On the Complexity of Computing Sparse Equilibria and Lower Bounds for No-Regret Learning in Games

Ioannis Anagnostides
Department of Computer Science, Carnegie Mellon University, Pittsburgh, PA, USA

Alkis Kalavasis
Department of Computer Science, Yale University, New Haven, CT, USA

Tuomas Sandholm
Department of Computer Science, Carnegie Mellon University, Pittsburgh, PA, USA

Manolis Zampetakis
Department of Computer Science, Yale University, New Haven, CT, USA

Abstract

Characterizing the performance of no-regret dynamics in multi-player games is a foundational problem at the interface of online learning and game theory. Recent results have revealed that when all players adopt specific learning algorithms, it is possible to improve exponentially over what is predicted by the overly pessimistic no-regret framework in the traditional adversarial regime, thereby leading to faster convergence to the set of *coarse correlated equilibria (CCE)* – a standard game-theoretic equilibrium concept. Yet, despite considerable recent progress, the fundamental complexity barriers for learning in normal- and extensive-form games are poorly understood. In this paper, we make a step towards closing this gap by first showing that – barring major complexity breakthroughs – any polynomial-time learning algorithms in extensive-form games need at least $2^{\log^{1/2-o(1)}|\mathcal{T}|}$ iterations for the average regret to reach below even an absolute constant, where $|\mathcal{T}|$ is the number of nodes in the game. This establishes a superpolynomial separation between no-regret learning in normal- and extensive-form games, as in the former class a logarithmic number of iterations suffices to achieve constant average regret. Furthermore, our results imply that algorithms such as multiplicative weights update, as well as its *optimistic* counterpart, require at least $2^{(\log\log m)^{1/2-o(1)}}$ iterations to attain an $O(1)$-CCE in m-action normal-form games under any parameterization. These are the first non-trivial – and dimension-dependent – lower bounds in that setting for the most well-studied algorithms in the literature. From a technical standpoint, we follow a beautiful connection recently made by Foster, Golowich, and Kakade (ICML '23) between *sparse* CCE and Nash equilibria in the context of Markov games. Consequently, our lower bounds rule out polynomial-time algorithms well beyond the traditional online learning framework, capturing techniques commonly used for accelerating centralized equilibrium computation.

2012 ACM Subject Classification Theory of computation → Convergence and learning in games

Keywords and phrases No-regret learning, extensive-form games, multiplicative weights update, optimism, lower bounds

Digital Object Identifier 10.4230/LIPIcs.ITCS.2024.5

Related Version *Full Version*: https://arxiv.org/pdf/2311.14869.pdf

Funding 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.

Acknowledgements We are grateful to the anonymous ITCS reviewers for their helpful feedback.

1 Introduction

At the heart of the intricate interplay between online learning and game theory, which can be traced all the way back to Blackwell's seminal approachability theorem [10] and Robinson's analysis of fictitious play [72], lies the fundamental *no-regret* framework. Here, a learner has to select round by round a sequence of actions so as to obtain a high cumulative reward; the crux presents itself in the online nature of the revealed information, in that each reward function is unbeknownst to the learner prior to the termination of that round. The canonical measure of performance in this online environment is the notion of *regret*, which contrasts the cumulative reward of the learner to that of the optimal fixed action in hindsight; a learner is said to incur *no-regret* if its regret grows sublinearly with the time horizon T. By now, it is well understood that when the sequence of rewards is produced adversarially, the minimax regret after T repetitions is precisely $\tilde{\Theta}(\sqrt{T \log m})$, where m denotes the number of available actions of the learner.[1] More broadly, online learnability in more general combinatorial domains, such as binary classification, can be characterized by a certain notion of dimension known as the Littlestone dimension [59, 9], painting a rather complete picture in the adversarial regime. In this context, the alluded nexus between the no-regret framework and game theory can be witnessed by the celebrated realization that players with sublinear regret converge to a certain game-theoretic equilibrium concept known as *coarse correlated equilibrium (CCE)* [46, 40, 17]. In particular, if players follow certain no-regret algorithms, such as the celebrated multiplicative weights update (`MWU`), the history of play induces an ϵ-CCE after merely $T = O(\frac{\log m}{\epsilon^2})$ repetitions of the game.

In light of our rather comprehensive understanding of online learning, one might expect that the fundamental barriers of no-regret learning in games have already been identified. However, as it turns out, this is not the case. Indeed, the regret incurred by each player when facing other learning agents can be remarkably smaller than what is predicted by the overly pessimistic no-regret framework. This is exemplified by the recent result of Daskalakis et al. [24], who proved that when players in an n-player m-action game follow the *optimistic* counterpart of `MWU` [70] (henceforth `OMWU`), each player's regret grows only as $\tilde{O}(n \log m)$, revealing an exponential separation compared to the lower bound in the adversarial regime. Another noteworthy example concerns the behavior of fictitious play: it is hopeless in the adversarial setting, where it can accumulate linear regret [17], but Julia Robinson famously proved that it converges to minimax equilibria when followed by both players in a (two-player) zero-sum game [72].

Despite the considerable interest recent work has devoted to understanding the problem of no-regret learning in games (Section 1.2 features several such results), little is known in terms of lower bounds. Daskalakis et al. [23] made an early effort by noting that incurring $\Omega(1)$ regret is – at least in some sense – inevitable; this boils down to the straightforward realization that even in a single-agent problem the first decision will likely be suboptimal, resulting in $\Omega(1)$ regret even if all the subsequent actions are optimal. Besides failing to provide a meaningful bound in terms of the dimensions of the game, another unsatisfactory feature of the lower bound of Daskalakis et al. [23] is that it can be bypassed by simply detaching the first iteration – in which case both players actually incur 0 regret. Can we hope to guarantee that each player will incur $\tilde{O}(1)$ regret, *independent* of the dimensions of the game, or are there fundamental barriers that circumscribe the performance of no-regret learners in games?

[1] We use the $\tilde{\Theta}(\cdot)$ notation to suppress $\mathsf{polylog}T$ factors.

1.1 Our results

Our primary contribution in this paper is to make a step towards filling the aforementioned knowledge gap by establishing the first non-trivial computational hardness results when multiple players are learning in games.

Our first lower bound concerns a class of no-regret dynamics that includes MWU, perhaps the most well-studied learning algorithm in the literature, as well its optimistic counterpart (OMWU). Before we proceed, we recall that the *exponential-time hypothesis (ETH)* for PPAD [4] postulates that there do no exist truly subexponential algorithms for solving EndOfALine – the prototypical PPAD-complete problem (Conjecture 3.7). By now, this is a fairly standard computational complexity assumption, which was – crucially for the purpose of this paper – famously invoked by Rubinstein [74] to settle the complexity of computing approximate Nash equilibria in two-player (normal-form) games.

▶ **Theorem 1.1.** *Consider the class of n-player m-action games in normal form. If each player $i \in [\![n]\!]$ follows* MWU *or* OMWU *and incurs (cumulative) regret Reg_i^T, there is a game $\mathcal{G}$ and an absolute constant $\epsilon > 0$ such that at least $T \geq 2^{(\log_2 \log_2 m)^{1/2-o(1)}}$ repetitions of the game are needed so that $\frac{1}{T} \max_{1 \leq i \leq n} \mathsf{Reg}_i^T \leq \epsilon$, unless ETH for* PPAD *(Conjecture 3.7) is false.*

This represents the first non-trivial lower bounds for no-regret learning in the fundamental setting of Theorem 1.1 under algorithms such as MWU and OMWU. Theorem 1.1 applies under any choice of learning rates (as specified in Corollary 3.9). For comparison, we have already alluded to the fact that $T = O(\log m) = O(2^{\log \log m})$ repetitions of MWU or OMWU suffice to guarantee that $\frac{1}{T} \max_{1 \leq i \leq n} \mathsf{Reg}_i^T \leq \epsilon$, for any absolute constant $\epsilon > 0$. As such, Theorem 1.1 leaves a certain gap compared to the best known upper bounds. In fact, Theorem 1.1 actually applies to a class of algorithms more general than MWU-type update rules, as we will make clear shortly.

Learning in extensive-form games

Although Theorem 1.1 concerns the behavior of (O)MWU under the standard normal-form representation of finite games, our approach actually revolves around proving hardness results for *extensive-form* games. The extensive-form representation is typically exponentially more compact – and thereby much more appropriate – when encoding games involving sequential moves as well as imperfect information. In light of their ubiquitous presence in real-world applications, there has been a considerable interest in understanding the performance of no-regret learning algorithms in the more challenging class of extensive-form games (see Section 1.2). Indeed, no-regret dynamics have been at the heart of recent landmark results in practical computation of strategies for large games [13, 14, 16, 7].

Sparse equilibria

To establish lower bounds for no-regret learning algorithms in extensive-form games, we follow a beautiful approach recently put forward by Foster et al. [41] in a different context, namely that of Markov games. Their idea is to use as a proxy a refinement of CCE in which a certain sparsity constraint is imposed. More precisely, a correlated distribution is said to be *k-sparse* if it can be expressed as the uniform mixture of k product distributions (Definition 2.2); as such, we clarify that a 1-sparse CCE is equivalent to a Nash equilibrium. The connection of this refinement with the no-regret framework is evident: any CCE derived from (independent) no-regret learners after T repetitions certainly satisfies the T-sparsity constraint [41]. The name of the game now is to establish hardness results for computing sparse CCE in a certain regime of sparsity, which in turn would readily impose barriers on the performance of (computationally efficient) no-regret algorithms.

As an aside, we argue that a k-sparse CCE is an important solution concept in its own right, besides the connection with no-regret learning. First, a correlated distribution is in general an exponential object; polynomial sparsity ensures that there exists a succinct representation of that distribution. Indeed, that refinement was central in the celebrated *ellipsoid against hope* algorithm of Papadimitriou and Roughgarden [66], the only known polynomial-time algorithm for computing (exact) CCE in succinct multi-player games [53]. Further, one important weakness of CCE compared to Nash equilibria is that the former has a much larger description complexity even if the sparsity is polynomial. This becomes especially relevant in some modern machine learning applications in which strategies are represented through massive neural networks, thereby necessitating storing a large sequence of such neural networks in order to simply represent a CCE, which can be prohibitive. In contrast, if a CCE with sparsity $k = 2$ was efficiently computable, that would effectively address such concerns.

Hardness of computing sparse CCE in extensive-form games

Having motivated the concept of a sparse CCE, we next state our main hardness result for computing such equilibria in extensive-form games under a certain sparsity regime. Below, for an extensive-form game described by a tree $\mathcal{T}$, we denote by $|\mathcal{T}|$ the number of nodes in $\mathcal{T}$ (the reader not familiar with the extensive-form representation can first turn to Section 2.2 for formal definitions).

▶ **Theorem 1.2.** *There is no algorithm that runs in time polynomial in the description of an extensive-form game $\mathcal{T}$ and can compute a $2^{\log_2^{1/2-o(1)}|\mathcal{T}|}$-sparse ϵ-CCE, even for an absolute constant $\epsilon > 0$, unless ETH for* PPAD *(Conjecture 3.7) is false.*

Prior to our work, even the complexity status of computing a 2-sparse $O(1)$-CCE in extensive-form games was open. Theorem 1.2 implies a superpolynomial separation for the problem of computing sparse CCE between normal- and extensive-form games. Indeed, we have seen that in normal-form games logarithmic – in the description of the game – sparsity is efficiently attainable; Theorem 1.2 precludes such a possibility in extensive-form games. To better contextualize Theorem 1.2, we remark that certain polynomial-time algorithms in extensive-form games attain roughly $2^{\log|\mathcal{T}|}$-sparsity (in the regime where $\epsilon = O(1)$), thereby leaving again a certain gap compared to Theorem 1.2. We further point out that Theorem 1.2, and implications thereof (Theorem 1.1 and Corollary 1.3), applies even for games with three players; it is open whether it extends to two-player games, a discrepancy explained in more detail in Section 3 (*cf.* [12, 41]).

Implications for no-regret dynamics

As a consequence, Theorem 1.2 circumscribes in extensive-form games the performance of any no-regret dynamics that have polynomial complexity per iteration.

▶ **Corollary 1.3.** *Suppose that each player follows an algorithm with polynomial iteration complexity in the description of an extensive-form game $\mathcal{T}$. If Reg_i^T is the regret incurred by player $i \in [\![n]\!]$, there is an extensive-form game $\mathcal{T}$ and an absolute constant $\epsilon > 0$ such that at least $T \geq 2^{\log_2^{1/2-o(1)}|\mathcal{T}|}$ repetitions are needed so that $\frac{1}{T}\max_{1\leq i\leq n}\mathsf{Reg}_i^T \leq \epsilon$, unless ETH for* PPAD *(Conjecture 3.7) is false.*

There are many compelling aspects of Corollary 1.3 worth stressing. First, it applies even in a centralized model well beyond the online and decentralized learning framework. As a concrete example, Corollary 1.3 applies even if the dynamics are *alternating* instead of

simultaneous. Alternation has been a remarkably successful ingredient in practical solvers [79], but it violates the online nature of the problem; indeed, the player who gets to play last has complete information about the current reward function. Nevertheless, even alternating dynamics are subject to the barriers imposed by Corollary 1.3. Furthermore, as we point out in Remark 3.6, Corollary 1.3 can be extended even if one considers a more general non-uniform notion of regret, which has been observed to lead to significantly faster convergence in practice [15]. Another interesting feature of Corollary 1.3 is that players are allowed to store a polynomial amount of information regarding past rewards. This is considerably stronger – in that designing lower bounds is harder – than the model of Daskalakis et al. [23] wherein only a constant number of prior rewards can be stored; that assumption was made by Daskalakis et al. [23] to preclude trivial exploration strategies in two-player zero-sum games whereby players first determine the entire payoff matrix, and then compute a minimax strategy with the information gathered. In contrast, such an exploration strategy is a legitimate possibility in the context of Corollary 1.3. As such, the model we consider here is so permissive that no hardness results can be established for no-regret learning in two-player zero-sum games, simply because there are polynomial-time algorithms for computing Nash equilibria in such games.

Returning to Theorem 1.1, the key connection is that algorithms such as (O)`MWU` can be efficiently simulated on the induced normal-form representation of the extensive-form game [35]. As such, Theorem 1.1 turns out to be a consequence of Corollary 1.3. As a result, Theorem 1.1 applies more broadly to any class of algorithms simulated on the induced normal form with per-iteration complexity polynomial in the representation of the underlying extensive-form game.

Technical approach

From a technical standpoint, we follow the approach of Foster et al. [41], who proved that computing sparse CCE in Markov (aka. stochastic) games is computationally hard even when targeting a polynomial sparsity. A crucial detail here is that Foster et al. [41] define CCE by allowing potentially non-Markovian deviations, for otherwise polynomial algorithms do exist [28]; this already separates regret minimization in Markov games from extensive-form games. The key observation of Foster et al. [41] is that sparse CCE in general-sum Markov games can be leveraged to efficiently compute *Nash equilibria* in general-sum (normal-form) games, thereby confronting immediate computational barriers [25, 18, 74]. Following this connection, we establish a similar reduction: we show that for any two-player m-action game $\mathcal{G}$ there is an extensive-form game $\mathcal{T} = \mathcal{T}(\mathcal{G})$ (Section 3.1) with the property that i) a T-sparse ϵ-CCE in $\mathcal{T}$ induces an $O(\epsilon)$-NE in $\mathcal{G}$ (Theorem 3.5), and ii) the description of $\mathcal{T}$ is of the order $m^{\log T/\epsilon^2}$. The key idea is that by repeating the underlying game $\mathcal{G}$ multiple times, a potentially deviating player could approximately discern the product distribution the rest of the players prescribe to, even though their randomization is unbeknownst to the deviator. In turn, this essentially forces a CCE in $\mathcal{T}$ to contain a Nash equilibrium strategy for $\mathcal{G}$ by virtue of a reduction due to Borgs et al. [12]; otherwise, there would exist a deviation with a significant profit in $\mathcal{T}$, contradicting the assumption that the original mixture of product distributions constitutes a CCE. This argument is the crux of the entire approach, and – following Foster et al. [41] – relies on some classical results on *online density estimation*, namely *Vovk's aggregating algorithm* [84]. In particular, Vovk's algorithm guarantees that a deviating player can identify, within ϵ total variation distance in expectation, the strategy of the rest of the players after $H = O(\frac{\log T}{\epsilon^2})$ repetitions of the game.

1.2 Further related work

The line of work endeavoring to characterize the performance of no-regret learners in games, beyond the adversarial regime [17, 11], was pioneered by Daskalakis et al. [23] in the context of two-player zero-sum games. Thereafter, it has attracted considerable interest in the literature [31, 68, 71, 77, 51, 50, 19, 34, 55, 42, 89, 86, 26], culminating in the breakthrough result of Daskalakis et al. [24] highlighted earlier in Section 1.

Yet, despite the significant progress, little is known in terms of lower bounds, with some notable exceptions. First, Syrgkanis et al. [77] showed that if one player follows `MWU` and the other player is best responding in the context of a two-player zero-sum game, one of the players must incur $\Omega(\sqrt{T})$ regret, no matter how the learning rate is set. With a more elaborate argument, Chen and Peng [19] established the same lower bound when both players follow `MWU` in a two-player game, again for any choice of learning rate. Both of those results were constructed based on binary-action games, and as such, they did not provide any meaningful lower bounds in terms of the dimensions of the game. Furthermore, Hadiji et al. [45] recently investigated the first-order query complexity of computing ϵ-Nash equilibria in $m \times m$ two-player zero-sum games (*cf.* [43, 37, 3, 36, 61]). They showed that $\Omega(m)$ (first-order) queries are needed when $\epsilon = 0$, and roughly $\Omega(\log(\frac{1}{m\epsilon}))$ when $\epsilon = O(\frac{1}{m^4})$, thereby leaving a substantial gap with the upper bound of $O(\frac{\log m}{\epsilon})$ attained via `OMWU`. The lower bounds we establish in this paper are quite different, being of computational nature. Indeed, we have already explained that in the more permissive (potentially centralized) model that we study here, there are no obstacles in attaining zero regret in a single iteration of a two-player zero-sum game.

Finally, one of our main results (Theorem 1.2) establishes a superpolynomial separation between no-regret learning in extensive- and normal-form games. It is worth stressing thus that learning in extensive-form games has been a particularly popular research topic in the literature (*e.g.*, [90, 33, 57, 38, 6, 5, 32, 62, 63, 31, 44, 48, 27, 80, 34, 30, 76, 87, 69, 39], and the numerous references therein). This emphasis stems to a large extent from the fact that the extensive-form representation is more suited to capture realistic settings that feature sequential moves and imperfect information.

2 Preliminaries

In this section, we provide the necessary background on normal- and extensive-form games, as well as the setting of online density estimation. Specifically, Section 2.1 formalizes the refinement of CCE we focus on; Section 2.2 introduces the extensive-form representation; and Section 2.3 describes Vovk's aggregation algorithm [84] in the context of online density estimation. For further background on learning in games, we refer to the excellent book of Cesa-Bianchi and Lugosi [17].

Conventions

We let $\mathbb{N} = \{1, 2, \dots, \}$ be the set of natural numbers. We oftentimes use the $O(\cdot), \Omega(\cdot), \Theta(\cdot)$ notation with a non-asymptotic semantic so as to suppress absolute constants. For a finite set S, we let $\mathsf{U}(S)$ denote the uniform distribution over S. $\log(\cdot)$ denotes the natural logarithmic (with base e). We generally use subscripts to indicate the player and superscripts (with parentheses) to specify the (discrete) time index.

2.1 Sparse coarse correlated equilibria

As we explained earlier in our introduction, our main focus here is on the problem of computing a refinement of the standard coarse correlated equilibrium (CCE) [64] satisfying a certain sparsity constraint. To formally introduce that refinement, let us first introduce the normal-form representation of finite games.

Normal-form games

Let $\mathcal{G}$ be a finite n-player game represented in normal form. The set of players will be denoted by $[\![n]\!] \triangleq \{1, 2, \dots, n\}$, and will be indexed by variables $i, i' \in [\![n]\!]$. In the normal-form representation, every player $i \in [\![n]\!]$ has a finite and nonempty set of available actions $\mathcal{A}_i$. For notational convenience, we will let $m \triangleq \max_{1 \leq i \leq n} |\mathcal{A}_i|$. For every possible combination of actions $\boldsymbol{a} \triangleq (a_1, \dots, a_n) \in \bigtimes_{i=1}^n \mathcal{A}_i$, there is a utility function $u_i : \bigtimes_{i=1}^n \mathcal{A}_i \to [-1, 1]$ that specifies the utility (or reward) $u_i(\boldsymbol{a})$ of player $i \in [\![n]\!]$ under that joint action; the range of the utilities here can be normalized to be in $[-1, 1]$ without any loss of generality. Each player $i \in [\![n]\!]$ is allowed to randomize by selecting a (mixed) strategy, a distribution over the available actions: $\boldsymbol{x}_i \in \Delta(\mathcal{A}_i) \triangleq \{\boldsymbol{x}_i \in \mathbb{R}_{\geq 0}^{\mathcal{A}_i} : \sum_{a_i \in \mathcal{A}_i} \boldsymbol{x}_i[a_i] = 1\}$. For a joint strategy $(\boldsymbol{x}_1, \dots, \boldsymbol{x}_n) \in \bigtimes_{i=1}^n \Delta(\mathcal{A}_i)$, we will denote by $\bigotimes_{i=1}^n \boldsymbol{x}_i$ the *product* distribution on $\Delta(\bigtimes_{i=1}^n \mathcal{A}_i)$ defined so that $(\bigotimes_{i=1}^n \boldsymbol{x}_i)[(a_1, \dots, a_n)] \triangleq \prod_{i=1}^n \boldsymbol{x}_i[a_i]$.

We are now ready to recall the standard concept of an approximate coarse correlated equilibrium (CCE) [64, 2]. Below, for a joint action $\boldsymbol{a} = (a_1, \dots, a_n) \in \bigtimes_{i=1}^n \mathcal{A}_i$, we use the usual shorthand notation $\boldsymbol{a}_{-i} \triangleq (a_1, \dots, a_{i-1}, a_{i+1}, \dots, a_n) \in \bigtimes_{i' \neq i} \mathcal{A}_{i'}$.

▶ **Definition 2.1** (Coarse correlated equilibrium). *Let $\mathcal{G}$ be an n-player game in normal form. A distribution over joint action profiles $\boldsymbol{\mu} \in \Delta\left(\bigtimes_{i=1}^n \mathcal{A}_i\right)$ is said to be an* ϵ-coarse correlated equilibrium (ϵ-CCE)*, with $\epsilon \in \mathbb{R}$, if for any player $i \in [\![n]\!]$ and any deviation $a_i' \in \mathcal{A}_i$,*[2]

$$\mathop{\mathbb{E}}_{\boldsymbol{a} \sim \boldsymbol{\mu}}[u_i(\boldsymbol{a})] \geq \mathop{\mathbb{E}}_{\boldsymbol{a} \sim \boldsymbol{\mu}}[u_i(a_i', \boldsymbol{a}_{-i})] - \epsilon. \tag{1}$$

For convenience, we will sometimes use the shorthand notation $u_i(\boldsymbol{\mu}) \triangleq \mathbb{E}_{\boldsymbol{a} \sim \boldsymbol{\mu}}[u_i(\boldsymbol{a})]$ and $u_i(a_i', \boldsymbol{\mu}_{-i}) \triangleq \mathbb{E}_{\boldsymbol{a} \sim \boldsymbol{\mu}}[u_i(a_i', \boldsymbol{a}_{-i})]$. A CCE is typically modeled via a trusted third party – a so-called *mediator* – who privately makes recommendations to each player; (1) guarantees that no player can gain more than an additive factor of ϵ through a (unilateral) deviation, *before* actually observing the mediator's recommendation. A 0-CCE will simply be referred to as a CCE. In this context, we will be concerned with a refinement of Definition 2.1 wherein a certain sparsity constraint is imposed, in the following formal sense.

▶ **Definition 2.2** (Sparse CCE). *Let $\mathcal{G}$ be an n-player game in normal form. A distribution over joint action profiles $\boldsymbol{\mu} \in \Delta\left(\bigtimes_{i=1}^n \mathcal{A}_i\right)$ satisfying Definition 2.1 is said to be* k-sparse *if it can be expressed as a uniform mixture of k product distributions; that is, there exist $(\boldsymbol{x}_1^{(1)}, \dots, \boldsymbol{x}_n^{(1)}), \dots, (\boldsymbol{x}_1^{(k)}, \dots, \boldsymbol{x}_n^{(k)}) \in \bigtimes_{i=1}^n \Delta(\mathcal{A}_i)$ such that $\boldsymbol{\mu} = \frac{1}{k} \sum_{\kappa=1}^k \bigotimes_{i=1}^n \boldsymbol{x}_i^{(\kappa)}$.*

From a computational standpoint, a $\mathsf{poly}(n, m)$-sparse CCE can be identified in polynomial time for any game of *polynomial type* – meaning that m is a polynomial with respect to the underlying description – satisfying the polynomial expectation property [66, 53]; the latter property postulates that for any product distribution $\boldsymbol{x}$ with a polynomial representation,

[2] As is standard, we abuse notation by parsing $u_i(a_i', \boldsymbol{a}_{-i})$ as $u_i(a_1, \dots, a_{i-1}, a_i', a_{i+1}, \dots, a_n)$; the same convention is adopted for the mixed extension of the utilities as well.

the expectation $\mathbb{E}_{\boldsymbol{a}\sim\boldsymbol{x}}[u_i(\boldsymbol{a})]$ can be computed in time $\mathsf{poly}(n, m)$, an assumption known to be satisfied in most succinct games of interest [66]. On the other end of the spectrum, a 1-sparse CCE is, by definition, a Nash equilibrium, thereby making the problem PPAD-hard [25]. Furthermore, specific no-regret learning algorithms, such as multiplicative weights update, yield an $O(\frac{\log m}{\epsilon^2})$-sparse ϵ-CCE in time $\mathsf{poly}(n, m, 1/\epsilon)$, again under the polynomial expectation property. As a result, the key question that arises is to characterize the threshold of computational tractability in terms of the sparsity parameter k.

Online learning in games

Before we proceed, we also point out the folklore connection between no-regret learning and CCE. In the online learning framework with full feedback, every repetition $t \in \mathbb{N}$ finds a player $i \in [\![n]\!]$ selecting an action $\boldsymbol{x}_i^{(t)} \in \Delta(\mathcal{A}_i)$, and subsequently observing as feedback from the environment the utility function $\boldsymbol{x}_i \mapsto \langle \boldsymbol{x}_i, \boldsymbol{u}_i^{(t)} \rangle$, where $\boldsymbol{u}_i^{(t)} \in [-1, 1]^{\mathcal{A}_i}$. The *regret* of player i under a time horizon $T \in \mathbb{N}$ is defined as

$$\mathsf{Reg}_i^T \triangleq \max_{\boldsymbol{x}_i^\star \in \Delta(\mathcal{A}_i)} \sum_{t=1}^{T} \langle \boldsymbol{x}_i^\star - \boldsymbol{x}_i^{(t)}, \boldsymbol{u}_i^{(t)} \rangle.$$

Specifically, in the setting of learning in games, the utility $\boldsymbol{u}_i^{(t)}$ observed by player $i \in [\![n]\!]$ is defined so that

$$\boldsymbol{u}_i^{(t)}[a_i] \triangleq \mathop{\mathbb{E}}_{\boldsymbol{a}_{-i}\sim\boldsymbol{x}_{-i}^{(t)}} [u_i(a_i, \boldsymbol{a}_{-i})], \tag{2}$$

where $\boldsymbol{x}_{-i}^{(t)}$ is the joint strategy of the other players at time t. In this context, the connection between CCE and no-regret learning is summarized in the following folklore fact [11, 17].

▶ **Proposition 2.3.** *Consider an n-player game in normal form, and suppose that each player $i \in [\![n]\!]$ produces the sequence of strategies $(\boldsymbol{x}_i^{(t)})$ under the sequence of utilities given by* (2)*. If each player i incurs regret Reg_i^T after $T \in \mathbb{N}$ repetitions, then the correlated distribution $\bar{\boldsymbol{\mu}} \triangleq \frac{1}{T}\sum_{t=1}^{T} \bigotimes_{i=1}^{n} \boldsymbol{x}_i^{(t)}$ is a $\frac{1}{T}\max_{1\le i\le n} \mathsf{Reg}_i^T$-CCE.*

2.2 Extensive-form games

While every finite game can be represented in normal form, such a representation can be dramatically inefficient in more structured classes of games. Specifically, in scenarios involving sequential moves and imperfect information, the canonical representation is the *extensive form* [75]. In such games, a rooted and directed tree $\mathcal{T}$ is explicitly given as part of the input. We let $\mathcal{H} = \mathcal{H}(\mathcal{T})$ denote the set of non-leaf nodes of $\mathcal{T}$. Each node $\tau \in \mathcal{H}$ that is not a leaf of $\mathcal{T}$ is uniquely associated with a player $i \in [\![n]\!]$ who selects an action from a finite and nonempty set of available actions $\mathcal{A}_\tau$;[3] the set of all nodes where player i acts will be denoted by $\mathcal{H}_i$. The leaves of the tree $\mathcal{T}$, which are also referred to as terminal nodes, are denoted by $\mathcal{Z}$. When the game transitions to a terminal node in $\mathcal{Z}$, utilities are assigned to each player i, as specified by an arbitrary utility function $u_i : \mathcal{Z} \to [-1, 1]$.

[3] In general, extensive-form games also feature *chance moves* (for example, the roll of a dice), which can be modeled via an additional fictitious "player;" our lower bounds in the sequel do not have to involve chance moves. Nevertheless, it is worth noting that the addition of chance moves is known to crucially affect the computational complexity of certain problems [83].

To model imperfect information, the set of nodes $\mathcal{H}_i$ belonging to player i is partitioned into *information sets* $\mathcal{J}_i$; each information set groups together nodes that player i cannot distinguish based on the information structure of the game. As is standard, we also tacitly assume throughout this paper that players have *perfect recall*, in that players never forget acquired information; in the absence of perfect recall, many natural problems immediately become NP-hard [56, 81, 88, 21]. In what follows, we will use $\mathcal{T}$ to represent the underlying extensive-form game, with the understanding that $\mathcal{T}$ indeed encodes all the information pertinent for its complete description.

▶ Remark 2.4 (Simultaneous moves). Although the aforedescribed standard formulation of extensive-form games features solely sequential moves, in that each node is associated with a single player, one can readily model simultaneous moves as well through the use of imperfect information. We will use this standard fact in the sequel to also incorporate simultaneous moves in order to simplify the exposition.

A strategy for a player $i \in [\![n]\!]$ is a mapping $\mathcal{J}_i \ni j \to \Delta(\mathcal{A}_j)$. It turns out that the set of each player's strategies can be represented compactly via a convex polytope [82, 73], namely the *sequence-form polytope* $\mathcal{X}_i$, so that the utility of each player can be expressed as a linear function (assuming that the rest of the players are fixed). This has been a crucial observation for designing efficient algorithms for a number of fundamental problems in extensive-form games. With a slight abuse of notation, for a sequence-form strategy $\boldsymbol{x}_i \in \mathcal{X}_i$ we will write $\boldsymbol{x}_{i,j}$ to denote the probability distribution over $\Delta(\mathcal{A}_j)$ induced by $\boldsymbol{x}_i$ at information set $j \in \mathcal{J}_i$; if $\boldsymbol{x}_{i,j}$ is not well-defined, in that $\boldsymbol{x}_i$ assigns zero probability to the subtree of j, we may take $\boldsymbol{x}_{i,j}$ to be an arbitrary distribution. For a joint strategy $(\boldsymbol{x}_1, \dots, \boldsymbol{x}_n) \in \bigtimes_{i=1}^n \mathcal{X}_i$, we use again the notation $\bigotimes_{i=1}^n \boldsymbol{x}_i$ the express the product distribution on the induced normal-form game, which is always represented implicitly.

Extensive-form games are not of polynomial type in that a player's number of pure stratagies is typically exponential in the description of the game, rendering the induced normal-form representation largely inefficient. Nevertheless, polynomial (in the size of the tree $\mathcal{T}$) algorithms for computing (exact) CCE are known to exist. In particular, Huang and von Stengel [52] have shown how to adapt the algorithm of Papadimitriou and Roughgarden [66] for certain correlated equilibrium concepts in extensive-form games. (CCE per Definition 2.2 is typically referred to as *normal-form* CCE (NFCCE) to differentiate with other notions of coarse correlation in extensive-form games [29].) We clarify that a CCE in extensive-form games can be indeed defined via Definition 2.1 through the induced normal-form representation. Our results here revolve around the complexity of the more refined concept introduced in Definition 2.2.

2.3 Online density estimation

We finally conclude this section by recalling some basic results regarding online density estimation, which will be useful in the sequel. Following Foster et al. [41], our proof will make use of *Vovk's aggregating algorithm* for online density estimation [84]. More precisely, the setting here is as follows. There are two players, the *nature* and the *learner*. There is also a set $\mathcal{O}$ called the *outcome space*, and a set $\mathcal{C}$ referred to as the *context space*; both can be assumed to be finite for our applications. The interaction between the learner and the nature proceeds for $h = 1, 2, \dots, H$ as follows.

1. Nature first reveals a context $c_h \in \mathcal{C}$;
2. the learner then predicts a distribution $\widehat{\boldsymbol{q}}_h \in \Delta(\mathcal{O})$ over outcomes based on the observed context c_h; and

3. Nature chooses an outcome $o_h \in \mathcal{O}$, and the learner incurs a logarithmic loss defined as $\ell_h(\widehat{\boldsymbol{q}}_h) \triangleq \log\left(\frac{1}{\widehat{\boldsymbol{q}}_h(o_h)}\right)$.

We measure the performance of the learner via the *regret* against a (finite) set of experts $\mathcal{E}$. In particular, every expert $e \in \mathcal{E}$ corresponds to a function $\boldsymbol{p}^{(e)} : \mathcal{C} \to \Delta(\mathcal{O})$, so that the regret of an algorithm with respect to the expert class $\mathcal{E}$ is defined as

$$\mathsf{Reg}_{\mathcal{E}}^{H} \triangleq \sum_{h=1}^{H} \ell_h(\widehat{\boldsymbol{q}}_h) - \min_{e\in\mathcal{E}} \left\{ \sum_{h=1}^{H} \ell_h\left(\boldsymbol{p}^{(e)}(c_h)\right) \right\}.$$

It is important to note here that the algorithm of the learner has access to the expert predictions $\{\boldsymbol{p}^{(e)}(c_h)\}_{e\in\mathcal{E}}$. In this context, Vovk's aggregating algorithm makes predictions for $h = 1, 2, \ldots, H$ via

$$\widehat{\boldsymbol{q}}_h \triangleq \mathop{\mathbb{E}}_{e\sim\widetilde{\boldsymbol{q}}_h}[\boldsymbol{p}^{(e)}(c_h)], \text{ where } \widetilde{\boldsymbol{q}}_h^{(e)} \triangleq \frac{\exp\left(-\sum_{v=1}^{h-1} \ell_v(\boldsymbol{p}^{(e)}(c_v))\right)}{\sum_{e'\in\mathcal{E}} \exp\left(-\sum_{v=1}^{h-1} \ell_v(\boldsymbol{p}^{(e')}(c_v))\right)} \quad \forall e \in \mathcal{E}. \tag{3}$$

The convention above is that a summation with no terms is defined as 0, so that $\widetilde{\boldsymbol{q}}_1$ is the uniform distribution over $\mathcal{E}$. We further take $\log(\frac{1}{0^+}) = +\infty$ and $\exp(-\infty) = 0$; under realizability (see Proposition 2.5), the denominator in (3) can never be 0, so (3) is indeed well-defined.

The main guarantee we will use for the aggregation algorithm (3) is summarized below. We recall first that the total variation distance between two discrete distributions $\boldsymbol{p}, \boldsymbol{q} \in \Delta(\mathcal{O})$ is defined as $D_{\mathrm{TV}}(\boldsymbol{p}, \boldsymbol{q}) \triangleq \frac{1}{2}\|\boldsymbol{p} - \boldsymbol{q}\|_1$.

▶ **Proposition 2.5** ([84]). *Suppose that the distribution of outcomes is realizable under some $e^\star \in \mathcal{E}$; that is, $o_h \sim \boldsymbol{p}^{(e^\star)}(c_h) \mid c_h$ for each $h \in [\![H]\!]$. Then, the predictions $(\widehat{\boldsymbol{q}}_h)_{1\le h\le H}$ produced by the aggregation algorithm* (3) *satisfy*

$$\frac{1}{H}\sum_{h=1}^{H} \mathbb{E}\left[D_{\mathrm{TV}}(\widehat{\boldsymbol{q}}_h, \boldsymbol{p}^{(e^\star)}(c_h))\right] \le \sqrt{\frac{\log|\mathcal{E}|}{H}},$$

where the expectation above is with respect to the underlying random process whereby nature selects the sequence of contexts $(c_1, \ldots, c_H) \in \mathcal{C}^H$.

3 Lower Bounds for No-Regret Learning in Games

In this section, we present our main results regarding the problem of computing sparse CCE in extensive-form games, as well as the implied lower bounds for no-regret learning in games.

To do so, we build on the reduction of Foster et al. [41] targeting Markov (aka. stochastic) games. In particular, we assume that we are given as input a two-player general-sum game $\mathcal{G}$ where each player has $m \in \mathbb{N}$ actions; that is, $|\mathcal{A}_1| = |\mathcal{A}_2| = m \ge 2$. We may also posit that every entry in the payoff matrices, say $\mathbf{M}_1, \mathbf{M}_2 \in \mathbb{Q}^{\mathcal{A}_1\times\mathcal{A}_2}$, can be represented with a number of bits polynomial in m. Further, we assume without any loss of generality that $|\mathbf{M}_1[a_1, a_2]|, |\mathbf{M}_2[a_1, a_2]| \le 1$, for any combination of actions $(a_1, a_2) \in \mathcal{A}_1 \times \mathcal{A}_2$. The key idea of the reduction is to show that a sparse CCE in a suitably constructed extensive-form game $\mathcal{T} = \mathcal{T}(\mathcal{G})$ (described in Section 3.1) can be used to obtain a Nash equilibrium in the original game $\mathcal{G}$; in turn, the computational hardness of Nash equilibria in two-player games [74] will preclude polynomial-time computation of sparse CCE under a certain sparsity regime (Theorem 1.2).

In what follows, Section 3.1 formally introduces the lifted extensive-form game $\mathcal{T}(\mathcal{G})$; Section 3.2 describes the process whereby a sparse CCE in $\mathcal{T}$ yields a Nash equilibrium in $\mathcal{G}$ (Algorithm 1); Section 3.3 establishes the correctness of Algorithm 1; and Section 3.4 provides the main implications for computing sparse CCE in extensive-form games, as well as no-regret learning in normal- and extensive-form games.

3.1 The lifted extensive-form game

Foster et al. [41] introduce two separate reductions in order to prove hardness results in Markov games, which differ depending on whether players' policies are allowed to be Markovian or not. In extensive-form games, strategies are of course not constrained to be Markovian since they can depend arbitrarily on the information available to that player. Accordingly, we will adapt the reduction of Foster et al. [41] that targets non-Markovian policies, which in turn is based on the reduction of Borgs et al. [12], leading to the lifted game described in this subsection.

As we explained in Remark 2.4, it will be convenient for our exposition to work with extensive-form games that include simultaneous moves; again, this comes without any essential loss since simultaneous moves can always be cast as sequential moves using imperfect information, a transformation that does not qualitatively alters our results.

Now, let $\mathcal{G}$ be the original two-player game in normal form. The basic idea is to construct an extensive-form game $\mathcal{T} = \mathcal{T}(\mathcal{G})$ consisting of H repetitions of $\mathcal{G}$, for a sufficiently large parameter $H \in \mathbb{N}$ to be specified later (Theorem 3.5). Following the approach of Foster et al. [41], a key ingredient is the addition of an auxiliary player, namely the *Kibitzer*, which is in turn based on the hardness result of Borgs et al. [12] pertaining the computation of Nash equilibria in repeated games. Specifically, Borgs et al. [12] reduced computing Nash equilibria in two-player normal-form games to computing Nash equilibria in three-player repeated games, thereby establishing that – the folk theorem notwithstanding – the latter problem is hard. Interestingly, this is not the case for two-player repeated games where polynomial-time algorithms do exist [60]; this suggests that proving hardness results for two-player extensive-form games could require a very different approach. So, returning to our reduction, $\mathcal{T}$ here is a three-player (extensive-form) game. By convention, player $\mathsf{K} \triangleq 3$ will represent the Kibitzer; we often use the symbol K instead of the index $i = 3$ for convenience in the presentation.

In each possible decision node (or simply state) $s \in \mathcal{S}$ of $\mathcal{T}$ each player simultaneously selects an action.[4] Specifically, each of the first two players select actions from $\mathcal{A}_1$ and $\mathcal{A}_2$, respectively (where those action sets are as given in the original game $\mathcal{G}$), while the action set of the Kibitzer, $\mathcal{A}_{\mathsf{K}}$, is defined as

$$\mathcal{A}_{\mathsf{K}} \triangleq \{(i, a_i) : i \in [2], a_i \in \mathcal{A}_i\}.$$

As such, each state $s \in \mathcal{S}$ is in bijective correspondence with a sequence of joint actions; it is critical in this construction that each player gets to observe the other players' actions from earlier rounds, for reasons that will become clear shortly. Further, the utilities of the players are then defined as follows. For a repetition $h \in [\![H]\!]$ and a joint action profile $(a_{1,h}, a_{2,h}, a_{\mathsf{K},h}) \in \mathcal{A}_1 \times \mathcal{A}_2 \times \mathcal{A}_{\mathsf{K}}$, with $a_{\mathsf{K},h} = (1, a'_{1,h})$, we define

[4] Here, we denote decision nodes with the symbol $\mathcal{S}$ instead of $\mathcal{H}$ as in Section 2.2 because $\mathcal{T}$ features simultaneous moves as well. We clarify that $\mathcal{S}$ contains precisely the information sets of each player in the sequential representation.

$$u_{i,h}(a_{1,h}, a_{2,h}, a_{\mathsf{K},h}) \triangleq \begin{cases} 0 & : i \neq 1, \mathsf{K}, \\ \frac{1}{H}\left(\mathbf{M}_1[a_{1,h}, a_{2,h}] - \mathbf{M}_1[a'_{1,h}, a_{2,h}]\right) & : i = 1, \\ \frac{1}{H}\left(\mathbf{M}_1[a'_{1,h}, a_{2,h}] - \mathbf{M}_1[a_{1,h}, a_{2,h}]\right) & : i = \mathsf{K}; \end{cases}$$

the utility functions are defined symmetrically when $a_{\mathsf{K},h} = (2, a'_{2,h})$. Specifically, we assume here that those rewards are given to the corresponding node in the game tree; while it is common – as we described earlier in Section 2.2 – to assign utilities only at leaf nodes, it is clear that one can always push all the utilities in the corresponding leaf nodes without altering the equilibria of the game. Indeed, for a sequence of joint actions $(\boldsymbol{a}_1, \dots, \boldsymbol{a}_H)$, which uniquely specifies a leaf node $z \in \mathcal{Z}$, the cumulative utility can be defined as $u_i : \mathcal{Z} \ni z \mapsto \sum_{h=1}^{H} u_{i,h}(a_{1,h}, a_{2,h}, a_{\mathsf{K},h})$. We note that normalizing by the factor H in $u_{i,h}(\cdot)$ above ensures that the cumulative payoffs in $\mathcal{T}$ are indeed in $[-1, 1]$. We further remark that $\mathcal{T}$ is a zero-sum game since $\sum_{i=1}^{3} u_i(z) = 0$, for any $z \in \mathcal{Z}$. Finally, it is evident that $\mathcal{T}$ is indeed a perfect-recall game.

We next state a straightforward fact, which follows directly from the definition of each utility function $u_{i,h}(\cdot)$.

▶ **Lemma 3.1.** *For any repetition $h \in [\![H]\!]$, player $i \in [\![3]\!]$, and strategies $\boldsymbol{x}_{-i,h} \in \bigtimes_{i' \neq i} \Delta(\mathcal{A}_{i'})$, it holds that $\max_{a_{i,h} \in \mathcal{A}_i} \mathbb{E}_{\boldsymbol{a}_{-i,h} \sim \boldsymbol{x}_{-i,h}}[u_{i,h}(a_{i,h}, \boldsymbol{a}_{-i,h})] \geq 0$.*

Another simple but important observation regarding the representation of $\mathcal{T}$ is the following bound on the number of nodes of $\mathcal{T}$, which will be represented as $|\mathcal{T}|$.

▷ Claim 3.2. Let $\mathcal{G}$ be a two-player m-action game. For the induced extensive-form game $\mathcal{T} = \mathcal{T}(\mathcal{G})$ it holds that $|\mathcal{T}| \leq 2^{H+1} m^{3H+3}$.

Proof. It is clear from our construction of the extensive-form game $\mathcal{T}$ that $|\mathcal{T}|$ can be expressed as $1 + 2m^3 + \cdots + (2m^3)^H \leq 2^{H+1} m^{3H+3}$. ◁

In particular, the description of the extensive-form game $\mathcal{T}$ is polynomial in the description of $\mathcal{G}$ when H is an absolute constant. In stark contrast, it is important to point out that the normal-form representation of $\mathcal{T}$ is exponential even if $H = 2$. Indeed, each player would have to specify an action in each of $1 + 2m^3$ decision nodes, which leads to at least m^{m^3} combinations in the normal-form representation for each player; this is why proving non-trivial hardness results for normal-form games appears to require a different approach. We also remark that the bound $m^{\Theta(H)}$ of Claim 3.2 clearly holds after we convert $\mathcal{T}$ into a sequential-move game, which suffices for our proof to carry over without simultaneous moves.

3.2 The algorithm

Based on the extensive-form game $\mathcal{T}$ described in Section 3.1, our main reduction is summarized in Algorithm 1. Before we proceed, let us make some clarifications. First, the function StateToSeq$(\cdot)$ in Line 7 takes as input a state $s_h \in \mathcal{S}_h$ corresponding to the hth repetition, and returns the unique sequence of joint actions $(\boldsymbol{a}_1, \dots, \boldsymbol{a}_{h-1})$ that leads to that state; if $h = 1$, we can assume that it returns the empty sequence. Further, the function PrevStates$(\cdot)$ in Line 8 takes again as input a state $s_h \in \mathcal{S}_h$ and returns the unique sequence of preceding states $(s_1, \dots, s_{h-1}) \in \mathcal{S}_1 \times \cdots \times \mathcal{S}_{h-1}$; if $h = 1$, this function is again assumed to return the empty sequence. With those semantics in mind, we point out that the condition in Line 10 is activated if and only if there exists $t \in [\![T]\!]$ such that

$\boldsymbol{x}^{(t)}_{i,s_v}[a_{i,v}] > 0$, for all $v = 1, 2, \dots h-1$; in the contrary case, the corresponding part of the tree is reached with probability 0 under the random process of interest (as defined in the proof of Theorem 3.5 below), in which case we may set $\widetilde{\boldsymbol{q}}_{i,s_h}$ to an arbitrary distribution over $[\![T]\!]$ (*e.g.*, the uniform as in Line 13).

Algorithm 1 Reduction for Theorem 3.5.

1 **Input**: Two-player m-action game $\mathcal{G}$ in normal form; accuracy $\epsilon > 0$; sparsity $T \in \mathbb{N}$
2 **Output**: A (9ϵ)-Nash equilibrium of $\mathcal{G}$
3 Construct the three-player extensive-form game $\mathcal{T}(\mathcal{G})$ with $H \geq \frac{\log T}{\epsilon^2}$ (Section 3.1)
4 Compute a T-sparse ϵ-CCE $\frac{1}{T}\sum_{t=1}^{T} \bigotimes_{i=1}^{3} \boldsymbol{x}_i^{(t)}$ (Definition 2.2) in $\mathcal{T}$
5 **for** $h \in [\![H]\!]$ **do**
6 **for** $s_h \in \mathcal{S}_h$ **do**
7 $(\boldsymbol{a}_1, \dots, \boldsymbol{a}_{h-1}) \triangleq \textsc{StateToSeq}(s_h)$
8 $(s_1, \dots, s_{h-1}) \triangleq \textsc{PrevStates}(s_h)$
9 **for** $i \in [\![2]\!]$ **do**
10 **if** $\sum_{t=1}^{T} \exp\left(-\sum_{v=1}^{h-1} \log\left(\frac{1}{\boldsymbol{x}^{(t)}_{i,s_v}[a_{i,v}]}\right)\right) > 0$ **then**
11 Let
$$\widetilde{\boldsymbol{q}}^{(t)}_{i,s_h} \triangleq \frac{\exp\left(-\sum_{v=1}^{h-1} \log\left(\frac{1}{\boldsymbol{x}^{(t)}_{i,s_v}[a_{i,v}]}\right)\right)}{\sum_{t'=1}^{T} \exp\left(-\sum_{v=1}^{h-1} \log\left(\frac{1}{\boldsymbol{x}^{(t')}_{i,s_v}[a_{i,v}]}\right)\right)} \quad \forall t \in [\![T]\!]$$
12 **else**
13 $\widetilde{\boldsymbol{q}}_{i,s_h} \triangleq \mathsf{U}([\![T]\!])$
14 $\widehat{\boldsymbol{q}}_{i,s_h} \triangleq \mathbb{E}_{t\sim\widetilde{\boldsymbol{q}}_{i,s_h}}[\boldsymbol{x}^{(t)}_{i,s_h}] \in \Delta(\mathcal{A}_i)$
15 **if** $(\widehat{\boldsymbol{q}}_{1,s_h}, \widehat{\boldsymbol{q}}_{2,s_h}) \in \Delta(\mathcal{A}_1) \times \Delta(\mathcal{A}_2)$ *is a* (9ϵ)*-Nash equilibrium of* $\mathcal{G}$ **then**
16 **return** $(\widehat{\boldsymbol{q}}_{1,s_h}, \widehat{\boldsymbol{q}}_{2,s_h}) \in \Delta(\mathcal{A}_1) \times \Delta(\mathcal{A}_2)$
17 **return** FAIL

It is evident that as long as $T = \mathsf{poly}(|\mathcal{T}|)$, all steps in Algorithm 1 can be implemented in time polynomial in the description of $\mathcal{T}$, with the exception of Line 4, which of course depends on the underlying algorithm used to compute a sparse CCE. Along with Claim 3.2, we arrive at the following conclusion.

▶ **Proposition 3.3.** *Let $\mathfrak{A}$ be an algorithm that takes as input an extensive-form game $\mathcal{T}$ and computes a T-sparse ϵ-CCE of $\mathcal{T}$ in time at most $Q(|\mathcal{T}|, T, 1/\epsilon)$. Then, Algorithm 1 instantiated with $\mathfrak{A}$ in Line 4 runs in time at most $Q(|\mathcal{T}|, T, 1/\epsilon) + Tm^{\Theta(H)}$.*

▶ Remark 3.4 (Bit complexity of exponential weights). Line 11 of Algorithm 1 updates $\widetilde{\boldsymbol{q}}^{(t)}_{i,s_h}$ using exponential weights (in accordance with the aggregation algorithm (3)), which could result in $\widetilde{\boldsymbol{q}}^{(t)}_{i,s_h}$ taking irrational values. This can be addressed by simply truncating those values to a sufficiently large polynomial number of bits, in which case the proof of Theorem 3.5 readily carries over. For simplicity, we assume in our analysis that $\widetilde{\boldsymbol{q}}^{(t)}_{i,s_h}$ is updated per Line 11, without taking into account the numerical imprecision.

It is worth commenting here on a couple of differences with [41, Algorithm 2]. First, Foster et al. [41] had to encode the joint action profile through the reward, so as to ensure that each player has observed the prior sequence of joint actions. This is not necessary in our setting since, by construction, the states of the extensive-form game $\mathcal{T}$ encode that information. Further, their algorithm is randomized since – among other steps – they sample a randomized trajectory under a certain random process. To obtain a deterministic algorithm, we instead essentially search over all possible trajectories – all states of the extensive-form game $\mathcal{T}$ – for a Nash equilibrium, which we can afford in our setting.

3.3 From sparse CCEin $\mathcal{T}$ to Nash equilibria in $\mathcal{G}$

We are now ready to proceed with the key proof of this section, which establishes the correctness of Algorithm 1.

▶ **Theorem 3.5.** *When $H \geq \frac{\log T}{\epsilon^2}$, Algorithm 1 returns a (9ϵ)-Nash equilibrium in the two-player m-action game $\mathcal{G}$.*

Proof. We consider a sequence of joint strategies $(\boldsymbol{x}_1^{(1)}, \boldsymbol{x}_2^{(1)}, \boldsymbol{x}_3^{(1)}), \dots, (\boldsymbol{x}_1^{(T)}, \boldsymbol{x}_2^{(T)}, \boldsymbol{x}_3^{(T)}) \in \bigtimes_{i=1}^3 \mathcal{X}_i$ in the extensive-form game $\mathcal{T}$ with the property that $\bar{\boldsymbol{\mu}} \triangleq \frac{1}{T}\sum_{t=1}^T \bigotimes_{i=1}^3 \boldsymbol{x}_i^{(t)}$ is an ϵ-CCE, and by construction T-sparse per Definition 2.2.

Let us fix a player $i \in [\![3]\!]$. For each state $s \in \mathcal{S}$, we define $\widetilde{\boldsymbol{q}}_{i,s} \in \Delta([\![T]\!])$ per Line 11; $\widehat{\boldsymbol{q}}_{i,s} \in \Delta(\mathcal{A}_i)$ per Line 14; and the deviation strategy $\boldsymbol{x}_i^\dagger \in \mathcal{X}_i$ so that for each state $s \in \mathcal{S}$ it holds that $\boldsymbol{x}_{i,s}^\dagger \triangleq \operatorname{argmax}_{a_{i,s} \in \mathcal{A}_i} \mathbb{E}_{\boldsymbol{a}_{-i,s} \sim \widehat{\boldsymbol{q}}_{-i,s}}[u_{i,h}(a_{i,s}, \boldsymbol{a}_{-i,s})]$. We will now make use of Proposition 2.5 regarding the aggregation algorithm (3) under a certain random process to be described shortly. In particular, under a different player $i' \neq i$, to relate our problem with the setup of online density estimation introduced earlier, we make the following correspondence:

- the context space $\mathcal{O}$ corresponds to the set of all possible states or decision nodes $\mathcal{S}$ of the extensive-form game $\mathcal{T}$;
- the set of experts $\mathcal{E}$ coincides with the set $\{\boldsymbol{x}_{i',s}^{(1)}, \dots, \boldsymbol{x}_{i',s}^{(T)}\}$, with outcome space $\mathcal{A}_{i'}$; and
- the time index $h \in [\![H]\!]$ in the context of online density estimation will now (fittingly) correspond to the repetition $h \in [\![H]\!]$.

We note that, by construction of the extensive-form game $\mathcal{T}$, player i observes the underlying state s_h at each repetition $h \in [\![H]\!]$, which fully specifies the sequence of joint actions leading up to that state. As a result, under a given random sequence of states $(s_1, \dots, s_H) \in \mathcal{S}^H$, we can apply the aggregation algorithm (3) with the aforementioned parameterization to obtain an estimate $\widehat{\boldsymbol{q}}_{i',s_h} \in \Delta(\mathcal{A}_{i'})$ for all repetitions $h \in [\![H]\!]$ and $i' \neq i$. Below, we overload the notation by letting $\widehat{\boldsymbol{q}}_{i',h} \triangleq \widehat{\boldsymbol{q}}_{i',s_h}$ and $\widetilde{\boldsymbol{q}}_{i',h} \triangleq \widetilde{\boldsymbol{q}}_{i',s_h}$ so as to be consistent with the notation of Section 2.3. Namely, we have that

$$\widehat{\boldsymbol{q}}_{i',h} \triangleq \mathop{\mathbb{E}}_{t \sim \widetilde{\boldsymbol{q}}_{i',h}}[\boldsymbol{x}_{i',s_h}^{(t)}], \text{ where } \widetilde{\boldsymbol{q}}_{i',h}^{(t)} \triangleq \frac{\exp\left(-\sum_{v=1}^{h-1} \ell_v(\boldsymbol{x}_{i',s_v}^{(t)})\right)}{\sum_{t'=1}^T \exp\left(-\sum_{v=1}^{h-1} \ell_v(\boldsymbol{x}_{i',s_v}^{(t')})\right)} \quad \forall t \in [\![T]\!].$$

Given that the sequence of states $(s_1, \dots, s_H)$ is produced by a certain random process (described next), $\widehat{\boldsymbol{q}}_{i',h}$ and $\widetilde{\boldsymbol{q}}_{i',h}$ are random variables. We also recall that $\ell_v(\boldsymbol{x}_{i',s_v}^{(t)}) = \log \frac{1}{\boldsymbol{x}_{i',s_v}^{(t)}[a_{i',v}]}$. Accordingly, the deviation $\boldsymbol{x}_{i,h}^\dagger \in \Delta(\mathcal{A}_i)$ for player $i \in [\![3]\!]$ is defined as follows:

$$\boldsymbol{x}_{i,h}^\dagger \triangleq \operatorname*{argmax}_{a_{i,h} \in \mathcal{A}_i} \mathop{\mathbb{E}}_{\boldsymbol{a}_{-i,h} \sim \widehat{\boldsymbol{q}}_{-i,h}}[u_{i,h}(a_{i,h}, \boldsymbol{a}_{-i,h})]. \tag{4}$$

We next argue about the deviation benefit of player i under the deviation strategy described above. In particular, we are interested in the payoff player i obtains under the random process wherein we first draw an index $t^\star$ uniformly at random from the set $[\![T]\!]$, player i plays according to the deviation strategy $\boldsymbol{x}_i^\dagger$, while the rest of the players play according to $\boldsymbol{x}_{-i}^{(t^\star)}$. By definition of this random process, realizability is satisfied: the observed distribution of outcomes obeys the law induced by $\boldsymbol{x}_{i'}^{(t^\star)} : \mathcal{S} \to \Delta(\mathcal{A}_{i'})$, conditioned on the time index $t^\star$. As a result, Proposition 2.5 implies that for each player $i \in [\![3]\!]$ and $i' \neq i$ it holds that

$$\mathop{\mathbb{E}}_{\boldsymbol{x}_i^\dagger \times \boldsymbol{x}_{-i}^{(t^\star)}} \left[\sum_{h=1}^{H} D_{\mathrm{TV}} \left(\widehat{\boldsymbol{q}}_{i',h}, \boldsymbol{x}_{i',s_h}^{(t^\star)} \right) \right] \leq \sqrt{H \log T}, \tag{5}$$

where the expectation and the sequence of states $(s_1, \dots, s_H) \in \mathcal{S}^H$ is taken with respect to the random process described above. As a result, we have that

$$\begin{aligned}
u_i(\boldsymbol{x}_i^\dagger, \bar{\boldsymbol{\mu}}_{-i}) &= \mathop{\mathbb{E}}_{t^\star \sim \mathsf{U}([\![T]\!])} [u_i(\boldsymbol{x}_i^\dagger, \boldsymbol{x}_{-i}^{(t^\star)})] = \mathop{\mathbb{E}}_{t^\star \sim \mathsf{U}([\![T]\!])} \mathop{\mathbb{E}}_{\boldsymbol{x}_i^\dagger, \boldsymbol{x}_{-i}^{(t^\star)}} \sum_{h=1}^{H} \mathop{\mathbb{E}}_{\boldsymbol{a}_{-i,h} \sim \boldsymbol{x}_{-i,s_h}^{(t^\star)}} [u_{i,h}(\boldsymbol{x}_{i,h}^\dagger, \boldsymbol{a}_{-i,h})] \\
&\geq -2\sqrt{\frac{\log T}{H}} + \mathop{\mathbb{E}}_{t^\star \sim \mathsf{U}([\![T]\!])} \mathop{\mathbb{E}}_{\boldsymbol{x}_i^\dagger, \boldsymbol{x}_{-i}^{(t^\star)}} \sum_{h=1}^{H} \mathop{\mathbb{E}}_{\boldsymbol{a}_{-i,h} \sim \widehat{\boldsymbol{q}}_{-i,h}} [u_{i,h}(\boldsymbol{x}_{i,h}^\dagger, \boldsymbol{a}_{-i,h})] \qquad (6) \\
&\geq -2\sqrt{\frac{\log T}{H}} + \mathop{\mathbb{E}}_{t^\star \sim \mathsf{U}([\![T]\!])} \mathop{\mathbb{E}}_{\boldsymbol{x}_i^\dagger, \boldsymbol{x}_{-i}^{(t^\star)}} \sum_{h=1}^{H} \max_{a_{i,h} \in \mathcal{A}_i} \mathop{\mathbb{E}}_{\boldsymbol{a}_{-i,h} \sim \widehat{\boldsymbol{q}}_{-i,h}} [u_{i,h}(a_{i,h}, \boldsymbol{a}_{-i,h})] \qquad (7) \\
&\geq -2\sqrt{\frac{\log T}{H}}, \qquad (8)
\end{aligned}$$

where (6) uses (5) along with the fact that

$$\begin{aligned}
\mathop{\mathbb{E}}_{\boldsymbol{a}_{-i,h} \sim \boldsymbol{x}_{-i,s_h}^{(t^\star)}} [u_{i,h}(\boldsymbol{x}_{i,h}^\dagger, \boldsymbol{a}_{-i,h})] &\geq \mathop{\mathbb{E}}_{\boldsymbol{a}_{-i,h} \sim \widehat{\boldsymbol{q}}_{-i,h}} [u_{i,h}(\boldsymbol{x}_{i,h}^\dagger, \boldsymbol{a}_{-i,h})] - \frac{1}{H} D_{\mathrm{TV}} \left(\boldsymbol{x}_{-i,s_h}^{(t^\star)}, \bigtimes_{i' \neq i} \widehat{\boldsymbol{q}}_{i',h} \right) \\
&\geq \mathop{\mathbb{E}}_{\boldsymbol{a}_{-i,h} \sim \widehat{\boldsymbol{q}}_{-i,h}} [u_{i,h}(\boldsymbol{x}_{i,h}^\dagger, \boldsymbol{a}_{-i,h})] - \frac{1}{H} \sum_{i' \neq i} D_{\mathrm{TV}} \left(\boldsymbol{x}_{i',s_h}^{(t^\star)}, \widehat{\boldsymbol{q}}_{i',h} \right),
\end{aligned}$$

since $|u_{i,h}(\cdot, \cdot)| \leq \frac{1}{H}$ (by construction) and the total variation distance between two product distributions is bounded by the sum of the total variation of the individual components [49]; (7) follows from the definition of $\boldsymbol{x}_{i,h}^\dagger$ in (4), which in particular implies that

$$\mathop{\mathbb{E}}_{\boldsymbol{a}_{-i,h} \sim \widehat{\boldsymbol{q}}_{-i,h}} [u_{i,h}(\boldsymbol{x}_{i,h}^\dagger, \boldsymbol{a}_{-i,h})] = \max_{a_{i,h} \in \mathcal{A}_i} \mathop{\mathbb{E}}_{\boldsymbol{a}_{-i,h} \sim \widehat{\boldsymbol{q}}_{-i,h}} [u_{i,h}(a_{i,h}, \boldsymbol{a}_{-i,h})];$$

and (8) follows from the fact that $\max_{a_{i,h} \in \mathcal{A}_i} \mathbb{E}_{\boldsymbol{a}_{-i,h} \sim \widehat{\boldsymbol{q}}_{-i,h}} [u_{i,h}(a_{i,h}, \boldsymbol{a}_{-i,h})] \geq 0$ (Lemma 3.1). Further, since $\bar{\boldsymbol{\mu}}$ is assumed to be an ϵ-CCE, (8) implies that for each player $i \in [\![3]\!]$,

$$u_i(\bar{\boldsymbol{\mu}}) \geq -2\sqrt{\frac{\log T}{H}} - \epsilon. \tag{9}$$

Given that $\mathcal{T}$ is zero-sum, we also have that $\sum_{i=1}^{3} u_i(\bar{\boldsymbol{\mu}}) = 0$; by (9), this in turn implies that $u_{\mathsf{K}}(\bar{\boldsymbol{\mu}}) = -u_1(\bar{\boldsymbol{\mu}}) - u_2(\bar{\boldsymbol{\mu}}) \leq 4\sqrt{\frac{\log T}{H}} + 2\epsilon$. We next focus on analyzing the deviation benefit of the Kibitzer. We define $\delta(\widehat{\boldsymbol{q}}_{1,h}, \widehat{\boldsymbol{q}}_{2,h})$ as

$$\max \left\{ \max_{a'_{1,h} \in \mathcal{A}_1} \left\{ \mathbf{M}_1[a'_{1,h}, \widehat{\boldsymbol{q}}_{2,h}] - \mathbf{M}_1[\widehat{\boldsymbol{q}}_{-\mathsf{K},h}] \right\}, \max_{a'_{2,h} \in \mathcal{A}_2} \left\{ \mathbf{M}_2[\widehat{\boldsymbol{q}}_{1,h}, a'_{2,h}] - \mathbf{M}_2[\widehat{\boldsymbol{q}}_{-\mathsf{K},h}] \right\} \right\}.$$

By (7), we have that

$$u_{\mathsf{K}}(\boldsymbol{x}_{\mathsf{K}}^{\dagger}, \bar{\boldsymbol{\mu}}_{-\mathsf{K}}) \geq \frac{1}{H} \mathop{\mathbb{E}}_{t^\star \sim \mathsf{U}(\llbracket T \rrbracket)} \mathop{\mathbb{E}}_{\boldsymbol{x}_{\mathsf{K}}^{\dagger}, \boldsymbol{x}_{-\mathsf{K}}^{(t^\star)}} \sum_{h=1}^{H} \delta(\widehat{\boldsymbol{q}}_{1,h}, \widehat{\boldsymbol{q}}_{2,h}) - 2\sqrt{\frac{\log T}{H}}.$$

Since $\bar{\boldsymbol{\mu}}$ is an ϵ-CCE, we also know that $u_{\mathsf{K}}(\boldsymbol{x}_{\mathsf{K}}^{\dagger}, \bar{\boldsymbol{\mu}}_{-\mathsf{K}}) \leq u_{\mathsf{K}}(\bar{\boldsymbol{\mu}}) + \epsilon \leq 3\epsilon + 4\sqrt{\frac{\log T}{H}}$, which in turn implies that

$$\mathop{\mathbb{E}}_{t^\star \sim \mathsf{U}(\llbracket T \rrbracket)} \mathop{\mathbb{E}}_{\boldsymbol{x}_{\mathsf{K}}^{\dagger}, \boldsymbol{x}_{-\mathsf{K}}^{(t^\star)}} \frac{1}{H} \sum_{h=1}^{H} \delta(\widehat{\boldsymbol{q}}_{1,h}, \widehat{\boldsymbol{q}}_{2,h}) \leq 9\epsilon,$$

where we used that $H \geq \frac{\log T}{\epsilon^2}$. Finally, given that $\delta(\widehat{\boldsymbol{q}}_{1,h}, \widehat{\boldsymbol{q}}_{2,h}) \geq 0$ (Lemma 3.1), we conclude that there exists some repetition $h \in \llbracket H \rrbracket$ and a state $s \in \mathcal{S}_h$ such that $\delta(\widehat{\boldsymbol{q}}_{1,s_h}, \widehat{\boldsymbol{q}}_{2,s_h}) \leq 9\epsilon$. That is, $(\widehat{\boldsymbol{q}}_{1,s_h}, \widehat{\boldsymbol{q}}_{2,s_h})$ is a (9ϵ)-Nash equilibrium, concluding the proof. ◀

▶ Remark 3.6. It is direct to see that the proof of Theorem 3.5 can be extended under a more general notion of sparse CCE, wherein $\bar{\boldsymbol{\mu}}$ is not necessarily a uniform mixture of product distributions; this notion of CCE is naturally associated with a weighted generalization of regret. This observation is important since in practice taking a non-uniform average can lead to significant gains in performance [15].

3.4 Implications

Having established Theorem 3.5, we are now ready to prove that computing approximate CCE in a certain regime of sparsity is hard, at least under some well-established complexity assumptions. In particular, we will leverage the hardness result of Rubinstein [74] (Theorem 3.8), which rests on the so-called *exponential-time hypothesis (ETH)* for PPAD [4]. That hypothesis pertains the complexity of solving EndOfALine, the prototypical PPAD-complete problem [65].

▶ **Conjecture 3.7** ([4]). *Solving EndOfALine on m-bit circuits with $\tilde{O}(m)$ gates requires time $2^{\tilde{\Omega}(m)}$.*

Assuming that this conjecture holds, Rubinstein [74] proved that the quasipolynomial algorithm of Lipton et al. [58] is essentially optimal.

▶ **Theorem 3.8** ([74]). *Assuming Conjecture 3.7, there is an absolute constant $\epsilon_0 > 0$ such that finding an ϵ_0-Nash equilibrium in two-player m-action games requires time $m^{\log_2^{1-o(1)} m}$.*

We now use Theorem 3.5 to prove the following hardness result. Below, we recall that we use the notation $|\mathcal{T}|$ to represent the number of nodes in the extensive-form game $\mathcal{T}$.

▶ **Theorem 1.2.** *There is no algorithm that runs in time polynomial in the description of an extensive-form game $\mathcal{T}$ and can compute a $2^{\log_2^{1/2-o(1)} |\mathcal{T}|}$-sparse ϵ-CCE, even for an absolute constant $\epsilon > 0$, unless ETH for PPAD (Conjecture 3.7) is false.*

Proof. Suppose that algorithm $\mathfrak{A}$ implementing Line 4 of Algorithm 1 runs in time polynomial in the description of $\mathcal{T}$ and computes a T-sparse $(\epsilon_0/9)$-CCE of $\mathcal{T}$, where $T = 2^{\log_2^{\gamma} |\mathcal{T}|}$ and $\gamma < \frac{1}{2}$. Then, by Proposition 3.3, Algorithm 1 runs in time $m^{\Theta(H)}$. As a result, it follows from Theorems 3.5 and 3.8 that for $H \geq \frac{81 \log T}{\epsilon_0^2}$, it must hold that $m^{\Theta(H)} \geq m^{\log_2^{1-o(1)} m}$. Further, Claim 3.2 implies that $T = 2^{\log_2^{\gamma} |\mathcal{T}|} \leq 2^{4H^{\gamma} \log_2^{\gamma} m}$. As a result, for a sufficiently large $H = O(\log_2^{\frac{\gamma}{1-\gamma}} m)$ it follows that $H \geq \frac{81 \log T}{\epsilon_0^2}$, which in turn implies that $H = \Omega(\log_2^{1-o(1)} m)$. As a result, we conclude that $\gamma \geq \frac{1}{2} - o(1)$, leading to the desired conclusion. ◀

A number of remarks regarding Theorem 1.2 are in order. First, Theorem 1.2 establishes a stark separation between extensive- and normal-form games. Indeed, as we have seen, in normal-form games there are polynomial-time algorithms, such as multiplicative weights update, that can compute an $O(\log m)$-sparse $O(1)$-CCE in polynomial time. In contrast, Theorem 1.2 precludes even computing a $\mathsf{polylog}|\mathcal{T}|$-sparse $O(1)$-CCE, unless the quasipolynomial-time algorithm of Lipton et al. [58] can be improved. It is also worth pointing out here that it is natural to expect that analogous lower bounds to Theorem 1.2 could be established under the more plausible conjecture $\mathsf{P} \neq \mathsf{PPAD}$ (instead of Conjecture 3.7). Yet, that seems to require a very different approach. Indeed, for the PPAD-hardness of ϵ-Nash equilibria in two-player games to kick in, one must take ϵ to be inversely polynomial to m [18]. In that case, the description of $\mathcal{T}$ becomes immediately $m^{\Omega(\mathsf{poly}(m))}$, which renders reductions analogous to Theorem 3.5 of little use. It is thus crucial for our approach to take $\epsilon > 0$ to be an (absolute) constant. Finally, we point out that Theorem 1.2 still applies by taking $T = 2^{C \log_2^{1/2-o(1)} |\mathcal{T}|}$, for any absolute constant $C > 0$.

To better contextualize Theorem 1.2, we remark that there are algorithms running in time polynomial in $\mathcal{T}$ that can compute an $O(k)$-sparse $O(1)$-CCE in every extensive-form game $\mathcal{T}$ with $k \leq \max_{1 \leq i \leq n} \log \left(\prod_{j \in \mathcal{J}_i} |\mathcal{A}_j| \right) = \max_{1 \leq i \leq n} \sum_{j \in \mathcal{J}_i} \log |\mathcal{A}_j|$; that is, k is at most nearly-linear in $|\mathcal{T}|$, and it can be much smaller depending on the information structure of $\mathcal{T}$. This is a direct consequence of the fact that algorithms such as multiplicative weights update can be implemented in polynomial time in extensive-form games [35], thereby implying that the regret of each player i will be bounded as $O(\sqrt{T \log m_i})$, where m_i represents the number of actions in the induced normal-form game: $m_i \triangleq \prod_{j \in \mathcal{J}_i} |\mathcal{A}_j|$. As a result, we see that there is a gap between our lower bound (Theorem 1.2) and the aforementioned best-known upper bound, which essentially amounts to improving the exponent $\frac{1}{2}$ in the term $2^{\log_2^{1/2-o(1)} |\mathcal{T}|}$ all the way up to 1.

Lower bounds for no-regret learning in extensive-form games

Relatedly, we next proceed by pointing out some important implications of Theorem 1.2 for bounding the regret incurred by no-regret learning algorithms in extensive-form games. In particular, a number of no-regret learning algorithms have been designed with iteration complexity polynomial in the description of the extensive-form game (Section 1.2). For example, one such broad class derives from the paradigm of *follow the perturbed leader* (FTPL) [1, 54, 47]; indeed, FTPL can be implemented efficiently under a linear optimization oracle, which can be in turn implemented in polynomial time in extensive-form games. Theorem 1.2 circumscribes the performance of any of those algorithms – and combinations thereof – when employed simultaneously by all players.

▶ **Corollary 1.3.** *Suppose that each player follows an algorithm with polynomial iteration complexity in the description of an extensive-form game $\mathcal{T}$. If Reg_i^T is the regret incurred by player $i \in [\![n]\!]$, there is an extensive-form game $\mathcal{T}$ and an absolute constant $\epsilon > 0$ such that at least $T \geq 2^{\log_2^{1/2-o(1)} |\mathcal{T}|}$ repetitions are needed so that $\frac{1}{T} \max_{1 \leq i \leq n} \mathsf{Reg}_i^T \leq \epsilon$, unless ETH for PPAD (Conjecture 3.7) is false.*

In words, obtaining an average regret below an absolute constant requires at least $2^{\log_2^{1/2-o(1)} |\mathcal{T}|}$ repetitions, which again stands in stark contrast to the performance of learning algorithms in normal-form games. Corollary 1.3 is an immediate consequence of Theorem 1.2, along with the folklore fact that the average product distribution after T repetitions – by definition T-sparse – constitutes a $\frac{1}{T} \max_{1 \leq i \leq n} \mathsf{Reg}_i^T$-CCE (Proposition 2.3). We clarify that

each regret Reg_i^T can be indeed computed in time $\mathsf{poly}(|\mathcal{T}|, T)$ since it amounts to computing a best response, in turn implying that one can efficiently determine for each repetition whether $\frac{1}{t}\max_{1\le i\le n}\mathsf{Reg}_i^t \le \epsilon$.

A noteworthy feature of Corollary 1.3 is that it applies even if players are following different no-regret algorithms, and the updates are not simultaneous. It is especially worth stressing that last point. One popular way of improving the performance of no-regret algorithms in games consists of *alternation* [79, 85], whereby players are updating their strategies in an alternating fashion. Alternation is, of course, not a legitimate choice within the framework of online learning as it trivializes the problem for the player who gets to play last; for example, that player could always just best respond, which would accumulate at most 0 regret for that player. Nevertheless, Corollary 1.3 still holds even under learning dynamics that are beyond the framework of online learning. Furthermore, unlike the paper of Daskalakis et al. [23], Corollary 1.3 does not limit the amount of memory players use, as long as the running time stays polynomial.

Lower bounds for (optimistic) MWU

Beyond extensive-form games, our results also turn out to have implications for the performance of certain no-regret learning algorithms in normal-form games. In particular, one can always cast an extensive-form game in normal form, and then use a no-regret learning algorithm on the induced normal-form game. In general, such an approach is not interesting algorithmically since the iteration complexity would be exponential, thereby rendering computational lower bounds of little use. However, it turns out there are certain algorithms for which each iteration can be indeed implemented in polynomial time, even though they operate over the – typically exponentially large – normal-form representation. This can be accomplished by leveraging the underlying structure of the extensive-form representation, and it is akin to the kernel trick [78, 8]. Perhaps most notably, such is the case for the celebrated multiplicative weights updates (`MWU`), as well as its *optimistic* counterpart (`OMWU`) [35] (see also [5, 20, 67]).

In particular, let us recall (`O`)`MWU` in the context of learning in games. In the vanilla `MWU` algorithm each player $i \in [\![n]\!]$ updates its strategy for $t = 1, 2, \dots$ as follows.

$$\boldsymbol{x}_i^{(t+1)}[a_i] = \frac{\boldsymbol{x}_i^{(t)}[a_i]\exp\left(\eta_i \boldsymbol{u}_i^{(t)}[a_i]\right)}{\sum_{a_i'\in\mathcal{A}_i}\boldsymbol{x}_i^{(t)}[a_i']\exp\left(\eta_i \boldsymbol{u}_i^{(t)}[a_i']\right)}, \quad \forall a_i \in \mathcal{A}_i. \tag{MWU}$$

Here, $\boldsymbol{u}_i^{(t)}[a_i] \triangleq \mathbb{E}_{\boldsymbol{a}_{-i}\sim\boldsymbol{x}_{-i}^{(t)}}[u_i(a_i, \boldsymbol{a}_{-i})]$; $\boldsymbol{x}_i^{(1)} \triangleq (1/|\mathcal{A}_i|, \dots, 1/|\mathcal{A}_i|)$; and $\eta_i > 0$ is the learning rate. Beyond the uniform distribution, one can also initialize `MWU` to any point in the (relative) interior of the simplex.[5] Similarly, `OMWU` [70] is defined via the following update rule.

$$\boldsymbol{x}_i^{(t+1)}[a_i] = \frac{\boldsymbol{x}_i^{(t)}[a_i]\exp\left(\eta_i(2\boldsymbol{u}_i^{(t)}[a_i] - \boldsymbol{u}_i^{(t-1)}[a_i])\right)}{\sum_{a_i'\in\mathcal{A}_i}\boldsymbol{x}_i^{(t)}[a_i']\exp\left(\eta_i(2\boldsymbol{u}_i^{(t)}[a_i'] - \boldsymbol{u}_i^{(t-1)}[a_i'])\right)}, \quad \forall a_i \in \mathcal{A}_i. \tag{OMWU}$$

`OMWU` can be seen as an variant of `MWU` in which a prediction term is incorporated into the update rule. In the definitions above, the player's strategies could take irrational values due to the exponential function, but this can be readily addressed by truncating to a sufficiently large number of bits, an operation that does not essentially alter any of the results.

[5] We should note that the analysis of Farina et al. [35] pertaining the iteration complexity of `MWU` in extensive-form games considers the uniform initialization, but can be directly generalized.

In general, it is evident that (O)MWU requires time $\Omega(|\mathcal{A}_i|)$ in each iteration, which is typically exponential in the context of extensive-form games. However, it turns out that each iteration of (O)MWU can be implicitly performed in time $\mathsf{poly}(|\mathcal{T}|)$ [35]; the analysis of Farina et al. [35] does not account for the numerical imprecision resulting from the exponential function, but their argument readily carries over by truncating to a sufficiently large number of bits. This leads to the following lower bound for the regret accumulated by such algorithms. Below, we tacitly assume that each learning rate is given by an efficiently computable function.

▶ **Corollary 3.9.** *Consider the class of n-player normal-form games where each player has m' actions. If each player $i \in [\![n]\!]$ follows* MWU *or* OMWU *with any learning rate $\eta_i = \eta_i(n, \log m', T)$ and incurs regret Reg_i^T after T repetitions, there is a game $\mathcal{G}$ and an absolute constant $\epsilon > 0$ such that at least $T \geq 2^{(\log_2 \log_2 m')^{1/2-o(1)}}$ repetitions of the game are needed so that $\frac{1}{T} \max_{1 \leq i \leq n} \mathsf{Reg}_i^T \leq \epsilon$, unless ETH for* PPAD *(Conjecture 3.7) fails.*

In proof, the extensive-form game $\mathcal{T}(\mathcal{G})$ described in Section 3.1 can be cast as a 3-player m'-action normal-form game $\mathcal{G}'$, where $m' = (2m)^{m^{\Theta(H)}}$. In this normal-form game $\mathcal{G}'$, each player can implement MWU or OMWU with per-iteration complexity polynomial in $|\mathcal{T}|$ [35], along with a representation of the iterates in $\mathsf{poly}(|\mathcal{T}|, T)$ space. As a result, Corollary 1.3 implies that, for any constant $C > 0$, at least $T \geq 2^{C \log_2^{1/2-o(1)} |\mathcal{T}|}$ repetitions are needed so that $\frac{1}{T} \max_{1 \leq i \leq n} \mathsf{Reg}_i^T \leq \epsilon$ for each game $\mathcal{T}$. The statement of Corollary 3.9 thus follows from the fact that $\log_2 |\mathcal{T}| \geq \Omega(\log_2 \log_2 m')$.

Let us point out another way to express the conclusion of Corollary 3.9. Suppose that the regret of each player, who has m' available actions, can be bounded as $\mathsf{Reg}_i^T \leq R(m')T^{1-\gamma}$, for some constant $\gamma \in (0, 1]$; this is the canonical form regret bounds assume. Corollary 3.9 then implies that $R(m') \geq 2^{\gamma(\log_2 \log_2 m')^{1/2-o(1)}}$.

We suspect that, at least under a specific parameterization, there should be a more elementary way of proving unconditional dimension-dependent lower bounds when multiple players follow algorithms such as MWU. The main advantage of our approach is that it applies to any parameterization (potentially game-specific) and a broad class of algorithms; as concrete examples, our approach is robust to considering alternating instead of simultaneous dynamics, different players following different variants of MWU, as well as using more general prediction mechanisms within the paradigm of optimistic MWU [77]. Beyond MWU-type update rules, we suspect that Corollary 3.9 applies more broadly to any member of follow the perturbed leader (FTPL).

4 Conclusions and Open Problems

In conclusion, we established the first dimension-dependent computational lower bounds for no-regret learning in extensive- and normal-form games, beyond the well-understood adversarial regime in online learning. A number of important questions remain open. Besides the obvious avenue of bridging the gaps between the current upper and lower bounds in extensive-form games, a fundamental question is to understand the complexity of computing sparse CCE (Definition 2.2) in normal-form games. Indeed, even the complexity status of that problem under a mixture of two product distributions is open, although the fact that algorithms such as MWU and OMWU require a superconstant number of iterations under any parameterization (Corollary 3.9) suggests that the problem is hard. Furthermore, our lower bounds apply to *coarse* correlated equilibria; while those naturally translate to stronger equilibrium concepts as well, such as correlated equilibria, it would be interesting to understand whether stronger hardness results can be obtained for such refinements. In

particular, in a surprising turn of events, it was recently shown that logarithmic sparsity is possible even for approximate correlated equilibria [22, 67]; are those guarantees for swap regret tight when learning in games? Interestingly, in a correlated equilibrium players observe additional information, which could potentially speed up the online density estimation procedure.

References

1 Jacob Abernethy, Chansoo Lee, and Ambuj Tewari. Perturbation techniques in online learning and optimization. *Perturbations, Optimization, and Statistics*, 233, 2016.

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

3 Yakov Babichenko. Query complexity of approximate nash equilibria. *Journal of the ACM*, 63(4):36:1–36:24, 2016.

4 Yakov Babichenko, Christos H. Papadimitriou, and Aviad Rubinstein. Can almost everybody be almost happy? In *Proceedings of the Conference on Innovations in Theoretical Computer Science*, pages 1–9. ACM, 2016.

5 Yu Bai, Chi Jin, Song Mei, Ziang Song, and Tiancheng Yu. Efficient phi-regret minimization in extensive-form games via online mirror descent. In *Proceedings of the Annual Conference on Neural Information Processing Systems (NeurIPS)*, 2022.

6 Yu Bai, Chi Jin, Song Mei, and Tiancheng Yu. Near-optimal learning of extensive-form games with imperfect information. In *International Conference on Machine Learning (ICML)*, pages 1337–1382. PMLR, 2022.

7 Anton Bakhtin, Noam Brown, Emily Dinan, Gabriele Farina, Colin Flaherty, Daniel Fried, Andrew Goff, Jonathan Gray, Hengyuan Hu, Athul Paul Jacob, Mojtaba Komeili, Karthik Konath, Minae Kwon, Adam Lerer, Mike Lewis, Alexander H. Miller, Sasha Mitts, Adithya Renduchintala, Stephen Roller, Dirk Rowe, Weiyan Shi, Joe Spisak, Alexander Wei, David Wu, Hugh Zhang, and Markus Zijlstra. Human-level play in the game of diplomacy by combining language models with strategic reasoning. *Science*, 378(6624):1067–1074, 2022.

8 Daniel Beaglehole, Max Hopkins, Daniel Kane, Sihan Liu, and Shachar Lovett. Sampling equilibria: Fast no-regret learning in structured games. In *Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 3817–3855. SIAM, 2023.

9 Shai Ben-David, Dávid Pál, and Shai Shalev-Shwartz. Agnostic online learning. In *Conference on Learning Theory (COLT)*, 2009.

10 David Blackwell. An analog of the minmax theorem for vector payoffs. *Pacific Journal of Mathematics*, 6:1–8, 1956.

11 Avrim Blum and Yishay Mansour. Learning, regret minimization, and equilibria, 2007.

12 Christian Borgs, Jennifer T. Chayes, Nicole Immorlica, Adam Tauman Kalai, Vahab S. Mirrokni, and Christos H. Papadimitriou. The myth of the folk theorem. *Games and Economic Behavior*, 70(1):34–43, 2010.

13 Michael Bowling, Neil Burch, Michael Johanson, and Oskari Tammelin. Heads-up limit hold'em poker is solved. *Science*, 347(6218), January 2015.

14 Noam Brown and Tuomas Sandholm. Superhuman AI for heads-up no-limit poker: Libratus beats top professionals. *Science*, pages 418–424, December 2018.

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

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

17 Nicolo Cesa-Bianchi and Gabor Lugosi. *Prediction, learning, and games*. Cambridge University Press, 2006.

18 Xi Chen, Xiaotie Deng, and Shang-Hua Teng. Settling the complexity of computing two-player Nash equilibria. *Journal of the ACM*, 2009.

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

20 Chirag Chhablani, Michael Sullins, and Ian A. Kash. Multiplicative weight updates for extensive form games. In *Autonomous Agents and Multi-Agent Systems*, pages 1071–1078. ACM, 2023.

21 Francis Chu and Joseph Halpern. On the NP-completeness of finding an optimal strategy in games with common payoffs. *International Journal of Game Theory*, 2001.

22 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, 2023.

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

24 Constantinos Daskalakis, Maxwell Fishelson, and Noah Golowich. Near-optimal no-regret learning in general games. In *Proceedings of the Annual Conference on Neural Information Processing Systems (NeurIPS)*, pages 27604–27616, 2021.

25 Constantinos Daskalakis, Paul W Goldberg, and Christos H Papadimitriou. The complexity of computing a nash equilibrium. *SIAM Journal on Computing*, 39(1), 2009.

26 Constantinos Daskalakis and Noah Golowich. Fast rates for nonparametric online learning: from realizability to learning in games. In *Proceedings of the Annual Symposium on Theory of Computing (STOC)*, pages 846–859. ACM, 2022.

27 Miroslav Dudík and Geoffrey J. Gordon. A sampling-based approach to computing equilibria in succinct extensive-form games. In *UAI 2009, Proceedings of the Twenty-Fifth Conference on Uncertainty in Artificial Intelligence, Montreal, QC, Canada, June 18-21, 2009*, pages 151–160. AUAI Press, 2009.

28 Liad Erez, Tal Lancewicki, Uri Sherman, Tomer Koren, and Yishay Mansour. Regret minimization and convergence to equilibria in general-sum markov games. In *International Conference on Machine Learning (ICML)*, volume 202 of *Proceedings of Machine Learning Research*, pages 9343–9373. PMLR, 2023.

29 Gabriele Farina, Tommaso Bianchi, and Tuomas Sandholm. Coarse correlation in extensive-form games. In *AAAI Conference on Artificial Intelligence (AAAI)*, volume 34, pages 1934–1941, 2020.

30 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):41:1–41:41, 2022.

31 Gabriele Farina, Christian Kroer, Noam Brown, and Tuomas Sandholm. Stable-predictive optimistic counterfactual regret minimization. In *International Conference on Machine Learning (ICML)*, 2019.

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

33 Gabriele Farina, Christian Kroer, and Tuomas Sandholm. Better regularization for sequential decision spaces: Fast convergence rates for nash, correlated, and team equilibria. In *Proceedings of the ACM Conference on Economics and Computation (EC)*, page 432. ACM, 2021.

34 Gabriele Farina, Christian Kroer, and Tuomas Sandholm. Faster game solving via predictive blackwell approachability: Connecting regret matching and mirror descent. In *AAAI Conference on Artificial Intelligence (AAAI)*, 2021.

35 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)*, volume 162 of *Proceedings of Machine Learning Research*, pages 6337–6357. PMLR, 2022.

36 John Fearnley, Martin Gairing, Paul W. Goldberg, and Rahul Savani. Learning equilibria of games via payoff queries. *Journal of Machine Learning Research*, 16:1305–1344, 2015.

37 John Fearnley and Rahul Savani. Finding approximate nash equilibria of bimatrix games via payoff queries. *ACM Trans. Economics and Comput.*, 4(4):25:1–25:19, 2016.

38 Côme Fiegel, Pierre Ménard, Tadashi Kozuno, Rémi Munos, Vianney Perchet, and Michal Valko. Adapting to game trees in zero-sum imperfect information games. In *International Conference on Machine Learning (ICML)*, volume 202 of *Proceedings of Machine Learning Research*, pages 10093–10135. PMLR, 2023.

39 Côme Fiegel, Pierre Ménard, Tadashi Kozuno, Rémi Munos, Vianney Perchet, and Michal Valko. Local and adaptive mirror descents in extensive-form games, 2023. `arXiv:2309.00656`.

40 Dean Foster and Rakesh Vohra. Calibrated learning and correlated equilibrium. *Games and Economic Behavior*, 21:40–55, 1997.

41 Dylan J. Foster, Noah Golowich, and Sham M. Kakade. Hardness of independent learning and sparse equilibrium computation in markov games. In *International Conference on Machine Learning (ICML)*, volume 202 of *Proceedings of Machine Learning Research*, pages 10188–10221. PMLR, 2023.

42 Dylan J. Foster, Zhiyuan Li, Thodoris Lykouris, Karthik Sridharan, and Éva Tardos. Learning in games: Robustness of fast convergence. In *Proceedings of the Annual Conference on Neural Information Processing Systems (NIPS)*, pages 4727–4735, 2016.

43 Paul W. Goldberg and Matthew J. Katzman. Lower bounds for the query complexity of equilibria in lipschitz games. *Theor. Comput. Sci.*, 962:113931, 2023.

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

45 Hédi Hadiji, Sarah Sachs, Tim van Erven, and Wouter M. Koolen. Towards characterizing the first-order query complexity of learning (approximate) nash equilibria in zero-sum matrix games, 2023.

46 Sergiu Hart and Andreu Mas-Colell. A simple adaptive procedure leading to correlated equilibrium. *Econometrica*, 68:1127–1150, 2000.

47 Elad Hazan. Introduction to online convex optimization. *Foundations and Trends in Optimization*, 2(3-4):157–325, 2016.

48 Johannes Heinrich, Marc Lanctot, and David Silver. Fictitious self-play in extensive-form games. In *International Conference on Machine Learning (ICML)*, volume 37 of *JMLR Workshop and Conference Proceedings*, pages 805–813. JMLR.org, 2015.

49 Wassily Hoeffding and J. Wolfowitz. Distinguishability of sets of distributions. *The Annals of Mathematical Statistics*, 29(3):700–718, 1958.

50 Yu-Guan Hsieh, Kimon Antonakopoulos, Volkan Cevher, and Panayotis Mertikopoulos. No-regret learning in games with noisy feedback: Faster rates and adaptivity via learning rate separation. In *Proceedings of the Annual Conference on Neural Information Processing Systems (NeurIPS)*, 2022.

51 Yu-Guan Hsieh, Kimon Antonakopoulos, and Panayotis Mertikopoulos. Adaptive learning in continuous games: Optimal regret bounds and convergence to nash equilibrium. In *Conference on Learning Theory (COLT)*, volume 134 of *Proceedings of Machine Learning Research*, pages 2388–2422. PMLR, 2021.

52 Wan Huang and Bernhard von Stengel. Computing an extensive-form correlated equilibrium in polynomial time. In *Internet and Network Economics, 4th International Workshop, WINE 2008*, volume 5385 of *Lecture Notes in Computer Science*, pages 506–513. Springer, 2008.

53 Albert Xin Jiang and Kevin Leyton-Brown. Polynomial-time computation of exact correlated equilibrium in compact games. *Games and Economic Behavior*, 91:347–359, 2015.

54 Adam Kalai and Santosh Vempala. Efficient algorithms for online decision problems. *Journal of Computer and System Sciences*, 71:291–307, 2005.

55 Ehsan Asadi Kangarshahi, Ya-Ping Hsieh, Mehmet Fatih Sahin, and Volkan Cevher. Let's be honest: An optimal no-regret framework for zero-sum games. In *International Conference on Machine Learning (ICML)*, volume 80 of *Proceedings of Machine Learning Research*, pages 2493–2501. PMLR, 2018.

56 Daphne Koller, Nimrod Megiddo, and Bernhard von Stengel. Fast algorithms for finding randomized strategies in game trees. In *Proceedings of the Annual Symposium on Theory of Computing (STOC)*, 1994.

57 Tadashi Kozuno, Pierre Ménard, Rémi Munos, and Michal Valko. Learning in two-player zero-sum partially observable markov games with perfect recall. In *Proceedings of the Annual Conference on Neural Information Processing Systems (NeurIPS)*, pages 11987–11998, 2021.

58 Richard Lipton, Evangelos Markakis, and Aranyak Mehta. Playing large games using simple strategies. In *Proceedings of the ACM Conference on Electronic Commerce (ACM-EC)*, pages 36–41, San Diego, CA, 2003. ACM.

59 Nick Littlestone. Learning quickly when irrelevant attributes abound: A new linear-threshold algorithm. *Machine learning*, 2:285–318, 1988.

60 Michael Littman and Peter Stone. A polynomial-time Nash equilibrium algorithm for repeated games. In *Proceedings of the ACM Conference on Electronic Commerce (ACM-EC)*, pages 48–54, San Diego, CA, 2003.

61 Arnab Maiti, Ross Boczar, Kevin G. Jamieson, and Lillian J. Ratliff. Query-efficient algorithms to find the unique nash equilibrium in a two-player zero-sum matrix game, 2023.

62 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 Marina Meila and Tong Zhang, editors, *International Conference on Machine Learning (ICML)*, volume 139 of *Proceedings of Machine Learning Research*, pages 7818–7828. PMLR, 2021.

63 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)*, pages 5584–5594. AAAI Press, 2021.

64 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.

65 Christos H. Papadimitriou. On the complexity of the parity argument and other inefficient proofs of existence. *Journal of Computer and system Sciences*, 48(3):498–532, 1994.

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

67 Binghui Peng and Aviad Rubinstein. Fast swap regret minimization and applications to approximate correlated equilibria, 2023.

68 Georgios Piliouras, Ryann Sim, and Stratis Skoulakis. Beyond time-average convergence: Near-optimal uncoupled online learning via clairvoyant multiplicative weights update. In *Proceedings of the Annual Conference on Neural Information Processing Systems (NeurIPS)*, 2022.

69 Ju Qi, Ting Feng, Falun Hei, Zhemei Fang, and Yunfeng Luo. Pure monte carlo counterfactual regret minimization, 2023. `arXiv:2309.03084`.

70 Alexander Rakhlin and Karthik Sridharan. Online learning with predictable sequences. In *Conference on Learning Theory*, pages 993–1019, 2013.

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

72 Julia Robinson. An iterative method of solving a game. *Annals of Mathematics*, 54:296–301, 1951.

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

74 Aviad Rubinstein. Inapproximability of nash equilibrium. *SIAM Journal on Computing*, 47(3):917–959, 2018.

75 Yoav Shoham and Kevin Leyton-Brown. *Multiagent systems: Algorithmic, game-theoretic, and logical foundations*. Cambridge University Press, 2008.

76 Ziang Song, Song Mei, and Yu Bai. Sample-efficient learning of correlated equilibria in extensive-form games. In *Proceedings of the Annual Conference on Neural Information Processing Systems (NeurIPS)*, 2022.

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

78 Eiji Takimoto and Manfred K. Warmuth. Path kernels and multiplicative updates. *Journal of Machine Learning Research*, 4:773–818, 2003.

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

80 Xiaohang Tang, Le Cong Dinh, Stephen Marcus McAleer, and Yaodong Yang. Regret-minimizing double oracle for extensive-form games. In *International Conference on Machine Learning (ICML)*, volume 202 of *Proceedings of Machine Learning Research*, pages 33599–33615. PMLR, 2023.

81 Emanuel Tewolde, Caspar Oesterheld, Vincent Conitzer, and Paul W. Goldberg. The computational complexity of single-player imperfect-recall games. In *Proceedings of the International Joint Conference on Artificial Intelligence (IJCAI)*, pages 2878–2887, 2023.

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

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

84 V. G. Vovk. Aggregating strategies. In *Conference on Learning Theory (COLT)*, pages 371–386. Morgan Kaufmann, 1990.

85 Andre Wibisono, Molei Tao, and Georgios Piliouras. Alternating mirror descent for constrained min-max games. In *Proceedings of the Annual Conference on Neural Information Processing Systems (NeurIPS)*, 2022.

86 Yuepeng Yang and Cong Ma. $O(T^{-1})$ convergence of optimistic-follow-the-regularized-leader in two-player zero-sum markov games. In *The Eleventh International Conference on Learning Representations, ICLR 2023*. OpenReview.net, 2023.

87 Brian Hu Zhang and Tuomas Sandholm. Finding and certifying (near-)optimal strategies in black-box extensive-form games. In *AAAI Conference on Artificial Intelligence (AAAI)*, pages 5779–5788. AAAI Press, 2021.

88 Brian Hu Zhang and Tuomas Sandholm. Team correlated equilibria in zero-sum extensive-form games via tree decompositions. In *AAAI Conference on Artificial Intelligence (AAAI)*, pages 5252–5259. AAAI Press, 2022.

89 Runyu Zhang, Qinghua Liu, Huan Wang, Caiming Xiong, Na Li, and Yu Bai. Policy optimization for markov games: Unified framework and faster convergence. In *Proceedings of the Annual Conference on Neural Information Processing Systems (NeurIPS)*, 2022.

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