A lower bound on swap regret in extensive-form games

May 20, 2024

Abstract

Recent simultaneous breakthroughs by Peng and Rubinstein [2024] and Dagan et al. [2024] have demonstrated the existence of $m^{\widetilde{\mathcal{O}}(1/\epsilon)}$-round algorithm for swap regret minimization in extensive-form games, where m is the number of nodes and ϵ is the desired (average) swap regret. However, the question of whether a $\mathsf{poly}(m, 1/\epsilon)$-round algorithm could exist remained open. In this paper, we show a lower bound that precludes the existence of such an algorithm. In particular, we show that achieving average swap regret ϵ in general extensive-form games requires at least $\exp\left(\min\left\{\widetilde{\Theta}(m^{1/12}), \Theta(\epsilon^{-1/5})\right\}\right)$ rounds.

1 Introduction

No-regret learning is a popular framework for modeling situations in which an agent faces an arbitrary, possibly adversarial environment. The agent seeks to minimize its *regret*, which is the difference between the utility it has earned and the maximum utility it could have earned by changing its strategy according to some *strategy transformation function*. The more strategy transformations are allowed, the tighter the resulting notion of regret. In sequential (extensive-form) imperfect-information games, especially adversarial games, algorithms based on no-regret learning have been pivotal in leading to superhuman performance in games ranging from poker [Moravčík et al., 2017, Brown and Sandholm, 2018, 2019] to *Diplomacy* [Bakhtin et al., 2022].

In games, there is a well-studied and tight connection between no-regret learning in games and solution concepts involving *correlation*. In particular, if all players in a game play according to a no-regret learning algorithm that minimizes a certain notion of regret, their average strategy profile will converge to a notion of correlated equilibrium that corresponds to the class of strategy transformations corresponding to that regret notion. Notable correlated solution concepts, and their corresponding sets of deviations (notions of regret), can be found in Table 1.

The tightest notion that can be defined in this manner is known as the *normal-form correlated equilibrium* (NFCE) [Aumann, 1974], which corresponds to measuring regret against the set of *all possible strategy transformations*—known as *swap regret* [Blum and Mansour, 2007]. It has long been believed that efficiently

Set of deviations	**Equilibrium concept**	**Best known no-regret guarantee**
Constant functions ("External regret")	Normal-form coarse correlated (NFCCE)	m/ϵ^2 [Farina et al., 2022]
"Trigger" functions Linear functions	Extensive-form correlated (EFCE) Linear-swap correlated (LCE)	m^2/ϵ^2 [Farina et al., 2022], [Farina and Pipis, 2023]
All functions ("Swap regret")	Normal-form correlated (NFCE)	$m^{\widetilde{\mathcal{O}}(1/\epsilon)}$ [Peng and Rubinstein, 2024], [Dagan et al., 2024]

Table 1: Some examples of notions of regret and corresponding notions of correlated equilibrium for extensive-form games, in increasing order of tightness. The "best no-regret guarantee" is the minimum number of iterations T after which an agent can guarantee (average) regret ϵ in that notion. m is the size of the game (number of nodes). All algorithms listed have guaranteed $\mathsf{poly}(m, 1/\epsilon)$ per-iteration time complexity.

computing an NFCE or minimizing swap regret is impossible. In fact, the believed hardness of computing NFCE has motivated the development of many more relaxed notions of equilibrium with nicer computational properties, including the extensive-form correlated equilibrium [von Stengel and Forges, 2008], linear-swap correlated equilibrium [Farina and Pipis, 2023], and behavioral correlated equilibrium [Morrill et al., 2021a,b]. Our focus in this paper, however, will be on NFCE itself.

Little was known about the feasibility of swap regret minimization (and, similarly, about computing NFCE) until a recent simultaneous breakthrough due to Peng and Rubinstein [2024] and Dagan et al. [2024] which implies that, for extensive-form games with m nodes, there is a no-regret algorithm that achieves (average) regret ϵ after $m^{\widetilde{\mathcal{O}}(1/\epsilon)}$ rounds. For extensive-form games, this was the first time that a polynomial-time swap regret minimization algorithm (even for constant ϵ) had been achieved, and resulted the first PTAS for NFCE. Both papers provide lower bounds, but these lower bounds are only for normal-form (*i.e.*, single-step) games, and do not preclude the existence of efficient swap-regret minimization algorithms for extensive-form games. Thus, there is an exponential gap (in terms of the dependence on ϵ) between the best regret bounds for swap regret and the best regret bounds for weaker notions of regret.

In this paper, we show that this exponential gap cannot be closed. In particular, we show the following result:

Theorem 1.1 (Main theorem, informal)**.** *There is no swap regret minimization algorithm for general extensive-form games that requires fewer than* $\exp\Big(\min\Big\{\widetilde{\Theta}(m^{1/12}), \Theta(\epsilon^{-1/5})\Big\}\Big)$ *rounds.*

Our result implies an exponential gap between swap regret and weaker notions of regret in the realm of extensive-form games. In particular, our result precludes the existence of $\mathsf{poly}(m, 1/\epsilon)$-time swap regret minimization algorithms.

2 Preliminaries

No-regret learning and Φ-regret. In *no-regret learning*, a learner with access to a *strategy set* $\mathcal{X} \subset \mathbb{R}^m$ faces an adversarial environment across many rounds. In each round $t = 1, \dots, T$, the learner outputs a distribution $\pi^t \in \Delta(\mathcal{X})$ and then the environment outputs an affine *utility function* $u^t : \mathcal{X} \to [-1, 1]$ which we will henceforth write for convenience as a vector $\boldsymbol{u}^t \in \mathbb{R}^n$. The utility function u^t may depend on all the past distributions $\pi^1, \dots, \pi^t$ selected by the learner. The learner then gets utility $\mathbb{E}_{\boldsymbol{x}\sim\pi^t} \langle \boldsymbol{u}^t, \boldsymbol{x}\rangle$.

In this paper, the notion of interest is *swap regret*. Intuitively, a learner has low swap regret if it could not have improved its utility by transforming its strategy according to any function $\phi : \mathcal{X} \to \mathcal{X}$. More formally, define

$$V(\phi) := \frac{1}{T}\sum_{t=1}^{T} \mathop{\mathbb{E}}_{\boldsymbol{x}\sim\pi^t} \left\langle \boldsymbol{u}^t, \phi(\boldsymbol{x})\right\rangle .$$

Thus in particular the total utility experienced by the learner is $V(\mathrm{Id})$, where $\mathrm{Id} : \mathcal{X} \to \mathcal{X}$ is the identity function. After T rounds, the *(average) swap-regret* is

$$\textsc{SwapRegret}(T) := \max_{\phi:\mathcal{X}\to\mathcal{X}} V(\phi) - V(\mathrm{Id}) = \max_{\phi:\mathcal{X}\to\mathcal{X}} \frac{1}{T}\sum_{t=1}^{T} \mathop{\mathbb{E}}_{\boldsymbol{x}\sim\pi^t} \left\langle \boldsymbol{u}^t, \phi(\boldsymbol{x}) - \boldsymbol{x}\right\rangle .$$

The goal of the learner is then to achieve small swap regret after a small number of rounds: for example, one may hope to achieve swap regret ϵ after $T = \mathsf{poly}(m, 1/\epsilon)$ rounds.

Other notions of regret, such as those mentioned in the introduction, can be defined by restricting the set of deviations to a set $\Phi \subset \mathcal{X}^{\mathcal{X}}$. Since for this paper we are only interested in swap regret, we will not explore this connection further.

Tree-form decision problems and extensive-form games. Tree-form decision problems describe *sequential* interactions between the player and the environment. In a tree-form decision problem, there is a rooted tree $\mathcal{T}$ of *nodes*. There are two types of nodes: *decision points*, at which the player selects an action, and *observation points*, at which the environment selects an observation. At each decision point j, actions are identified with outgoing edges, and we use A_j to denote this set. Leaves of $\mathcal{T}$ are called *terminal nodes*. We will use m to denote the number of terminal nodes.

A *pure strategy* is a choice of one action at each decision point. The *tree-form* or *realization-form representation* of the pure strategy is the vector $\boldsymbol{x} \in \{0,1\}^m$ for which $\boldsymbol{x}[z] = 1$ if and only if the player plays *all* actions on the path from the root node to terminal node z. The set of tree-form pure strategies will be denoted $\mathcal{X} \subset \{0,1\}^m$. Different pure strategies may have the same tree-form representation, but for our purposes we will only require the tree-form representation of strategies, and therefore we will not distinguish between strategies with the same tree-form representation. A *mixed strategy* $\pi \in \Delta(\mathcal{X})$ is a distribution over pure strategies.

Tree-form decision problems naturally model the decision problems faced by players in an *extensive-form game*. For our purposes, a (perfect-recall) extensive-form game with n players is defined by n tree-form strategy sets $\mathcal{X}_1, \dots, \mathcal{X}_n$, and one n-linear *utility function* $u_i : \mathcal{X}_1 \times \cdots \times \mathcal{X}_n \to [-1, 1]$ for each player $i \in [n]$, which defines the utility of player i when each player $j \in [n]$ plays strategy $\boldsymbol{x}_j \in \mathcal{X}_j$.

The solution concepts of interest is the *(normal-form) correlated equilibrium* (NFCE) [Aumann, 1974]. An *ϵ-NFCE* is a distribution $\pi \in \Delta(\mathcal{X}_1 \times \cdots \times \mathcal{X}_n$ with the property that no player can profit by applying any function $\phi : \mathcal{X}_i \to \mathcal{X}_i$ to their strategy. That is, π is an ϵ-NFCE if

$$\mathop{\mathbb{E}}_{(\boldsymbol{x}_1,\dots,\boldsymbol{x}_n)\sim\pi} \left[u_i(\phi(\boldsymbol{x}_i), \boldsymbol{x}_{-i}) - u_i(\boldsymbol{x}_i, \boldsymbol{x}_{-i}^t)\right] \le \epsilon$$

for all players i and functions $\phi : \mathcal{X}_i \to \mathcal{X}_i$.

suppose we have an n-player game played repeatedly over T rounds, and on each round $t \in [T]$ let $\pi_i^t \in \Delta(\mathcal{X}_i)$ be the mixed strategy played by player i.

Proposition 2.1. *Suppose that each player plays according to a no-swap-regret learning algorithm using utility maps $u_i^t : \mathcal{X}_i \to [-1, 1]$ given by*

$$u_i^t(\boldsymbol{x}_i) = \mathop{\mathbb{E}}_{\boldsymbol{x}_{-i}\sim\pi_{-i}^t} u_i(\boldsymbol{x}_i, \boldsymbol{x}_{-i}^t).$$

Let $\pi^t \in \Delta(\mathcal{X}_1 \times \cdots \times \mathcal{X}_n)$ be the product distribution whose marginal on $\mathcal{X}_i$ is π_i^t. If the swap-regret of every player i is bounded by ϵ, then the average profile $\pi := \frac{1}{T}\sum_{i=1}^T \pi^t$ is an ϵ-NFCE.

The proof follows immediately by comparing the definition of swap regret and the definition of NFCE.

3 Swap Regret Minimization Algorithms

We now review previously-known results about no-swap-regret learning algorithms.

Simplices (Normal-form games). For our purposes, a *normal-form game* is a game in which every player's decision problem consists of a single decision point with m actions, that is, $\mathcal{X} = \{\boldsymbol{e}_1, \dots, \boldsymbol{e}_m\} \subset \{0,1\}^m$ where $\boldsymbol{e}_k$ is the kth standard basis vector in $\mathbb{R}^m$. We will abuse language slightly and refer to $\mathcal{X}$ as the m-simplex, even though it is actually the convex hull of $\mathcal{X}$ that is the m-simplex. Blum and Mansour [2007] showed that efficient algorithms exist for minimizing swap regret over the simplex:

Theorem 3.1 (Blum and Mansour, 2007)**.** *There exists a no-regret learning algorithm for simplices that achieves swap regret ϵ after $T = \widetilde{\mathcal{O}}(m/\epsilon^2)$ rounds.*

One may wonder whether this is optimal, *e.g.*, whether it is possible to achieve a logarithmic dependence on m. Recent simultaneous breakthrough work by Dagan et al. [2024] and Peng and Rubinstein [2024] has essentially completely answered this question for normal-form games:

Theorem 3.2 (Dagan et al., 2024, Peng and Rubinstein, 2024, upper bound). *There exists a no-regret learning algorithm for simplices that achieves swap regret ϵ after $T = (\log m)^{\widetilde{\mathcal{O}}(1/\epsilon)}$ rounds.*

Both papers also provided (nearly-)matching lower bounds. Here we state a particularly simple-to-state lower bound due to Dagan et al. [2024]:

Theorem 3.3 (Dagan et al., 2024, lower bound). *Let $T < m/4$. Then there exists an oblivious*[1] *adversary such that the swap regret of any learner is $\Omega(\log^{-5} T)$.*

Tree-form strategy sets (Extensive-form games). For more general extensive-form games, the picture is less clear. For an upper bound, one can consider a tree-form decision problem with M pure strategies (*i.e.*, $|\mathcal{X}| = M$) as simply an "easier version" of a normal-form decision problem where each pure strategy is treated as a different action, *i.e.*, where the strategy set is the M-simplex. Theorem 3.2 therefore implies a similar bound on swap regret for tree-form decision problems[2].

Corollary 3.4 (Dagan et al., 2024, Peng and Rubinstein, 2024, tree-form upper bound). *Let $\mathcal{X} \subset \{0,1\}^m$ be a tree-form strategy set. There exists a no-regret learning algorithm for tree-form decision problems that achieves swap regret ϵ after $T = (\log |\mathcal{X}|)^{\widetilde{\mathcal{O}}(1/\epsilon)} \le m^{\widetilde{\mathcal{O}}(1/\epsilon)}$ rounds.*

Showing a matching lower bound for extensive form, however, remained open. The main difficulty is that the adversary is restricted to *linear* utility functions $u^t : \mathcal{X} \to \mathbb{R}$; the adversary in Theorem 3.3 does not use linear utility functions. The purpose of this paper is to close this discrepancy, by showing a lower bound that almost matches Corollary 3.4. Our main result is the following.

Theorem 3.5 (Main theorem). *There exist arbitrarily large tree-from strategy sets $\mathcal{X} \subset \{0,1\}^m$ with the following property. Let $\epsilon > 0$ and suppose $T \le \exp\Big(\min\Big\{\widetilde{\Theta}(m^{1/12}), \Theta(\epsilon^{-1/5})\Big\}\Big)$. Then there exists an oblivious adversary running for T iterations against which no learner can achieve expected swap regret better than ϵ.*

Intuitively, the proof of Theorem 3.5 works by "embedding" the adversary of Theorem 3.3 into a tree-form decision problem such that the utility functions u^t can remain linear. This works by choosing random vectors in $\{-1,1\}^n$ (for some appropriately-chosen dimension n) to simulate the "actions" in the normal-form decision problem, and then exploiting the concentration property that an exponentially-large number of such vectors $\{\boldsymbol{a}_i\}_{i=1}^M$ can be chosen such that $\langle \boldsymbol{a}_i, \boldsymbol{a}_j \rangle \approx 0$ for all $i \ne j$.

Like Theorem 3.3, our lower bound is *information-theoretic*: it does not rely on computational hardness results, and thus applies to *any* no-regret learning algorithm no matter how computationally powerful.

4 Proof of Theorem 3.5

Before proving Theorem 3.5, we first state a more detailed version of the normal-form lower bound (Theorem 3.3). This restatement changes notation so as to avoid mix notation between tree-form and normal-form decision problems, and extracts some useful properties of the adversary.

Theorem 4.1 (Dagan et al. 2024, expanded version of Theorem 3.3). *Let $\mathcal{A}$ be the M-simplex, and let $T < M/4$. Then there exists an adversary on $\mathcal{A}$ with the following properties:*

1. *The adversary selects a sequence $(\boldsymbol{u}^1, \dots, \boldsymbol{u}^T) \sim \mathcal{D}$ from some distribution $\mathcal{D} \in \Delta(\mathcal{A}^T)$, and then outputs utility vector $\boldsymbol{u}^t$ at time t regardless of the sequence of distributions played by the learner.*
2. *There exists a strategy $\boldsymbol{a}^* \in \mathcal{A}$ that is never used by the adversary.*
3. *There exists a partition $\mathcal{A} = \mathcal{A}_1 \sqcup \dots \sqcup \mathcal{A}_d$ where $d \le \mathcal{O}(\log T)$ with the following property. Within each set $\mathcal{A}_i$, number the actions $\mathcal{A}_i = \{\boldsymbol{a}_{i1}, \dots, \boldsymbol{a}_{iM_i}\}$. For any sequence $(\boldsymbol{u}^1, \dots, \boldsymbol{u}^T) \in \operatorname{supp} \mathcal{D}$, the*

[1] An adversary is *oblivious* if its choices of utility vectors $\boldsymbol{u}^t$ do not depend on the learner's chosen distributions π^t.

[2] The below bound, as stated, is only information-theoretic. However, the information-theoretic bound is implementable by an efficient (*i.e.*, $\mathsf{poly}(m, 1/\epsilon)$-time-per-iteration) algorithm, which is described by Dagan et al. [2024] and Peng and Rubinstein [2024], and beyond the scope of this paper.

adversary plays actions in $\mathcal{A}_i$ only in increasing order. That is, if $\boldsymbol{u}^t = \boldsymbol{a}_{ij}$ and $\boldsymbol{u}^{t'} = \boldsymbol{a}_{ij'}$ and $t \le t'$, then $j \le j'$.

4. *The swap regret of any learner against this adversary is $\Omega(d^{-5}) = \Omega(\log^{-5} T)$.*

We now prove Theorem 3.5.

Consider the following family of tree-from strategy sets, parameterized by natural numbers d and n. First the learner picks an index $i \in [d]$. Then the environment picks $j \in [n]$, and finally the learner picks a binary action. A strategy is identified (up to linear transformations) by a $\boldsymbol{x} \in \mathbb{R}^{d \times n}$ where $\boldsymbol{x}[i, \cdot] \in \{-\frac{1}{\sqrt{n}}, \frac{1}{\sqrt{n}}\}^n$ and $\boldsymbol{x}[i', \cdot] = \mathbf{0}$ if $i' \neq i$ (*i.e.*, $\boldsymbol{x}$ as a matrix has exactly one nonzero row). For convenienence we will use $\mathcal{X}_i \subset \mathcal{X}$ to denote the set of pure strategies where the learner plays i at the root. Let C be an absolute constant large enough to make the asymptotic bounds in Theorem 4.1 true.

The adversary works as follows. First, for each $i \in [d]$, it populates $\mathcal{A}_i$ with uniformly randomly chosen strategies $\boldsymbol{a}_{i1}, \dots, \boldsymbol{a}_{iM_i} \in \mathcal{X}_i$. Then the adversary plays as in Theorem 4.1.

We will claim that, for any learner against this adversary, there exists a learner against the adversary of Theorem 4.1 that achieves a similar swap regret—and thus the swap regret of the former learner must be large. First, we will construct the latter adversary.

Let $\pi^1, \dots, \pi^T \in \Delta(\mathcal{X})$ be the sequence of distributions played by the learner. Note that π^t can depend on the utilities $\boldsymbol{u}^{1:t-1} \in \mathcal{A}$ that are played by the adversary. Consider the sequence $\bar{\pi}^1, \dots, \bar{\pi}^T \in \Delta(\mathcal{A})$, where $\bar{\pi}^t$ is the distribution that samples $\boldsymbol{x} \sim \pi^t$ and plays according to $p_{\boldsymbol{x}} \in \Delta(\mathcal{A})$, defined as follows. Let $\boldsymbol{x} \in \mathcal{X}_i$ be any strategy. Let ϵ be a parameter to be selected later. There are two cases.

1. $\langle \boldsymbol{x}, \boldsymbol{a}_{ij} \rangle \le \epsilon$ for every $\boldsymbol{a}_{ij} \in \mathcal{A}_i$. Then define $p_{\boldsymbol{x}} = \boldsymbol{a}^*$ deterministically.
2. $\langle \boldsymbol{x}, \boldsymbol{a}_{ij} \rangle > \epsilon$ for some $\boldsymbol{a}_{ij} \in \mathcal{A}_i$. Let j be the *largest* such index, let $\beta = \langle \boldsymbol{x}, \boldsymbol{a}_{ij} \rangle$, and define $p_{\boldsymbol{x}}$ as the distribution that is $\boldsymbol{a}^*$ with probability $1 - \beta$ and $\boldsymbol{a}_{ij}$ with probability β.

A critical property for us will be that the learner cannot "guess in advance" what future unobserved $\boldsymbol{a}_{ij}$s will be, since these are sampled uniformly at random. That is, in Case 2, $\boldsymbol{x}$ can only be played with large probability once the adversary has played $\boldsymbol{a}_{ij}$.

To be more formal, we first define some notation. For every $\boldsymbol{a}_{ij} \in \mathcal{A}_i$ let t_{ij} be the first iteration on which the adversary plays $\boldsymbol{a}_{ij}$ (or $t_{ij} = T$ if this never happens). For $\boldsymbol{x} \in \mathcal{X}_i$, if $\boldsymbol{x}$ is in Case 1 above then define $t_{\boldsymbol{x}} = 0$, and otherwise define $t_{\boldsymbol{x}} = t_{ij}$, where j is as in Case 2.

There are two properties that we will critically need to use about $t_{\boldsymbol{x}}$. The first states that the learner cannot place large mass on $\boldsymbol{x}$ until after $t_{\boldsymbol{x}}$, because doing so would require the learner to guess a vector heavily correlated with $\boldsymbol{a}_{ij}$ before the learner observes $\boldsymbol{a}_{ij}$.

Lemma 4.2. *Let $\delta := \dfrac{1}{T} \displaystyle\sum_{\boldsymbol{x} \in \mathcal{X}} \sum_{t=1}^{t_{\boldsymbol{x}}} \pi^t(\boldsymbol{x})$. Then $\mathbb{E}\, \delta \le M e^{-n\epsilon^2/2}$.*

Proof. Since the learner has not yet observed $\boldsymbol{a}_{ij}$ at time t_{ij}, its prior strategy sequence $\pi^{1:t_{ij}}(\boldsymbol{x})$ must be independent of $\boldsymbol{a}_{ij}$. Moreover, if $t \le t_{\boldsymbol{x}}$ then there must exist some j with $t_{ij} \ge t$ and $\langle \boldsymbol{x}, \boldsymbol{a}_{ij} \rangle \ge \epsilon$—namely, the j defining Case 2. Thus we have:

$$\begin{aligned}
\mathbb{E}\,\delta &= \mathbb{E}\,\frac{1}{T}\sum_{\boldsymbol{x}\in\mathcal{X}}\sum_{t=1}^{t_{\boldsymbol{x}}}\pi^t(\boldsymbol{x}) \\
&\le \mathbb{E}\,\frac{1}{T}\sum_{i=1}^{d}\sum_{\boldsymbol{x}\in\mathcal{X}_i}\sum_{t=1}^{T}\pi^t(\boldsymbol{x})\sum_{j:t_{ij}\ge t}\mathbb{1}\{\langle\boldsymbol{x},\boldsymbol{a}_{ij}\rangle\ge\epsilon\} \\
&= \underbrace{\frac{1}{T}\sum_{i=1}^{d}\sum_{\boldsymbol{x}\in\mathcal{X}_i}\sum_{j=1}^{M_i}\mathbb{E}\left[\sum_{t\le t_{ij}}\pi^t(\boldsymbol{x})\right]}_{\le M}\underbrace{\mathbb{E}\left[\mathbb{1}\{\langle\boldsymbol{x},\boldsymbol{a}_{ij}\rangle\ge\epsilon\}\right]}_{\le e^{-n\epsilon^2/2}}\le Me^{-n\epsilon^2/2}.
\end{aligned}$$

where in the last line we use the fact that $\boldsymbol{a}_{ij}$ is independent of $\pi^{1:t_{ij}}(\boldsymbol{x})$ and then Hoeffding's inequality. □

The second property is that, for $t > t_{\boldsymbol{x}}$, utilities of $\boldsymbol{x}$ under $\boldsymbol{u}^t$ are approximately the same as those of of $p_{\boldsymbol{x}}$ under the losses in Theorem 4.1.

Lemma 4.3. *For $t > t_{\boldsymbol{x}}$, we have $\langle\boldsymbol{x},\boldsymbol{u}^t\rangle \le p_{\boldsymbol{x}}(\boldsymbol{u}^t)+\epsilon$.*

Proof. Let $\boldsymbol{x}\in\mathcal{X}_i$. There are two cases. First, if $\langle\boldsymbol{x},\boldsymbol{a}_{ij}\rangle\le\epsilon$ for every $\boldsymbol{a}_{ij}\in\mathcal{A}_i$. Then for every t, we have $\boldsymbol{u}^t\notin\operatorname{supp}p_{\boldsymbol{x}}=\{\boldsymbol{a}^*\}$ (because the adversary never plays $\boldsymbol{a}^*$), and $\langle\boldsymbol{x},\boldsymbol{u}^t\rangle\le\epsilon$ by definition, so we are done.

Otherwise, let j be the largest index for which $\langle\boldsymbol{x},\boldsymbol{a}_{ij}\rangle>\epsilon$. Then $t_{\boldsymbol{x}}=t_{ij}$ by definition, and since $t>t_{ij}$, by Property 4 the adversary is no longer allowed to play $\boldsymbol{a}_{ij'}$ for $j'<j$. Thus, either $\boldsymbol{u}^t\notin\operatorname{supp}p_{\boldsymbol{x}}$ and $\langle\boldsymbol{x},\boldsymbol{u}^t\rangle\le\epsilon$, or $\boldsymbol{u}^t=\boldsymbol{a}_{ij}$. The former case reduces to the previous paragraph. In the latter case, we have $\langle\boldsymbol{x},\boldsymbol{u}^t\rangle=\beta=p_{\boldsymbol{x}}(\boldsymbol{u}^t)$ by construction of f. □

For the rest of this proof we will use $\bar{V}(\phi)$ to denote the utilities experienced by $\bar{\pi}^t$ under the utilities in Theorem 4.1. That is,

$$\bar{V}(\phi)=\frac{1}{T}\sum_{t=1}^{T}\sum_{\boldsymbol{a}\in\mathcal{A}}\bar{\pi}^t(\boldsymbol{a})\mathbb{1}\{\phi(\boldsymbol{a})=\boldsymbol{u}^t\}=\frac{1}{T}\sum_{t=1}^{T}\sum_{\boldsymbol{x}\in\mathcal{X}}\pi^t(\boldsymbol{x})\Pr_{\boldsymbol{a}\sim p_{\boldsymbol{x}}}[\phi(\boldsymbol{a})=\boldsymbol{u}^t]$$

By Theorem 4.1, there exists a function $\bar{\phi}:\mathcal{A}\to\mathcal{A}$ such that[3] $\mathbb{E}[\bar{V}(\bar{\phi})-\bar{V}(\mathrm{Id})]\ge 1/Cd^5$. It suffices to show that $\mathbb{E}[V(\phi)-V(\mathrm{Id})]$ is large. To do this, we will show that, up to small errors, $V(\mathrm{Id})\le\bar{V}(\mathrm{Id})$ and $V(\phi)\approx\bar{V}(\bar{\phi})$.

For the first approximation, we have

$$\begin{aligned}
V(\mathrm{Id}) &= \frac{1}{T}\sum_{\boldsymbol{x}\in\mathcal{X}}\sum_{t=1}^{T}\pi^t(\boldsymbol{x})\langle\boldsymbol{x},\boldsymbol{u}^t\rangle \\
&\le \frac{1}{T}\sum_{\boldsymbol{x}\in\mathcal{X}}\sum_{t>t_{\boldsymbol{x}}}\pi^t(\boldsymbol{x})\langle\boldsymbol{x},\boldsymbol{u}^t\rangle+\delta \\
&\le \frac{1}{T}\sum_{\boldsymbol{x}\in\mathcal{X}}\sum_{t>t_{\boldsymbol{x}}}\pi^t(\boldsymbol{x})p_{\boldsymbol{x}}(\boldsymbol{u}^t)+\epsilon+\delta \\
&\le \frac{1}{T}\sum_{\boldsymbol{x}\in\mathcal{X}}\sum_{t=1}^{T}\pi^t(\boldsymbol{x})p_{\boldsymbol{x}}(\boldsymbol{u}^t)+\epsilon+2\delta=\bar{V}(\mathrm{Id})+\epsilon+2\delta.
\end{aligned}$$

[3]Technically ϕ is a random variable dependent on $\boldsymbol{u}^1,\ldots,\boldsymbol{u}^T$.

For the second, we have

$$
\begin{aligned}
V(\phi) &= \frac{1}{T} \sum_{\boldsymbol{x} \in \mathcal{X}} \sum_{t=1}^{T} \pi^t(\boldsymbol{x}) \left\langle \phi(\boldsymbol{x}), \boldsymbol{u}^t \right\rangle \\
&\geq \frac{1}{T} \sum_{\boldsymbol{x} \in \mathcal{X}} \sum_{t > t_{\boldsymbol{x}}} \pi^t(\boldsymbol{x}) \mathop{\mathbb{E}}_{\boldsymbol{a} \sim p_{\boldsymbol{x}}} \left\langle \bar{\phi}(\boldsymbol{a}), \boldsymbol{u}^t \right\rangle - \delta \\
&\geq \frac{1}{T} \sum_{\boldsymbol{x} \in \mathcal{X}} \sum_{t > t_{\boldsymbol{x}}} \pi^t(\boldsymbol{x}) \mathop{\mathbb{E}}_{\boldsymbol{a} \sim p_{\boldsymbol{x}}} \left\langle \bar{\phi}(\boldsymbol{a}), \boldsymbol{u}^t \right\rangle - \epsilon - \delta \\
&\geq \sum_{\boldsymbol{x} \in \mathcal{X}} \frac{1}{T} \sum_{t=1}^{T} \pi^t(\boldsymbol{x}) \mathop{\mathbb{E}}_{\boldsymbol{a} \sim p_{\boldsymbol{x}}} \left\langle \bar{\phi}(\boldsymbol{a}), \boldsymbol{u}^t \right\rangle - \epsilon - 2\delta = \bar{V}(\bar{\phi}) - \epsilon - 2\delta.
\end{aligned}
$$

Thus,

$$
\mathbb{E}[V(\phi) - V(\mathrm{Id})] \geq \mathbb{E}\left[\bar{V}(\bar{\phi}) - \bar{V}(\mathrm{Id}) - 2\epsilon - 4\delta\right] \geq \frac{1}{Cd^5} - 2\epsilon - 4\delta \geq \frac{1}{3Cd^5} = \epsilon
$$

by taking

$$
\epsilon = \frac{1}{3Cd^5} \quad \text{and} \quad n = \frac{2 \log 12CMd^5}{\epsilon^2} = \widetilde{\Theta}(d^{11}).
$$

The resulting tree-form decision problem hence has dimension $m = d \cdot n = \widetilde{\mathcal{O}}(d^{12})$, and since $d = \Theta(\epsilon^{-1/5}) \leq \Theta(\log T)$ we have that the swap regret is at least ϵ until $T \geq \min\left\{M/4, \exp\bigl(\Theta(\epsilon^{-1/5})\bigr)\right\}$, where

$$
M = \frac{1}{12Cd^5} \exp\left(\frac{n\epsilon^2}{2}\right) = \exp\left(\widetilde{\Theta}(m^{1/12})\right)
$$

as desired.

5 Conclusion and Future Research

By extending a recent lower bound for normal-form games, we have established a lower bound that precludes the existence of fully polynomial-time (*i.e.*, $\mathsf{poly}(m, 1/\epsilon)$) algorithms for swap regret minimization in tree-form decision problems.

Our result leaves open several natural questions for future research.

1. Our counterexample applies only for extensive-form games with a particular structure. Is swap regret minimization also information-theoretically impossible for simpler structures, such as *single-stage* Bayesian games, in which the strategy set is a product of simplices?
2. Are there uncoupled learning dynamics that yield $\mathsf{poly}(m, 1/\epsilon)$-time convergence to NFCE when applied in games? Our result does not preclude this possibility, since the behavior of the adversary in Theorem 3.5 is likely not the behavior of any learning agent in a game.
3. What is the complexity of *computing* one NFCE in an extensive-form game (by any method, not just limited to independent learning dynamics? This is a problem that was stated as early as Papadimitriou and Roughgarden [2005] and remains open.

The majority of correlated notions of equilibrium that *are* known to be efficiently computable have corresponding efficient no-regret learning algorithms with convergence guarantee of the form $\mathsf{poly}(m, 1/\epsilon)$. For example, the notions of correlated equilibrium mentioned in the introduction (up to and including linear correlated equilibria) admit both efficient no-regret algorithms [Zhang et al., 2024] and efficient algorithms for exact computation [Farina and Pipis, 2024]. We thus believe that our main result, which precludes efficient no-regret learning for swap regret, is evidence against the existence of an efficient algorithm (learning dynamics or otherwise) for computing an NFCE in an extensive-form game. Proving or disproving this claim is an important open question.

References

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

Anton Bakhtin, Noam Brown, Emily Dinan, Gabriele Farina, Colin Flaherty, Daniel Fried, Andrew Goff, Jonathan Gray, Hengyuan Hu, et al. Human-level play in the game of diplomacy by combining language models with strategic reasoning. *Science*, 378(6624):1067–1074, 2022.

Avrim Blum and Yishay Mansour. From external to internal regret. *Journal of Machine Learning Research*, 8(6), 2007.

Noam Brown and Tuomas Sandholm. Superhuman AI for heads-up no-limit poker: Libratus beats top professionals. *Science*, 359(6374):418–424, 2018.

Noam Brown and Tuomas Sandholm. Superhuman AI for multiplayer poker. *Science*, 365(6456):885–890, 2019.

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. *Symposium on Theory of Computing (STOC)*, 2024.

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

Gabriele Farina and Charilaos Pipis. Polynomial-time computation of exact phi-equilibria in polyhedral games. *arXiv preprint arXiv:2402.16316*, 2024.

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, 2022.

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

Dustin Morrill, Ryan D'Orazio, Reca Sarfati, Marc Lanctot, James R Wright, Amy R Greenwald, and Michael Bowling. Hindsight and sequential rationality of correlated play. In *AAAI Conference on Artificial Intelligence (AAAI)*, volume 35, pages 5584–5594, 2021a.

Dustin Morrill, Ryan D'Orazio, Marc Lanctot, James R Wright, Michael Bowling, and Amy R Greenwald. Efficient deviation types and learning for hindsight rationality in extensive-form games. In *International Conference on Machine Learning (ICML)*, pages 7818–7828. PMLR, 2021b.

Christos Papadimitriou and Tim Roughgarden. Computing equilibria in multi-player games. In *Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 82–91, Vancouver, BC, Canada, 2005. SIAM.

Binghui Peng and Aviad Rubinstein. Fast swap regret minimization and applications to approximate correlated equilibria. *Symposium on Theory of Computing (STOC)*, 2024.

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, Gabriele Farina, and Tuomas Sandholm. Mediator interpretation and faster learning algorithms for linear correlated equilibria in general sequential games. In *International Conference on Learning Representations (ICLR)*, 2024.