induction that $\boldsymbol{q}_i^{(t)}[\sigma]$ is such that $\boldsymbol{q}_i^{(t)}[\sigma] \leq (1+\kappa)^{3\mathfrak{D}_i[\sigma]-2}\boldsymbol{q}_i^{(t-1)}[\sigma]$ and $\boldsymbol{q}_i^{(t-1)}[\sigma] \leq (1+\kappa)^{3\mathfrak{D}_i[\sigma]-2}\boldsymbol{q}_i^{(t)}[\sigma]$, 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 $\sigma = (j,a)$ be a sequence such that $j \in \mathcal{J}_i$ corresponds to a root information set of player i. Then, it follows from Algorithm 1 that $d_\sigma = \boldsymbol{\lambda}_i[j]$, in turn implying that $\boldsymbol{q}_i^{(t)}[\sigma] = \boldsymbol{q}_j^{(t)}[\sigma]$. Thus, $\boldsymbol{q}_i^{(t)}[\sigma]$ is indeed κ-multiplicative-stable.

Next, consider some sequence $\sigma = (j, a)$ at depth $\mathfrak{D}_i[\sigma] \geq 2$ such that all ancestor sequences—*i.e.* all $\sigma_{j'}$ for $j' \preceq j$—satisfy the inductive hypothesis. Then, we have that

$$\begin{aligned}
\boldsymbol{q}_i^{(t)}[\sigma] &= \frac{\sum_{j' \preceq j} \boldsymbol{\lambda}_i^{(t)}[j'] \boldsymbol{q}_{j'}^{(t)}[\sigma] \boldsymbol{q}_i^{(t)}[\sigma_{j'}]}{\sum_{j' \preceq j} \boldsymbol{\lambda}_i^{(t)}[j']} && (29)\\
&\leq (1+\kappa)^3 \frac{\sum_{j' \preceq j} \boldsymbol{\lambda}_i^{(t-1)}[j'] \boldsymbol{q}_{j'}^{(t-1)}[\sigma] \boldsymbol{q}_i^{(t)}[\sigma_{j'}]}{\sum_{j' \preceq j} \boldsymbol{\lambda}_i^{(t-1)}[j']} && (30)\\
&\leq (1+\kappa)^3 (1+\kappa)^{3\mathfrak{D}_i[\sigma]-5} \boldsymbol{q}_i^{(t-1)}[\sigma] && (31)\\
&= (1+\kappa)^{3\mathfrak{D}_i[\sigma]-2} \boldsymbol{q}_i^{(t-1)}[\sigma],
\end{aligned}$$

where (29) derives from the formula of Algorithm 1; (30) uses the κ-multiplicative-stability of the sequences $(\boldsymbol{\lambda}_i^{(t)})$ and $(\boldsymbol{q}_j^{(t)})$, for any $j \in \mathcal{J}_i$; and (31) leverages the inductive hypothesis. Similar reasoning yields:

$$\begin{aligned}
\boldsymbol{q}_i^{(t)}[\sigma] &= \frac{\sum_{j' \preceq j} \boldsymbol{\lambda}_i^{(t)}[j'] \boldsymbol{q}_{j'}^{(t)}[\sigma] \boldsymbol{q}_i^{(t)}[\sigma_{j'}]}{\sum_{j' \preceq j} \boldsymbol{\lambda}_i^{(t)}[j']}\\
&\geq \frac{1}{(1+\kappa)^3} \frac{\sum_{j' \preceq j} \boldsymbol{\lambda}_i^{(t-1)}[j'] \boldsymbol{q}_{j'}^{(t-1)}[\sigma] \boldsymbol{q}_i^{(t)}[\sigma_{j'}]}{\sum_{j' \preceq j} \boldsymbol{\lambda}_i^{(t-1)}[j']}\\
&\geq \frac{1}{(1+\kappa)^3} \frac{1}{(1+\kappa)^{3\mathfrak{D}_i[\sigma]-5}} \boldsymbol{q}_i^{(t-1)}[\sigma]\\
&\geq \frac{1}{(1+\kappa)^{3\mathfrak{D}_i[\sigma]-2}} \boldsymbol{q}_i^{(t-1)}[\sigma].
\end{aligned}$$

Thus, if $\mathfrak{D}_i$ is the depth of $\mathcal{J}_i$, we conclude that $\boldsymbol{q}_i^{(t)}[\sigma] \leq (1+\kappa)^{3\mathfrak{D}_i-2} \boldsymbol{q}_i^{(t-1)}[\sigma] \leq e^{3\mathfrak{D}_i\kappa-2\kappa} \boldsymbol{q}_i^{(t-1)}[\sigma] \leq (1+6\mathfrak{D}_i\kappa) \boldsymbol{q}_i^{(t-1)}[\sigma]$, where we used the inequalities $1+x \leq e^x$ for all $x \in \mathbb{R}$, and $e^x \leq 1+2x$ for $x \in [0, 1/2]$, applicable as long as $\kappa \leq 1/(12\mathfrak{D}_i)$. Similarly, we obtain that $\boldsymbol{q}_i^{(t)} \geq (1+12\mathfrak{D}_i\kappa)^{-1} \boldsymbol{q}_i^{(t-1)}$, concluding the proof. □

B 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 in [Farina et al., 2019c]. The main result is given in Theorem B.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, at every observation point $k \in \mathcal{K}$ the agent may receive a feedback in the form of a signal in the 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 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 decision point j. It is important to point out that in all extensive-form games of perfect recall the agents face a sequential decision process. A central ingredient for our construction of stable-predictive CFR is a decomposition of the strategy space, described in detail below.

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

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

$$\mathcal{X}_k^{\triangle} := \bigtimes_{j \in C_k} \mathcal{X}_j^{\triangle}; \tag{32}$$

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

$$\mathcal{X}_j^{\triangle} := \left\{ \begin{pmatrix} \boldsymbol{\lambda}[1] \\ \vdots \\ \boldsymbol{\lambda}[m_j] \\ \boldsymbol{\lambda}[1]\boldsymbol{x}_1 \\ \vdots \\ \boldsymbol{\lambda}[m_j]\boldsymbol{x}_{m_j} \end{pmatrix} : (\boldsymbol{\lambda}[1], \dots, \boldsymbol{\lambda}[m_j]) \in \Delta^{m_j}, \boldsymbol{x}_1 \in \mathcal{X}_{k_1}^{\triangle}, \dots, \boldsymbol{x}_{m_j} \in \mathcal{X}_{k_{m_j}}^{\triangle} \right\}. \tag{33}$$

In view of this decomposition, the basic ingredients for the overall construction are given in Proposition B.1 and Proposition B.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 B.1. *Consider an observation node $k \in \mathcal{K}$, and assume access to a κ_j-multiplicative-stable (α_j, β_j)-predictive regret minimizer $\mathcal{R}_j^{\triangle}$ over the sequence-form strategy space $\mathcal{X}_j^{\triangle}$, for each $j \in C_k$. 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 $A = \sum_{j \in C_k} \alpha_j$ and $B = \sum_{j \in C_k} \beta_j$.*

Proof. Given the decomposition of (32), the composite regret minimizer can be constructed using a regret circuit for the Cartesian product [Farina et al., 2019b]. In particular, it is direct to verify that $\mathrm{Reg}_k^{\triangle,T} = \sum_{j \in C_k} \mathrm{Reg}_j^{\triangle,T}$, where $\mathrm{Reg}_k^{\triangle,T}$ is the regret accumulated by the composite regret minimizer, and $\mathrm{Reg}_j^{\triangle,T}$ the regret of each individual regret minimizer $\mathcal{R}_j^{\triangle}$. In particular, by assumption we know that

$$\mathrm{Reg}_j^{\triangle,T} \le \alpha_j + \beta_j \sum_{t=1}^{T} \|\boldsymbol{\ell}_j^{\triangle,(t)} - \boldsymbol{\ell}_j^{\triangle,(t-1)}\|_\infty^2.$$

As a result, we can conclude that

$$\mathrm{Reg}_k^{\triangle,T} \le \left(\sum_{j \in C_k} \alpha_j \right) + \left(\sum_{j \in C_k} \beta_j \right) \sum_{t=1}^{T} \|\boldsymbol{\ell}_k^{\triangle,(t)} - \boldsymbol{\ell}_k^{\triangle,(t-1)}\|_\infty^2,$$

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

In the following construction the regret circuit for the convex hull uses an advanced prediction mechanism, analogously to that we explained in Remark A.2.

Proposition B.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 $k \in C_j$. Moreover, assume access to a κ-multiplicative-stable (α, β)-predictive regret minimizer $\mathcal{R}_\Delta$ over the*

simplex $\Delta(\mathcal{A}_j)$. Then, we can construct a $(\kappa + \kappa K + K)$-multiplicative-stable (A, B)-predictive regret minimizer $\mathcal{R}_j^{\triangle}$ for the sequence-form strategy space $\mathcal{X}_j^{\triangle}$, where

$$A = \alpha + \max_{k \in C_j}\{\alpha_k\};$$
$$B = \max_{k \in C_j}\{\beta_k\} + \beta \|Q\|_1^2,$$

where $\|Q\|_1$ 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 4.4. First, we have that, by assumption, the regret $\mathrm{Reg}_k^{\triangle,T}$ accumulated by each regret minimizer $\mathcal{R}_k^{\triangle}$ can be bounded as

$$\mathrm{Reg}_k^{\triangle,T} \le \alpha_k + \beta_k \sum_{t=1}^{T} \|\boldsymbol{\ell}_k^{\triangle,(t)} - \boldsymbol{\ell}_k^{\triangle,(t-1)}\|_\infty^2.$$

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 4.4, imply that

$$\mathrm{Reg}_j^{\triangle,T} \le \alpha + \max_{k \in C_j}\{\alpha_k\} + \max_{k \in C_j}\{\beta_k\} \sum_{t=1}^{T} \|\boldsymbol{\ell}_j^{\triangle,(t)} - \boldsymbol{\ell}_j^{\triangle,(t-1)}\|_\infty^2 + \beta \sum_{t=1}^{T} \|\boldsymbol{\ell}_\lambda^{(t)} - \boldsymbol{\ell}_\lambda^{(t-1)}\|_\infty^2, \tag{34}$$

where $\boldsymbol{\ell}_\lambda^{(t)}$ represents the utility function received as input by $\mathcal{R}_\Delta$. Next, similarly to the analysis of Proposition A.1, we can infer that for some $k \in C_j$,

$$\|\boldsymbol{\ell}_\lambda^{(t)} - \boldsymbol{\ell}_\lambda^{(t-1)}\|_\infty = |\langle \boldsymbol{\ell}_j^{\triangle,(t)} - \boldsymbol{\ell}_j^{\triangle,(t-1)}, \boldsymbol{x}_k^{(t)} \rangle| \le \|\boldsymbol{\ell}_j^{\triangle,(t)} - \boldsymbol{\ell}_j^{\triangle,(t-1)}\|_\infty \|\boldsymbol{x}_k^{(t)}\|_1 \le \|\boldsymbol{\ell}_j^{\triangle,(t)} - \boldsymbol{\ell}_j^{\triangle,(t-1)}\|_\infty \|Q\|_1$$

where we used that $\|\boldsymbol{x}_k^{(t)}\|_1 \le \|Q\|_1$. As a result, if we plug-in this bound to (34) we can conclude that

$$\mathrm{Reg}_j^{\triangle,T} \le \left(\alpha + \max_{k \in C_j}\{\alpha_k\}\right) + \left(\max_{k \in C_j}\{\beta_k\} + \beta \|Q\|_1^2\right) \sum_{t=1}^{T} \|\boldsymbol{\ell}_j^{\triangle,(t)} - \boldsymbol{\ell}_j^{\triangle,(t-1)}\|_\infty^2.$$

Finally, the $(\kappa+\kappa K+K)$-multiplicative-stability of $\mathcal{R}_j^{\triangle}$ can be directly verified from the decomposition given in (33). □

Remark B.3. Given the decomposition provided in Equation (33), the regret circuit for the convex hull should operate on the appropriate "lifted" space, but this does not essentially alter the analysis of the regret since the augmented entries in the lifted space remain invariant; this is illustrated and further explained in [Farina et al., 2019b, Figure 7].

Finally, we inductively combine Proposition B.1 and Proposition B.2 in order to establish the main result of this section: a stable-predictive variant of CFR.

Theorem B.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 (A, B)-predictive regret minimizer $\mathcal{R}^{\triangle}$ for the space of sequence-form strategies $\mathcal{X}^{\triangle}$, such that*

$$A = O\left(\frac{\log|\mathcal{A}|}{\eta} \|Q\|_1\right);$$
$$B = O(\eta \|Q\|_1^3), \tag{35}$$

where $|\mathcal{A}| := \max_{j \in \mathcal{J}} |\mathcal{A}_j|$; $\|\boldsymbol{\ell}\|_\infty$ is an upper bound on the ℓ_∞ norm of the utilities observed by $\mathcal{R}^{\triangle}$; $\|Q\|_1$ is an upper bound on the ℓ_1 norm of any $\boldsymbol{x} \in \mathcal{X}^{\triangle}$; and $\mathfrak{D}$ is the depth of the decision process. Moreover, the sequence of strategies produced by $\mathcal{R}^{\triangle}$ is $O(\eta \mathfrak{D} \|Q\|_1 \|\boldsymbol{\ell}\|_\infty)$-multiplicative-stable.

Proof. First of all, it is easy to see that all losses observed by the "local" regret minimizers—*i.e.*, the *counterfactual losses* [Farina et al., 2019c, Section 4]—have ℓ_∞ bounded by $O(\|Q\|_1\|\ell\|_\infty)$. As a result, we can conclude from Lemma 4.6 that the output of each local regret minimizer $\mathcal{R}_j^\triangle$ under OMWU with a sufficiently small learning rate η is $O(\eta\|Q\|_1\|\ell\|_\infty)$-multiplicative-stable. Along with Proposition B.2, we can inductively infer that the output of $\mathcal{R}^\triangle$ is $O(\eta\mathfrak{D}\|Q\|_1\|\ell\|_\infty)$-multiplicative-stable, for a sufficiently small $\eta = O(1/(\mathfrak{D}\|Q\|_1\|\ell\|_\infty))$. This established the claimed bound for the multiplicative stability.

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 information set $j \in \mathcal{J}$ instantiated with OMWU with learning rate η will be $(\log(|\mathcal{A}_j|/\eta, \eta)$-predictive. As a result, the predictive bound in (35) follows inductively from Proposition B.2. □

Naturally, the same bounds apply for constructing a regret minimizer for the subspace $\mathcal{X}_j^\triangle$, for any decision point $j \in \mathcal{J}$, as required in Proposition 4.1.

C DESCRIPTION OF GAME INSTANCES USED IN THE EXPERIMENTS

In this section we give a description of the game instances used in our experiments. The parameters associated with each game are summarized in Table 2.

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., 2019d]. 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 et al. [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 dices 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.

Goofspiel. This game was introduced—in its current form—by Ross [1971]. In Goofspiel every player has a hand of cards numbered from 1 to r, where r is the *rank* of the game. An additional stack of r cards is shuffled and singled out as winning the current prize. In each turn a prize card is revealed, and each player privately chooses one of its cards to bid. The player with the highest card wins the current prize; in case of a tie, the prize card is discarded. After r turns have been completed, all the prizes have been dealt out and players obtain the sum of the values of the prize cards they have won. It is worth noting that, due to the tie-breaking mechanism, even two-player instances are general-sum. We also consider instances with *limited information*—the actions of the other players are observed only at the end of the game. This makes the game strategically more involved as players have less information regarding previous opponents' actions.

Game	Players	Decision points	Sequences	Leaves
Kuhn poker	3	36	75	78
Sheriff	2	73	222	256
Goofspiel	3	837	934	1296
Liar's dice	3	1536	3069	13797

Table 2. The parameters of each game.