Mediator Interpretation and Faster Learning Algorithms for Linear Correlated Equilibria in General Extensive-Form Games

Brian Hu Zhang
Carnegie Mellon University
bhzhang@cs.cmu.edu

Gabriele Farina
MIT
gfarina@mit.edu

Tuomas Sandholm
Carnegie Mellon University
Strategic Machine, Inc.
Strategy Robot, Inc.
Optimized Markets, Inc.
sandholm@cs.cmu.edu

March 19, 2024

Abstract

A recent paper by Farina & Pipis (2023) established the existence of uncoupled no-linear-swap regret dynamics with polynomial-time iterations in extensive-form games. The equilibrium points reached by these dynamics, known as *linear correlated equilibria*, are currently the tightest known relaxation of correlated equilibrium that can be learned in polynomial time in any finite extensive-form game. However, their properties remain vastly unexplored, and their computation is onerous. In this paper, we provide several contributions shedding light on the fundamental nature of linear-swap regret. First, we show a connection between linear deviations and a generalization of *communication deviations* in which the player can make queries to a "mediator" who replies with action recommendations, and, critically, the player is not constrained to match the timing of the game as would be the case for communication deviations. We coin this latter set the *untimed communication (UTC) deviations*. We show that the UTC deviations coincide precisely with the linear deviations, and therefore that any player minimizing UTC regret also minimizes linear-swap regret. We then leverage this connection to develop state-of-the-art no-regret algorithms for computing linear correlated equilibria, both in theory and in practice. In theory, our algorithms achieve polynomially better per-iteration runtimes; in practice, our algorithms represent the state of the art by several orders of magnitude.

1 Introduction

In no-regret learning, a player repeatedly interacts with a possibly adversarial environment. The task of the player is to minimize its *regret*, which is defined to be the difference between the utility experienced by the player, and the largest utility that it could have achieved in hindsight if it had played other strategies instead, according to some *strategy transformation* or *deviation*. The set of allowable deviations defines the notion of regret, with larger sets corresponding to tighter notions. Two extremes are *external deviations*, which are the set of all constant transformations, and *swap deviations*, which are all possible functions.

In games, no-regret learning has a very tight connection to notions of *correlated equilibrium*. Each notion of regret has its corresponding notion of equilibrium, which will be reached by a set of players that independently run no-regret algorithms for that notion of regret. External and swap deviations, respectively, correspond to the well-known *normal-form coarse correlated* equilibrium (Moulin & Vial, 1978) (NFCCE) and *normal-form correlated* equilibrium (Aumann, 1974) (NFCE). For extensive-form games specifically, other sets of deviations include the *trigger deviations* (Gordon et al., 2008; Farina et al., 2022a), which correspond to *extensive-form correlated equilibrium* (von Stengel & Forges, 2008), and the *communication deviations* (Zhang & Sandholm, 2022; Fujii, 2023), which correspond to *communication equilibrium*[1] (Myerson, 1986; Forges, 1986).

[1]Technically, communication equilibria are more broad than (Φ_i)-equilibria where Φ_i is the set of communication deviations: in a communication equilibrium, there is an explicit *mediator* who has the power not only to sample a strategy profile, but

In this paper, we consider a notion of regret first studied for extensive-form games by Farina & Pipis (2023), namely, regret with respect to the set of *linear functions* from the strategy set to itself. This notion is a natural stepping stone between external regret, which is very well studied, and swap regret, for which achieving $\text{poly}(d) \cdot T^c$ regret, where d is the size of the decision problem and $c < 1$, is a long-standing open problem. We make two main contributions.

The first contribution is conceptual: we give, for extensive-form games, an *interpretation* of the set of linear deviations. More specifically, we will first introduce a set of deviations, which we will call the *untimed communication (UTC) deviations* that, a priori, seems very different from the set of linear deviations at least on a conceptual level. The deviation set, rather than being defined *algebraically* (linear functions), will be defined in terms of an interaction between a *deviator*, who wishes to evaluate the deviation function at a particular input, and a *mediator*, who answers queries about the input. We will show the following result, which is our first main theorem:

Theorem. *The untimed communication deviations are precisely the linear deviations.*

The mediator-based framework is more in line with other extensive-form deviation sets—indeed, all prior notions of regret for extensive form, to our knowledge, including all the notions discussed above, can be expressed in terms of the framework. As such, the above theorem places linear deviations firmly within the same framework usually used to study deviations in extensive form.

We will then demonstrate that the set of UTC deviations is expressible in terms of *scaled extensions* (Farina et al., 2019c), opening up access to a wide range of extremely fast algorithms for regret minimization, both theoretically and practically, for UTC deviations and thus also for linear deviations. Our second main theorem is as follows.

Theorem (Faster linear-swap regret minimization)**.** *There exists a regret minimizer with regret $O(d^2\sqrt{T})$ against all linear deviations, and whose per-iteration complexity is dominated by the complexity of computing a fixed point of a linear map $\phi^{(t)} : \text{co}\,\mathcal{X} \to \text{co}\,\mathcal{X}$.*

In particular, using the algorithm of Cohen et al. (2021) to solve the linear program of finding a fixed point, our per-iteration complexity is $\widetilde{O}(d^\omega)$, where $\omega \approx 2.37$ is the current matrix multiplication constant and $\widetilde{O}$ hides logarithmic factors. We elaborate on the fixed-point computation in Section 6. This improves substantially on the result of Farina & Pipis (2023), which has the same regret bound but whose per-iteration computation involved a *quadratic* program (namely, an ℓ_2 projection), which has higher complexity than a linear program (they give a bound of $\widetilde{O}(d^{10})$). Finally, we demonstrate via experiments that our method is also empirically faster than the prior method.

2 Preliminaries

Here, we review fundamentals of tree-form decision making, extensive-form games, and online convex optimization. Our exposition and notation mostly follows Farina & Pipis (2023).

2.1 Tree-Form Decision Making

A *tree-form decision problem* is a rooted tree where every path alternates between two types of nodes: *decision points* ($j \in \mathcal{J}$) and *observation points* (or *sequences*) ($\sigma \in \Sigma$). The root node $\varnothing \in \Sigma$ is always an observation point. At decision points, the edges are called *actions*, and the player must select one of the legal actions. At observation points, the edges are called *observations* or *signals*, and the player observes one of the signals before continuing. The number of sequences is denoted $d = |\Sigma|$. The parent of a node s is denoted p_s. The set of actions available at a decision point j is denoted A_j. The set of decision point following an observation

also to *pass private information* between players—so a communication equilibrium is not necessarily a correlated profile at all. This distinction is fairly fundamental: it is the reason why polynomial-time algorithms for optimal communication equilibrium can exist for extensive-form games (Zhang & Sandholm, 2022). We may call a communication equilibrium that also happens to be a correlated profile a *private communication equilibrium*, where *private* denotes that the mediator is not allowed to pass information between players. However, since this paper focuses on no-regret learning, we have no reason to make this distinction, so we largely ignore it.

point σ will be denoted C_σ. An observation node $\sigma \in \Sigma$ is uniquely identified by its parent decision point j and the action a taken at j. We will hence use ja as an alternative notation for the same observation point.

A *sequence-form pure strategy* for the player is a vector $\boldsymbol{x} \in \{0,1\}^d$, indexed by sequences in Σ, where $\boldsymbol{x}(\sigma) = 1$ if the player selects *every* action on the $\varnothing \to \sigma$ path. A *sequence-form mixed strategy* is a convex combination of sequence-form pure strategies. We will use $\mathcal{X}$ to denote the set of sequence-form pure strategies. An important property (Romanovskii, 1962; von Stengel, 1996) is that the convex hull of $\mathcal{X}$, which we will denote $\operatorname{co}\mathcal{X}$, is described by a system of linear constraints:

$$\boldsymbol{x}(\varnothing) = 1, \qquad \boldsymbol{x}(p_j) = \sum_{a \in A_j} \boldsymbol{x}(ja) \ \forall j \in \mathcal{J}. \tag{1}$$

Tree-form decision problems naturally encode the decision problems faced by a player with perfect recall in an *extensive-form game*. An extensive-form game with n players is a game of incomplete information played on a tree of nodes $\mathcal{H}$. At every non-leaf node $h \in H$, the children of h are labeled with actions $a \in A_h$. Each nonterminal node is assigned to a different player, and the player to whom a node is assigned selects the action at that node. The nodes assigned to a given player are partitioned into *information sets*, or *infosets*; a player cannot distinguish among the nodes in a given infoset, and therefore a pure strategy must play the same action at every node in an infoset. Finally, each player has a *utility function* $u_i : \mathcal{Z} \to \mathbb{R}$, where $\mathcal{Z}$ is the set of terminal nodes. We will assume *perfect recall*, that is, we will assume that players never forget information.

In an extensive-form game, a perfect-recall player's decision problem is a tree-form decision problem whose size (number of nodes) is linear in the size of the game tree, and the utility functions are linear in every player's strategy. We will use $\mathcal{X}_i$ to denote the tree-form decision problem faced by player i. Then the utility functions $u_i : \operatorname{co}\mathcal{X}_1 \times \cdots \times \operatorname{co}\mathcal{X}_n \to \mathbb{R}$ are linear in each player's strategy.

2.2 Online Convex Optimization and Φ-Regret

In *online convex optimization* (Zinkevich, 2003), a player (or "learner") has a strategy set $\mathcal{X} \subseteq \mathbb{R}^d$, and repeatedly faces a possibly-adversarial environment. More formally, at every iteration $t = 1, \dots, T$, the player selects a distribution $\pi^{(t)} \in \Delta(\mathcal{X})$, and an adversary simultaneously selects a *utility vector* $\boldsymbol{u}^{(t)} \in [0,1]^d$. The player then observes the utility $\boldsymbol{u}^{(t)}$, selects a new strategy $\boldsymbol{x}^{(t+1)} \in \mathcal{X}$, and so on. Our metric of performance will be the notion of Φ-*regret* (Greenwald & Jafari, 2003). Given a set of transformations[2] $\Phi \subseteq (\operatorname{co}\mathcal{X})^{\mathcal{X}}$:

Definition 2.1. The Φ-*regret* of the player after T timesteps is given by

$$\operatorname{Reg}_\Phi(T) := \max_{\phi \in \Phi} \sum_{t=1}^{T} \mathop{\mathbb{E}}_{\boldsymbol{x} \sim \pi^{(t)}} \left\langle \boldsymbol{u}^{(t)}, \phi(\boldsymbol{x}) - \boldsymbol{x} \right\rangle.$$

Various choices of Φ correspond to various notions of regret, with larger sets resulting in stronger notions of regret. In an extensive-form game, notions of Φ-regret correspond to notions of equilibrium. For each player $i \in [n]$, let $\Phi_i \subseteq (\operatorname{co}\mathcal{X}_i)^{\mathcal{X}_i}$ be a set of transformations for player i.

A distribution $\pi \in \Delta(\mathcal{X}_1 \times \cdots \times \mathcal{X}_n)$ is called a *correlated profile*. A ε-(Φ_i)-equilibrium is a correlated profile such that $\mathbb{E}_{\boldsymbol{x} \sim \pi}\left[u_i(\phi(\boldsymbol{x}_i), \boldsymbol{x}_{-i}) - u_i(\boldsymbol{x}_i, \boldsymbol{x}_{-i})\right] \le \varepsilon$. for every player i and deviation $\phi \in \Phi_i$. If all players independently run Φ_i-regret minimizers over their strategy sets $\mathcal{X}_i$, the empirical frequency of play $\pi = \operatorname{Unif}(\pi^{(1)}, \dots, \pi^{(T)})$ will be an ε-(Φ_i)-equilibrium for $\varepsilon = \max_i \operatorname{Reg}_{\Phi_i}(T)/T$. Thus, Φ-regret minimizers immediately imply no-regret learning algorithms converging to (Φ_i)-equilibria. Some common choices of Φ, and corresponding equilibrium notions, are in Table 1. In this paper, our focus will be on *linear-swap regret*, which is the regret against the set Φ_{LIN} of all linear[3] maps $\phi : \mathcal{X} \to \operatorname{co}\mathcal{X}$. To our knowledge, linear-swap regret was first proposed by Gordon et al. (2008) for general convex spaces. They developed a general framework

[2] A^B is the set of functions from B to A

[3] For sets $\mathcal{X}$ whose affine hull excludes the origin, there is no point in distinguishing affine maps from linear maps. Sequence-form strategy sets $\mathcal{X}$ are such sets, because $\boldsymbol{x}(\varnothing) = 1$ is always a constraint. So, throughout this paper, we will not distinguish between linear and affine maps.

Deviations Φ	Equilibrium concept	References
Constant (external), $\Phi = \{\phi : \boldsymbol{x} \mapsto \boldsymbol{x}_0 \mid \boldsymbol{x}_0 \in \mathcal{X}\}$	Normal-form coarse correlated	Moulin & Vial (1978)
Trigger (see Section 3)	Extensive-form correlated	von Stengel & Forges (2008)
Communication (see Section 3)	Communication	Forges (1986); Myerson (1986)
Linear / Untimed communication	Linear correlated	Farina & Pipis (2023); **this paper**
Swap, $\Phi = \mathcal{X}^{\mathcal{X}}$	Normal-form correlated	Aumann (1974)

Table 1: Some examples of deviation sets Φ and corresponding notions of correlated equilibrium, in increasing order of size of Φ (and thus increasing tightness of the equilibrium concept)

for Φ-regret minimization, which we now review. We start by observing that any linear $\phi : \mathcal{X} \to \operatorname{co}\mathcal{X}$ is naturally extended to a function $\phi : \operatorname{co}\mathcal{X} \to \operatorname{co}\mathcal{X}$ by setting $\phi : \operatorname{co}\mathcal{X} \ni \boldsymbol{x} \mapsto \mathbb{E}_{\boldsymbol{x}' \sim \pi} \phi(\boldsymbol{x}')$, where $\pi \in \Delta(\mathcal{X})$ is any distribution for which $\mathbb{E}_{\boldsymbol{x}' \sim \pi} \boldsymbol{x}' = \boldsymbol{x}$ (The choice of distribution is irrelevant because of linearity of expectation, and thus $\phi : \operatorname{co}\mathcal{X} \to \operatorname{co}\mathcal{X}$ is uniquely defined.)

Theorem 2.2 (Gordon et al., 2008)**.** *Let $\Phi \subseteq \Phi_{\text{LIN}}$ be a convex set of transformations, and let $\mathcal{R}_\Phi$ be a deterministic*[4] *external regret minimizer over Φ, whose regret after T timesteps is R. Then the following algorithm achieves Φ-regret R on $\mathcal{X}$ after T timesteps: At every timestep t, the player queries $\mathcal{R}_\Phi$ for a strategy (transformation) $\phi^{(t)} \in \Phi$, and the player selects a strategy $\boldsymbol{x}^{(t)} \in \operatorname{co}\mathcal{X}$ that is a fixed point of $\phi^{(t)}$, that is, $\phi^{(t)}(\boldsymbol{x}^{(t)}) = \boldsymbol{x}^{(t)}$. Upon observing utility $\boldsymbol{u}^{(t)}$, the player forwards the utility $\phi \mapsto \langle \boldsymbol{u}^{(t)}, \phi(\boldsymbol{x}^{(t)}) \rangle$ to $\mathcal{R}_\Phi$.*

Therefore, to construct a Φ_{LIN}-regret minimizer over $\mathcal{X}$, it suffices to be able to (1) minimize *external* regret over Φ_{LIN}, and (2) compute fixed points of transformations $\phi^{(t)}$. For linear $\phi^{(t)} : \boldsymbol{x} \mapsto \mathbf{A}\boldsymbol{x}$, computing a fixed point amounts to solving a linear program. Therefore, the focus of this paper will be on external regret minimizers over the set Φ_{LIN}.

For extensive-form games, linear-swap regret was recently studied in detail by Farina & Pipis (2023): they provide a characterization of the set Φ_{LIN} when $\mathcal{X}$ is a sequence-form polytope, and thus derive an algorithm for minimizing Φ_{LIN}-regret over $\mathcal{X}$. Their paper is the starting point of ours.

3 Mediators and UTC Deviations

With the notable exception of linear deviations, most sets of deviations Φ for extensive-form games are defined by interactions between a *mediator* who holds a strategy $\boldsymbol{x} \in \mathcal{X}$, and a *deviator*, who should compute the function $\phi(\boldsymbol{x})$ by making queries to the mediator. The set of deviations is then defined by what queries that the player is allowed to make. Before continuing, we will first formulate the sets Φ mentioned in Section 2.2 in this paradigm, for intuition. For a given decision point j, call an action $a \in A_j$ the *recommended action at* j, denoted $a(\boldsymbol{x}, j)$, if $\boldsymbol{x}(ja) = 1$. Since $\boldsymbol{x}$ is a sequence-form strategy, it is possible for a decision point to have no recommended action if its parent p_j is itself not recommended.

- Constant (NFCCE): The deviator cannot to make any queries to the mediator.
- Trigger (EFCE): The deviator, upon reaching a decision point j, learns the recommended action (if any) at j before selecting its own action.
- Communication: The deviator maintains a *state* with the mediator, which is a sequence σ, initially $\varnothing$. Upon reaching a decision point j, the deviator selects a decision point $j' \in C_\sigma$ (possibly $j' \neq j$) at which to query the mediator, the deviator observes the recommendation $a' = a(\boldsymbol{x}, j')$, then the deviator must pick an action $a \in A_j$. The state is updated to $j'a'$.
- Swap (NFCE): The deviator learns the whole strategy $\boldsymbol{x}$ before selecting its strategy.

[4] A deterministic regret minimizer is one that uses no randomness internally to compute its strategies. When the strategy set (here Φ) is convex, and the notion is external regret, the learner need not randomize: since utilities are linear, picking a distribution $\pi^{(t)} \in \Delta(\Phi)$ is equivalent to deterministically selecting the point $\phi^{(t)} := \mathbb{E}_{\phi \sim \pi^{(t)}} \phi$. Thus, we allow the $\mathcal{R}_\Phi$-adversary (here, the player itself) to set a utility $\phi \mapsto \langle \boldsymbol{u}^{(t)}, \phi(\boldsymbol{x}^{(t)}) \rangle$ that depends on the learner's choice of ϕ.

An example of a communication deviation can be found in Section 5, and further discussion of these solution concepts can be found in Appendix B. Of these, the closest notion to ours is the notion of communication deviation, and that is the starting point of our construction. One critical property of communication deviations is that the mediator and deviator "share a clock": for every decision point reached, the deviator must make exactly one query to the mediator. As the name suggests, our set of *untimed* deviations results from removing this timing restriction, and therefore allowing the deviator to make *any number* (zero, one, or more than one) of queries to the mediator for every decision point reached. We formally define the decision problem faced by an untimed deviator as follows.

Definition 3.1. The *UTC decision problem* corresponding to a given tree-form decision problem is defined as follows. Nodes are identified with pairs $(s, \widetilde{s})$ where $s, \widetilde{s} \in \Sigma \cup \mathcal{J}$. s represents the state of the real decision problem, and $\widetilde{s}$ represents the state of the mediator. The root is $(\varnothing, \varnothing) \in \Sigma \times \Sigma$.

1. $(\sigma, \widetilde{\sigma}) \in \Sigma \times \Sigma$ is an observation point. The deviator observes the next decision point $j \in C_\sigma$, and the resulting decision point is $(j, \widetilde{\sigma})$
2. $(j, \widetilde{j}) \in \mathcal{J} \times \mathcal{J}$ is an observation point. The deviator observes the recommendation $a = a(\boldsymbol{x}, \widetilde{j})$, and the resulting decision point is $(j, \widetilde{j}a)$.
3. $(j, \widetilde{\sigma}) \in \mathcal{J} \times \Sigma$ is a decision point. The deviator can choose to either play an action $a \in A_j$, or to query a decision point $\widetilde{j} \in C_{\widetilde{\sigma}}$. In the former case, the resulting observation point is $(ja, \widetilde{\sigma})$ for $a \in A_j$; in the latter case, the resulting observation point is $(j, \widetilde{j})$.

Any mixed strategy of the deviator in this decision problem defines a function $\phi : \mathcal{X} \to \operatorname{co} \mathcal{X}$, where $\phi(\boldsymbol{x})(\sigma)$ is the probability that an untimed deviator plays all the actions on the path to σ when the mediator recommends according to pure strategy $\boldsymbol{x}$. We thus define:

Definition 3.2. An *UTC deviation* is any function $\phi : \mathcal{X} \to \operatorname{co} \mathcal{X}$ induced by a mixed strategy of the deviator in the UTC decision problem.

Clearly, the set of UTC deviations is at least as large as the set of communication deviations, and at most as large as the set of swap deviations. In the next section, we will discuss how to represent UTC deviations, and show that UTC deviations coincide precisely with linear deviations.

4 Representation of UTC Deviations and Equivalence between UTC and Linear Deviations

Since UTC deviations are defined by a decision problem, one method of representing such deviations is to express it as a tree-form decision problem and use the sequence-form representation. However, the UTC decision problem is not a tree—it is a DAG, since there are multiple ways of reaching any given decision point $(j, \widetilde{\sigma})$ depending on the ordering of the player's past actions and queries. Converting it to a tree by considering the tree of paths through the DAG would result in an exponential blowup: a decision point $(j, \widetilde{\sigma})$, where j is at depth k and $\widetilde{\sigma}$ is at depth ℓ, can be reached in roughly $\binom{k+\ell}{k}$ ways, so the total number of paths can be exponential in the depth of the decision problem even when the number of sequences, $d = |\Sigma|$, is not.

However, it is still possible to define the "sequence form" of a pure deviation in our UTC decision problem as follows[5]: it is a pair of matrices $(\mathbf{A}, \mathbf{B})$ where $\mathbf{A} \in \{0,1\}^{\Sigma \times \Sigma}$ encodes the part corresponding to sequences $(\sigma, \widetilde{\sigma})$, and $\mathbf{B} \in \{0,1\}^{\mathcal{J} \times \mathcal{J}}$ encodes the part corresponding to decision points $(j, \widetilde{j})$. $\mathbf{A}(\sigma, \widetilde{\sigma}) = 1$ if the deviator plays all the actions on *some* path to observation point $(\sigma, \widetilde{\sigma})$, and similarly $\mathbf{B}(j, \widetilde{j}) = 1$ if the deviator plays all the actions on some path to observation node $(j, \widetilde{j})$. Since the only possible way for two paths to end at the same observation point is for the deviator to have changed the order of actions and queries, for any given pure strategy of the deviator, at most one path can exist for both cases. Therefore, the set of mixed

[5]This construction is a special case of the more general construction of sequence forms for DAG decision problems explored by Zhang et al. (2023) in the case of team games.

sequence-form deviations can be expressed using the following set of constraints:

$$\begin{aligned} \mathbf{A}(p_j, \widetilde{\sigma}) + \mathbf{B}(j, p_{\widetilde{\sigma}}) &= \sum_{a \in A_j} \mathbf{A}(ja, \widetilde{\sigma}) + \sum_{\widetilde{j} \in C_{\widetilde{\sigma}}} \mathbf{B}(j, \widetilde{j}) && \forall j \in \mathcal{J}, \widetilde{\sigma} \in \Sigma \\ \mathbf{A}(\varnothing, \varnothing) &= 1 \\ \mathbf{A}(\varnothing, \widetilde{\sigma}) &= 0 && \forall \widetilde{\sigma} \neq \varnothing \\ \mathbf{A}, \mathbf{B} &\geq 0 \end{aligned} \tag{2}$$

where, in a slight abuse of notation, we define $\mathbf{B}(j, p_\varnothing) := 0$ for every $j \in \mathcal{J}$. Moreover, for any pair of matrices $(\mathbf{A}, \mathbf{B})$ satisfying the constraint system and therefore defining some deviation $\phi : \mathcal{X} \to \operatorname{co} \mathcal{X}$, it is easy to compute how ϕ acts on any $\boldsymbol{x} \in \mathcal{X}$: the probability that the deviator plays all the actions on the $\varnothing \to \sigma$ path is simply given by

$$\sum_{\widetilde{\sigma} \in \Sigma} \boldsymbol{x}(\widetilde{\sigma}) \mathbf{A}(\sigma, \widetilde{\sigma}) = (\mathbf{A}\boldsymbol{x})(\sigma),$$

and therefore ϕ is nothing more than a matrix multiplication with $\mathbf{A}$, that is, $\phi(\boldsymbol{x}) = \mathbf{A}\boldsymbol{x}$. We have thus shown that every UTC deviation is linear, that is, $\Phi_{\text{UTC}} \subseteq \Phi_{\text{LIN}}$. In fact, the reverse inclusion holds too:

Theorem 4.1. *The UTC deviations are precisely the linear deviations. That is,* $\Phi_{\text{UTC}} = \Phi_{\text{LIN}}$.

The proof is deferred to Appendix D. Since the two sets are equivalent, in the remainder of the paper, we will use the terms *UTC deviation* and *linear deviation* (similarly, *UTC regret* and *linear-swap regret*) interchangeably.

5 Example

In this section, we provide an example in which the UTC deviations are strictly more expressive than the communication deviations. Consider the game in Figure 1. The subgames rooted at **D** and **E** are guessing games, where ▲ must guess ▼'s action, with a large penalty for guessing wrong. Consider the correlated profile that mixes uniformly among the four pure profiles ($\mathbf{a}_i$, $\mathbf{b}_j$, $\mathbf{c}_1$, $\mathbf{f}_i$, $\mathbf{g}_j$) for $i, j \in \{1, 2\}$. In this profile, the information that ▲ needs to guess perfectly is contained in the recommendations: the recommendation at **A** tells it how to guess at **D**, and the recommendation at **B** tells it how to guess at **E**. With a communication deviation, ▲ cannot access this information in a profitable way, since upon reaching **C**, ▲ must immediately make its first mediator query. Hence, this profile is a communication equilibrium. However, with an *untimed* communication deviation, ▲ can profit: it should, upon reaching[6] **C**, play action $\mathbf{c}_2$ *without making a mediator query*, and then query **A** if it observes **D**, and **B** if it observes **E**. This deviation is allowed only due to the untimed nature of UTC deviations allows the deviating player to *delay* its query to the mediator until it reaches either **D** or **E**. In a *timed* communication deviation, this deviation is impossible, because the player must make its first query (**A**, **B**, or **C**) *before* reaching **D** or **E**, and thus that query cannot be conditioned on which one of **D** or **E** will be reached.

Another example, where the player can profit from making *more than one* query, and untimed deviations affects the set of possible equilibrium outcomes, can be found in Appendix C.

6 Regret Minimization on Φ_{UTC}

In this section, we discuss how Theorem 4.1 can be used to construct very efficient Φ_{LIN}-regret minimizers, both in theory and in practice. The key observation we use here is due to Zhang et al. (2023): they observed that DAG decision problems have a structure that allows them to be expressed as *scaled extensions*, allowing the application of the *counterfactual regret minimization* (CFR) framework (Zinkevich et al., 2007; Farina et al., 2019a):

[6] The actions/queries ▲ makes at **A** and **B** are irrelevant, because ▲ only cares about maximizing utility, and it always gets utility 0 regardless of what it does. In the depiction of this deviation in Figure 2, the deviator always plays action 1 at **A** and **B**.

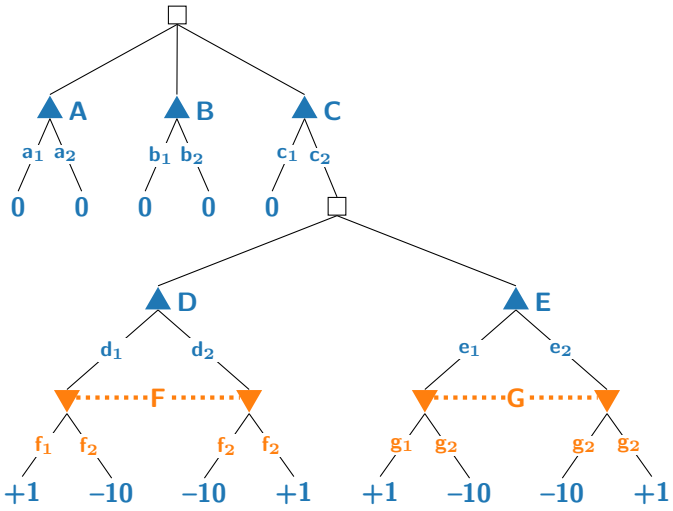
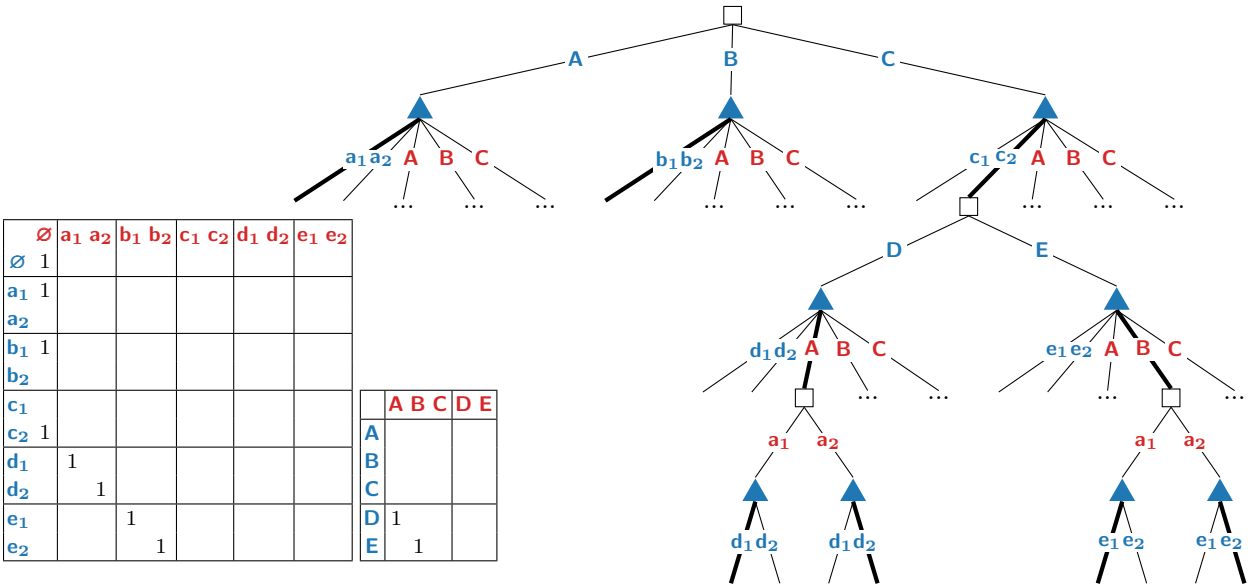


Figure 1: An example extensive-form game in which communication deviations are a strict subset of UTC deviations. There are two players, P1 (▲) and P2 (▼). Infosets for both players are labeled with capital letters (*e.g.*, **A**) and joined by dotted lines. Actions are labeled with lowercase letters and subscripts (*e.g.*, $\mathbf{a_1}$). P1's utility is labeled on each terminal node. P2's utility is zero everywhere (not labeled). Boxes are chance nodes, at which chance plays uniformly at random.



	∅	a1 a2	b1 b2	c1 c2	d1 d2	e1 e2
∅	1					
a1 a2	1					
b1 b2	1					
c1 c2						
d1 d2		1 1				
e1 e2			1 1			

	A B C	D E
A B C		
D E	1 1	

Figure 2: A part of the UTC decision problem for ▲ corresponding to the same game. Nodes labeled ▲ are decision points for ▲; boxes are observation points. "..." denotes that the part of the decision problem following that edge has been omitted. Terminal nodes are unmarked. Red edge labels indicate interactions with the mediator; blue edge labels indicate interactions with the game. The profitable untimed deviation discussed in Section 5 is indicated by the thick lines. The first action taken in that profiable deviation, $\mathbf{c_2}$, is not legal for a timed deviator, because a timed deviator must query the mediator once before taking its first action. The matrices (lower-left corner) are the pair of matrices $(\mathbf{A}, \mathbf{B})$ corresponding to that same deviation. All blank entries are 0.

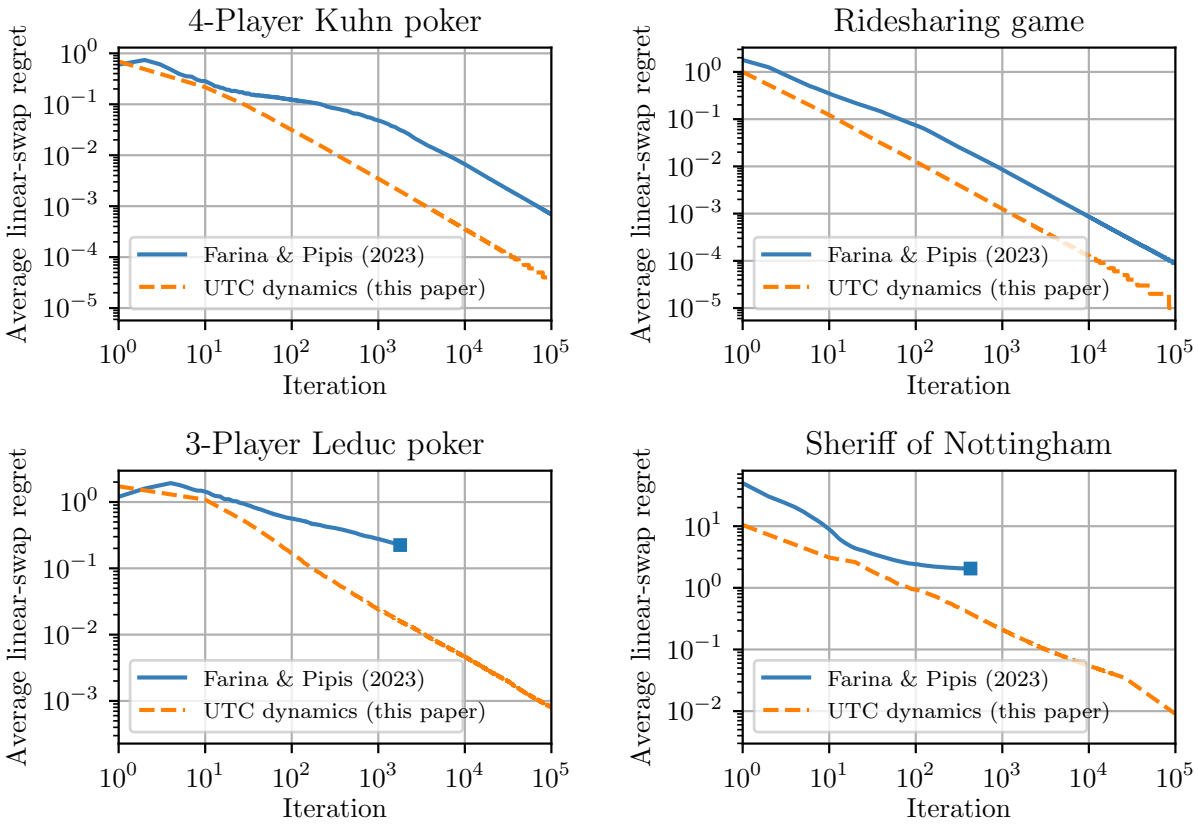


Figure 3: Experimental comparison between our dynamics and those of Farina & Pipis (2023) for approximating a linear correlated equilibrium in extensive-form games. Each algorithm was run for a maximum of 100,000 iterations or 6 hours, whichever was hit first. Runs that were terminated due to the time limit are marked with a square ■.

Theorem 6.1 (CFR for Φ_{LIN}, special case of Zhang et al., 2023)**.** *CFR-based algorithms can be used to construct an external regret minimizer on Φ_{UTC} (and thus also on Φ_{LIN}) with $O(d^2\sqrt{T})$ regret and $O(d^2)$ per-iteration complexity.*

Applying Theorem 2.2 now yields:

Theorem 6.2. *CFR-based algorithms can be used to construct a Φ_{LIN}-regret minimizer with $O(d^2\sqrt{T})$ regret, and per-iteration complexity dominated by the complexity of computing a fixed point of a linear transformation $\phi^{(t)} : \operatorname{co}\mathcal{X} \to \operatorname{co}\mathcal{X}$.*

As mentioned in the introduction, this significantly improves the per-iteration complexity of linear-swap regret minimization. Fixed points can be computed by finding a feasible solution to the constraint system $\{\boldsymbol{x} \in \mathcal{X}, \mathbf{A}\boldsymbol{x} = \boldsymbol{x}\}$, where $\boldsymbol{x} \in \mathcal{X}$ is expressed using the sequence-form constraints (1). This is a linear program with $O(d)$ variables and constraints, so the LP algorithm of Cohen et al. (2021) yields a fixed-point computation algorithm with runtime $\widetilde{O}(d^\omega)$.

For comparison, the algorithm of Farina & Pipis (2023) requires an ℓ_2 projection onto $\mathcal{X}$ on every iteration, which requires solving a convex quadratic program; the authors of that paper derive a bound of $\widetilde{O}(d^{10})$, which, although polynomial, is much slower than our algorithm. CFR-based algorithms are currently the fastest practical regret minimizers (Brown & Sandholm, 2019; Farina et al., 2021)—therefore, showing that our method allows such algorithms to be applied is also a significant practical step. In Section 7, we will show empirically that the resulting algorithm is significantly better than the previously-known state of the art, in terms of both per-iteration time complexity and number of iterations.

Game	Our algorithm	Farina & Pipis (2023)	Speedup
4-Player Kuhn poker	5.65ms ± 0.30ms	195ms ± 7ms	35×
Ridesharing game	676µs ± 80µs	160ms ± 7ms	237×
3-Player Leduc poker	42.0ms ± 0.7ms	12.1s ± 1.0s	287×
Sheriff of Nottingham	114ms ± 16ms	50.2s ± 9.6s	442×

Table 2: Comparison of average time per iteration. For each combination of game instance and algorithm, the mean and standard deviation of the iteration runtime are noted.

Game	Target gap	Our algorithm	Farina & Pipis (2023)	Speedup
4-Player Kuhn poker	7×10^{-4}	32.8s	5h 25m	595×
Ridesharing game	9×10^{-5}	8.89s	4h 07m	1667×
3-Player Leduc poker	0.224	2.12s	6h 00m	10179×
Sheriff of Nottingham	2.06	2.00s	6h 00m	10800×

Table 3: Comparison of time taken to achieve a particular linear swap equilibrium gap. The gap is whatever gap was achieved by the algorithm of Farina & Pipis (2023) before termination.

7 Experimental Evaluation

We empirically investigate the performance of our learning dynamics for linear correlated equilibrium, compared to the recent algorithm by Farina & Pipis (2023). We test on four benchmark games:

- 4-player Kuhn poker, a multiplayer variant of the classic benchmark game introduced by Kuhn (1950). The deck has 5 cards. This game has 3,960 terminal states.
- A ridesharing game, a two-player general-sum game introduced as a benchmark for welfare-maximizing equilibria by Zhang et al. (2022). This game has 484 terminal states.
- 3-player Leduc poker, a three-player variant of the classic Leduc poker introduced by Southey et al. (2005). Only one bet per round is allowed, and the deck has 6 cards (3 ranks, 2 suits). The game has 4,500 terminal states.
- Sheriff of Nottingham, a two-player general-sum game introduced by Farina et al. (2019b) for its richness of equilibrium points. The smuggler has 10 items, a maxmimum bribe of 2, and 2 rounds to bargain. The game has 2,376 terminal states.

We run our algorithm based on the UTC polytope, and that of Farina & Pipis (2023) (with the learning rate $\eta = 0.1$ as used by the authors), for a limit of 100,000 iterations or 6 hours, whichever is hit first. Instead of solving linear programs to find the fixed points, we use power iteration, which is faster in practice. All experiments were run on the same machine with 32GB of RAM and a processor running at a nominal speed of 2.4GHz. For our learning dynamics, we employed the CFR algorithm instantiated with the regret matching$^+$ (Tammelin, 2014) regret minimizer at each decision point (see Theorem 6.1). Experimental results are shown in Figure 3.

One of the most appealing features of our algorithm is that allows CFR-based methods to apply. CFR-based methods are the fastest regret minimizers in practice, so it is unsurprising that using them results in better convergence as seen in Figure 3. Another appealing feature is that our method sidesteps the need of projecting onto the set of transformations. This is in contrast with the algorithm of Farina & Pipis (2023), which requires an expensive projection at every iteration. We observe that this difference results in a dramatic reduction in iteration runtime between the two algorithms, which we quantify in Table 2. So, we remark that when accounting for *time* instead of iterations on the x-axis of the plots in Figure 3, the difference in performance between the algorithms appears even stronger. Such a plot is available in Appendix E.

8 Conclusion and Future Research

In this paper, we have introduced a new representation for the set of linear deviations when the strategy space is sequence form. Our representation connects linear deviations to the mediator-based framework that is more typically used for correlation concepts in extensive-form games, and therefore gives a reasonable game-theoretic interpretation of what linear equilibria represent. It also leads to state-of-the-art no-linear-regret algorithms, both in theory and in practice. Several natural questions remain open:

1. Is there an algorithm whose swap regret is $\text{poly}(d) \cdot T^c$ for $c < 1$ in extensive-form games? (See also Appendix B for some recent progress on this problem.)
2. What would be a reasonable definition of *untimed communication equilibrium*, as a refinement of *communication equilibrium* (see also Appendix A.6)?
3. For extensive-form correlated equilibrium, it is possible to achieve $\text{poly}(d) \cdot \log(T)$ regret (Anagnostides et al., 2023), and to compute exact equilibria in polynomial time (Huang & von Stengel, 2008). Can one extend these results to linear equilibria?

Acknowledgements

This material is based on work supported by the Vannevar Bush Faculty Fellowship ONR N00014-23-1-2876, National Science Foundation grants RI-2312342 and RI-1901403, ARO award W911NF2210266, and NIH award A240108S001.

References

Ioannis Anagnostides, Gabriele Farina, and Tuomas Sandholm. Near-optimal phi-regret learning in extensive-form games. In *International Conference on Machine Learning (ICML)*, 2023.

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

Noam Brown and Tuomas Sandholm. Solving imperfect-information games via discounted regret minimization. In *AAAI Conference on Artificial Intelligence (AAAI)*, 2019.

Michael B Cohen, Yin Tat Lee, and Zhao Song. Solving linear programs in the current matrix multiplication time. *Journal of the ACM (JACM)*, 68(1):1–39, 2021.

Yuval Dagan, Constantinos Daskalakis, Maxwell Fishelson, and Noah Golowich. From external to swap regret 2.0: An efficient reduction and oblivious adversary for large action spaces. *arXiv preprint arXiv:2310.19786*, 2023.

Gabriele Farina and Charilaos Pipis. Polynomial-time linear-swap regret minimization in imperfect-information sequential games. *arXiv preprint arXiv:2307.05448*, 2023.

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

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

Gabriele Farina, Chun Kai Ling, Fei Fang, and Tuomas Sandholm. Efficient regret minimization algorithm for extensive-form correlated equilibrium. In *Conference on Neural Information Processing Systems (NeurIPS)*, 2019c.

Gabriele Farina, Andrea Celli, Nicola Gatti, and Tuomas Sandholm. Connecting optimal ex-ante collusion in teams to extensive-form correlation: Faster algorithms and positive complexity results. In *International Conference on Machine Learning*, 2021.

Gabriele Farina, Andrea Celli, Alberto Marchesi, and Nicola Gatti. Simple uncoupled no-regret learning dynamics for extensive-form correlated equilibrium. *Journal of the ACM*, 69(6):1–41, 2022a.

Gabriele Farina, Chung-Wei Lee, Haipeng Luo, and Christian Kroer. Kernelized multiplicative weights for 0/1-polyhedral games: Bridging the gap between learning in extensive-form and normal-form games. In *International Conference on Machine Learning (ICML)*, pp. 6337–6357. PMLR, 2022b.

Francoise Forges. An approach to communication equilibria. *Econometrica: Journal of the Econometric Society*, pp. 1375–1385, 1986.

Françoise Forges and Frédéric Koessler. Communication equilibria with partially verifiable types. *Journal of Mathematical Economics*, 41(7):793–811, 2005.

Kaito Fujii. Bayes correlated equilibria and no-regret dynamics. *arXiv preprint arXiv:2304.05005*, 2023.

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

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

Wan Huang and Bernhard von Stengel. Computing an extensive-form correlated equilibrium in polynomial time. In *International Workshop on Internet and Network Economics*, pp. 506–513. Springer, 2008.

Andrew Kephart and Vincent Conitzer. The revelation principle for mechanism design with signaling costs. *ACM Transaction on Economics and Computation (TEAC)*, 9(1):1–35, 2021.

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

H. Moulin and J.-P. Vial. Strategically zero-sum games: The class of games whose completely mixed equilibria cannot be improved upon. *International Journal of Game Theory*, 7(3-4):201–221, 1978.

Roger B Myerson. Multistage games with communication. *Econometrica: Journal of the Econometric Society*, pp. 323–358, 1986.

Binghui Peng and Aviad Rubinstein. Fast swap regret minimization and applications to approximate correlated equilibria. *arXiv preprint arXiv:2310.19647*, 2023.

I. Romanovskii. Reduction of a game with complete memory to a matrix game. *Soviet Mathematics*, 3, 1962.

Finnegan Southey, Michael Bowling, Bryce Larson, Carmelo Piccione, Neil Burch, Darse Billings, and Chris Rayner. Bayes' bluff: Opponent modelling in poker. In *Conference on Uncertainty in Artificial Intelligence (UAI)*, July 2005.

Oskari Tammelin. Solving large imperfect information games using cfr+. *arXiv preprint arXiv:1407.5042*, 2014.

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

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

Brian Hu Zhang and Tuomas Sandholm. Polynomial-time optimal equilibria with a mediator in extensive-form games. In *NeurIPS*, 2022.

Brian Hu Zhang, Gabriele Farina, Andrea Celli, and Tuomas Sandholm. Optimal correlated equilibria in general-sum extensive-form games: Fixed-parameter algorithms, hardness, and two-sided column-generation. *arXiv preprint arXiv:2203.07181*, 2022.

Brian Hu Zhang, Gabriele Farina, and Tuomas Sandholm. Team belief dag: Generalizing the sequence form to team games for fast computation of correlated team max-min equilibria via regret minimization. In *International Conference on Machine Learning (ICML)*, 2023.

Martin Zinkevich. Online convex programming and generalized infinitesimal gradient ascent. In *International Conference on Machine Learning (ICML)*, pp. 928–936, Washington, DC, USA, 2003.

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

A Discussion

Here, we discuss a few points of interest about UTC and linear deviations.

A.1 The Convex Hull of Pure Deviations

In our definitions, we were careful to allow transformations $\phi \in \Phi$ to map from the set of pure strategies, $\mathcal{X}$, to its convex hull $\operatorname{co}\mathcal{X}$, instead of insisting that every pure strategy map onto another pure strategy. One might ask whether this makes a difference in our definitions. For example, if in Definition 2.1 we restrict our attention to $\phi : \mathcal{X} \to \mathcal{X}$, does the definition change? In symbols, for a given set of transformations Φ, is $\Phi \subseteq \operatorname{co}\widehat{\Phi}$ where $\widehat{\Phi} = \{\phi \in \Phi : \phi(\mathcal{X}) \in \mathcal{X}\ \forall \boldsymbol{x} \in \mathcal{X}\}$? For the other sets of deviations mentioned in this paper (external, swap, trigger, and communication), the answer is already known to be positive.

Our equivalence theorem between UTC and linear deviations gives an answer to this question for the set of linear deviations as well. Since the UTC deviations are defined by a decision problem, every mixed UTC deviation is by definition equivalent to a distribution over pure UTC deviations. That is, the vertices of Φ_{UTC} are the pure UTC deviations: they map pure strategies $\mathcal{X}$ to pure strategies. Since $\Phi_{\text{UTC}} = \Phi_{\text{LIN}}$, this proves:

Corollary A.1. *When $\mathcal{X}$ is a sequence-form polytope, the extreme points of Φ_{LIN} are the linear maps $\phi : \mathcal{X} \to \mathcal{X}$,* i.e.*, the linear maps that map pure strategies to pure strategies. Thus, $\Phi_{\text{LIN}} = \widehat{\Phi}_{\text{LIN}}$*

This result is not obvious *a priori*. For example, it fails to generalize to other sets of functions Φ, or to Φ_{LIN} for polytopes $\mathcal{X}$ that are not sequence-form polytopes:

- Other sets of functions Φ: Let $\Phi = \{\phi\}$ consist of a single constant function $\phi : \boldsymbol{x} \mapsto \boldsymbol{x}^*$, where $\boldsymbol{x}^* \in (\operatorname{co}\mathcal{X}) \setminus \mathcal{X}$. Then $\widehat{\Phi}$ is empty, so $\Phi \not\subseteq \operatorname{co}\widehat{\Phi}$.
- Non-sequence-form polytopes: Take $\mathcal{X}$ to be a trapezoid $ABCD$ where AB is the longer of the two bases and consider the linear map ϕ with $\phi(A) = D, \phi(D) = A$, and $\phi(B) = C$. Then ϕ is an extreme point of Φ_{LIN}, but $\phi(C)$ will lie somewhere along segment AB, but at neither endpoint—that is, not at a vertex. Figure 4 has a visual depiction.

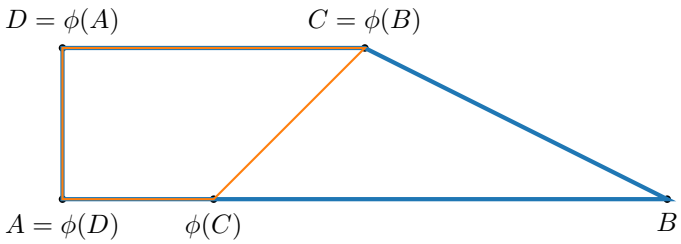


Figure 4: A visual depiction of the argument that Corollary A.1 cannot generalize to all polytopes. The affine map ϕ maps the large blue polygon onto the small orange polygon, and ϕ is a vertex of the set of linear maps from polygon $ABCD$ to itself, yet $\phi(C)$ is not a vertex of $ABCD$.

A.2 Generalization to Arbitrary Pairs of Polytopes

Our main result characterizes the set of linear maps $\phi : \mathcal{X} \to \mathcal{X}$ for sequence-form polytopes $\mathcal{X}$. However, it is actually more general than this: an identical proof works to characterize the set of linear maps $\phi : \mathcal{Y} \to \operatorname{co}\mathcal{X}$ for any (possibly different!) sequence-form polytopes $\mathcal{X}$ and $\mathcal{Y}$. Hence, we have shown:

Theorem A.2. *Let $\mathcal{X}, \mathcal{Y}$ be sequence-form strategy sets. The linear maps $\phi : \mathcal{Y} \to \operatorname{co}\mathcal{X}$ are precisely the functions induced by strategies in the DAG decision problem whose nodes are identified with pairs $(s, \widetilde{s})$, where $s \in \Sigma_{\mathcal{X}} \cup \mathcal{J}_{\mathcal{X}}$ and $\widetilde{s} \in \Sigma_{\mathcal{Y}} \cup \mathcal{J}_{\mathcal{Y}}$, and which behaves analogously to Definition 3.1.*

Although we are mostly concerned with the case $\mathcal{X} = \mathcal{Y}$ in this paper, we state this extension in the hope that it may be of independent interest. We will also use it in the proof of the revelation principle (Theorem A.6).

A.3 Uniqueness of Representation

The statement of Theorem 4.1 discusses Φ_{UTC} and Φ_{LIN} as *sets of functions* $\Phi \subseteq (\operatorname{co} \mathcal{X})^{\mathcal{X}}$. It does *not* imply that for every linear map $\phi : \mathcal{X} \to \operatorname{co} \mathcal{X}$ there is *exactly* one representation of ϕ as a deviator strategy in the UTC decision problem, only that there is *at least* one representation. Indeed, the external deviations (constant functions $\phi : \boldsymbol{x} \mapsto \boldsymbol{x}^*$ for fixed $\boldsymbol{x}^* \in \operatorname{co} \mathcal{X}$ can be represented via a large number of different strategies in the UTC decision problem: the deviator may send any number of queries to the mediator, before eventually deciding to ignore the queries and play according to $\boldsymbol{x}^*$, and such a deviator would still represent the external deviation ϕ.

Similarly, Theorem 4.1 also does not imply that every matrix $\mathbf{A} \in \mathbb{R}^{\Sigma \times \Sigma}$ representing a linear map $\phi_{\mathbf{A}} : \mathcal{X} \to \operatorname{co} \mathcal{X}$ is part of a pair $(\mathbf{A}, \mathbf{B})$ satisfying the system of equations (2). Indeed, the proof of Theorem 4.1 only shows that for every linear $\phi : \mathcal{X} \to \operatorname{co} \mathcal{X}$, there is *at least one* pair $(\mathbf{A}, \mathbf{B})$ satisfying (2) where $\mathbf{A}$ represents ϕ. It is easy to construct matrices $\mathbf{A}$ that represent linear maps, yet cannot satisfy (2), by changing the first row of $\mathbf{A}$ to some other vector $\boldsymbol{c}$ with $\boldsymbol{c}^\top \boldsymbol{x} = 1$ for all $\boldsymbol{x} \in \mathcal{X}$.

A.4 When Untimed and Timed Communication Deviations Coincide

If all players have only one layer of decision nodes, the game is a single-stage *Bayesian game*—in that special case, the communication deviations and linear deviations will coincide[7]. This property was also proven by Fujii (2023), but our framework gives a particularly simple proof via Theorem 4.1: for any UTC deviation in a single-stage game, the deviator makes either no queries or one query to the mediator. A communication deviator can simulate the same function by making the same query (if any), or, if the UTC deviator makes no query, by making an arbitrary query and ignoring the reply. It turns out that the converse is also essentially true:

Theorem A.3. *Consider any decision problem with no nontrivial decision points—that is, the player has at least two legal actions at every decision point. The (timed) communication deviations coincide with the untimed communication deviations (and hence also the linear deviations)* if and only if *every path through the decision problem contains at most one decision point.*

Proof. The *if* direction was shown above and by Fujii (2023), so it suffices to show the *only if* direction. Suppose there are two decision points, A and B, such that B is a child of action a_1 at A. Let a_2 be another action at A, and let b_1 and b_2 be two actions at B. (The game in Figure 5 has such a structure). Consider any deviation ϕ that plays action a_i if it is recommended b_i, for $i \in \{1, 2\}$. It is easy to construct untimed deviations with this behavior, but timed deviations cannot have this behavior, because a timed deviation cannot know the recommendation at B while still at decision point A. □

A.5 Relation between Our Representation and That of Farina & Pipis (2023)

Our paper and the paper of Farina & Pipis (2023) both take similar approaches to minimizing Φ_{LIN}-regret: both papers use the framework of Gordon et al. (2008) to reduce the problem to minimizing external regret over the set of linear maps, and then derive a system of constraints for the set of matrices $\mathbf{A} \in \mathbb{R}^{\Sigma \times \Sigma}$ that represent linear maps. The representations are, however, significantly different:

- The representation of Farina & Pipis (2023) cannot be expressed in scaled extensions. As such, that paper was forced to resort to less efficient regret minimization techniques. This difference is what allows us to improve upon their results.
- As a technical note, the representation of Farina & Pipis (2023) will always result in a matrix $\mathbf{A} \in \mathbb{R}^{\Sigma \times \Sigma}$ where the columns of $\mathbf{A}$ corresponding to nonterminal sequences are filled with zeros. While this is without loss of generality due to the constraints defining the sequence form, it sometimes results in intuitively-strange representations: for example, their representation does not represent the identity map $\text{Id} : \mathcal{X} \to \mathcal{X}$ as the identity matrix $\mathbf{I} \in \mathbb{R}^{\Sigma \times \Sigma}$, whereas our representation will.

[7] Here, by two sets of deviations *coinciding*, we mean that the same set of deviation functions $\phi : \mathcal{X} \to \operatorname{co} \mathcal{X}$ is available to both deviators.

- While our representation generalizes to arbitrary pairs of sequence-form polytopes according to Appendix A.2, theirs generalizes even further, to functions $\phi : \mathcal{Y} \to \mathcal{X}$ as long as $\mathcal{Y}$ is sequence form and $\mathcal{X}$ has *some* small set of linear constraints (not necessarily sequence form) describing it. We likely cannot hope for our representation to generalize as far: our proof of equivalence relies fundamentally on both input and output being sequence-form strategy sets.

A.6 Untimed Communication Equilibria

The UTC deviations, like all sets of deviations, give rise to a notion of equilibrium. We define:

Definition A.4. In an extensive-form game, an *untimed private communication equilibrium* is a correlated profile that is a (Φ_i)-equilibrium where Φ_i is player i's set of UTC deviations.

We add the word "private" here in the name to emphasize the fact that the mediator must have a separate interaction with each player—that is, the mediator cannot use its interactions with one player to inform how it gives recommendations to another player. This is enforced by the fact that the equilibrium is a correlated profile. See Footnote 1 regarding why this distinction is important.

Defining untimed communication equilibrium without such a privacy restriction seems to be a subtle task, and is orthogonal to and beyond the scope of the present work. However, we will make a few informal comments here. Untimed communication equilibria (without the privacy constraint) are difficult to define in a way that does not quickly collapse to the regular notion of communication equilibrium. In games with three or more players, the mediator is always guaranteed that two of the players have not deviated, and those two players will have messages synchronized with the game clock. Therefore, under reasonable assumptions on how often each player makes moves, the mediator will immediately know if the deviating player is sending out-of-order messages, and this concept would reduce immediately to the regular communication equilibrium. It is entirely unclear how to define a notion of untimed (non-private) communication equilibrium that does not exhibit such a collapse.

In two-player games, it is possible that there is a reasonable way to define untimed communication equilibria. The above collapse does not apply, because the mediator will not know which player is the one sending out-of-timing messages. However, this definition would still be rather subtle—for example, when do the out-of-order messages arrive to the mediator, relative to the *other player's* messages? We leave these issues to future work.

A.7 Revelation Principle for Untimed Private Commnunication Equilibrium

All the other notions of equilibrium involving a mediator, discussed in Section 3, obey a *revelation principle*, which we now discuss using the example of normal-form correlated equilibrium. The original definition of Aumann (1974) did not initially refer to correlated profiles; instead, the definition posited an arbitrary joint distribution of correlated signals. In this section, we break down the notions of equilibrium that we have defined so far, and reconstruct them from the perspective of this arbitrary set of signals, and show that the resulting notions are equivalent for our notion of untimed private communication equilibrium.

We start with NFCE as an illustrative example. Let $\pi \in \Delta(\mathcal{S}_1 \times \cdots \times \mathcal{S}_n)$, where $\mathcal{S}_i$ is an arbitrary set of signals for player i. The mediator samples a joint signal $(s_1, \dots, s_n) \sim \pi$, and then each player privately observes its own signal s_i and selects (possibly at random) a strategy $\boldsymbol{x} \in \mathcal{X}$. An NFCE is then a tuple $(\pi, \phi_1, \dots, \phi_n)$, where $\phi_i : \mathcal{S}_i \to \operatorname{co} \mathcal{X}$ is the function by which player i selects its strategy given a signal, such that each player's choice of ϕ_i maximizes that player's utility given π and the other ϕ_is, that is,

$$\mathop{\mathbb{E}}_{(s_1,\dots,s_n)\sim\pi}[u_i(\phi_i'(s_i), \phi_{-i}(s_{-i}) - u_i(\phi_i(s_i), \phi_{-i}(s_{-i})] \le 0$$

for every player i and other possible function $\phi_i' : \mathcal{S}_i \to \operatorname{co} \mathcal{X}$. An NFCE is *direct* if $\mathcal{S}_i = \mathcal{X}_i$ and $\phi_i : \mathcal{X}_i \to \operatorname{co} \mathcal{X}_i$ is the identity function. The *revelation principle* states that every NFCE is outcome-equivalent[8] to a direct equilibrium.

[8] By *outcome-equivalent*, we mean that the distribution over terminal nodes in the extensive-form game is the same in both equilibria.

To generalize this to extensive-form and communication equilibria (timed and untimed), we follow the approach of Myerson (1986); Forges (1986). In their approach, a mediator of player i is a map $M_i : \mathcal{S}_i^{\leq H} \to \mathcal{S}_i$ (where $\mathcal{S}_i^{\leq H}$ is the set of all sequences over $\mathcal{S}_i$ of length $\leq H$, and H is some large but finite number (at least the depth of player i's decision problem.) that determines what message the mediator sends in reply to a player whose message history with the mediator is a finite sequence $\boldsymbol{s} = \{s_i\}$. We will assume that $\mathcal{S}_i$ at least is expressive enough to send an empty message, a decision point, or an observation point: $\mathcal{S}_i \supseteq \mathcal{J}_i \sqcup \Sigma_i \sqcup \{\bot\}$. The three notions of extensive-form correlated equilibrium, private communication equilibrium, and untimed private communication equilibrium will differ in how the player interacts with the mediator. We will describe player i's interactions by a set of functions $\Phi_i \subseteq (\operatorname{co} \mathcal{X}_i)^{\mathcal{M}_i}$ where $\mathcal{M}_i$ is a set of mediators: each function $\phi_i \in \Phi_i$ represents the player i choosing how it interacts with the mediator and how it uses those interactions to inform its choices of action. Then, as before, an equilibrium is a tuple $(\pi, \phi_1, \dots, \phi_n)$ where $\phi \in \Delta(\mathcal{M}_1 \times \dots \mathcal{M}_n)$ is a distribution over mediators and no player i can profit by switching to a different $\phi_i' \in \Phi_i$. The three notions above then differ in the choice of set Φ_i:

- Extensive-form correlated equilibria are equilibria where Φ_i is the set of interactions in which the player, upon reaching a decision point j, must send that decision point to the mediator.
- Private communication equilibria are equilibria where Φ_i is the set of interactions in which the player, upon reaching a decision point j, must send a single message (which may or may not be the decision point j) to the mediator.
- Untimed private communication equilibria are equilibria where Φ_i is the set of interactions in which the player, upon reaching a decision point j, may send any number of messages to the mediator.

The *direct mediator* $M_i^{\boldsymbol{x_i}}$ for a pure strategy $\boldsymbol{x}_i \in \mathcal{X}_i$ is the mediator who acts by sending the recommendation a at infoset j if and only if the message history matches the $\varnothing_i \to j$ path, otherwise $\bot$. Formally, $M_i^{\boldsymbol{x}_i}(\boldsymbol{s}) = a(\boldsymbol{x}_i, j)$ if $\boldsymbol{s} = (j^{(1)}, a^{(1)}, j^{(2)}, \dots, j)$ is the path to j in player i's decision tree, and $\bot$ otherwise. We write $\mathcal{M}_i^* := \{M_i^{\boldsymbol{x}_i} \mid \boldsymbol{x}_i \in \mathcal{X}_i\}$ for the set of direct mediators on $\mathcal{X}_i$. Notice that, for direct M_i, the sets of interactions valid for each of the three equilibrium notions reduces to the sets of deviations defined in Section 3. Analogous to the NFCE case, an equilibrium $(\pi, \phi_1^*, \dots, \phi_n^*)$ (in any of the previous three notions) is called *direct* if π is a distribution over direct mediators, and ϕ_i^* is the map $M_i^{\boldsymbol{x}_i} \mapsto \boldsymbol{x}_i$ (which is the analogy of the identity map). We are now ready to state the revelation principle for these notions.

Theorem A.5 (Revelation principle: for EFCE, proven by von Stengel & Forges (2008); for communication equilibrium, proven by Myerson (1986); Forges (1986) and refined by Zhang & Sandholm (2022))**.** *For EFCE and (private) communication equilibrium, every equilibrium is outcome-equivalent to a direct equilibrium.*

Our main result in this section is that the same holds for untimed private communication equilibrium:

Theorem A.6 (Revelation principle for untimed private communication equilibrium)**.** *Every untimed private communication equilibrium is outcome-equivalent to a direct untimed private communication equilibrium.*

Proof. Let $(\pi, \phi_1, \dots, \phi_n)$ be some (possibly indirect) equilibrium. Observe that we can view the mediator as holding a strategy $\boldsymbol{y} \in \mathcal{Y}_i$, where $\mathcal{Y}_i$ is the decision problem whose nodes correspond to sequences $\boldsymbol{s} \in \mathcal{S}_i^{\leq H}$, *i.e.*, to message histories. Notice that, by construction of the message set $\mathcal{S}_i$, $\mathcal{Y}$ contains a copy of each $\mathcal{X}_i$ within it, and that direct mediators $M_i^{\boldsymbol{x}_i}$ constrain themselves to states within this copy of $\mathcal{X}_i$ by terminating the interaction (sending $\bot$ forever) if the history of communication fails to match a state in player i's decision problem. We will use this fact later.

By Theorem A.2, each player's strategy set Φ_i is the set of linear maps $\mathcal{Y}_i \to \operatorname{co} \mathcal{X}_i$. Now, consider the direct profile $(\pi^*, \phi_i^*, \dots, \phi_n^*)$ where $\pi^* \in \Delta(\mathcal{M}_1^* \times \dots \times \mathcal{M}_n^*)$ is given by sampling $(M_1, \dots, M_n) \sim \pi$, sampling $\boldsymbol{x}_i \in \mathcal{X}_i$ from any distribution whose expectation is $\phi_i(M_i)$ for every player i, and finally outputting $(M_1^{\boldsymbol{x}_1}, \dots, M_n^{\boldsymbol{x}_n})$. Clearly, this profile is outcome-equivalent to the original profile, so it only remains to show that it is also an equilibrium. Consider any deviation ϕ_i' of player i from the direct equilibrium.

We proceed by contrapositive. Suppose that $(\pi^*, \phi_i^*, \dots, \phi_n^*)$ is not an equilibrium: player i has profitable deviation ϕ_i'. Since a direct mediator is constrained, as above, to act within player i's decision problem, ϕ_i' can be expressed as a UTC deviation $\phi_i' : \mathcal{X}_i \to \operatorname{co} \mathcal{X}_i$. Since all UTC deviations are linear, ϕ_i' is itself linear, and

can also be extended to a function $\phi_i' : \operatorname{co} \mathcal{X}_i \to \operatorname{co} \mathcal{X}_i$. Now let $\psi_i : \mathcal{Y}_i \to \operatorname{co} \mathcal{X}_i$ be given by $\psi_i = \phi_i' \circ \phi_i$, and observe that, since the composition of linear functions is linear, ψ_i is a linear map, that is, $\psi_i \in \Phi_i$. Moreover, by construction, the profiles (π, ψ_i, ϕ_{-i}) and $(\pi^*, \phi_i', \phi_{-i}^*)$ must induce the same outcome distributions—and therefore, ψ_i is a profitable deviation against the original equilibrium $(\pi, \phi_1, \dots, \phi_n)$. □

This result justifies the definitions of equilibrium we have been using throughout the paper before reaching this point. We remark that, although the proof is usually not difficult, the revelation principle is not a given or automatic fact that can be assumed without proof: there are other settings where it fails, such as when the deviator's set of allowable messages depends on its true type in a nontrivial manner (*e.g.*, Forges & Koessler, 2005; Kephart & Conitzer, 2021).

B Previous Φ-Regret Algorithms

Citation	**Deviation set** (Φ)	**Regret bound**	**Complexity**	**CFR?**
Zinkevich et al. (2007)	External	$O(d\sqrt{T})$	$O(d)$	Yes
Farina et al. (2022b)	External	$O(\sqrt{dT})$	$O(d)$	No
Farina et al. (2022a)	Trigger	$O(d\sqrt{T})$	FP	Yes
Fujii (2023) ‡	Communication	$\widetilde{O}(\sqrt{dT})$	FP	No
Farina & Pipis (2023)	Linear	$O(d^2\sqrt{T})$	QP	No
This paper	Linear	$O(d^2\sqrt{T})$	FP	Yes
Peng & Rubinstein (2023) Dagan et al. (2023)	Swap	$T \cdot \widetilde{O}\left(\frac{\log d}{\log T}\right)$	$O(d \log T)$	Yes

Table 4: Comparison of Φ-regret minimizing algorithms for extensive-form games. **Complexity** is per-iteration. "QP" and "FP" denote solving a quadratic program and a fixed-point problem, respectively. **CFR?** denotes whether the algorithm is based on the CFR framework, which is important in practice because, as stated in the body, CFR-based methods are the best practical regret minimizers. ‡: The algorithm and analysis of Fujii (2023) only applies to single-step Bayesian games, not general extensive-form games.

Table 4 summarizes the various Φ-regret algorithms known for extensive-form games. As suggested by the table, the main improvements of our algorithm compared to that of Farina & Pipis (2023) in the same setting are (1) the faster per-iteration complexity (fixed-point computation versus quadratic program), and (2) the enabling of the use of the CFR framework, which leads to faster practical performance. The remarkable result of Peng & Rubinstein (2023) and Dagan et al. (2023) is very recent and in parallel with our work, and shows that there exists a PTAS for NFCE (with runtime roughly $d^{\widetilde{O}(1/\varepsilon)}$). It remains an open problem whether there exists a swap-regret minimizing algorithm whose regret is $\text{poly}(d) \cdot T^c$ for some $c < 1$, which would imply an FPTAS for NFCE in extensive-form games.

C Another Example

In this section, we provide another example of untimed communication deviations, especially as they differ from (timed) communication deviations. Consider the game in Figure 5. Consider the correlated profile that mixes uniformly between the pure profiles ($\mathbf{a_1}$, $\mathbf{b_1}$, $\mathbf{c_1}$) and ($\mathbf{a_1}$, $\mathbf{b_2}$, $\mathbf{c_2}$). This is a communication equilibrium: ▲ cannot profitably deviate, because its utility in the **B** subgame is always 0, and if it chooses to disobey the recommendation $\mathbf{a_1}$ its expected utility will be also 0, because it cannot ask for another recommendation before choosing what action to play. However, ▲ has the following profitable UTC deviation: ask for the recommendation at **B** *before* deciding which action to play at **A**. If the recommendation is $\mathbf{b_1}$, play $\mathbf{a_2}$; if the recommendation is $\mathbf{b_2}$, play $\mathbf{a_3}$.

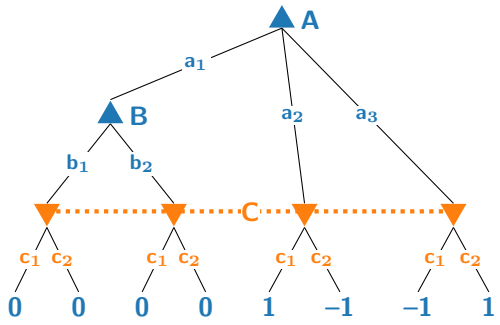


Figure 5: Another example. The notation is shared with Figure 1. In this example, ▲'s strategy set is equivalent to a simplex, so the linear deviations coincide with its swap deviations. As such, we will not bother to depict the UTC decision problem or matrices.

Notice that, in this example, ▲'s decision problem is essentially that of a normal-form game; therefore, its linear deviations coincide with its swap deviations. However, due to the timing restriction on communication deviations, the communication deviations are more restricted than the swap deviations.

This example also shows that the untimed private communication equilibria (see Appendix A.6) are not outcome-equivalent to the timed private communication equilibria: in this game, every correlated profile is a distribution over terminal nodes (outcomes), so the fact that there exists a private communication equilibrium with a profitable UTC deviation is enough to disprove outcome equivalence.

D Proof of Theorem 4.1

Theorem 4.1. *The UTC deviations are precisely the linear deviations. That is,* $\Phi_{\text{UTC}} = \Phi_{\text{Lin}}$.

We start with a lemma.

Lemma D.1. *Let* $f : \mathcal{X} \to \mathbb{R}_{\geq 0}$ *be a linear map, where* $\mathcal{X}$ *is a sequence-form strategy space. Then there exists a unique vector* $\boldsymbol{c}$ *such that:*

1. $f(\boldsymbol{x}) = \boldsymbol{c}^\top \boldsymbol{x}$ *for all* $\boldsymbol{x} \in \mathcal{X}$,
2. $\boldsymbol{c}$ *has all nonnegative entries, and*
3. *for every decision point* I, *there is at least one action* a *such that* $\boldsymbol{c}(ja) = 0$.

Proof. Let $f(\boldsymbol{x}) = \boldsymbol{c}^\top \boldsymbol{x}$, where $\boldsymbol{c}$ is currently arbitrary (*i.e.*, it may not satisfy (2) and (3)). Then, for each decision point j in bottom-up order, let $\boldsymbol{c}^*(j) := \min_a \boldsymbol{c}(ja)$. Subtract $\boldsymbol{c}^*(j)$ from $\boldsymbol{c}(ja)$ for every action a, and add $\boldsymbol{c}^*(j)$ to $\boldsymbol{c}(p_j)$. Since $\boldsymbol{x}$ satisfies the constraint $\boldsymbol{x}(p_j) = \sum_a x(ja)$, this does not change the validity of $\boldsymbol{c}$, and by the end of the algorithm, (2) and (3) will be satisfied except possibly that $\boldsymbol{c}(\varnothing) \geq 0$. To see that $\boldsymbol{c}(\varnothing) \geq 0$, let $\boldsymbol{x}$ be the pure strategy that plays the zeroing action a specified by (3) at every decision point. Then, by construction, $\boldsymbol{c}^\top \boldsymbol{x} = \boldsymbol{c}(\varnothing) \geq 0$. To see that $\boldsymbol{c}$ is unique, note that there was no choice at any step in the above process: the transformation performed at each decision point is the only way to satisfy conditions (2) and (3) without changing the linear map. □

Now let $\mathbf{A}$ represent a linear map $\mathcal{X} \to \mathcal{X}$, where the rows of $\mathbf{A}$ are represented according to the above lemma. That is, $\mathbf{A}$ has all nonnegative entries, and moreover for any $\widetilde{j} \in \mathcal{J}$ and $\sigma \in \Sigma$, we have $\mathbf{A}(\sigma, \widetilde{j}a) = 0$ for some action a. It remains only to show:

Lemma D.2. *There exists a matrix* $\mathbf{B}$ *such that* $(\mathbf{A}, \mathbf{B})$ *satisfies all constraints in the constraint system* (2).

Proof. $\mathbf{A}(\varnothing, \varnothing) = 1$ and $\mathbf{A}(\varnothing, \widetilde{\sigma}) = 0$ for $\widetilde{\sigma} \neq \varnothing$ follow from the fact that $(\mathbf{A}\boldsymbol{x})(\varnothing) = 1$ for all $\boldsymbol{x}$, that is $\mathbf{A}(\varnothing, \cdot) : \mathcal{X} \to [0, 1]$ is the identically-1 function, which by Lemma D.1 has the above form. We are thus left

with the main constraint,

$$\mathbf{A}(p_j, \widetilde{\sigma}) + \mathbf{B}(j, p_{\widetilde{\sigma}}) = \sum_{a \in A_j} \mathbf{A}(ja, \widetilde{\sigma}) + \sum_{\widetilde{\jmath} \in C_{\widetilde{\sigma}}} \mathbf{B}(j, \widetilde{\jmath}) \tag{3}$$

for every $(j, \widetilde{\sigma}) \in \mathcal{J} \times \Sigma$.

Define $\mathbf{B}$ and another matrix $\widetilde{\mathbf{B}} \in \mathbb{R}^{\mathcal{J} \times \Sigma}$ as follows:

$$\widetilde{\mathbf{B}}(j, \widetilde{\sigma}) = \sum_{a \in A_j} \mathbf{A}(ja, \widetilde{\sigma}) + \sum_{\widetilde{\jmath} \in C_{\widetilde{\sigma}}} \mathbf{B}(j, \widetilde{\jmath}) - \mathbf{A}(p_j, \widetilde{\sigma})$$
$$\mathbf{B}(j, \widetilde{\jmath}) = \min_{a \in A_{\widetilde{\jmath}}} \widetilde{\mathbf{B}}(j, \widetilde{\jmath}a)$$

To see that $\mathbf{B}$ satisfies all the constraints (3), let $\boldsymbol{x}$ be any fully-mixed strategy, and I be any decision point. Then:

$$\begin{aligned}
0 &= \sum_{a \in A_j} (\mathbf{A}\boldsymbol{x})(ja) - (\mathbf{A}\boldsymbol{x})(p_j) \\
&= \sum_{\widetilde{\sigma} \in \Sigma} \boldsymbol{x}(\widetilde{\sigma}) \left(\sum_{a \in A_j} \mathbf{A}(ja, \widetilde{\sigma}) - \mathbf{A}(p_j, \widetilde{\sigma}) \right) \\
&= \sum_{\widetilde{\sigma} \in \Sigma} \boldsymbol{x}(\widetilde{\sigma}) \left(\widetilde{\mathbf{B}}(j, \widetilde{\sigma}) - \sum_{\widetilde{\jmath} \in C_{\widetilde{\sigma}}} \mathbf{B}(j, \widetilde{\jmath}) \right) \\
&= \sum_{\widetilde{\jmath}} \left(\sum_{a \in A_{\widetilde{\jmath}}} \boldsymbol{x}(\widetilde{\jmath}a) \widetilde{\mathbf{B}}(j, \widetilde{\jmath}a) - \boldsymbol{x}(p_{\widetilde{\jmath}}) \mathbf{B}(j, \widetilde{\jmath}) \right) \\
&\geq \sum_{\widetilde{\jmath}} \mathbf{B}(j, \widetilde{\jmath}) \left(\sum_{a \in A_{\widetilde{\jmath}}} \boldsymbol{x}(\widetilde{\jmath}a) - \boldsymbol{x}(p_{\widetilde{\jmath}}) \right) = 0
\end{aligned}$$

Thus, the inequality must in fact be an equality, and since all its terms are nonnegative, we thus have $\widetilde{\mathbf{B}}(j, \widetilde{\jmath}a) = \mathbf{B}(j, \widetilde{\jmath})$ for all $a \in A_{\widetilde{\jmath}}$, so the constraints (3) are satisfied by definition of $\widetilde{\mathbf{B}}$.

To see that $\mathbf{B} \geq 0$, suppose not. Let $(j, \widetilde{\jmath})$ be a last (*i.e.*, farthest from the root, with respect to the ordering of the DAG) pair for which $\mathbf{B}(j, \widetilde{\jmath}) < 0$. Then, for any action $\widetilde{a} \in A_{\widetilde{\jmath}}$, we have

$$\mathbf{A}(p_j, \widetilde{\jmath}\widetilde{a}) + \mathbf{B}(j, \widetilde{\jmath}) = \sum_{a \in A_j} \mathbf{A}(ja, \widetilde{\jmath}\widetilde{a}) + \sum_{\widetilde{\jmath}' \in C_{\widetilde{\jmath}\widetilde{a}}} \mathbf{B}(j, \widetilde{\jmath}') \geq 0$$

where the inequality is because $(j, \widetilde{\jmath})$ is farthest from the root so all the terms on the right-hand side are nonnegative. Therefore, $\mathbf{A}(p_j, \widetilde{\jmath}\widetilde{a}) > 0$. But this should hold for every action $\widetilde{a}$, contradicting the construction of $\mathbf{A}$, which includes the condition that there must exist a $\widetilde{a}$ for which $\mathbf{A}(p_j, \widetilde{\jmath}\widetilde{a}) = 0$. □

E Additional Plots

Below we present a variant of the plots in Figure 3, in which *time* (and not iterations) is reported on the x-axis.

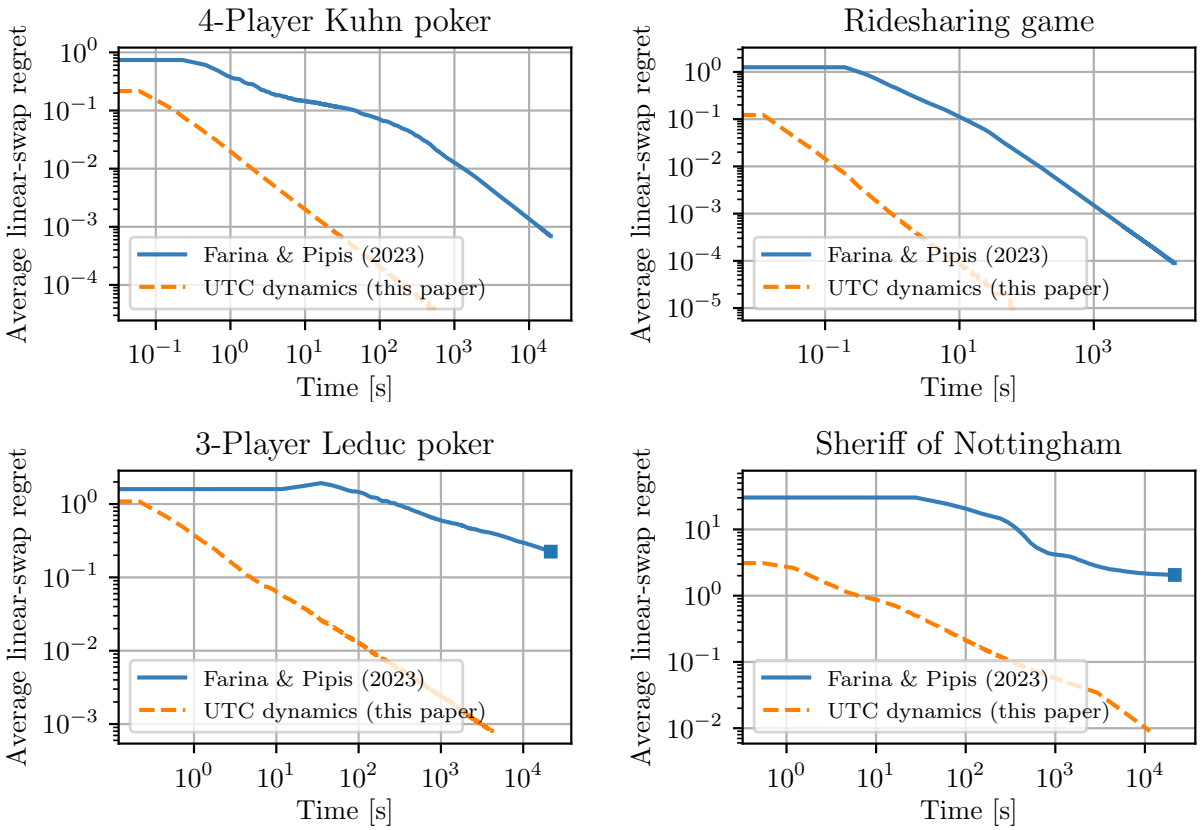


Figure 6: Experimental comparison between our dynamics and those of Farina & Pipis (2023) for approximating a linear correlated equilibrium in extensive-form games. Each algorithm was run for a maximum of 100,000 iterations or 6 hours, whichever was hit first. Runs that were terminated due to the time limit are marked with a square ■. Compared to Figure 3, the plots in this figure have *time* on the x-axis.