Efficient Decentralized Learning Dynamics for Extensive-Form Coarse Correlated Equilibrium: No Expensive Computation of Stationary Distributions Required

Gabriele Farina*,[1] Andrea Celli*,[2] Tuomas Sandholm[1,3,4,5]

[1]Carnegie Mellon University, [2]Bocconi University, [3]Strategic Machine, Inc., [4]Strategy Robot, Inc., [5]Optimized Markets, Inc.
gfarina@cs.cmu.edu, andrea.celli2@unibocconi.it, sandholm@cs.cmu.edu

Abstract

While in two-player zero-sum games the Nash equilibrium is a well-established prescriptive notion of optimal play, its applicability as a prescriptive tool beyond that setting is limited. Consequently, the study of decentralized learning dynamics that guarantee convergence to *correlated* solution concepts in multiplayer, general-sum extensive-form (*i.e.*, tree-form) games has become an important topic of active research. The per-iteration complexity of the currently known learning dynamics depends on the specific correlated solution concept considered. For example, in the case of extensive-form correlated equilibrium (EFCE), all known dynamics require, as an intermediate step at each iteration, to compute the stationary distribution of multiple Markov chains, an expensive operation in practice. Oppositely, in the case of normal-form coarse correlated equilibrium (NFCCE), simple no-external-regret learning dynamics that amount to a linear-time traversal of the tree-form decision space of each agent suffice to guarantee convergence. This paper focuses on extensive-form coarse correlated equilibrium (EFCCE), an intermediate solution concept that is a subset of NFCCE and a superset of EFCE. Being a superset of EFCE, any learning dynamics for EFCE automatically guarantees convergence to EFCCE. However, since EFCCE is a simpler solution concept, this begs the question: *do learning dynamics for EFCCE that avoid the expensive computation of stationary distributions exist?* This paper answers the previous question in the positive. Our learning dynamics only require the orchestration of no-external-regret minimizers, thus showing that EFCCE is more akin to NFCCE than to EFCE from a learning perspective. Our dynamics guarantees that the empirical frequency of play after T iteration is a $O(1/\sqrt{T})$-approximate EFCCE with high probability, and an EFCCE almost surely in the limit.

1 Introduction

In a normal-form game (i.e., a game with simultaneous moves), a *correlated strategy* is defined as a probability distribution over joint action profiles, and it is customarily modeled via a trusted external mediator that draws an action profile from this distribution, and privately recommends to each player their component. A correlated strategy is a *correlated equilibrium* (CE) if, for each player, the mediator's recommendation is the best action in expectation, as-

suming all the other players follow their recommended actions (Aumann 1974). CE is an appealing solution concept in real-world strategic interactions involving more than two players with arbitrary (i.e., general-sum) utilities. Indeed, in those settings, the notion of CE overcomes several weaknesses of the *Nash equilibrium* (NE) (Nash 1950). In particular, in settings beyond two-players zero-sum games, the NE is prone to equilibrium selection issues, it is computationally intractable (being PPAD-complete even in two-player games (Chen and Deng 2006; Daskalakis, Goldberg, and Papadimitriou 2009)), and the social welfare that can be attained at an NE may be arbitrarily lower than what can be achieved through a CE (Koutsoupias and Papadimitriou 1999; Roughgarden and Tardos 2002; Celli and Gatti 2018). In contrast, a CE explicitly models synchronization between players, and it is computable in polynomial time in normal-form games. Moreover, in arbitrary normal-form games, the notion of CE arises naturally from simple decentralized learning dynamics (Foster and Vohra 1997; Hart and Mas-Colell 2000). Decentralized learning dynamics offer a parallel, scalable avenue for computing equilibria, and allow players to circumvent the—often unreasonable—assumption that they have perfect knowledge of other players' payoff functions. In particular, players can adjust their strategies on the basis of their own private payoff function, and on the observed behavior of the other players. In the case of NE, decentralized learning dynamics are only known in the two-player zero-sum setting (see, e.g., Cesa-Bianchi and Lugosi (2006); Hart and Mas-Colell (2003)).

Extensive-form games generalize normal-form games by modeling both sequential and simultaneous moves, as well as imperfect information. Because of their sequential nature, extensive-form games admit various notions of correlated equilibrium, which essentially differ in the time at which each player can decide whether to deviate or to follow recommendations. 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, Bianchi, and Sandholm (2020), and the *normal-form coarse correlated equilibrium* (NFCCE) by Celli, Coniglio, and Gatti (2019). The set of those equilibria are such that, for any extensive-form game, EFCE $\subseteq$ EFCCE $\subseteq$ NFCCE. Decentralized no-regret learning dynamics are known for the

set of EFCE (Celli et al. 2020; Farina et al. 2021; Morrill et al. 2021), and they require, as an intermediate step at each iteration, to compute the stationary distribution of multiple Markov chains, which can be an expensive operation in practice. On the other hand, the set of NFCCE admits simple no-external-regret learning dynamics that amount to a linear-time traversal of the tree-form decision space of each agent (Celli et al. 2019). This paper studies decentralized learning dynamics converging to the set of EFCCE. In an EFCCE, before the beginning of the game, the mediator draws a recommended action for each of the possible information sets that players may encounter in the game, according to some known probability distribution defined over joint deterministic strategies. These recommendations are not immediately revealed to each player. Instead, the mediator incrementally reveals relevant action recommendations as players reach new information sets. At each information set the acting player has to commit to following the recommended move before it is revealed to them, by only knowing the mediator's policy used to draw recommendations and the past recommendations issued from the root of the game tree down to the current information set (Farina, Bianchi, and Sandholm 2020). If the acting player decides to deviate (i.e., commits to *not following* the recommendation), their recommendations will no longer be issued by the mediator. Since the set of EFCEs is a subset of the set of EFCCEs (Farina, Bianchi, and Sandholm 2020), learning dynamics for EFCE automatically guarantees convergence to EFCCE. However, since EFCCE is a simpler solution concept, the following natural question arises: *do learning dynamics for EFCCE that avoid the expensive computation of stationary distributions exist?* This paper answers the previous question in the positive. In particular, we define the notion of *coarse trigger regret* as a particular instantiation of the *phi-regret minimization* framework (Greenwald and Jafari 2003; Stoltz and Lugosi 2007; Gordon, Greenwald, and Marks 2008), and we show that if each player behaves according to a no-coarse-trigger-regret algorithm, then the empirical frequency of play approaches the set of EFCCEs. Then, we provide an efficient algorithm for minimizing coarse trigger regret based on the general template for constructing phi-regret minimizers by Gordon, Greenwald, and Marks (2008). We show that, in contrast to EFCE, any convex combination of coarse trigger deviation functions admits a fixed point strategy which can be computed in closed form, without requiring to compute the stationary distribution of any Markov chain. In particular, our learning dynamics only require the orchestration of no-external-regret minimizers, thus showing that EFCCE is more akin to NFCCE than to EFCE from a learning perspective. Our algorithm guarantees that the empirical frequency of play after T iteration is a $O(1/\sqrt{T})$-approximate EFCCE with high probability, and an EFCCE almost surely in the limit.

Related work. The study of adaptive procedures converging to a CE in normal-form games dates back to the works by Foster and Vohra (1997), Fudenberg and Levine (1995, 1999), and Hart and Mas-Colell (2000, 2001). In more recent years, a growing effort has been devoted to understanding the relationships between no-regret learning dynam-

ics and equilibria in extensive-form games. While in two-player zero-sum extensive-form games it is widely known that no-regret learning dynamics converge to an NE (see, e.g., (Zinkevich et al. 2008; Tammelin et al. 2015; Lanctot et al. 2009; Brown and Sandholm 2019)) the general case of multi-player general-sum games is less understood. Celli et al. (2019) provide variations of the classical CFR algorithm, showing that they provably converge to the set of NFCCEs. Celli et al. (2020) describe learning dynamics that converge to the set of EFCE almost surely in the limit. Their algorithm requires to instantiate and manage a number of *internal regret minimizers* growing linearly in the number of information sets in the game. Each internal regret minimizer internally requires the computation of a stationary distribution of a Markov chain (Cesa-Bianchi and Lugosi 2006; Blum and Mansour 2007). Farina et al. (2021) extend the work by Celli et al. (2020), giving convergence guarantees to the set of EFCEs at finite time in high probability. The latter paper operates within the phi-regret minimization framework of Gordon, Greenwald, and Marks (2008), and requires the computation of the stationary distribution of multiple Markov chains at each iteration. The recent work by Morrill et al. (2021) presents a general framework for achieving hindsight rational learning (Morrill et al. 2020) in extensive-form games for various types of behavioral deviations. It is known that, when framework by Morrill et al. (2021) (EFR) is instantiated with different choices of sets of behavioral deviations, EFR leads to different solution concepts (including EFCCE in the case of *blind causal deviations*). Just like the other mentioned approaches, the EFR framework requires the computation of fixed points of linear transformations at each iteration. We conjecture that a similar result as this paper (*i.e.*, the existence of a fixed point that can computed in closed form without the need to compute any stationary distribution of a Markov chain) could also be derived within the EFR framework, when blind causal deviations are considered, though we leave exploration of that direction open.

2 Preliminaries

The set $\{1, \ldots, n\}$, with $n \in \mathbb{N}_{>0}$, is compactly denoted as $[n]$. Given a set S, we denote its convex hull with the symbol $\operatorname{co} S$.

2.1 Extensive-Form Games

An extensive-form game is usually defined by means of an oriented rooted game tree. The set of nodes that are not a leaf of the game tree is denoted by $\mathcal{H}$. Each node $h \in \mathcal{H}$ is called a *decision node* and has associated a player that acts at that node by choosing one action from the set of available actions at h, which we denote by $\mathcal{A}(h)$. In an n-player extensive-form game, the set of players is the set $[n] \cup \{c\}$, where c denotes the chance player, which is a fictitious player that selects actions according to fixed probability distributions representing exogenous stochasticity of the environment (e.g., a roll of the dice). Leaves of the game tree are called *terminal nodes*, and represent the outcomes of the game; their set of available actions is conventionally set to $\emptyset$ and they are not assigned to an acting player. The set of such nodes is

denoted by $\mathcal{Z}$. When the game transitions to a terminal node $z \in \mathcal{Z}$, payoffs are assigned to each non-chance player according to the set of payoff functions $\{u^{(i)} : \mathcal{Z} \to \mathbb{R}\}_{i \in [n]}$. Moreover, we let $p^{(c)} : \mathcal{Z} \to (0, 1)$ denote the function assigning to each terminal node z the product of probabilities of chance moves encountered on the path from the root of the game tree to z.

Imperfect information. The set of decision nodes of each player $i \in [n]$ is partitioned into a collection $\mathcal{I}^{(i)}$ of sets of nodes, called *information sets*. Each information set $I \in \mathcal{I}^{(i)}$ groups together nodes that Player i cannot distinguish between when Player i acts. Therefore, we have that $A(h) = A(h')$ for any pair of nodes $h, h' \in I$. Then, we can safely write $A(I)$ to indicate the set of actions available at any decision node belonging to $I \in \mathcal{I}^{(i)}$. As it is customary in the literature, we assume that the extensive-form game has *perfect recall*, that is, information sets are such that no player forgets information once acquired. This means that, for any player $i \in [n]$ and any two nodes $h, h' \in I$, with $I \in \mathcal{I}^{(i)}$, the sequence of Player i's actions from the root to h must coincide with the sequence of Player i's actions from the root to h'. Therefore, for any $i \in [n]$, we can define a partial ordering $\prec$ on $\mathcal{I}^{(i)}$ as follows: for any $I, I' \in \mathcal{I}^{(i)}$, $I \prec I'$ if there exist nodes $h' \in I'$ and $h \in I$ such that the path from the root of the game to h' passes through h. An immediate consequence of perfect recall is that for any $i \in [n]$, $\mathcal{I}^{(i)}$ is well-ordered by $\prec$ (i.e., given $I \in \mathcal{I}^{(i)}$, the set of its predecessors forms a chain).

Sequences. For any player $i \in [n]$, information set $I \in \mathcal{I}^{(i)}$, and action $a \in A(I)$, we denote by $\sigma = (I, a)$ the sequence of Player i's actions on the path from the root of the game tree down to action a (included) taken at any decision node in information set I. We denote by $\varnothing$ the *empty sequence* of Player i. Then, the set of Player i's sequences is defined as $\Sigma^{(i)} := \{(I, a) : I \in \mathcal{I}^{(i)}, a \in A(I)\} \cup \{\varnothing\}$. Given an information set $I \in \mathcal{I}^{(i)}$, we denote by $\sigma^{(i)}(I) \in \Sigma^{(i)}$ the *parent sequence* of I, that is, the last sequence encountered by Player i on the path from the root of the game tree to any node in I. Whenever $\sigma^{(i)}(I) = (I', a)$, we say that I is *immediately reachable* from sequence $\sigma^{(i)}(I)$. If Player i never acts before I, then $\sigma^{(i)}(I) = \varnothing$, and we say that information set I is a *root information set* of Player i. Moreover, for any $z \in \mathcal{Z}$, $\sigma^{(i)}(z) \in \Sigma^{(i)}$ is the last sequence of Player i's actions encountered on the path from the root of the game tree to terminal node z. We let $\sigma^{(i)}(z) = \varnothing$ if Player i never plays on the path from the root to z. Analogously to what we did for information sets, we introduce a partial ordering on sequences: for every $i \in [n]$, and any pair $\sigma, \sigma' \in \Sigma^{(i)}$, the relation $\sigma \prec \sigma'$ holds if $\sigma = \varnothing \neq \sigma'$, or if the sequences are such that $\sigma = (I, a)$, $\sigma' = (I', a')$, and the set of Player i's actions on the path from the root to I' include playing action a at one node belonging to I. For any $i \in [n]$, $\sigma \in \Sigma^{(i)}$, and $I \in \mathcal{I}^{(i)}$, we write $\sigma \succ I$ to mean that the sequence of Player i's actions σ must lead the player to pass through I, formally $\sigma = (I', a') \in \Sigma^{(i)} \setminus \{\varnothing\} \wedge I' \succeq I$. Moreover, for $\sigma \in \Sigma^{(i)}$ and $I \in \mathcal{I}^{(i)}$, we write $\sigma \succeq I$ when

$\sigma \succ I$ or $\sigma = (I, a)$. Then, we let $\Sigma_I^{(i)} := \{\sigma \in \Sigma^{(i)} : \sigma \succeq I\} \subseteq \Sigma^{(i)}$ be the set of Player i's sequences that terminate at I or any of its descendant information sets, and $\mathcal{Z}_I^{(i)} := \{z \in \mathcal{Z} : \sigma^{(i)}(z) \succeq I\}$ be the set of terminal nodes reachable from information set $I \in \mathcal{I}^{(i)}$.

Sequence-form strategies. A *sequence-form strategy* for Player $i \in [n]$ is a vector $\boldsymbol{q} \in \mathbb{R}_{\geq 0}^{|\Sigma^{(i)}|}$ such that each entry $\boldsymbol{q}[(I, a)]$ specifies the product of the probabilities of playing all of Player i's actions on the path from the root down to action a at information set I (included) (Koller, Megiddo, and von Stengel 1996; Romanovskii 1962; von Stengel 1996). The set of valid sequence-form strategies for Player i is defined by some linear probability-mass-conservation constraints. Formally,

Definition 1. *The* sequence-form strategy polytope *for Player $i \in [n]$ is the convex polytope $Q^{(i)} := \{\boldsymbol{q} \in \mathbb{R}_{\geq 0}^{|\Sigma^{(i)}|} : \boldsymbol{q}[\varnothing] = 1, \text{ and } \boldsymbol{q}[\sigma^{(i)}(I)] = \sum_{a \in A(I)} \boldsymbol{q}[(I, a)], \forall I \in \mathcal{I}^{(i)}\}$.*

We let $Q_I^{(i)} := \{\boldsymbol{q} \in \mathbb{R}_{\geq 0}^{|\Sigma_I^{(i)}|} : \sum_{a \in A(j)} \boldsymbol{q}[(I, a)] = 1, \text{ and } \boldsymbol{q}[\sigma^{(i)}(I')] = \sum_{a \in A(I')} \boldsymbol{q}[(I', a)], \forall I' \succ I\}$ be the set of sequence form strategies only specifying Player i's behavior at information set I and all of its descendant. The set of *deterministic sequence-form strategies* for Player $i \in [n]$ is defined as $\Pi^{(i)} := Q^{(i)} \cap \{0, 1\}^{|\Sigma^{(i)}|}$, and the set of *deterministic sequence-form strategies for the subtree rooted at I* is $\Pi_I^{(i)} := Q_I^{(i)} \cap \{0, 1\}^{|\Sigma_I^{(i)}|}$. Kuhn's Theorem implies that, for any $i \in [n]$, $Q^{(i)} = \text{co}\,\Pi^{(i)}$, and $Q_I^{(i)} = \text{co}\,\Pi_I^{(i)}$ for any $I \in \mathcal{I}^{(i)}$ (Kuhn 1953). We denote as $\Pi := \times_{i \in [n]} \Pi^{(i)}$ the set of *joint* deterministic sequence-form strategies of all the players. Moreover, $\boldsymbol{\pi}^{(-i)} \in \times_{j \neq i} \Pi^{(j)}$ is a tuple specifying one deterministic sequence form strategy for each player other than i. It is often useful to express Player i's payoff function as a function of joint deterministic sequence-form strategy profiles belonging to Π. With a slight abuse of notation let $u^{(i)} : \Pi \to \mathbb{R}$ be such that, for each $\boldsymbol{\pi} = (\boldsymbol{\pi}^{(1)}, \dots, \boldsymbol{\pi}^{(n)}) \in \Pi$,

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

2.2 Regret Minimization and Phi-Regret Minimization

A *regret minimizer* for a set $\mathcal{X}$ is an abstract model for a decision maker that repeatedly interacts with a black-box environment. At each time t, a regret minimizer provides two operations: (i) NEXTELEMENT will make the regret minimizer output an element $\boldsymbol{x}^t \in \mathcal{X}$; (ii) OBSERVEUTILITY$(\ell^t)$ will inform the regret minimizer of the environment's feedback in the form of a linear utility function $\ell^t : \mathcal{X} \to \mathbb{R}$ which may depend adversarially on past choices $\boldsymbol{x}^1, \dots, \boldsymbol{x}^{t-1}$ of the regret minimizer. At each t, the regret minimizer will output a decision $\boldsymbol{x}^t$ on the basis of previous outputs $\boldsymbol{x}^1, \dots, \boldsymbol{x}^{t-1}$ and corresponding observed utility functions $\ell^1, \dots, \ell^{t-1}$. However, no information about

future losses is available to the decision maker. The performance of a regret minimizer is usually evaluated in terms of its *cumulative regret*

$$R^T := \max_{\boldsymbol{x}^\star \in \mathcal{X}} \sum_{t=1}^{T} \left(\ell^t(\boldsymbol{x}^\star) - \ell^t(\boldsymbol{x}^t) \right). \qquad (1)$$

The cumulative regret represents how much Player i would have gained by always playing the best action in hindsight, given the history of utility functions observed up to iteration T. Then, the objective is to guarantee a cumulative regret growing asymptotically sublinearly in the time T. For example, various regret minimizers guarantee a cumulative regret $R^T = O(\sqrt{T})$ at all times T for any convex and compact set $\mathcal{X}$ (see, e.g., Cesa-Bianchi and Lugosi (2006)).

A *phi-regret minimizer* (Stoltz and Lugosi 2007; Greenwald and Jafari 2003) is a generalization of the notion of regret minimizer which can be defined as follows.

Definition 2. *Given a set $\mathcal{X}$ of points and a set Φ of linear transformations $\phi : \mathcal{X} \to \mathcal{X}$, a phi-regret minimizer relative to Φ for the set $\mathcal{X}$—abbreviated "Φ-regret minimizer"—is an object with the same semantics and operations of a regret minimizer, but whose quality metric is its* cumulative phi-regret relative to Φ *(or Φ-regret for short)*

$$R^T := \max_{\phi^\star \in \Phi} \sum_{t=1}^{T} \left(\ell^t(\phi^\star(\boldsymbol{x}^t)) - \ell^t(\boldsymbol{x}^t) \right). \qquad (2)$$

The goal for a phi-regret minimizer is to guarantee that its phi-regret grows asymptotically sublinearly in T.

We observe that a regret minimizer is a special case of a phi-regret minimizer as the cumulative regret defined in Equation (1) can be obtained from Equation (2) by setting $\Phi = \{\mathcal{X} \ni \boldsymbol{x} \mapsto \hat{\boldsymbol{x}} : \hat{\boldsymbol{x}} \in \mathcal{X}\}$.

A general construction by Gordon, Greenwald, and Marks (2008) gives a way to construct a Φ-regret minimizer for $\mathcal{X}$ starting from any standard regret minimizer for the set of functions Φ. Specifically, let $\mathcal{R}_\Phi$ be a deterministic regret minimizer for the set of transformations Φ whose cumulative regret grows sublinearly, and assume that every $\phi \in \Phi$ admits a fixed point $\phi(\boldsymbol{x}) = \boldsymbol{x} \in \mathcal{X}$. Then, a Φ-regret minimizer $\mathcal{R}$ can be constructed starting from $\mathcal{R}_\Phi$ as follows:

- Each call to $\mathcal{R}$.NEXTELEMENT first calls NEXTELEMENT on $\mathcal{R}_\Phi$ to obtain the next transformation ϕ^t. Then, a fixed point $\boldsymbol{x}^t = \phi^t(\boldsymbol{x}^t)$ is computed and output.
- Each call to $\mathcal{R}$.OBSERVEUTILITY(ℓ^t) with linear utility function ℓ^t constructs the linear utility function $L^t : \phi \mapsto \ell^t(\phi(\boldsymbol{x}^t))$, where $\boldsymbol{x}^t$ is the last-output strategy, and passes it to $\mathcal{R}_\Phi$ by calling $\mathcal{R}_\Phi$.OBSERVEUTILITY(L^t).

3 Coarse Trigger Regret and Relationship with EFCCE

In this section we describe the notion of *coarse trigger deviation function* building on an idea by Gordon, Greenwald, and Marks (2008, Section 3). Then, we use this notion to formally characterize the set of EFCCEs, and to define the notion of *coarse trigger regret minimizer* as an instance of a

phi-regret minimizer. Finally, we establish a formal connection between the set of EFCCEs and the behavior of agents minimizing their coarse trigger regret.

3.1 Coarse Trigger Deviation Functions

For any $i \in [n]$, information set $\hat{I} \in \mathcal{I}^{(i)}$, and $\hat{\boldsymbol{\pi}} \in \Pi_{\hat{I}}^{(i)}$, a *coarse trigger deviation function for $\hat{I}$ and $\hat{\boldsymbol{\pi}}$* is a linear function which manipulates $|\Sigma^{(i)}|$-dimensional vectors so that any deterministic sequence form strategy that do not lead Player i down to $\hat{I}$ is left unmodified. On the other hand, if a deterministic sequence form strategy prescribes Player i to pass through $\hat{I}$, then its behavior at $\hat{I}$ and all of its descendant information sets is replaced with the behavior specified by the continuation strategy $\hat{\boldsymbol{\pi}}$.[1]

Definition 3 (Coarse Trigger Deviation Function). *Given an information set $\hat{I} \in \mathcal{I}^{(i)}$, and a continuation strategy $\hat{\boldsymbol{\pi}} \in \Pi_{\hat{I}}^{(i)}$, we say that a linear function $f : \mathbb{R}^{|\Sigma^{(i)}|} \to \mathbb{R}^{|\Sigma^{(i)}|}$ is a coarse trigger deviation function corresponding to information set $\hat{I}$ and continuation strategy $\hat{\boldsymbol{\pi}}$ if the following two conditions hold:*

- $f(\boldsymbol{\pi}) = \boldsymbol{\pi}$, *for all $\boldsymbol{\pi} \in \Pi^{(i)} : \boldsymbol{\pi}\left[\sigma^{(i)}(\hat{I})\right] = 0$;*

- *for any $\sigma \in \Sigma^{(i)}$, and $\boldsymbol{\pi} \in \Pi^{(i)} : \boldsymbol{\pi}\left[\sigma^{(i)}(\hat{I})\right] = 1$,*

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

For any $\hat{I} \in \mathcal{I}^{(i)}$ and $\hat{\boldsymbol{\pi}} \in \Pi_{\hat{I}}^{(i)}$, it is useful to instantiate a coarse trigger deviation function in the form of a linear map $\phi_{\hat{I} \to \hat{\boldsymbol{\pi}}} : \mathbb{R}^{|\Sigma^{(i)}|} \ni \boldsymbol{\pi} \mapsto M_{\hat{I} \to \hat{\boldsymbol{\pi}}}\boldsymbol{\pi}$, where $M_{\hat{I} \to \hat{\boldsymbol{\pi}}} \in \mathbb{R}_{\geq 0}^{|\Sigma^{(i)}| \times |\Sigma^{(i)}|}$ is the matrix such that, for any $\sigma_r, \sigma_c \in \Sigma^{(i)}$,

$$M_{\hat{I} \to \hat{\boldsymbol{\pi}}}[\sigma_r, \sigma_c] = \begin{cases} 1 & \text{if } \sigma_c \not\succeq \hat{I} \text{ and } \sigma_r = \sigma_c \\ \hat{\boldsymbol{\pi}}[\sigma_r] & \text{if } \sigma_c = \sigma^{(i)}(\hat{I}) \text{ and } \sigma_r \succeq \hat{I} \\ 0 & \text{otherwise} \end{cases}.$$

As a simple example of how such linear mappings are built is given in in Figure 1, where it is reported the matrix corresponding to $\phi_{\text{B} \to \hat{\boldsymbol{\pi}}}$ with $\hat{\boldsymbol{\pi}}$ being the continuation strategy corresponding to always playing 4 at information set B.

Let $\Psi^{(i)} := \left\{ \phi_{\hat{I} \to \hat{\boldsymbol{\pi}}} : \hat{I} \in \mathcal{I}^{(i)}, \hat{\boldsymbol{\pi}} \in \Pi_{\hat{I}}^{(i)} \right\}$ be the set of all possible linear mappings defining coarse trigger deviation functions for Player i. Then, we define the concept of *coarse trigger regret minimizer*. This notion will be shown to have a close connection with extensive-form coarse correlated equilibria.

Definition 4 (Coarse Trigger Regret Minimizer). *For every $i \in [n]$, we call coarse trigger regret minimizer for player i any $\Psi^{(i)}$-regret minimizer for the set of deterministic sequence-form strategies $\Pi^{(i)}$.*

---

[1] Our definition of coarse trigger deviation function can be seen as the sequence-form counterpart to the *blind causal behavioral deviations* defined by Morrill et al. (2021).
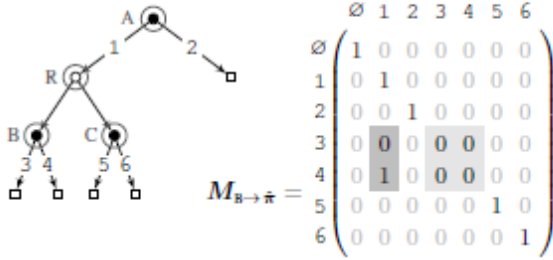
Figure 1: (Left) A simple sequential game with two players. Black round nodes are decision nodes of Player 1, white round nodes are decision nodes of Player 2. White square nodes represent terminal nodes. The set of sequences of Player 1 is $\Sigma^{(1)} = \{\varnothing, 1, \dots, 6\}$. (Right) Example of the matrix defining a coarse trigger deviation.

3.2 Extensive-Form Coarse Correlated Equilibria

Equipped with the notion of coarse trigger deviation function we can provide the following definition of *extensive-form coarse correlated equilibria* (EFCCE).

Definition 5 (ε-EFCCE, EFCCE). *Given $\epsilon \geq 0$, a probability distribution $\mu \in \Delta^{|\Pi|}$ is an ϵ-approximate EFCCE (ϵ-EFCCE for short) if, by letting $\pi = (\pi^{(1)}, \dots, \pi^{(n)}) \in \Pi$, for every player $i \in [n]$, and every coarse trigger deviation function $\phi_{\hat{I} \to \hat{\pi}} \in \Psi^{(i)}$, it holds*

$$\mathbb{E}_{\pi \sim \mu} \left[u^{(i)}\left(\phi_{\hat{I} \to \hat{\pi}}\left(\pi^{(i)}\right), \pi^{(-i)}\right) - u^{(i)}(\pi) \right] \leq \epsilon. \quad (3)$$

A probability distribution $\mu \in \Delta^{|\Pi|}$ is an EFCCE if it is a 0-EFCCE.

The above definition can be easily interpreted as the canonical definition of EFCCE by Farina, Bianchi, and Sandholm (2020). In their terminology, Equation (3) means that the expected utility of any trigger agent $(\hat{I}, \hat{\pi})$ is never larger than the expected utility that Player i would obtain by following recommendations by more than the amount ϵ.

We can now prove one of the central results of the paper, which shows that if each player $i \in [n]$ in the game plays according to a $\Psi^{(i)}$-regret minimizer, then the empirical frequency of play approaches the set of EFCCEs (all omitted proofs are reported in Appendix A).

Theorem 1. *For each player $i \in [n]$, let $\pi^{(i),1}, \dots, \pi^{(i),T}$ be a sequence of deterministic sequence form strategies with cumulative $\Psi^{(i)}$-regret $R^{(i),T}$ with respect to the sequence of linear functions*

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

and let $\mu \in \Delta^{|\Pi|}$ be the corresponding empirical frequency of play, defined as the probability distribution such that, for each $\pi = (\pi^{(1)}, \dots, \pi^{(n)}) \in \Pi$,

$$\mu[\pi] := \frac{1}{T} \sum_{t=1}^{T} \mathbb{1}_{\{\pi = (\pi^{(1),t}, \dots, \pi^{(n),t})\}}.$$

Then, the empirical frequency of play μ is an ϵ-EFCCE, with $\epsilon := \frac{1}{T} \max_{i \in [n]} R^{(i),T}$.

4 An Efficient No-Coarse-Trigger-Regret Algorithm

In this section we describe our efficient no-coarse-trigger regret minimizer. Our approach follows the framework by Gordon, Greenwald, and Marks (2008) (see Section 2.2). In order to apply the framework by Gordon, Greenwald, and Marks (2008) we need (i) to provide a regret minimizer for the set of coarse trigger deviation functions (Section 4.1), and (ii) to show that for any $\phi \in \Psi^{(i)}$ it is possible to compute in poly-time a sequence-form strategy $q \in Q^{(i)}$ such that $\phi(q) = q$ (Section 4.2). We will provide an efficient closed form solution for the latter problem by exploiting the particular structure of coarse trigger deviation functions.

4.1 Regret Minimization for the Set $\mathrm{co}\,\Psi^{(t)}$

Fix any player i. In this subsection, we discuss how one can construct an efficient regret minimizer for the *convex hull* $\mathrm{co}\,\Psi^{(i)}$ of the set of coarse trigger deviation functions. Since $\mathrm{co}\,\Psi^{(i)} \supseteq \Psi^{(i)}$, any such regret minimizer for $\mathrm{co}\,\Psi^{(i)}$ is in particular a regret minimizer for $\Psi^{(i)}$. At a high level, our construction decomposes the problem of minimizing regret on $\Psi^{(i)}$ into $|\mathcal{I}^{(i)}|$ subproblems, one for each possible trigger information set. Intuitively, given any $\hat{I} \in \mathcal{I}^{(i)}$, the subproblem for $\hat{I}$ corresponds to learning a continuation strategies for the trigger $\hat{I}$. In particular, each subproblem is itself a regret minimization problem, on the set of all continuation strategies $q \in Q_{\hat{I}}^{(i)}$, for which efficient regret minimization algorithms are known.

To make the above regret decomposition formal, we operate within the framework of *regret circuits* (Farina, Kroer, and Sandholm 2019). Regret circuits provide ways to decompose the problem of minimizing regret over a composite set into the problem of minimizing regret over the individual sets. Once regret minimizers for the individual sets have been constructed, an appropriate regret circuit will combine their outputs to guarantee low regret over the original composite set. For our purposes we will only need two regret circuit constructions: one for the convex hull, and one for affine transformations. We recall their main properties next.

Proposition 1. *Let $\mathcal{X}_1, \dots, \mathcal{X}_m$ be a finite collection of sets, $\mathcal{R}_1, \dots, \mathcal{R}_m$ be any regret minimizers for them, and $\mathcal{R}_{\Delta^m}$ be a regret minimizers for the m-simplex. A regret minimizer $\mathcal{R}_{\mathrm{co}}$ for the set $\mathrm{co}\{\mathcal{X}_1, \dots, \mathcal{X}_m\}$ can be built as follows:*

- *$\mathcal{R}_{\mathrm{co}}$.NEXTELEMENT outputs the element $\lambda_1^t x_1^t + \dots + \lambda_m^t x_m^t \in \mathrm{co}\{\mathcal{X}_1, \dots, \mathcal{X}_m\}$, where, for each $j \in [m]$, x_j^t is obtained by calling NEXTELEMENT on $\mathcal{R}_j$, and the vector $(\lambda_1^t, \dots, \lambda_m^t)$ is obtained by calling NEXTELEMENT on $\mathcal{R}_{\Delta^m}$.*
- *$\mathcal{R}_{\mathrm{co}}$.OBSERVEUTILITY($L^t$) forwards the linear utility function L^t to each $\mathcal{R}_j$, $j \in [m]$, and then calls $\mathcal{R}_{\Delta^m}$.OBSERVEUTILITY with the linear utility function $(\lambda_1, \dots, \lambda_m) \mapsto L^t(x_1^t)\lambda_1 + \dots + L^t(x_m^t)\lambda_m$.*

Furthermore, the composite regret minimizer $\mathcal{R}_{\mathrm{co}}$ satisfies the regret bound $R_{\mathrm{co}}^T \leq R_{\Delta^m}^T + \max_{j \in [m]} R_j^T$ for all T.

Proposition 2. *Let $\mathcal{X}$ be a convex and compact set, $g : \mathcal{X} \to \mathcal{Y}$ be an affine map, and $\mathcal{R}_{\mathcal{X}}$ any regret minimizer for the set*

$\mathcal{X}$. Then, a regret minimizer $\mathcal{R}_{g(\mathcal{X})}$ for the set $g(\mathcal{X})$ can be obtained as follows:

- $\mathcal{R}_{g(\mathcal{X})}$.NEXTELEMENT outputs $g(x^t)$, where x^t is obtained by calling NEXTELEMENT on $\mathcal{R}_{\mathcal{X}}$.
- $\mathcal{R}_{g(\mathcal{X})}$.OBSERVEUTILITY($L^t$) forwards the linear utility $x \mapsto L^t(g(x)) - L^t(g(0))$ to the $\mathcal{R}_{\mathcal{X}}$ regret minimizer.

Furthermore, the composite regret minimizer $\mathcal{R}_{g(\mathcal{X})}$ satisfies the regret bound $R_{g(\mathcal{X})}^T = R_{\mathcal{X}}^T$ at all times T.

We start from the following observation:

$$\text{co } \Psi^{(i)} = \text{co}\left\{ \Psi_{\hat{I}}^{(i)} : \hat{I} \in \mathcal{I}^{(i)} \right\}, \text{ where}$$

$$\Psi_{\hat{I}}^{(i)} := \text{co}\left\{ \phi_{\hat{I} \to \hat{\pi}} : \hat{\pi} \in \Pi_{\hat{I}}^{(i)} \right\}.$$

Hence, by virtue of the convex hull regret circuit (Proposition 1), in order to construct a regret minimizer for co $\Psi^{(i)}$ it suffices to construct regret minimizers for each of the $\Psi_{\hat{I}}^{(i)}$.

Consider now the mapping $g_{\hat{I}} : \mathbb{R}^{|\Sigma_{\hat{I}}^{(i)}|} \ni x \mapsto \phi_{\hat{I} \to x}$, which is promptly verified to be affine by definition of coarse trigger deviation function (see Section 3.1). Then,

$$\Psi_{\hat{I}}^{(i)} = g_{\hat{I}}\left(\text{co } \Pi_{\hat{I}}^{(i)}\right) = g_{\hat{I}}\left(\mathcal{Q}_{\hat{I}}^{(i)}\right) \qquad \forall \hat{I} \in \mathcal{I}^{(i)}.$$

In other words, the set $\Psi_{\hat{I}}^{(i)}$ is the image of $\mathcal{Q}_{\hat{I}}^{(i)}$ under an affine transformation. By using the affine transformation regret circuit (Proposition 2), a regret minimizer for $\Psi_{\hat{I}}^{(i)}$ can be constructed from any regret minimizer for $\mathcal{Q}_{\hat{I}}^{(i)}$. Because efficient regret minimizers for $\mathcal{Q}_{\hat{I}}^{(i)}$ are well known in the literature (e.g., the CFR algorithm by Zinkevich et al. (2008)), our approach based on regret circuits yields an efficient regret minimizer for co $\Psi^{(i)}$, summarized in Algorithm 1.

Theorem 2. *Let $\mathcal{R}$ be a regret minimizer instantiated as specified in Algorithm 1, and employing the CFR algorithm for each $\mathcal{R}_{\hat{I}}$ and the regret matching algorithm for $\mathcal{R}_{\Delta^{|\mathcal{I}^{(i)}|}}$. After observing a sequence of linear utility functions $L^1, \ldots, L^T : \text{co } \Psi^{(i)} \to \mathbb{R}$ with range upper bounded by U (that is, for all $t \in T$ $\max_{\phi, \phi' \in \text{co } \Psi^{(i)}}\{L^t(\phi) - L^t(\phi')\} \le U$), the regret cumulated by the transformations $\phi^1, \ldots, \phi^T \in \text{co } \Psi^{(i)}$ produced by $\mathcal{R}$ is such that*

$$R^T = \max_{\phi^* \in \text{co } \Psi^{(i)}} \sum_{t=1}^{T} \left(L^t(\phi^*) - L^t(\phi^t) \right) \le 2U|\Sigma^{(i)}|\sqrt{T}.$$

4.2 Closed-Form Fixed Point Computation

We complete the construction of our $\Psi^{(i)}$-regret minimizer by showing that, for any player $i \in [n]$ and $\phi \in \text{co } \Psi^{(i)}$, it is possible to compute a sequence-form strategy $q^* \in \mathcal{Q}^{(i)}$ such that $\phi(q^*) = q^*$ in linear time in $|\Sigma^{(i)}|D^{(i)}$, where $D^{(i)}$ denotes the maximum depth of the game tree considering only actions of Player i (i.e., $D^{(i)} = \max_{z \in Z}|\{\sigma \in \Sigma^{(i)} : \sigma \prec \sigma^{(i)}(z)\}|$). Unlike all known results for EFCE, our result does not rely—as an intermediate step—on the

---

Algorithm 1: Regret minimizer for the set co $\Psi^{(i)}$

Data: player $i \in [n]$, one regret minimizer $\mathcal{R}_{\hat{I}}$ for the set $\Psi_{\hat{I}}^{(i)}$ for each $\hat{I} \in \mathcal{I}^{(i)}$, one regret minimizer $\mathcal{R}_{\Delta^{|\mathcal{I}^{(i)}|}}$ for the $|\mathcal{I}^{(i)}|$-simplex.

function NEXTELEMENT()
1: $\lambda^t \leftarrow \mathcal{R}_{\Delta^{|\mathcal{I}^{(i)}|}}$.NEXTELEMENT()
2: **for** $\hat{I} \in \mathcal{I}^{(i)}$ **do** $\phi_{\hat{I} \to q_{\hat{I}}^t} \leftarrow \mathcal{R}_{\hat{I}}$.NEXTELEMENT()
3: **return** $\sum_{\hat{I} \in \mathcal{I}^{(i)}} \lambda^t[\hat{I}] \phi_{\hat{I} \to q_{\hat{I}}^t}$

function OBSERVEUTILITY(L^t)
1: **for** $\hat{I} \in \mathcal{I}^{(i)}$ **do** $\mathcal{R}_{\hat{I}}$.OBSERVEUTILITY(L^t)
2: $\ell_\lambda^t \leftarrow \Delta^{|\mathcal{I}^{(i)}|} \ni \lambda \mapsto \sum_{\hat{I} \in \mathcal{I}^{(i)}} \lambda[\hat{I}] L^t(\phi_{\hat{I} \to q_{\hat{I}}^t})$
3: $\mathcal{R}_{\Delta^{|\mathcal{I}^{(i)}|}}$.OBSERVEUTILITY($\ell_\lambda^t$)

Algorithm 2: FIXEDPOINT(ϕ)

Input: $\phi = \sum_{\hat{I} \in \mathcal{I}^{(i)}} \lambda[\hat{I}] \phi_{\hat{I} \to q_{\hat{I}}} \in \text{co } \Psi^{(i)}$
Output: $q \in \mathcal{Q}^{(i)}$ fixed point of ϕ
1: $q^* \leftarrow 0 \in \mathbb{R}_{\ge 0}^{|\Sigma^{(i)}|}$, $\quad q^*[\varnothing] \leftarrow 1$
2: **for** $\sigma = (I, a) \in \Sigma^{(i)} \setminus \{\varnothing\}$ in top-down ($\prec$) order **do**
3: $\quad d_\sigma \leftarrow \sum_{I' \in \mathcal{I}^{(i)} : I' \preceq I} \lambda[I']$
4: $\quad$ **if** $d_\sigma = 0$ **then** $q^*[\sigma] \leftarrow \dfrac{q^*[\sigma^{(i)}(I)]}{|\mathcal{A}(I)|}$
5: $\quad$ **else** $q^*[\sigma] \leftarrow \dfrac{1}{d_\sigma}\left(\displaystyle\sum_{I' \in \mathcal{I}^{(i)} : I' \preceq I} \lambda[I'] \, q_{I'}[\sigma] \, q^*[\sigma^{(i)}(I')] \right)$
6: **return** q^*

---

computation the stationary distributions of some stochastic matrices (Farina et al. 2021; Morrill et al. 2021), and does not require to manage any internal-regret minimizer as in Celli et al. (2020). Instead, we show that the fixed-point strategy q^* can be found in closed form for any $\phi \in \text{co } \Psi^{(i)}$.

Let $\phi = \sum_{\hat{I} \in \mathcal{I}^{(i)}} \lambda[\hat{I}] \phi_{\hat{I} \to q_{\hat{I}}}$ be any deviation function returned by Algorithm 1, where $\lambda \in \Delta^{|\mathcal{I}^{(i)}|}$, and $q_{\hat{I}} \in \mathcal{Q}_{\hat{I}}^{(i)}$ for each $\hat{I} \in \mathcal{I}^{(i)}$. Equipped with this additional notation, Algorithm 2 describes an efficient procedure to compute a fixed point of a given transformation $\phi \in \text{co } \Psi^{(i)}$. The algorithm iterates over sequences of Player i according to their partial ordering $\prec$ (i.e., it is never the case that a sequence $\sigma = (I, a)$ is considered before $\sigma^{(i)}(I)$). For each sequence $\sigma = (I, a)$, the algorithm computes $d_\sigma \in \mathbb{R}_{\ge 0}$ as the sum of the weights of the convex combination corresponding to information sets preceding I (Line 3). If $d_\sigma = 0$, then the matrix M corresponding to the transformation ϕ must be such that $M[\sigma, \sigma] = 1$, and $M[\sigma, \sigma'] = 0$ for all $\sigma' \ne \sigma$. Therefore, the choice we make at σ is indifferent as long as the resulting q^* is a well-formed sequence-form strategy. We set q^* so that the probability-mass flow is evenly divided among

Algorithm 3: Regret minimizer for the set co $\Psi^{(i)}$

Data: player $i \in [n]$, $\mathcal{R}$ regret minimizer for the set co $\Psi^{(i)}$ defined in Algorithm 1.

---

function NEXTSTRATEGY()

 1: $\phi^t = \sum_{\hat{I} \in \mathcal{J}^{(i)}} \lambda^t[\hat{I}] \, \phi_{\hat{I} \to q_{\hat{I}}^t} \leftarrow \mathcal{R}.\text{NEXTELEMENT}()$

 2: **return** $x^t \leftarrow \text{FIXEDPOINT}(\phi^t)$

function OBSERVEUTILITY(ℓ^t)

 1: Define the linear function $L^t : \phi \mapsto \ell^t(\phi(x^t))$

 2: $\mathcal{R}.\text{OBSERVEUTILITY}(L^t)$

---

sequences originating in I (Line 4). Finally, when $d_\sigma \neq 0$, Line 5 assigns to $q^\star[\sigma]$ a value equal to a weighted sum of $q_{I'}[\sigma]q^\star[\sigma']$ for sequences $\sigma' = (I', a')$ preceding information set I. The following theorem shows that Algorithm 2 is indeed correct and behaves as desired.

Theorem 3. *For any player $i \in [n]$, and transformation $\phi = \sum_{\hat{I} \in \mathcal{J}^{(i)}} \lambda[\hat{I}]\phi_{\hat{I} \to q_{\hat{I}}} \in \text{co } \Psi^{(i)}$, the vector $q^\star \in \mathbb{R}^{|\Sigma^{(i)}|}$ obtained through Algorithm 2 is such that $q^\star \in \mathcal{Q}^{(i)}$, and $\phi(q^\star) = q^\star$. Algorithm 2 runs in linear time in $|\Sigma^{(i)}|D^{(i)}$.*

As a direct consequence of the correctness (Theorem 2 and Theorem 3) of the two steps required by Gordon, Greenwald, and Marks (2008)'s construction (recalled in Section 2.2) we have the following.

Corollary 1. *Algorithm 3 is a co $\Psi^{(i)}$-regret minimizer for the set $\mathcal{Q}^{(i)}$. Thus, when all player play according to Algorithm 3 where at all t the utility ℓ^t of each player is set to their own linear utility function given the opponents' actions, the empirical frequency of play in the game after T iterations converges to a $O(1/\sqrt{T})$-EFCCE with high probability, and converges almost surely to an EFCCE in the limit.*

5 Experimental Evaluation

We experimentally investigate the convergence of our no-regret learning dynamics for EFCCE on four standard benchmark games: 3-player Kuhn poker, 3-player Leduc poker, 3-player Goofspiel, 2-player (general-sum) Battleship. See Appendix B for a description of each game. For each game we investigate the rate of convergence to EFCCE, measured as the maximum expected increase in utility that any player can obtain by optimally responding to the mediator that recommends strategies based on the correlated distribution of play induced by the learning dynamics, of our EFCCE learning dynamics and of the very related EFCE learning dynamics in Farina et al. (2021), which were obtained using the same framework as this paper. Experimental results are available in Figure 2. In each game, we ran our EFCCE dynamics and the EFCE dynamics for the same number of iterations (1000 iterations in Leduc poker and Goofspiel, 5000 in Kuhn poker, 10000 in Battleship). Each fixed-point computation in the EFCE dynamics was performed through an optimized implementation of the power iteration method. The power iteration was interrupted when
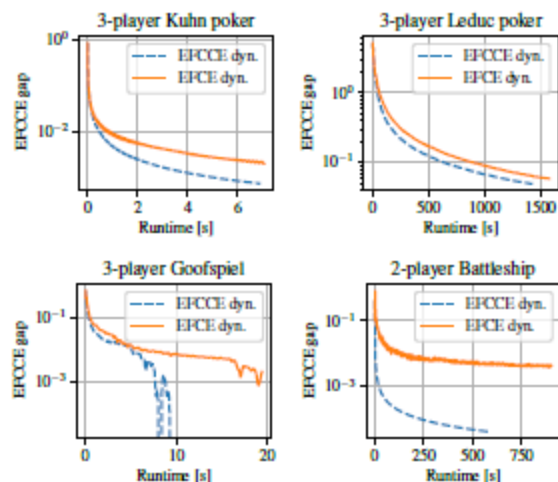

Figure 2: Rate of convergence to EFCCE (measured through the EFCCE gap) of the EFCCE learning dynamics introduced in this paper and of the related EFCE learning dynamics by Farina et al. (2021) on four standard benchmark games. The y-axis is on a log scale.

the Euclidean norm of the fixed-point residual got below the threshold value of 10^{-6} (this typically was achieved in less than 10 iterations of the method).

In all four games, we observe that our EFCCE dynamics outperform the EFCE dynamics, often by a significant margin. This is consistent with intuition, as the EFCCE dynamics are solving a strictly harder problem (minimizing the EFCE gap, instead of the EFCCE gap). In Goofspiel, the EFCCE dynamics induce a correlated distribution of play that is an exact EFCCE after roughly 500 iterations.

In Kuhn, Leduc, and Goofspiel the runtime of each iteration is comparable between EFCE and EFCCE dynamics, while in the Battleship game, the EFCCE dyanmics are roughly 30% faster per iteration. This is consistent with the observation that the amout of work necessary to find a stationary distribution grows approximately cubically with the maximum number of actions at any decision point in the game, which is higher in Battleship compared to the other games.

6 Conclusions and Future Research

We showed that—at least when analyzed through the phi-regret minimization framework—the computation of the fixed point required at each iteration in EFCCE learning dynamics can be carried out in closed form using a simple formula that avoids the computation of stationary distributions of Markov chains. This is contrast to all known learning dynamics for EFCE.

We conjecture that a similar result could be derived within the EFR framework, when blind causal deviations are considered, though we leave exploration of that direction open.

Acknowledgments

References

Aumann, R. J. 1974. Subjectivity and correlation in randomized strategies. *Journal of mathematical Economics*, 1(1): 67–96.

Blum, A.; and Mansour, Y. 2007. From external to internal regret. *Journal of Machine Learning Research*, 8(Jun): 1307–1324.

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

Celli, A.; Coniglio, S.; and Gatti, N. 2019. 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.

Celli, A.; and Gatti, N. 2018. Computational Results for Extensive-Form Adversarial Team Games. In *AAAI Conference on Artificial Intelligence (AAAI)*.

Celli, A.; Marchesi, A.; Bianchi, T.; and Gatti, N. 2019. Learning to Correlate in Multi-Player General-Sum Sequential Games. In *Advances in Neural Information Processing Systems*, 13055–13065.

Celli, A.; Marchesi, A.; Farina, G.; and Gatti, N. 2020. No-regret learning dynamics for extensive-form correlated equilibrium. In *Advances in Neural Information Processing Systems*, volume 33.

Cesa-Bianchi, N.; and Lugosi, G. 2006. *Prediction, learning, and games*. Cambridge University Press.

Chen, X.; and Deng, X. 2006. Settling the complexity of two-player Nash equilibrium. In *2006 47th Annual IEEE Symposium on Foundations of Computer Science*, 261–272. IEEE.

Daskalakis, C.; Goldberg, P. W.; and Papadimitriou, C. H. 2009. The complexity of computing a Nash equilibrium. *SIAM Journal on Computing*, 39(1): 195–259.

Farina, G.; Bianchi, T.; and Sandholm, T. 2020. Coarse Correlation in Extensive-Form Games. In *The Thirty-Fourth AAAI Conference on Artificial Intelligence (AAAI-20)*.

Farina, G.; Celli, A.; Gatti, N.; and Sandholm, T. 2018. Ex ante coordination and collusion in zero-sum multi-player extensive-form games. In *Advances in Neural Information Processing Systems (NeurIPS)*.

Farina, G.; Celli, A.; Marchesi, A.; and Gatti, N. 2021. Simple Uncoupled No-Regret Learning Dynamics for Extensive-Form Correlated Equilibrium. *arXiv preprint arXiv:2104.01520*.

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

Farina, G.; Ling, C. K.; Fang, F.; and Sandholm, T. 2019. Correlation in extensive-form games: Saddle-point formulation and benchmarks. In *Advances in Neural Information Processing Systems*, 9229–9239.

Foster, D. P.; and Vohra, R. V. 1997. Calibrated learning and correlated equilibrium. *Games and Economic Behavior*, 21(1-2): 40.

Fudenberg, D.; and Levine, D. K. 1995. Consistency and cautious fictitious play. *Journal of Economic Dynamics and Control*, 19(5-7): 1065–1089.

Fudenberg, D.; and Levine, D. K. 1999. Conditional universal consistency. *Games and Economic Behavior*, 29(1-2): 104–130.

Gordon, G. J.; Greenwald, A.; and Marks, C. 2008. No-regret learning in convex games. In *International Conference on Machine learning*, 360–367.

Greenwald, A.; and Jafari, A. 2003. A general class of no-regret learning algorithms and game-theoretic equilibria. In *Learning theory and kernel machines*, 2–12. Springer.

Guennebaud, G.; Jacob, B.; et al. 2010. Eigen v3. http://eigen.tuxfamily.org.

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

Hart, S.; and Mas-Colell, A. 2001. A general class of adaptive strategies. *Journal of Economic Theory*, 98(1): 26–54.

Hart, S.; and Mas-Colell, A. 2003. Uncoupled Dynamics Do Not Lead to Nash Equilibrium. *The American Economic Review*, 93(5): 1830–1836.

Koller, D.; Megiddo, N.; and von Stengel, B. 1996. Efficient Computation of Equilibria for Extensive Two-Person Games. *Games and Economic Behavior*, 14(2): 247–259.

Koutsoupias, E.; and Papadimitriou, C. 1999. Worst-case equilibria. In *Annual Symposium on Theoretical Aspects of Computer Science*, 404–413. Springer.

Kuhn, H. W. 1950. A simplified two-person poker. *Contributions to the Theory of Games*, 1: 97–103.

Kuhn, H. W. 1953. *Extensive Games and the Problem of Information*, 193–216. Princeton University Press.

Lanctot, M.; Waugh, K.; Zinkevich, M.; and Bowling, M. H. 2009. Monte Carlo Sampling for Regret Minimization in Extensive Games. In *Advances in Neural Information Processing Systems*, 1078–1086.

Morrill, D.; D'Orazio, R.; Lanctot, M.; Wright, J. R.; Bowling, M.; and Greenwald, A. 2021. Efficient Deviation Types and Learning for Hindsight Rationality in Extensive-Form Games. *arXiv preprint arXiv:2102.06973*.

Morrill, D.; D'Orazio, R.; Sarfati, R.; Lanctot, M.; Wright, J.; Greenwald, A.; and Bowling, M. 2020. Hindsight and Sequential Rationality of Correlated Play. In *The Thirty-Fourth AAAI Conference on Artificial Intelligence (AAAI-20)*.

Nash, J. F. 1950. Equilibrium points in n-person games. *Proceedings of the national academy of sciences*, 36(1): 48–49.

Romanovskii, I. 1962. Reduction of a Game with Complete Memory to a Matrix Game. *Soviet Mathematics*, 3.

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

Roughgarden, T.; and Tardos, É. 2002. How bad is selfish routing? *Journal of the ACM*, 49(2): 236–259.

Southey, F.; Bowling, M. H.; Larson, B.; Piccione, C.; Burch, N.; Billings, D.; and Rayner, D. C. 2005. Bayes' Bluff: Opponent Modelling in Poker. In *Conference on Uncertainty in Artificial Intelligence (UAI)*, 550–558.

Stoltz, G.; and Lugosi, G. 2007. Learning correlated equilibria in games with compact sets of strategies. *Games and Economic Behavior*, 59(1): 187–208.

Tammelin, O.; Burch, N.; Johanson, M.; and Bowling, M. 2015. Solving Heads-Up Limit Texas Hold'em. In *International Joint Conferences on Artificial Intelligence*, 645–652.

von Stengel, B. 1996. Efficient Computation of Behavior Strategies. *Games and Economic Behavior*, 14(2): 220–246.

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

Zinkevich, M.; Johanson, M.; Bowling, M.; and Piccione, C. 2008. Regret minimization in games with incomplete information. In *Advances in Neural Information Processing Systems*, 1729–1736.

A Proofs

A.1 Proof for Section 3

Theorem 1. *For each player $i \in [n]$, let $\pi^{(i),1}, \ldots, \pi^{(i),T}$ be a sequence of deterministic sequence form strategies with cumulative $\Psi^{(i)}$-regret $R^{(i),T}$ with respect to the sequence of linear functions*

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

and let $\mu \in \Delta^{|\Pi|}$ be the corresponding empirical frequency of play, defined as the probability distribution such that, for each $\pi = (\pi^{(1)}, \ldots, \pi^{(n)}) \in \Pi$,

$$\mu[\pi] := \frac{1}{T} \sum_{t=1}^{T} \mathbf{1}_{\{\pi = (\pi^{(1),t}, \ldots, \pi^{(n),t})\}}.$$

Then, the empirical frequency of play μ is an ϵ-EFCCE, with $\epsilon := \frac{1}{T} \max_{i \in [n]} R^{(i),T}$.

Proof. For any $i \in [n]$, we have $R^{(i),T} \le \epsilon T$. Then, by definition of $R^{(i),T}$ as per Equation (2), for any coarse trigger deviation function $\phi \in \Psi^{(i)}$ is must hold

$$T\epsilon \ge \sum_{t=1}^{T}\left(\ell^{(i),t}\left(\phi(\pi^{(i),t})\right) - \ell^{(i),t}\left(\pi^{(i),t}\right)\right) = \sum_{t=1}^{T}\left(u^{(i)}\left(\phi(\pi^{(i),t}), \pi^{(-i),t}\right) - u^{(i)}(\pi^t)\right).$$

This yields the following

$$T\epsilon \ge \sum_{t=1}^{T} \sum_{\pi \in \Pi} \mathbf{1}_{\{\pi = \pi^t\}}\left(u^{(i)}\left(\phi(\pi^{(i)}), \pi^{(-i)}\right) - u^{(i)}(\pi)\right)$$

$$= \sum_{\pi \in \Pi}\left(\sum_{t=1}^{T} \mathbf{1}_{\{\pi = \pi^t\}}\right)\left(u^{(i)}\left(\phi(\pi^{(i)}), \pi^{(-i)}\right) - u^{(i)}(\pi)\right)$$

$$= T \sum_{\pi \in \Pi} \mu[\pi]\left(u^{(i)}\left(\phi(\pi^{(i)}), \pi^{(-i)}\right) - u^{(i)}(\pi)\right).$$

This is precisely the definition of ϵ-EFCCE (Definition 5), as we wanted to show. □

A.2 Proof for Section 4.1

Theorem 2. *Let $\mathcal{R}$ be a regret minimizer instantiated as specified in Algorithm 1, and employing the CFR algorithm for each $\mathcal{R}_j$ and the regret matching algorithm for $\mathcal{R}_{\Delta^{|\mathcal{I}^{(i)}|}}$. After observing a sequence of linear utility functions $L^1, \ldots, L^T : \mathrm{co}\,\Psi^{(i)} \to \mathbb{R}$ with range upper bounded by U (that is, for all $t \in T$ $\max_{\phi, \phi' \in \mathrm{co}\,\Psi^{(i)}}\{L^t(\phi) - L^t(\phi')\} \le U$), the regret cumulated by the transformations $\phi^1, \ldots, \phi^T \in \mathrm{co}\,\Psi^{(i)}$ produced by $\mathcal{R}$ is such that*

$$R^T = \max_{\phi^\star \in \mathrm{co}\,\Psi^{(i)}} \sum_{t=1}^{T}\left(L^t(\phi^\star) - L^t(\phi^t)\right) \le 2U|\Sigma^{(i)}|\sqrt{T}.$$

Proof. We start by recalling the known regret bounds for CFR (Zinkevich et al. 2008, Theorems 3 and 4) and regret matching (Hart and Mas-Colell 2000).

Lemma 1 (Known regret bound for CFR). *Let $i \in [n]$ and $I \in \mathcal{I}^{(i)}$, and consider the CFR algorithm run on the set $\mathcal{Q}_I^{(i)}$. Then, for any sequence of linear utility functions $\bar{\ell}^1, \ldots, \bar{\ell}^T : \mathcal{Q}^{(i)} \to \mathbb{R}$ with range upper bounded as $\max_{q,q' \in \mathcal{Q}_I^{(i)}}(\bar{\ell}^t(q) - \bar{\ell}^t(q')) \le U$ at all t, the regret cumulated by the CFR algorithm satisfies the inequality*

$$R_{\mathrm{CFR}}^T \le U|\Sigma_I^{(i)}|\sqrt{T}.$$

Lemma 2 (Known regret bound for regret matching). *Consider the regret matching algorithm applied to a simplex domain Δ^m. For any sequence of linear utility functions $\bar{\ell}^1, \ldots, \bar{\ell}^T : \Delta^m \to \mathbb{R}$ with range upper bounded as $\max_{x,x' \in \Delta^m}(\bar{\ell}^t(x) - \bar{\ell}^t(x')) \le U$ at all t, the regret cumulated by the regret matching algorithm satisfies the inequality*

$$R_{\mathrm{RM}}^T \le Um\sqrt{T}.$$

By Proposition 1 and by construction of Algorithm 2 we have that

$$R^T \leq R^T_{\Delta^{|\mathcal{J}^{(i)}|}} + \max_{\hat{I} \in \mathcal{J}^{(i)}} R^T_{\hat{I}}. \tag{4}$$

The loss function observed by the regret minimizer $\mathcal{R}_{\Delta^m}$ at time t is $\ell^t_\lambda \leftarrow \Delta^{|\mathcal{J}^{(i)}|} \ni \lambda \mapsto \sum_{\hat{I} \in \mathcal{J}^{(i)}} \lambda[\hat{I}] L^t(\phi_{\hat{I} \to q^t_{\hat{I}}})$ (Line 2).
Since $\phi_{\hat{I} \to \hat{\pi}}(q) \in \mathcal{Q}^{(i)}$ for any $q \in \mathcal{Q}^{(i)}$, we have that the maximum range of ℓ^t_λ is at most equal to the maximum range
of L^t. Therefore, by Lemma 2 we have $R^T_{\Delta^{|\mathcal{J}^{(i)}|}} \leq U|\mathcal{J}^{(i)}|\sqrt{T}$. We not turn our attention to the regret minimizers $\mathcal{R}_{\hat{I}}$. Fix
$\hat{I} \in \mathcal{J}^{(i)}$, by Proposition 2 the loss observed at time t by the CFR algorithm running on the set $Q^T_{\hat{I}}$ is $L^t(g_{\hat{I}}(\cdot)) - L^t(g_{\hat{I}}(0))$.
Then, the range of this linear function is equal to $\max_{q,q' \in \mathcal{Q}^{(i)}} L^t(g_{\hat{I}}(q)) - L^t(g_{\hat{I}}(q'))$, which is upper bounded by U since $g_{\hat{I}}$
maps sequence-form strategies into valid sequence-form strategies. By Proposition 2 we have that $R^T_{\hat{I}}$ is at most equal to the
regret cumulated by CFR run on $\mathcal{Q}^{(i)}_{\hat{I}}$. This, together with Lemma 1, yields that, for any $\hat{I} \in \mathcal{J}^{(i)}$, $R^T_{\hat{I}} \leq U|\Sigma^{(i)}_{\hat{I}}|\sqrt{T}$. Then, by
substituting into (4),

$$R^T \leq U|\mathcal{J}^{(i)}|\sqrt{T} + \max_{\hat{I} \in \mathcal{J}^{(i)}} U|\Sigma^{(i)}_{\hat{I}}|\sqrt{T} \leq U|\mathcal{J}^{(i)}|\sqrt{T} + U|\Sigma^{(i)}|\sqrt{T} \leq 2|\Sigma^{(i)}|\sqrt{T}.$$

This concludes the proof. □

Corollary 1. *Algorithm 3 is a co $\Psi^{(i)}$-regret minimizer for the set $\mathcal{Q}^{(i)}$. Thus, when all player play according to Algorithm 3
where at all t the utility ℓ^t of each player is set to their own linear utility function given the opponents' actions, the empirical
frequency of play in the game after T iterations converges to a $O(1/\sqrt{T})$-EFCCE with high probability, and converges almost
surely to an EFCCE in the limit.*

Proof. Theorem 2 establishes that Algorithm 1 is a regret minimizer for the set co $\Psi^{(i)} \supseteq \Psi^{(i)}$. Theorem 3 establishes that
Algorithm 2 returns a fixed point $q \in \mathcal{Q}^{(i)}$ for any $\phi \in$ co $\Psi^{(i)}$. Hence, by using the result by Gordon, Greenwald, and
Marks (2008), Algorithm 3 is a $\Psi^{(i)}$-regret minimizer for the set $\mathcal{Q}^{(i)}$ for each player i. At each time t, Algorithm 3 returns
a randomized strategy $q^{(i),t} \in \mathcal{Q}^{(i)}$ that Player i should play. A standard application of the Azuma-Hoeffding inequality
shows that by sampling actions according to $q^{(i),t}$, the $\Psi^{(i)}$-regret incurred by Player i grows by an amount bounded above by
$O(\sqrt{T \log(1/\delta)})$ with probability at least $1 - \delta$, for any $\delta \in (0,1)$. Hence, by invoking Theorem 1, with probability at least
$1 - \delta$, after any T iterations the empirical frequency of play is an ϵ-EFCCE where

$$\epsilon = O\left(\frac{1}{T}(\sqrt{T} + \sqrt{T \log(1/\delta)})\right) = O\left(\frac{1}{\sqrt{T}}(1 + \log(1/\delta))\right). \tag{5}$$

This concludes the proof of the first part of the statement. Going from the high-probability regret guarantee for any $\delta \in (0,1)$
and T given in (5) to almost-sure convergence in the limit as $T \to \infty$ is a direct application of the classic Borel-Cantelli
lemma. □

A.3   Proof for Section 4.2

The following result will be useful when proving Theorem 3.

Lemma 3. *For any $\phi = \sum_{I' \in \mathcal{J}^{(i)}} \lambda[I']\phi_{I' \to q_{I'}} \in$ co $\Psi^{(i)}$, $q \in \mathcal{Q}^{(i)}$, and $\sigma = (I,a) \in \Sigma^{(i)}$, it holds*

$$\phi(q)[\sigma] - q[\sigma] = \left(\sum_{I' \preceq I} \lambda[I']q_{I'}[\sigma]q[\sigma^{(i)}(I')]\right) - d_\sigma q[\sigma].$$

Proof. Fix a sequence $\sigma \in \Sigma^{(i)}$. Then, by definition of the linear mappings $\phi_{I' \to q_{I'}}$, we have

$$\begin{aligned}
\phi(q)[\sigma] &= \sum_{I' \in \mathcal{J}^{(i)}} \lambda[I']\phi_{I' \to q_{I'}}(q)[\sigma] \\
&= \sum_{I' \in \mathcal{J}^{(i)}} \lambda[I'] \begin{cases} q_{I'}[\sigma]q[\sigma^{(i)}(I')] & \text{if } \sigma \succeq I' \\ q[\sigma] & \text{otherwise} \end{cases} \\
&= \left(1 - \sum_{I' \preceq \sigma} \lambda[I']\right) q[\sigma] + \sum_{I' \preceq \sigma} \lambda[I']q_{I'}[\sigma]q[\sigma^{(i)}(I')].
\end{aligned}$$

By re-arranging the above equation we obtain the statement. □

Theorem 3. *For any player $i \in [n]$, and transformation $\phi = \sum_{\hat{I} \in \mathcal{I}^{(i)}} \lambda[\hat{I}]\phi_{\hat{I} \to q_{\hat{I}}} \in \text{co}\,\Psi^{(i)}$, the vector $q^* \in \mathbb{R}^{|\Sigma^{(i)}|}$ obtained through Algorithm 2 is such that $q^* \in \mathcal{Q}^{(i)}$, and $\phi(q^*) = q^*$. Algorithm 2 runs in linear time in $|\Sigma^{(i)}|D^{(i)}$.*

Proof. The proof is divided into three parts: (i) we show that, for any $\phi \in \text{co}\,\Psi^{(i)}$, the vector $q^* \in \mathbb{R}^{|\Sigma^{(i)}|}$ obtained through Algorithm 2 is such that $q^* \in \mathcal{Q}^{(i)}$ (i.e., it is a valid sequence-form strategy); (ii) we show that, for any $\phi \in \text{co}\,\Psi^{(i)}$, the sequence-form strategy q^* obtained via Algorithm 2 is such that $\phi(q^*) = q^*$; (iii) finally, we show that Algorithm 2 runs in time $O(|\Sigma^{(i)}|D^{(i)})$.

Part 1: q^* is a sequence-form strategy. By construction (Line 1), $q^*[\varnothing] = 1$. Then, we need to show that, for each $I \in \mathcal{I}^{(i)}$, $\sum_{a \in A(I)} q^*[(I,a)] = q^*[\sigma^{(i)}(I)]$ (see Definition 1). For any $I \in \mathcal{I}^{(i)}$ such that $d_\sigma = 0$ it is immediate to see that the above constraint holds by construction (Line 4). For each $I \in \mathcal{I}^{(i)}$ such that $d_\sigma \neq 0$ we have that

$$
\sum_{a \in A(I)} q^*[(I,a)] = \frac{1}{d_\sigma}\left(\sum_{a \in A(I)} \sum_{I' \preceq I} \lambda[I']q_{I'}[(I,a)]q^*[\sigma^{(i)}(I')]\right)
$$

$$
= \frac{1}{d_\sigma}\left(\sum_{I' \preceq I} \lambda[I']q^*[\sigma^{(i)}(I')]\left(\sum_{a \in A(I)} q_{I'}[(I,a)]\right)\right)
$$

$$
= \frac{1}{d_\sigma}\left(\sum_{I' \preceq I} \lambda[I']q^*[\sigma^{(i)}(I')] \cdot \begin{cases} q_{I'}[\sigma^{(i)}(I)] & \text{if } I' \prec I \\ 1 & \text{otherwise} \end{cases}\right),
$$

where the first equality holds by Line 5 in Algorithm 2, and the last equality holds because $q_{I'} \in \mathcal{Q}_{I'}^{(i)}$. We distinguish between two cases: if $d_{\sigma^{(i)}(I)} = 0$, then $\lambda[I'] = 0$ for each $I' \prec I$. Therefore, since we are assuming $d_\sigma \neq 0$, it must be the case that $d_\sigma = \lambda[I] \neq 0$. This yields the following

$$
\sum_{a \in A(I)} q^*[(I,a)] = \frac{1}{d_\sigma}\left(\sum_{I' \preceq I} \lambda[I']q^*[\sigma^{(i)}(I')] \cdot \begin{cases} q_{I'}[\sigma^{(i)}(I)] & \text{if } I' \prec I \\ 1 & \text{otherwise} \end{cases}\right)
$$

$$
= \frac{1}{\lambda[I]}\left(\lambda[I]q^*[\sigma^{(i)}(I)]\right) = q^*[\sigma^{(i)}(I)].
$$

Contrarily, if $d_{\sigma^{(i)}(I)} \neq 0$, then $q^*[\sigma^{(i)}(I)]$ was set according to Line 5, and thus

$$
q^*[\sigma^{(i)}(I)] = \frac{1}{d_{\sigma^{(i)}(I)}}\left(\sum_{I' \prec I} \lambda[I']q^*[\sigma^{(i)}(I')]q_{I'}[\sigma^{(i)}(I)]\right). \tag{6}
$$

By definition of d_σ (Line 3), $d_\sigma = d_{\sigma^{(i)}(I)} + \lambda[I]$. Then,

$$
\sum_{a \in A(I)} q^*[(I,a)] = \frac{1}{d_\sigma}\left(\sum_{I' \preceq I} \lambda[I']q^*[\sigma^{(i)}(I')] \cdot \begin{cases} q_{I'}[\sigma^{(i)}(I)] & \text{if } I' \prec I \\ 1 & \text{otherwise} \end{cases}\right)
$$

$$
= \frac{1}{d_{\sigma^{(i)}(I)} + \lambda[I]}\left(\lambda[I]q^*[\sigma^{(i)}(I)] + \sum_{I' \prec I} \lambda[I']q^*[\sigma^{(i)}(I')]q_{I'}[\sigma^{(i)}(I)]\right)
$$

$$
= \frac{1}{d_{\sigma^{(i)}(I)} + \lambda[I]}\left(\lambda[I]q^*[\sigma^{(i)}(I)] + d_{\sigma^{(i)}(I)}q^*[\sigma^{(i)}(I)]\right) = q^*[\sigma^{(i)}(I)],
$$

where the second to last equality is obtained by Equation (6). This concludes the first part of the proof.

Part 2: q^* is a fixed point of ϕ. Fix a sequence $\sigma = (I,a) \in \Sigma^{(i)}$. We want to show that $\phi(q^*)[\sigma] - q^*[\sigma] = 0$. If $\sum_{I' \preceq I} \lambda[I'] = 0$, then it immediately holds that $\phi(q^*)[\sigma] = q^*[\sigma]$. Otherwise, if $\sum_{I' \preceq I} \lambda[I'] \neq 0$, by applying Lemma 3 and by subsequently substituting $q^*[\sigma]$ according to Line 5, we obtain

$$
\phi(q^*)[\sigma] - q^*[\sigma] = \left(\sum_{I' \preceq I} \lambda[I']q_{I'}[\sigma]q^*[\sigma^{(i)}(I')]\right) - d_\sigma q^*[\sigma]
$$

$$
= \left(\sum_{I' \preceq I} \lambda[I']q_{I'}[\sigma]q^*[\sigma^{(i)}(I')]\right) - \frac{d_\sigma}{d_\sigma}\left(\sum_{I' \preceq I} \lambda[I']q_{I'}[\sigma]q^*[\sigma^{(i)}(I')]\right) = 0.
$$

This concludes this part of the proof.

Part 3: time complexity. For each sequence in $\Sigma^{(i)} \setminus \{\varnothing\}$ (Line 2), Algorithm 2 has to visit at most $D^{(i)}$ information sets as part of the sums required on Lines 3 and 5. This completes the proof. □

B   Experimental Evaluation

B.1   Description of game instances

The size (in terms on number of infosets and sequences) of the parametric instances we use as benchmark is described in Figure 3. In the following, we provide a detailed explanation of the rules of the games.

	Players	Ranks	Player	Infosets	Sequences
Kuhn	3	4	Player 1	16	33
			Player 2	16	33
			Player 3	16	33
Goofspiel	3	3	Player 1	837	934
			Player 2	837	934
			Player 3	837	934
Leduc	3	3	Player 1	3294	7687
			Player 2	3294	7687
			Player 3	3294	7687

	Grid	Rounds	Player	Infosets	Sequences
Battleship	$(3, 2)$	3	Player 1	1413	2965
			Player 2	1873	4101

Figure 3: Size of our game instances in terms of number of sequences and infosets for each player of the game.

Kuhn poker   The two-player version of the game was originally proposed by (Kuhn 1950), while the three-player variation is due to (Farina et al. 2018). In a three-player Kuhn poker game with rank r, there are r possible cards. Each player initially pays one chip to the pot, and she/he is dealt a single private card. The first player may *check* or *bet* (*i.e.*, put an additional chip in the pot). Then, the second player can check or bet after a first player's check, or *fold/call* the first player's bet. If no bet was previously made, the third player can either check or bet. Otherwise, she/he has to fold or call. After a bet of the second player (resp., third player), the first player (resp., the first and the second players) still has to decide whether to fold or to call the bet. At the showdown, the player with the highest card who has not folded wins all the chips in the pot.

Goofspiel   This game was originally introduced by (Ross 1971). Goofspiel is essentially a bidding game where each player has a hand of cards numbered from 1 to r (*i.e.*, the rank of the game). A third stack of r cards is shuffled and singled out as prizes. Each turn, a prize card is revealed, and each player privately chooses one of her/his cards to bid, with the highest card winning the current prize. In case of a tie, the prize card is discarded. After r turns, all the prizes have been dealt out and the payoff of each player is computed as follows: each prize card's value is equal to its face value and the players' scores are computed as the sum of the values of the prize cards they have won. We remark that due to the tie-breaking rule that we employ, even two-player instances of the game are general-sum. All the Goofspiel instances have *limited information, i.e.,* actions of the other players are observed only at the end of the game. This makes the game strategically more challenging, as players have less information regarding previous opponents' actions.

Leduc   We use a three-player version of the classical Leduc hold'em poker introduced by Southey et al. (2005). In a Leduc game instance with r ranks the deck consists of three suits with r cards each. As the game starts players pay one chip to the pot. There are two betting rounds. In the first one a single private card is dealt to each player while in the second round a single board card is revealed. The maximum number of raise per round is set to two, with raise amounts of 2 and 4 in the first and second round, respectively.

Battleship   Battleship is a parametric version of the classic board game, where two competing fleets take turns at shooting at each other. For a detailed explanation of the Battleship game see the work by (Farina et al. 2019) that introduced it. Our instance has loss multiplier equal to 2, and one ship of length 2 and value 1 for each player

B.2   Details about experimental setup

All experiments were run on a machine with a 16-core 2.80GHz CPU and 32GB of RAM. Fixed points for EFCE dynamics were computed via the Eigen library version 3.3.7 (Guennebaud, Jacob et al. 2010).

At each time t, we let all players in the game pick their *mixed* strategy $q^{(i),t} \in Q^{(i)}$ according to Algorithm 3. Each player then observed their own linear utility function.

Note that no randomization is used in the experiments. Indeed, while it would technically be possible to have all players sample and play a deterministic strategy $\pi^{(i),t}$ from $q^{(i),t}$ and later compute the empirical frequency of the $\pi^{(i),t}$, we instead compute the EFCCE gap of the expected empirical frequency directly. This greatly improves the convergence rate in practice.