NORMAL-FORM CORRELATION IN MARKOV GAMES

Anonymous authors
Paper under double-blind review

ABSTRACT

There has been a surge of recent work on correlated equilibrium concepts in Markov games. However, existing results focus on concepts weaker than *normal-form correlated equilibria (NFCEs)*, leaving open the more challenging question of computing such equilibria, which goes back to the seminal work of Papadimitriou and Roughgarden (JACM'08). Here, we establish the first efficient algorithm for NFCEs in finite-horizon Markov games with a fixed number of players n. In particular, with S states, horizon H, and at most A actions per player, it computes an ϵ-NFCE in time $S(AH/\epsilon)^{O(n)}$. Moreover, under the assumption that recommendations are independent across states, we complement our positive result by showing PPAD-completeness—that is, computational equivalence to Nash equilibria—either in many-player games or when the precision is exponentially small.

The key idea behind our approach is to run backward induction on a sequence of auxiliary stage games, but with the twist that in each step we compute a *constant-expectation correlated equilibrium*. This is a natural refinement of correlated equilibrium in which the conditional expected payoff from obeying is independent of the recommendation. In fact, our reduction goes both ways, establishing an equivalence between constant-expectation CEs and NFCEs in Markov games. For a fixed number of players, we observe that a constant-expectation CE can be computed approximately by combining linear programming with suitable discretization. In contrast, it is PPAD-hard in i) polymatrix (many-player) games at constant precision, and ii) two-player games at exponentially small precision. The latter result follows from an unexpected connection to rank-2 two-player games.

1 INTRODUCTION

The correlated equilibrium, introduced by Aumann (1974), is a foundational notion in game theory. Unlike Nash equilibria (Nash, 1951), the standard solution concept in noncooperative games, the set of correlated equilibria is the feasible region of a linear program, thereby enjoying more favorable computational properties. Indeed, a seminal result by Papadimitriou & Roughgarden (2008) shows that a correlated equilibrium can be computed in polynomial time even in certain multi-player games of *polynomial type*, meaning that the number of pure strategies of each player is polynomial in the description of the input. However, in most realistic settings, the number of pure strategies of each player is exponentially large, since a pure strategy should prescribe how to act across a range of different contingencies. The complexity of correlated equilibria in such settings has been a long-standing open problem in this area, first highlighted by Papadimitriou & Roughgarden (2008).

An important class of games with a large number of contingency plans is *Markov* (aka. stochastic) games. There has been much recent work examining correlated equilibrium concepts in (fully observable) Markov games (discussed further in Section A). However, existing algorithmic results focus on relaxations of the *normal-form correlated equilibrium (NFCE)*, most notably notions akin to extensive-form correlated equilibrium (von Stengel & Forges, 2008), wherein recommendations are revealed only upon reaching an information set along the realized path. In the stronger normal-form notion, a mediator samples a complete policy profile beforehand and (privately) reveals to each player the entire policy recommendation. This means that a player may condition its deviation on recommendations at states that are never reached. In this paper, we focus on the following question.

Are there efficient algorithms for computing normal-form correlated equilibria (NFCEs) in Markov games?

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

NFCE is a suitable solution concept in systems where a mediator—a trusted third party—specifies complete policy prescriptions before deployment, with no need to conceal future recommendations or remain involved throughout execution. For example, a traffic coordinator could recommend complete routing policies to autonomous vehicles before departure, allowing each vehicle to inspect its full policy and decide whether to follow it without requiring further guidance during the trip. Moreover, determining whether efficient algorithms exist for NFCE calls for fundamentally new technical insights, which will deepen our understanding of correlated equilibria in dynamic environments.

1.1 Our results

Our main positive result is an efficient algorithm for this problem in finite-horizon Markov games with a fixed number of players n.

Main Theorem 1. *Let S be the number of states, H the horizon, and A the maximum number of actions per player. There is an algorithm that computes an ϵ-NFCE whose running time is $S(AH/\epsilon)^{O(n)}\mathsf{poly}(L)$, where L is the bit complexity of the input.*

This shows that, for a fixed number of players, there is a fully polynomial-time approximation scheme (FPTAS) for normal-form correlated equilibrium. More broadly, beyond polynomial-type games, the best known result is a PTAS (Dagan et al., 2024; Peng & Rubinstein, 2024). Main Theorem 1 gives an exponential improvement in the dependence on $1/\epsilon$ over those results, and is the first FPTAS for computing NFCE in an interesting class of games beyond the normal-form setting.

One limitation of our positive result is that it scales poorly with the number of players; the presence of multiple players is a well-known challenge in multiagent reinforcement learning (*e.g.* Cui et al., 2023; Wang et al., 2023). We complement Main Theorem 1 by showing that, under *state independent recommendations*, it cannot be significantly improved.

Main Theorem 2. *Computing an ϵ-NFCE with independent state recommendations is* PPAD*-hard even with horizon $H = 2$, in each of the following cases:*

(i) $(n+1)$-player succinct games, two actions per player at each state, and $\epsilon = \Theta(1/n)$.

(ii) Three-player games, a constant number of states, and exponentially small ϵ.

That is, computing such an NFCE is as hard as finding a Nash equilibrium in general-sum games.

1.2 Technical overview

We now highlight the key technical insights of our approach.

Failure of backward induction. The most natural approach that comes to mind is backward induction, whereby each continuation state is assigned its expected correlated equilibrium payoff over the local recommendations. However, the crux is that in an NFCE, each player observes its entire recommended policy in advance, so recommendations at future states may reveal conditional continuation payoffs that differ from the average over the local recommendations. This means that a player can profitably deviate by altering an earlier action to avoid a continuation state where its recommendation there signals a lower payoff.

More concretely, we consider a two-stage game in which Player 1 initially chooses between a terminal payoff of $5/4$ and a continuation state where the players play the game "Bach or Stravinsky," with payoffs $(2, 1)$ and $(1, 2)$ on the diagonal and $(0, 0)$ otherwise (Figure 1). Mixing equally between the profiles on the diagonal is a CE of the continuation game and gives Player 1 an expected payoff of $3/2$. As a result, backward induction based on this CE prescribes entering that state, because $3/2 > 5/4$. However, the full policy recommendation reveals which action profile should be played in the "Bach or Stravinsky" game. In particular, when Player 1 is recommended the action yielding payoff 2 (B), following the prescribed policy is optimal. On the other hand, when recommended the action yielding payoff 1 (S), it can instead choose the terminal payoff of $5/4$, which is a profitable deviation. So, the correlated distribution resulting from backward induction is not an NFCE.

Constant-expectation correlated equilibria. To address this, the key ingredient is a refinement of correlated equilibria in normal-form games, known as *constant-expectation correlated equilibrium*.

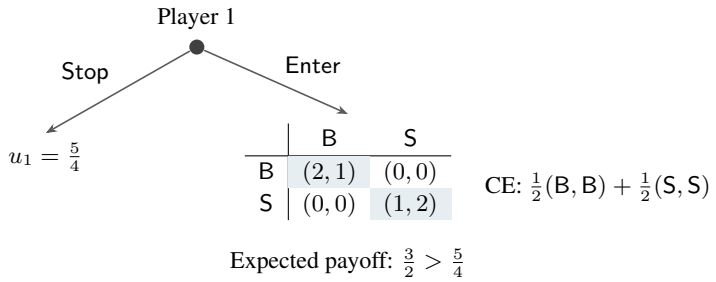


Figure 1: Backward induction prescribes Enter using the continuation CE's average payoff. The CE mixes equally between the two shaded profiles. Observing its full policy lets Player 1 stop whenever S is recommended, so the resulting distribution is not an NFCE.

To our knowledge, this was first examined by Sorin (1998), and subsequently used by Heller (2012).[1] In a one-shot game, a constant-expectation correlated equilibrium is a correlated equilibrium for which the conditional expected payoff from obedience *does not depend on the recommended action*. Even though this is a natural notion, its complexity has not been examined before. For example, the correlated equilibrium constructed in Figure 1 which breaks backward induction does *not* satisfy this property: for Player 1, the payoff conditional on observing B is 2, whereas the payoff conditional on observing S is 1. On the other hand, a Nash equilibrium clearly satisfies this requirement. However, unlike Nash equilibria, which are hard to compute (Daskalakis et al., 2009), our key observation is that constant-expectation CEs admit an efficient algorithm (FPTAS) for a fixed number of players: we exhaustively search over each player's constant obedience payoff and solve a linear feasibility program for each candidate (Proposition 3.3).

Rectifying backward induction. We now return to Markov games. The basic idea is to run backward induction by computing in each stage a constant-expectation correlated equilibrium. Then, the final distribution results from sampling the local recommendations independently across the states, which form pure Markovian policy profiles. This independence is crucial for the argument: after conditioning on a player's entire recommended policy, the other players' actions remain independent across states, and the environment induced by conditioning is a usual Markov decision process. The crucial role of constant expectation is that it makes conditional obedience payoffs approximately independent of the local recommendation. Together with independent sampling, this provides continuation values that remain valid after conditioning on the entire recommended policy, making backward induction sound (Lemma 4.6).

Complexity of constant-expectation CEs and implications for Markov NFCE. We finally circle back to the proof of our main hardness result, Main Theorem 2. The upshot is that, under independent state recommendations, our previous *reduction goes both ways*: a constant-expectation CE reduces to an NFCE of a suitable Markov game (Theorem 5.11). Moreover, and perhaps surprisingly, we show that computing a constant-expectation CE is PPAD-hard in the following cases.

Main Theorem 3. *Computing a constant-expectation correlated equilibrium is* PPAD*-hard in*

(i) binary-action n-player polymatrix games at constant precision, and

(ii) two-player normal-form games at exponentially small precision.

Together with our positive result, this fully characterizes the complexity of computing constant-expectation correlated equilibria, which is a natural solution concept in its own right. Main Theorem 3 marks a rare problem that admits an FPTAS and yet is PPAD-hard at exponentially small approximation error. Another notable example is *constant-rank* bimatrix games (Kannan & Theobald, 2010; Mehta, 2014). Our reduction in fact starts from a rank-2 bimatrix game, establishing an unexpected connection between constant-expectation CEs and constant-rank bimatrix games. We defer the details of our hardness results to Section 5 and additional related work to Section A.

[1] Sorin (1998) used the term "distribution equilibrium," but we use the more descriptive nomenclature of Heller (2012). Heller (2012) examined existence in the infinite-horizon setting, whereas we focus on computation.

2 PRELIMINARIES

We first describe the basic models of normal-form and Markov games.

2.1 NORMAL-FORM GAMES

A finite *normal-form game* is specified by $\mathcal{G} = ([n], (\mathcal{A}_i)_{i=1}^n, (u_i)_{i=1}^n)$. We make the usual normalization assumption $u_i : \mathcal{A} \to [0,1]$, where $\mathcal{A} := \prod_i \mathcal{A}_i$ is the joint action space. For a (potentially correlated) distribution $\mu \in \Delta(\mathcal{A})$, we use the shorthand notation

$$p_i^\mu(a_i) := \sum_{\boldsymbol{a}_{-i}} \mu(a_i, \boldsymbol{a}_{-i}) \text{ and } w_i^\mu(a_i) := \sum_{\boldsymbol{a}_{-i}} \mu(a_i, \boldsymbol{a}_{-i}) u_i(a_i, \boldsymbol{a}_{-i}).$$

Our main notion of equilibrium is the *correlated equilibrium*. We will be interested in two different notions of *approximate* correlated equilibrium, which we now define.

Definition 2.1 (Correlated equilibrium; Aumann, 1974)**.** A distribution $\mu \in \Delta(\mathcal{A})$ is a δ-*correlated equilibrium (CE)* if, for every player $i \in [n]$ and deviation function $\phi_i : \mathcal{A}_i \to \mathcal{A}_i$,

$$\mathop{\mathbb{E}}_{\boldsymbol{a}\sim\mu}\big[u_i(\phi_i(a_i), \boldsymbol{a}_{-i}) - u_i(\boldsymbol{a})\big] \le \delta.$$

It is a δ-*conditional correlated equilibrium* if, for every i and every pair $a_i, a_i' \in \mathcal{A}_i$ with $p_i^\mu(a_i) > 0$,

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

The latter notion of approximation is stronger: every δ-conditional CE is also a δ-CE, but not vice-versa. Our upper bounds will use conditional CE, and our lower bounds will use CE.

2.2 MARKOV GAMES

An n-player *Markov game* is played over stages $t = 1, \dots, H$. At stage t, the state belongs to a finite set $\mathcal{S}_t$. At state $s \in \mathcal{S}_t$, player i has a finite action set $\mathcal{A}_{i,t,s}$. The joint action set is denoted by $\mathcal{A}_{t,s} := \prod_{i=1}^n \mathcal{A}_{i,t,s}$. We will write $\boldsymbol{a}_t \in \mathcal{A}_{t,s_t}$. Player i receives reward $r_{i,t}(s_t, \boldsymbol{a}_t)$, and the next state is drawn from $\mathbb{P}_t(\cdot \mid s_t, \boldsymbol{a}_t) \in \Delta(\mathcal{S}_{t+1})$. (This randomization is independent conditioned on the current state and joint action.) The initial state is drawn from a distribution $\lambda \in \Delta(\mathcal{S}_1)$.

A *history* before stage t is $h_t = (s_1, \boldsymbol{a}_1, s_2, \boldsymbol{a}_2, \dots, s_{t-1}, \boldsymbol{a}_{t-1}, s_t)$. A *pure (history-dependent) policy* of player i maps every feasible history h_t to an action in $\mathcal{A}_{i,t,s_t}$. We will use Π_i to denote the set of pure policies of player i, and $\Pi := \prod_{i=1}^n \Pi_i$ to denote the set of pure policy profiles. A policy is *Markov* if it ignores the history before (t, s). In particular, a Markov policy is given by a vector $\boldsymbol{\pi}_i = \big(\pi_{i,t}(s)\big)_{t\in[H],\, s\in\mathcal{S}_t}$, where $\pi_{i,t}(s) \in \mathcal{A}_{i,t,s}$. For a *joint policy* $\boldsymbol{\pi} = (\pi_1, \dots, \pi_n)$, we let $u_i(\boldsymbol{\pi}) := \mathbb{E}[\sum_{t=1}^H r_{i,t}(s_t, \boldsymbol{a}_t)]$ be the expected cumulative reward. For convenience, we make the following normalization assumption.

Assumption 2.2 (Normalization)**.** For every $i, t, s, \boldsymbol{a}$, we have $0 \le r_{i,t}(s, \boldsymbol{a}) \le 1/H$.

In particular, every cumulative payoff lies in $[0, 1]$. This normalization is without loss after rescaling.

Normal-form correlation. A finite-horizon Markov game induces a finite normal-form game, in which every (possibly history-dependent) policy π_i is an action. Our main solution concept for Markov games will be the *normal-form correlated equilibrium* (NFCE), which is simply Definition 2.1 applied to this induced normal-form game. As with normal-form games, we will use both the standard and conditional notions of approximate NFCE; our upper bounds will apply to the latter and lower bounds to the former. This equilibrium requirement is significantly stronger than those guaranteed by existing algorithms, which revolve around notions akin to *extensive-form correlated equilibrium*, where recommendations are revealed only when the corresponding information sets are reached (von Stengel & Forges, 2008).

3 COMPUTING CONSTANT-EXPECTATION CORRELATED EQUILIBRIUM

We first focus on approximately computing *constant-expectation correlated equilibria*—a refinement of the usual notion—in normal-form games. This refinement of correlated equilibrium will be key to our results about Markov games.

Definition 3.1 (Constant-expectation CE)**.** A distribution μ is an (ϵ, δ)-*constant-expectation correlated equilibrium* if it is a δ-CE and, for every player $i \in [n]$, there exists $v_i \in \mathbb{R}$ such that

$$\sum_{a_i \in \mathrm{supp}(p_i^\mu)} p_i^\mu(a_i) \, |\mathbb{E}_{\boldsymbol{a}\sim\mu}[u_i(\boldsymbol{a}) \mid a_i] - v_i| \le \epsilon.$$

Equivalently, $\sum_{a_i \in \mathcal{A}_i} |w_i^\mu(a_i) - v_i p_i^\mu(a_i)| \le \epsilon$. Our algorithm will in fact satisfy exact obedience and the stronger condition $\ell_i \le \mathbb{E}_{\boldsymbol{a}\sim\mu}[u_i(\boldsymbol{a}) \mid a_i] \le \ell_i + \epsilon$ for every supported a_i, whereas our hardness results hold for the more permissive averaged condition.

To clarify this concept, we next provide an illustrative example.

Example 3.2. *We consider the following bimatrix game $\mathcal{G}$, where $\mathcal{A}_1 = \{\mathsf{U}, \mathsf{D}\}$ and $\mathcal{A}_2 = \{\mathsf{L}, \mathsf{R}\}$:*

	L	R
U	$(1,1)$	$(0,0)$
D	$(0,0)$	$(1,1)$

Let μ be the uniform distribution over the pure profiles (U, L) and (D, R). This is not a Nash equilibrium, because μ is not a product distribution. It is, however, a constant-expectation CE. On the other hand, the same distribution is not a constant-expectation CE in the following bimatrix game.

	L	R
U	$(1,1)$	$(0,0)$
D	$(0,0)$	$(1/2, 1/2)$

Nevertheless, μ is a correlated equilibrium in both games.

Linear programming-based algorithm. The basic idea is to exhaustively search over all possible values in Definition 3.1, up to some small granularity. In particular, for a rational $\epsilon > 0$, we let $G_\epsilon := \{k\epsilon : k = 0, 1, \ldots, \lfloor 1/\epsilon \rfloor\}$ be a discretization of the $[0,1]$ interval. For a fixed $\boldsymbol{\ell} = (\ell_i)_{i=1}^n \in G_\epsilon^n$, we consider the following linear feasibility program in μ:

$$\begin{aligned} &\mu \in \mathrm{CE}(\mathcal{G}), \\ &\ell_i p_i^\mu(a_i) \le w_i^\mu(a_i) \le (\ell_i + \epsilon) p_i^\mu(a_i) \quad \text{for every } i \in [n] \text{ and } a_i \in \mathcal{A}_i. \end{aligned} \tag{2}$$

Above, $\mathrm{CE}(\mathcal{G})$ denotes the set of correlated equilibria of $\mathcal{G}$. We use the notation $A := \max_{1 \le i \le n} |\mathcal{A}_i|$.

Proposition 3.3. *There exists $\boldsymbol{\ell} \in G_\epsilon^n$ for which the linear program (2) is feasible, and every feasible solution is an $(\epsilon/2, 0)$-constant-expectation correlated equilibrium. Moreover, $\ell_i \le \mathbb{E}_{\boldsymbol{a}\sim\mu}[u_i(\boldsymbol{a}) \mid a_i] \le \ell_i + \epsilon$ for every player i and supported action a_i. Therefore, if L is an upper bound on the bit complexity of the payoff tensor of $\mathcal{G}$, the running time is $(A/\epsilon)^{O(n)}\mathsf{poly}(L)$.*

The running time of this algorithm scales exponentially with the number of players. In Section 5, we will show that this obstacle is fundamental.

4 Efficient backward induction for NFCE in Markov games

Let $\mathcal{S} := \{(t, s) : t \in [H],\ s \in \mathcal{S}_t\}$ and $S := |\mathcal{S}| = \sum_t |\mathcal{S}_t|$. We initialize $\ell_{i,H+1}(s) := 0$ for every player i and terminal state $s \in \mathcal{S}_{H+1}$. We then process stages in reverse order. At a state $s \in \mathcal{S}_t$, we define an auxiliary normal-form game with payoff

$$g_{i,t,s}(\boldsymbol{a}) := r_{i,t}(s, \boldsymbol{a}) + \sum_{s' \in \mathcal{S}_{t+1}} \mathbb{P}_t(s' \mid s, \boldsymbol{a}) \ell_{i,t+1}(s'). \tag{3}$$

Using Proposition 3.3, we can find a correlated equilibrium $\mu_{t,s} \in \Delta(\mathcal{A}_{t,s})$ and values $\ell_{i,t}(s)$ satisfying

$$\ell_{i,t}(s) \le \mathbb{E}_{\boldsymbol{a}\sim\mu_{t,s}}\big[g_{i,t,s}(\boldsymbol{a}) \mid a_i\big] \le \ell_{i,t}(s) + \epsilon \tag{4}$$

for every i and every $a_i \in \mathcal{A}_{i,t,s}$ with positive marginal probability under $\mu_{t,s}$. After all states have been processed, we independently sample

$$\boldsymbol{a}_{t,s} = (a_{1,t,s}, \ldots, a_{n,t,s}) \sim \mu_{t,s} \tag{5}$$

for every $(t, s) \in \mathcal{S}$. The sampled pure Markov policy of player i is

$$\pi_{i,t}(s) := a_{i,t,s}. \tag{6}$$

We define ρ to be the resulting distribution over pure Markov policy profiles, succinctly represented through $(\mu_{t,s})_{(t,s)\in\mathcal{S}}$.

Lemma 4.1. *For every t, s, i, $0 \le \ell_{i,t}(s) \le (H - t + 1)/H$, and the auxiliary payoffs at stage t lie in $[0, (H - t + 1)/H]$.*

This shows that the payoffs lie in $[0, 1]$, so we can apply Proposition 3.3. Now, by virtue of independence, the constructed distribution satisfies the following key property.

Lemma 4.2. *Consider a player i and a pure Markov policy $\boldsymbol{\pi}_i$ with $\rho_i(\boldsymbol{\pi}_i) > 0$. Conditional on $\boldsymbol{\pi}_i$, the random vectors $\{\boldsymbol{\pi}_{-i,t}(s) : (t, s) \in \mathcal{S}\}$ are independent. At (t, s), their* conditional law *is*

$$\nu^{\boldsymbol{\pi}_i}_{i,t,s}(\boldsymbol{a}_{-i}) := \mathbb{P}[\boldsymbol{\pi}_{-i,t}(s) = \boldsymbol{a}_{-i} \mid \boldsymbol{\pi}_i] = \frac{\mu_{t,s}(\pi_{i,t}(s), \boldsymbol{a}_{-i})}{\mu_{i,t,s}(\pi_{i,t}(s))}, \tag{7}$$

where $\mu_{i,t,s}$ is the ith marginal of $\mu_{t,s}$.

This will allow us to argue about the deviation benefit of a player i upon receiving its recommendation. In particular, conditioning on $\boldsymbol{\pi}_i$, the other players' policy components are independent across states. Conditioning further on the history retains the law at (t, s) as $\nu^{\boldsymbol{\pi}_i}_{i,t,s}$ per (7).

Corollary 4.3. *For a player i, consider any possibly randomized, history-dependent deviation conditioned on $\boldsymbol{\pi}_i$. At any history h_t at state s having positive probability under this deviation, the conditional distribution of the opponents' prescribed actions at (t, s) is $\nu^{\boldsymbol{\pi}_i}_{i,t,s}$.*

4.1 The induced Markov decision process

We now leverage this notion of constant-expectation correlated equilibrium for Markov games.

For a player i and $\boldsymbol{\pi}_i \in \operatorname{supp}(\rho_i)$, we will analyze the *conditional MDP* $\mathcal{M}_i(\boldsymbol{\pi}_i)$ with state spaces $\mathcal{S}_t$, action sets $\mathcal{A}_{i,t,s}$, expected rewards

$$\bar{r}^{\boldsymbol{\pi}_i}_{i,t}(s, a'_i) := \sum_{\boldsymbol{a}_{-i}} \nu^{\boldsymbol{\pi}_i}_{i,t,s}(\boldsymbol{a}_{-i}) r_{i,t}(s, a'_i, \boldsymbol{a}_{-i}), \tag{8}$$

and transition kernels

$$\overline{\mathbb{P}}^{\boldsymbol{\pi}_i}_{i,t}(s' \mid s, a'_i) := \sum_{\boldsymbol{a}_{-i}} \nu^{\boldsymbol{\pi}_i}_{i,t,s}(\boldsymbol{a}_{-i}) \mathbb{P}_t(s' \mid s, a'_i, \boldsymbol{a}_{-i}). \tag{9}$$

Lemma 4.4 (Reduction to an MDP)**.** *Conditional on $\boldsymbol{\pi}_i$, the expected payoff of every possibly randomized, history-dependent deviation of player i in the Markov game equals its expected payoff in the MDP $\mathcal{M}_i(\boldsymbol{\pi}_i)$.*

This in particular implies that an optimal deviation may be chosen as deterministic and Markovian.

We define the Bellman values of this MDP by $V^{*,\boldsymbol{\pi}_i}_{i,H+1}(s) := 0$,

$$Q^{*,\boldsymbol{\pi}_i}_{i,t}(s, a'_i) := \bar{r}^{\boldsymbol{\pi}_i}_{i,t}(s, a'_i) + \sum_{s'} \overline{\mathbb{P}}^{\boldsymbol{\pi}_i}_{i,t}(s' \mid s, a'_i) V^{*,\boldsymbol{\pi}_i}_{i,t+1}(s'),$$

and $V^{*,\boldsymbol{\pi}_i}_{i,t}(s) := \max_{a'_i \in \mathcal{A}_{i,t,s}} Q^{*,\boldsymbol{\pi}_i}_{i,t}(s, a'_i)$. Now, let $V^{\boldsymbol{\pi}_i}_{i,t}(s)$ denote player i's value in $\mathcal{M}_i(\boldsymbol{\pi}_i)$ when following the recommended Markov policy $\boldsymbol{\pi}_i$ from stage t onward. We have

$$\begin{aligned} V^{\boldsymbol{\pi}_i}_{i,t}(s) = \sum_{\boldsymbol{a}_{-i}} \nu^{\boldsymbol{\pi}_i}_{i,t,s}(\boldsymbol{a}_{-i}) \Big(& r_{i,t}(s, \pi_{i,t}(s), \boldsymbol{a}_{-i}) \\ & + \sum_{s'} \mathbb{P}_t(s' \mid s, \pi_{i,t}(s), \boldsymbol{a}_{-i}) V^{\boldsymbol{\pi}_i}_{i,t+1}(s') \Big). \end{aligned} \tag{10}$$

Having computed a constant-expectation CE in each auxiliary game, we can derive the following.

Lemma 4.5. *Let $a_i := \pi_{i,t}(s)$. For every action $a'_i \in \mathcal{A}_{i,t,s}$,*

$$\mathbb{E}_{\boldsymbol{a}_{-i} \sim \nu^{\boldsymbol{\pi}_i}_{i,t,s}}[g_{i,t,s}(a'_i, \boldsymbol{a}_{-i})] \le \ell_{i,t}(s) + \epsilon \text{ and } \mathbb{E}_{\boldsymbol{a}_{-i} \sim \nu^{\boldsymbol{\pi}_i}_{i,t,s}}[g_{i,t,s}(a_i, \boldsymbol{a}_{-i})] \ge \ell_{i,t}(s).$$

Lemma 4.6. *For every i, every $\boldsymbol{\pi}_i \in \operatorname{supp}(\rho_i)$, every stage t, and every state $s \in \mathcal{S}_t$,*

$$V^{\boldsymbol{\pi}_i}_{i,t}(s) \ge \ell_{i,t}(s) \text{ and } V^{*,\boldsymbol{\pi}_i}_{i,t}(s) \le \ell_{i,t}(s) + (H - t + 1)\epsilon.$$

Algorithm 1: Backward induction for NFCE in Markov games

Input: Finite-horizon Markov game satisfying Assumption 2.2; accuracy $\epsilon \in (0, 1]$.
Output: An ϵ-conditional NFCE represented by $(\mu_{t,s})_{(t,s)\in\mathcal{S}}$, sampled independently.

$\eta \leftarrow \epsilon/H$
$\ell_{i,H+1}(s) \leftarrow 0$ for every $i \in [n]$ and $s \in \mathcal{S}_{H+1}$
for $t = H, H-1, \ldots, 1$ **do**
  foreach $s \in \mathcal{S}_t$ **do**
    Define, for every $i \in [n]$ and $\boldsymbol{a} \in \mathcal{A}_{t,s}$,

$$g_{i,t,s}(\boldsymbol{a}) \leftarrow r_{i,t}(s, \boldsymbol{a}) + \sum_{s' \in \mathcal{S}_{t+1}} \mathbb{P}_t(s' \mid s, \boldsymbol{a}) \ell_{i,t+1}(s').$$

    foreach $\boldsymbol{\ell} \in G_\eta^n$ **do**
      Solve (2) with payoffs $(g_{i,t,s})_{i=1}^n$, vector $\boldsymbol{\ell}$, and width η
      if *feasible, with solution* μ **then**
        $\mu_{t,s} \leftarrow \mu$ and $\ell_{i,t}(s) \leftarrow \ell_i$ for every $i \in [n]$
        break
return $(\mu_{t,s})_{(t,s)\in\mathcal{S}}$, *inducing* ρ *via* (5)–(6)

4.2 Main result

We are now ready to establish our main result.

Theorem 4.7 (Backward-induction conditional NFCE guarantee)**.** *The distribution ρ over pure Markov policies produced by Algorithm 1 is an ϵ-conditional NFCE. Moreover, assuming the stage games and transitions are given explicitly with bit complexity at most L, and $A := \max_{i,t,s} |\mathcal{A}_{i,t,s}|$, it runs in time*

$$S \left(\frac{AH}{\epsilon} \right)^{O(n)} \mathsf{poly}(L).$$

Proof. We consider a player i and a recommendation $\boldsymbol{\pi}_i \in \mathrm{supp}(\rho_i)$. By Lemma 4.4, the optimal conditional deviation value from the initial state distribution λ is $\mathbb{E}_{s_1\sim\lambda}[V_{i,1}^{*,\boldsymbol{\pi}_i}(s_1)]$. Following the recommendation yields $\mathbb{E}_{s_1\sim\lambda}[V_{i,1}^{\boldsymbol{\pi}_i}(s_1)]$. Thus, Lemma 4.6 implies

$$\mathbb{E}_{s_1\sim\lambda}[V_{i,1}^{*,\boldsymbol{\pi}_i}(s_1) - V_{i,1}^{\boldsymbol{\pi}_i}(s_1)] \le H\eta = \epsilon.$$

The output is represented by the S distributions $\mu_{t,s}$, which are to be sampled independently. □

5 Hardness results

We now turn to our hardness results. We begin by establishing PPAD-hardness for constant-expectation CEs, and then translate those results to NFCEs in Markov games.

5.1 Succinct multi-player games

We first examine succinct many-player games. Here, we require an output representation from which the marginal probabilities can be computed in polynomial time, such as a mixture of product distributions (Papadimitriou & Roughgarden, 2008).

Our hardness result holds even in *polymatrix games*. A *polymatrix game* is specified by a graph $([n], E)$ and a payoff matrix $\mathbf{M}_{ii'}$ for every edge $(i, i') \in E$. Let N_i be the *neighbor set* of player i. At a pure action profile $\boldsymbol{a}$, player i receives $u_i(\boldsymbol{a}) = \sum_{i'\in N_i} \mathbf{M}_{ii'}(a_i, a_{i'})$. We say that the game is normalized if $u_i(\boldsymbol{a}) \in [0, 1]$ for every player and action profile. Our reduction will use the hardness result of Deligkas et al. (2024).

Theorem 5.1 (Deligkas et al., 2024)**.** *For every constant $0 < \delta < 1/12$, computing a δ-Nash equilibrium in a normalized polymatrix game of maximum degree three with two actions per player is* PPAD*-hard.*

Let $\mathcal{G}$ be such a polymatrix game with $\mathcal{A}_i = \{0, 1\}$ for every player. We may assume that the source game has no isolated players. We define the payoff difference

$$d_i(\boldsymbol{a}_{-i}) := u_i(1, \boldsymbol{a}_{-i}) - u_i(0, \boldsymbol{a}_{-i}) = \sum_{i' \in N_i} \big(\mathbf{M}_{ii'}(1, a_{i'}) - \mathbf{M}_{ii'}(0, a_{i'})\big).$$

Normalization implies $d_i(\boldsymbol{a}_{-i}) \in [-1, 1]$. For a constant $0 < \epsilon < 1/24$, we set $\alpha := 1/4 + 6\epsilon$, so $12\epsilon < \alpha < 1/2$. We construct a transformed game $\mathcal{G}'$ on the same graph and action sets, with payoffs

$$u_i'(0, \boldsymbol{a}_{-i}) := \frac{1}{2} + \alpha d_i(\boldsymbol{a}_{-i}) \text{ and } u_i'(1, \boldsymbol{a}_{-i}) := \frac{1}{2} + \frac{1}{2} d_i(\boldsymbol{a}_{-i}).$$

This game is still a polymatrix game, as we point out below.

Lemma 5.2. *The game $\mathcal{G}'$ is a normalized polymatrix game of maximum degree three with two actions per player.*

For polymatrix games, the utilities can be expressed in terms of the marginal strategies.

Lemma 5.3. *Let $\mu \in \Delta(\mathcal{A})$, and let $\boldsymbol{x}$ be its marginal strategy profile. Then, for every player i,*

$$\mathbb{E}_{\boldsymbol{a} \sim \mu}[d_i(\boldsymbol{a}_{-i})] = u_i(1, \boldsymbol{x}_{-i}) - u_i(0, \boldsymbol{x}_{-i}).$$

The following lemma now makes crucial use of the constant expectation property.

Lemma 5.4. *Let μ be an (ϵ, δ)-constant-expectation CE of $\mathcal{G}'$. For any player i,*

$$\left(\frac{1}{2} - \alpha\right) \sum_{a_i \in \mathrm{supp}(p_i^\mu)} p_i^\mu(a_i) \max\left\{(1 - 2a_i)\mathbb{E}_\mu[d_i(\boldsymbol{a}_{-i}) \mid a_i], 0\right\} \leq \delta.$$

If both actions have positive marginal probability, then

$$\min\{p_i^\mu(0), p_i^\mu(1)\} \left| \frac{1}{2}\mathbb{E}_\mu[d_i(\boldsymbol{a}_{-i}) \mid a_i = 1] - \alpha \mathbb{E}_\mu[d_i(\boldsymbol{a}_{-i}) \mid a_i = 0] \right| \leq \epsilon.$$

This lemma, together with simple algebra, implies a certain equilibrium collapse: the marginals of any constant-expectation CE reveal a Nash equilibrium of the original game.

Lemma 5.5. *If μ is an (ϵ, δ)-constant-expectation correlated equilibrium of $\mathcal{G}'$ and $\boldsymbol{x}$ is its marginal strategy profile, then $\boldsymbol{x}$ is an η-Nash equilibrium of $\mathcal{G}$, where $\eta = \epsilon/\alpha + \delta/(\alpha(1 - 2\alpha))$.*

Combining with the hardness result of Deligkas et al. (2024) (Theorem 5.1), we arrive at the following.

Theorem 5.6 (PPAD-hardness)**.** *For every pair of constants such that $0 < \epsilon < 1/24$ and $0 \leq \delta < (1 - 24\epsilon)^2/96$, computing an (ϵ, δ)-constant-expectation correlated equilibrium in a normalized polymatrix game of maximum degree three with two actions per player is* PPAD*-hard.*

5.2 Two-player games

We now turn our attention to two-player (bimatrix) games. We reduce from the problem of computing an exact Nash equilibrium of a rank-2 bimatrix game, which is PPAD-hard (Chen et al., 2016). Let $\mathcal{G}$ be such a game, so the sum of its payoff matrices has rank at most two. We write

$$u_1(\boldsymbol{a}) + u_2(\boldsymbol{a}) = f_1(a_1)g_1(a_2) + f_2(a_1)g_2(a_2), \tag{11}$$

where $f_1, f_2 : \mathcal{A}_1 \to [-1, 1]$ and $g_1, g_2 : \mathcal{A}_2 \to [-1, 1]$. For a parameter $K \geq 3$, we consider the game $\mathcal{G}'$ on the same action sets with payoffs

$$\begin{aligned} u_1'(\boldsymbol{a}) &:= \frac{u_1(\boldsymbol{a}) + 2g_2(a_2) + K(g_1(a_2) + 2)}{4K}, \\ u_2'(\boldsymbol{a}) &:= \frac{u_2(\boldsymbol{a}) + 2f_1(a_1) + K(f_2(a_1) + 2)}{4K}. \end{aligned} \tag{12}$$

The strategic equivalence claimed below uses a standard invariance (Adsul et al., 2021).

Lemma 5.7. *$\mathcal{G}'$ has payoffs in $[0,1]$, has rank at most two, and has the same set of CEs as $\mathcal{G}$. Moreover, μ is a δ-CE of $\mathcal{G}'$ if and only if it is a $(4K\delta)$-CE of $\mathcal{G}$.*

The role of constant expectation is to make recommendations nearly uninformative about opponent-dependent terms, on average. For Player 1, when $K \gg 1$, the term $Kg_1(a_2)$ dominates the remaining payoff. So, we can control the expected absolute deviation of the conditional expectation of $g_1(a_2)$ from its unconditional expectation. Player 2 similarly controls $f_2(a_1)$. These are precisely the two correlations that can separate a correlated equilibrium from a Nash equilibrium, as we establish in Lemma 5.9. In what follows, we use the notation $\mathrm{Cov}_\mu(X,Y) := \mathbb{E}_\mu[XY] - \mathbb{E}_\mu[X]\mathbb{E}_\mu[Y]$.

Lemma 5.8. *If μ is an η-CE of $\mathcal{G}$ with marginal strategy profile $\boldsymbol{x}$, then*

$$\sum_{i=1}^{2}\left(\max_{a_i'\in\mathcal{A}_i} u_i(a_i', \boldsymbol{x}_{-i}) - u_i(\boldsymbol{x})\right) \le 2\eta + \sum_{\ell=1}^{2}\mathrm{Cov}_\mu(f_\ell(a_1), g_\ell(a_2)).$$

As discussed, the averaged constant-expectation property suffices to control the covariance terms.

Lemma 5.9. *If μ is an (ϵ,δ)-constant-expectation correlated equilibrium of $\mathcal{G}'$, then, for $\ell \in \{1,2\}$,*

$$|\mathrm{Cov}_\mu(f_\ell(a_1), g_\ell(a_2))| \le 8\epsilon + \frac{6}{K}.$$

Combining, we arrive at the following consequence, where we assume that ϵ and δ are given as part of the input in binary, so that $\log(1/\epsilon)$ and $\log(1/\delta)$ are polynomial.

Theorem 5.10. *Computing an (ϵ,δ)-constant-expectation correlated equilibrium is* PPAD*-hard in bimatrix games when ϵ and δ are exponentially small.*

5.3 Hardness for NFCE in Markov games

We finally circle back to Markov games. Our main hardness result establishes that relying on constant-expectation CEs in backward induction is fundamental:

Theorem 5.11. *Let $\mathcal{G}$ be an n-player normal-form game with utilities bounded in $[0,1]$, and $0 \le \epsilon \le 1/(16n)$. There is a Markov game $\mathcal{M}$ with $n+1$ players, horizon $H=2$, $S=n+1$ states, and stage rewards bounded in $[0,1/2]$, such that every ϵ-NFCE with independent state recommendations yields a $(16n\epsilon, 16\epsilon)$-constant-expectation CE of $\mathcal{G}$.*

By independent state recommendations, we mean that the joint action recommendations are sampled independently across pairs (t,s); recommendations to different players at the same state may be correlated. Combining the above result with Theorem 5.6 and Theorem 5.10, we therefore immediately derive PPAD-hardness of computing NFCE with independent state recommendations.

Corollary 5.12. *Computing an ϵ-NFCE with independent state recommendations is* PPAD*-hard for Markov games satisfying Assumption 2.2, even with horizon $H=2$, in each of the following cases:*

1. *$(n+1)$-player succinct games, two actions per player at each state, and $\epsilon = 1/(3072n)$.*
2. *Three-player games, a constant number of states, and exponentially small ϵ.*

6 Conclusion

We established an efficient algorithm for computing conditional normal-form correlated equilibria in finite-horizon Markov games with a fixed number of players. Our key insight is that constant-expectation correlated equilibria, combined with independent sampling across states, make backward induction sound even when players observe their entire recommended policies before play. Our results suggest several interesting questions for future research. First, our algorithm is neither decentralized nor model free. Addressing this limitation is an important challenge for future work, but we believe our work lays the groundwork for such extensions. It would also be interesting to understand whether there are efficient algorithms when the number of players is part of the input. Our PPAD-hardness rules this out under state independent recommendations, but perhaps one can circumvent that result by introducing correlation between states. Finally, the complexity of computing NFCEs in other game representations, most notably the extensive form, remains a challenging open question.

AI USE STATEMENT

The initial idea for this paper, and the main idea for the proof of Main Theorem 1, are due to the authors. Various versions of GPT were used to assist in proofs of some of the lower bounds. All proofs were checked, understood, and rewritten by the authors.

ETHICS STATEMENT

This paper examines the computational complexity of equilibrium concepts in Markov games. It presents theoretical results which do not raise any immediate ethical concerns.

REPRODUCIBILITY STATEMENT

This is a purely theoretical paper. Proofs of all claims are included, either in the main body or in the appendix, so that theoretical claims can be fully verified.

REFERENCES

Bharat Adsul, Jugal Garg, Ruta Mehta, and Milind Sohoni. Rank-1 bimatrix games: a homeomorphism and a polynomial time algorithm. In *Symposium on Theory of computing (STOC)*, 2011.

Bharat Adsul, Jugal Garg, Ruta Mehta, Milind Sohoni, and Bernhard von Stengel. Fast algorithms for rank-1 bimatrix games. *Operations Research*, 69(2):613–631, 2021.

Noga Alon, Troy Lee, Adi Shraibman, and Santosh Vempala. The approximate rank of a matrix and its algorithmic applications: approximate rank. In *Symposium on Theory of computing (STOC)*, 2013.

Ioannis Anagnostides and Weiqiang Zheng. Algorithmic collusion and the complexity of information-value-free equilibria. *arXiv:2609.22757*, 2026.

Ioannis Anagnostides, Constantinos Daskalakis, Gabriele Farina, Noah Golowich, Tuomas Sandholm, and Brian Hu Zhang. On the complexity of correlated equilibria beyond normal-form games. *arXiv:2605.17665*, 2026.

Eshwar Ram Arunachaleswaran, Natalie Collina, Yishay Mansour, Mehryar Mohri, Jon Schneider, and Balasubramanian Sivan. Swap regret and correlated equilibria beyond normal-form games. In *Conference on Economics and Computation (EC)*, 2025.

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

Yang Cai, Haipeng Luo, Chen-Yu Wei, and Weiqiang Zheng. Near-optimal policy optimization for correlated equilibrium in general-sum Markov games. In *International Conference on Artificial Intelligence and Statistics (AISTATS)*, 2024.

Xi Chen, Xiaotie Deng, and Shang-Hua Teng. Settling the complexity of computing two-player Nash equilibria. *Journal of the ACM (JACM)*, 56(3):1–57, 2009.

Xi Chen, Anthi Orfanou, Dimitris Paparas, and Mihalis Yannakakis. Rank 2 bimatrix games are PPAD-hard. Unpublished manuscript, 2016.

Qiwen Cui, Kaiqing Zhang, and Simon S. Du. Breaking the curse of multiagents in a large state space: RL in Markov games with independent linear function approximation. In *Conference on Learning Theory (COLT)*, 2023.

Yuval Dagan, Constantinos Daskalakis, Maxwell Fishelson, and Noah Golowich. From external to swap regret 2.0: An efficient reduction for large action spaces. In *Symposium on Theory of Computing (STOC)*, 2024.

Constantinos Daskalakis, Paul W Goldberg, and Christos H Papadimitriou. The complexity of computing a Nash equilibrium. *Communications of the ACM*, 52(2):89–97, 2009.

Constantinos Daskalakis, Noah Golowich, and Kaiqing Zhang. The complexity of Markov equilibrium in stochastic games. In *Conference on Learning Theory (COLT)*, 2023.

Constantinos Daskalakis, Gabriele Farina, Maxwell Fishelson, Charilaos Pipis, and Jon Schneider. Efficient learning and computation of linear correlated equilibrium in general convex games. In *Symposium on Theory of Computing (STOC)*, 2025.

Argyrios Deligkas, John Fearnley, Alexandros Hollender, and Themistoklis Melissourgos. Pure-Circuit: Tight inapproximability for PPAD. *Journal of the ACM*, 71(5):1–48, 2024.

Kousha Etessami and Mihalis Yannakakis. On the complexity of Nash equilibria and other fixed points. *SIAM Journal on Computing*, 39(6):2531–2597, 2010.

Gabriele Farina, Andreas Kontogiannis, Ioannis Panageas, and Vasilis Pollatos. The complexity of computing coarse correlated equilibria in Markov games with a single controller. *arXiv:2607.10897*, 2026.

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)*, 2023.

Geoffrey J Gordon, Amy Greenwald, and Casey Marks. No-regret learning in convex games. In *International Conference on Machine learning (ICML)*, 2008.

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

Jason D Hartline, Chang Wang, and Chenhao Zhang. The computable but not learnable information-value-free equilibria and regulation of algorithmic collusion. In *Conference on Economics and Computation (EC)*, 2026.

Yuval Heller. Sequential correlated equilibria in stopping games. *Operations Research*, 60(1): 209–224, 2012.

Chi Jin, Qinghua Liu, Yuanhao Wang, and Tiancheng Yu. V-learning—a simple, efficient, decentralized algorithm for multiagent reinforcement learning. *Mathematics of Operations Research*, 49(4): 2295–2322, 2024.

Ravi Kannan and Thorsten Theobald. Games of fixed rank: A hierarchy of bimatrix games. *Economic Theory*, 42(1):157–173, 2010.

Xiangyu Liu and Kaiqing Zhang. Partially observable multiagent reinforcement learning with information sharing. *SIAM Journal on Control and Optimization*, 64(2):673–697, 2026.

Weichao Mao, Haoran Qiu, Chen Wang, Hubertus Franke, Zbigniew Kalbarczyk, and Tamer Başar. $\widetilde{\mathcal{O}}(T^{-1})$ convergence to (coarse) correlated equilibria in full-information general-sum Markov games. In *Learning for Dynamics & Control Conference (L4DC)*, 2024.

Ruta Mehta. Constant rank bimatrix games are PPAD-hard. In *Symposium on Theory of Computing (STOC)*, 2014.

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

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 *Conference on Artificial Intelligence (AAAI)*, 2021b.

John Nash. Non-cooperative games. *Annals of Mathematics*, 54(2):286–295, 1951.

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

Nikolas Patris and Ioannis Panageas. Learning Nash equilibria in rank-1 games. In *International Conference on Learning Representations (ICLR)*, 2024.

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

Ziang Song, Song Mei, and Yu Bai. When can we learn general-sum Markov games with a large number of players sample-efficiently? In *International Conference on Learning Representations (ICLR)*, 2022.

Sylvain Sorin. Distribution equilibrium I: Definition and examples. Working Paper 98-35, THEMA, Université Paris X–Nanterre, 1998.

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

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

Yuanhao Wang, Qinghua Liu, Yu Bai, and Chi Jin. Breaking the curse of multiagency: Provably efficient decentralized multi-agent RL with function approximation. In *Conference on Learning Theory (COLT)*, 2023.

Asrin Efe Yorulmaz, Ugur Aydin, and Tamer Basar. Independent reinforcement learning in discounted Markov games. *arXiv:2609.00504*, 2026.

Brian Hu Zhang, Ioannis Anagnostides, Emanuel Tewolde, Ratip Emin Berker, Gabriele Farina, Vincent Conitzer, and Tuomas Sandholm. Learning and computation of Φ-equilibria at the frontier of tractability. In *Conference on Economics and Computation (EC)*, 2025.

A FURTHER RELATED WORK

Markov games. There has been much work examining correlated equilibrium concepts in Markov games. Song et al. (2022); Jin et al. (2024) developed decentralized algorithms for learning (coarse) correlated equilibria with sample complexity that avoids an exponential dependence on the number of players. Wang et al. (2023); Cui et al. (2023) extended such guarantees to settings with function approximation, while Cai et al. (2024); Mao et al. (2024) obtained near-optimal convergence rates in the full-feedback setting. These results concern equilibrium notions weaker than NFCE and do not guarantee obedience when each player observes its entire recommended policy before play. Our results allow arbitrary history-dependent deviations conditioned on the entire policy recommendation. On the flip side, our algorithm is centralized and model based, and is polynomial only for a fixed number of players.

A complementary line of work has established computational barriers. Daskalakis et al. (2023) showed hardness for stationary Markov coarse correlated equilibria in discounted stochastic games, while also giving efficient algorithms for nonstationary Markov coarse correlated equilibria. Farina et al. (2026) showed that hardness persists even in two-player Markov games with a single controller. These hardness results concern stationary equilibria in the discounted setting, whereas our algorithm leverages the finite horizon through backward induction. Separately, Foster et al. (2023) established hardness for computing coarse correlated equilibria represented as sparse mixtures of product policies, which rules out computationally efficient independent no-regret learning. More recently, Yorulmaz et al. (2026) obtained computational lower bounds for independent learning in discounted games.

Beyond fully observable Markov games, Liu & Zhang (2026) examined partially observable games with information sharing, obtaining quasi-polynomial guarantees under suitable information-sharing and observability assumptions. Our approach relies on full observability to solve the continuation problems separately.

Beyond Markov games. As discussed earlier, existing algorithmic results focus on relaxations of the normal-form correlated equilibrium. Such relaxations can be understood through the framework of Φ-*equilibria* (Greenwald & Jafari, 2003; Stoltz & Lugosi, 2007; Gordon et al., 2008), which has engendered much recent work (Morrill et al., 2021a;b; Arunachaleswaran et al., 2025; Daskalakis et al., 2025; Zhang et al., 2025). Anagnostides et al. (2026) showed that computing a normal-form correlated equilibrium in concave (quadratic) games is as hard as finding fixed points of contractions, which is believed to be a hard problem. The complexity, for example, in extensive-form games remains an open question. In particular, it is open whether there is an FPTAS for NFCE even in two-player (imperfect-information) extensive-form games. It is worth noting that our approach relies crucially on full observability and does not directly extend to extensive-form games. The basic challenge in extensive-form games is that continuation problems cannot be solved independently, so there are additional obstacles in applying backward induction.

Constant-rank bimatrix games. An unexpected connection we uncover in our paper is between constant-expectation CEs and Nash equilibria in constant-rank bimatrix games. This class of games was introduced by Kannan & Theobald (2010), motivated by applications in economics. Unlike general bimatrix games, where there is no FPTAS unless $\mathsf{P} = \mathsf{PPAD}$ (Chen et al., 2009), Kannan & Theobald (2010) showed that constant-rank bimatrix games admit an FPTAS (*cf.* Alon et al., 2013). Mehta (2014) later established that, even in rank-3 bimatrix games, finding a Nash equilibrium is PPAD-complete when the precision is exponentially small. In contrast, rank-1 games admit a polynomial-time algorithm (Adsul et al., 2011) and are amenable to learning dynamics (Patris & Panageas, 2024).

Other refinements of correlated equilibria. A related refinement is *information-value-free correlated equilibrium (IVFCE)*, introduced recently by Hartline et al. (2026), which requires that each player's (unconditional) expected payoff can be matched by a fixed action chosen without observing the recommendation. In contrast, constant-expectation CE requires each player's conditional expected payoff from obedience to be independent of its recommendation. Neither property implies the other, although both hold at any Nash equilibrium. The two notions also exhibit similarities in their computational complexity. In succinct many-player games, both are PPAD-hard to compute even at constant precision (Anagnostides & Zheng, 2026). In explicit normal-form games, however,

an exact IVFCE can be computed in polynomial time (Hartline et al., 2026), whereas we show that computing a constant-expectation CE is PPAD-hard at exponentially small precision even with two players, despite admitting an FPTAS for a fixed number of players.

B OMITTED PROOFS

This section contains the omitted proofs.

B.1 PROOFS FROM SECTION 3

We begin with Proposition 3.3.

Proposition 3.3. *There exists $\boldsymbol{\ell} \in G_\epsilon^n$ for which the linear program* (2) *is feasible, and every feasible solution is an $(\epsilon/2, 0)$-constant-expectation correlated equilibrium. Moreover, $\ell_i \le \mathbb{E}_{\boldsymbol{a}\sim\mu}[u_i(\boldsymbol{a}) \mid a_i] \le \ell_i + \epsilon$ for every player i and supported action a_i. Therefore, if L is an upper bound on the bit complexity of the payoff tensor of $\mathcal{G}$, the running time is $(A/\epsilon)^{O(n)}\mathsf{poly}(L)$.*

Proof. By Nash's theorem (Nash, 1951), $\mathcal{G}$ admits a Nash equilibrium $\boldsymbol{x} = (\boldsymbol{x}_i)_{i=1}^n$. Let $\mu = \bigotimes_i \boldsymbol{x}_i$ be the induced product distribution, which is clearly a correlated equilibrium. Every action $a_i \in \mathrm{supp}(\boldsymbol{x}_i)$ is a best response to $\boldsymbol{x}_{-i}$, so $\mathbb{E}_{\boldsymbol{a}\sim\mu}[u_i(\boldsymbol{a}) \mid a_i] = u_i(a_i, \boldsymbol{x}_{-i})$. For $a_i \notin \mathrm{supp}(\boldsymbol{x}_i)$, both $p_i^\mu(a_i)$ and $w_i^\mu(a_i)$ are zero. For $a_i \in \mathrm{supp}(\boldsymbol{x}_i)$, let $\ell_i := \epsilon\lfloor u_i(a_i, \boldsymbol{x}_{-i})/\epsilon \rfloor$. Since $u_i(a_i, \boldsymbol{x}_{-i}) \in [0, 1]$, we have $\ell_i \in G_\epsilon$ and $\ell_i \le u_i(a_i, \boldsymbol{x}_{-i}) < \ell_i + \epsilon$. This shows that μ must be a feasible point of the LP induced by $\boldsymbol{\ell}$.

Conversely, every feasible solution to (2) is a correlated equilibrium. Whenever $p_i^\mu(a_i) > 0$, dividing the last line by $p_i^\mu(a_i)$ gives $\ell_i \le \mathbb{E}_{\boldsymbol{a}\sim\mu}[u_i(\boldsymbol{a}) \mid a_i] \le \ell_i + \epsilon$. Thus, choosing $v_i = \ell_i + \epsilon/2$ bounds the absolute deviation by $\epsilon/2$ for every supported recommendation, and hence also after averaging over recommendations, giving an $(\epsilon/2, 0)$-constant-expectation correlated equilibrium as required by Definition 3.1.

Finally, the claimed running time follows because G_ϵ^n contains $(1/\epsilon)^{O(n)}$ points, and each induced linear program can be solved in time polynomial in $A^n(L + \log(1/\epsilon))$. □

B.2 PROOFS FROM SECTION 4

We continue with the proofs from Section 4.

Lemma 4.1. *For every t, s, i, $0 \le \ell_{i,t}(s) \le (H - t + 1)/H$, and the auxiliary payoffs at stage t lie in $[0, (H - t + 1)/H]$.*

Proof. At stage $H + 1$, the claim holds because the continuation values are zero. Suppose it holds at stage $t + 1$. By the assumed normalization (Assumption 2.2) and the definition of the auxiliary normal-form games, we have

$$0 \le g_{i,t,s}(\boldsymbol{a}) \le \frac{1}{H} + \frac{H - t}{H} = \frac{H - t + 1}{H}.$$

Since $\ell_{i,t}(s) \ge 0$ belongs to the grid and is bounded above by a supported action's conditional expected auxiliary payoff, the same upper bound holds for $\ell_{i,t}(s)$. □

Lemma 4.2. *Consider a player i and a pure Markov policy $\boldsymbol{\pi}_i$ with $\rho_i(\boldsymbol{\pi}_i) > 0$. Conditional on $\boldsymbol{\pi}_i$, the random vectors $\{\boldsymbol{\pi}_{-i,t}(s) : (t, s) \in \mathcal{S}\}$ are independent. At (t, s), their* conditional law *is*

$$\nu_{i,t,s}^{\boldsymbol{\pi}_i}(\boldsymbol{a}_{-i}) := \mathbb{P}[\boldsymbol{\pi}_{-i,t}(s) = \boldsymbol{a}_{-i} \mid \boldsymbol{\pi}_i] = \frac{\mu_{t,s}(\pi_{i,t}(s), \boldsymbol{a}_{-i})}{\mu_{i,t,s}(\pi_{i,t}(s))}, \tag{7}$$

where $\mu_{i,t,s}$ is the ith marginal of $\mu_{t,s}$.

Proof. Let $\rho_i(\boldsymbol{\pi}_i) > 0$. Independence in the constructed distribution (5) gives $\rho_i(\boldsymbol{\pi}_i) = \prod_{(t,s)\in\mathcal{S}} \mu_{i,t,s}(\pi_{i,t}(s))$. Moreover, for any $(\boldsymbol{a}_{-i,t,s})_{(t,s)\in\mathcal{S}}$,

$$\mathbb{P}\big[\boldsymbol{\pi}_{-i,t}(s) = \boldsymbol{a}_{-i,t,s} \text{ for all } (t, s) \mid \boldsymbol{\pi}_i\big] = \frac{\prod_{(t,s)\in\mathcal{S}} \mu_{t,s}(\pi_{i,t}(s), \boldsymbol{a}_{-i,t,s})}{\prod_{(t,s)\in\mathcal{S}} \mu_{i,t,s}(\pi_{i,t}(s))} = \prod_{(t,s)\in\mathcal{S}} \nu_{i,t,s}^{\boldsymbol{\pi}_i}(\boldsymbol{a}_{-i,t,s}),$$

as claimed. □

Lemma 4.4 (Reduction to an MDP)**.** *Conditional on $\boldsymbol{\pi}_i$, the expected payoff of every possibly randomized, history-dependent deviation of player i in the Markov game equals its expected payoff in the MDP $\mathcal{M}_i(\boldsymbol{\pi}_i)$.*

Proof. Corollary 4.3 shows that after every history at (t, s), and for every deviating action a_i', the conditional expected immediate reward and transition probabilities are given by (8) and (9), respectively. Applying iterated expectation over the stages, the expected cumulative reward in the Markov game conditional on $\boldsymbol{\pi}_i$ is equal to that in $\mathcal{M}_i(\boldsymbol{\pi}_i)$. □

Lemma 4.5. *Let $a_i := \pi_{i,t}(s)$. For every action $a_i' \in \mathcal{A}_{i,t,s}$,*

$$\mathbb{E}_{\boldsymbol{a}_{-i}\sim\nu^{\boldsymbol{\pi}_i}_{i,t,s}}[g_{i,t,s}(a_i', \boldsymbol{a}_{-i})] \le \ell_{i,t}(s) + \epsilon \text{ and } \mathbb{E}_{\boldsymbol{a}_{-i}\sim\nu^{\boldsymbol{\pi}_i}_{i,t,s}}[g_{i,t,s}(a_i, \boldsymbol{a}_{-i})] \ge \ell_{i,t}(s).$$

Proof. Because $\boldsymbol{\pi}_i \in \operatorname{supp}(\rho_i)$, the recommendation a_i has positive marginal probability under $\mu_{t,s}$. Thus,

$$\mathbb{E}_{\boldsymbol{a}_{-i}\sim\nu^{\boldsymbol{\pi}_i}_{i,t,s}}[g_{i,t,s}(a_i, \boldsymbol{a}_{-i})] \ge \mathbb{E}_{\boldsymbol{a}_{-i}\sim\nu^{\boldsymbol{\pi}_i}_{i,t,s}}[g_{i,t,s}(a_i', \boldsymbol{a}_{-i})].$$

Combining this with (4) gives the claim. □

Lemma 4.6. *For every i, every $\boldsymbol{\pi}_i \in \operatorname{supp}(\rho_i)$, every stage t, and every state $s \in \mathcal{S}_t$,*

$$V^{\boldsymbol{\pi}_i}_{i,t}(s) \ge \ell_{i,t}(s) \text{ and } V^{*,\boldsymbol{\pi}_i}_{i,t}(s) \le \ell_{i,t}(s) + (H - t + 1)\epsilon.$$

Proof. We proceed by backward induction, starting from the lower bound. At $t = H + 1$, the claim is immediate. Suppose that it holds at stage $t + 1$, and let $a_i := \pi_{i,t}(s)$. Using (10), the induction hypothesis $V^{\boldsymbol{\pi}_i}_{i,t+1}(s') \ge \ell_{i,t+1}(s')$, and Lemma 4.5, we obtain

$$\begin{aligned} V^{\boldsymbol{\pi}_i}_{i,t}(s) &\ge \mathbb{E}_{\boldsymbol{a}_{-i}\sim\nu^{\boldsymbol{\pi}_i}_{i,t,s}}\left[r_{i,t}(s, a_i, \boldsymbol{a}_{-i}) + \sum_{s'} \mathbb{P}_t(s' \mid s, a_i, \boldsymbol{a}_{-i})\ell_{i,t+1}(s')\right] \\ &= \mathbb{E}_{\boldsymbol{a}_{-i}\sim\nu^{\boldsymbol{\pi}_i}_{i,t,s}}[g_{i,t,s}(a_i, \boldsymbol{a}_{-i})] \ge \ell_{i,t}(s). \end{aligned}$$

For the upper bound, the Bellman recursion, the induction hypothesis $V^{*,\boldsymbol{\pi}_i}_{i,t+1}(s') \le \ell_{i,t+1}(s') + (H - t)\epsilon$, and Lemma 4.5 give

$$\begin{aligned} V^{*,\boldsymbol{\pi}_i}_{i,t}(s) &\le \max_{a_i'\in\mathcal{A}_{i,t,s}} \mathbb{E}_{\boldsymbol{a}_{-i}\sim\nu^{\boldsymbol{\pi}_i}_{i,t,s}}[g_{i,t,s}(a_i', \boldsymbol{a}_{-i}) + (H - t)\epsilon] \\ &\le \ell_{i,t}(s) + \epsilon + (H - t)\epsilon \\ &= \ell_{i,t}(s) + (H - t + 1)\epsilon. \end{aligned}$$

This completes the induction. □

B.3 Proofs for Section 5

We conclude with the proofs from Section 5, starting from our hardness result for polymatrix games.

Lemma 5.2. *The game $\mathcal{G}'$ is a normalized polymatrix game of maximum degree three with two actions per player.*

Proof. For each edge (i, i'), we define

$$\begin{aligned} \mathbf{M}'_{ii'}(0, a_{i'}) &:= \frac{1}{2|N_i|} + \alpha\big(\mathbf{M}_{ii'}(1, a_{i'}) - \mathbf{M}_{ii'}(0, a_{i'})\big), \\ \mathbf{M}'_{ii'}(1, a_{i'}) &:= \frac{1}{2|N_i|} + \frac{1}{2}\big(\mathbf{M}_{ii'}(1, a_{i'}) - \mathbf{M}_{ii'}(0, a_{i'})\big). \end{aligned}$$

Summing over $i' \in N_i$ gives the specified payoffs. Because $d_i \in [-1, 1]$ and $0 < \alpha < 1/2$, both payoffs lie in $[0, 1]$. The interaction graph and action sets remain the same, and the construction has polynomial bit complexity. □

Lemma 5.3. *Let $\mu \in \Delta(\mathcal{A})$, and let $\boldsymbol{x}$ be its marginal strategy profile. Then, for every player i,*

$$\mathbb{E}_{\boldsymbol{a}\sim\mu}[d_i(\boldsymbol{a}_{-i})] = u_i(1, \boldsymbol{x}_{-i}) - u_i(0, \boldsymbol{x}_{-i}).$$

Proof. Linearity and the polymatrix structure give

$$\begin{aligned}\mathbb{E}_{\boldsymbol{a}\sim\mu}[d_i(\boldsymbol{a}_{-i})] &= \sum_{i'\in N_i} \mathbb{E}_{\boldsymbol{a}\sim\mu}[\mathbf{M}_{ii'}(1, a_{i'}) - \mathbf{M}_{ii'}(0, a_{i'})] \\ &= \sum_{i'\in N_i} \sum_{a_{i'}\in\{0,1\}} x_{i'}(a_{i'})\big(\mathbf{M}_{ii'}(1, a_{i'}) - \mathbf{M}_{ii'}(0, a_{i'})\big) \\ &= u_i(1, \boldsymbol{x}_{-i}) - u_i(0, \boldsymbol{x}_{-i}).\end{aligned}$$

□

Lemma 5.4. *Let μ be an (ϵ, δ)-constant-expectation CE of $\mathcal{G}'$. For any player i,*

$$\left(\frac{1}{2} - \alpha\right) \sum_{a_i\in\mathrm{supp}(p_i^\mu)} p_i^\mu(a_i) \max\left\{(1 - 2a_i)\mathbb{E}_\mu[d_i(\boldsymbol{a}_{-i}) \mid a_i], 0\right\} \le \delta.$$

If both actions have positive marginal probability, then

$$\min\{p_i^\mu(0), p_i^\mu(1)\} \left|\frac{1}{2}\mathbb{E}_\mu[d_i(\boldsymbol{a}_{-i}) \mid a_i = 1] - \alpha\mathbb{E}_\mu[d_i(\boldsymbol{a}_{-i}) \mid a_i = 0]\right| \le \epsilon.$$

Proof. Since $1/2 - \alpha > 0$, we consider the deviation map that switches supported recommendation 0 precisely when $\mathbb{E}_\mu[d_i(\boldsymbol{a}_{-i}) \mid a_i = 0] > 0$, switches supported recommendation 1 precisely when $\mathbb{E}_\mu[d_i(\boldsymbol{a}_{-i}) \mid a_i = 1] < 0$, and otherwise obeys. Its expected gain is the sum of the positive gains from the two recommendations, so δ-CE obedience gives

$$\left(\frac{1}{2} - \alpha\right) \sum_{a_i\in\mathrm{supp}(p_i^\mu)} p_i^\mu(a_i) \max\left\{(1 - 2a_i)\mathbb{E}_\mu[d_i(\boldsymbol{a}_{-i}) \mid a_i], 0\right\} \le \delta.$$

Suppose now that $p_i^\mu(0), p_i^\mu(1) > 0$. By Definition 3.1, there exists v_i such that

$$\begin{aligned}\epsilon &\ge p_i^\mu(0)\,|\mathbb{E}_\mu[u_i'(\boldsymbol{a}) \mid a_i = 0] - v_i| + p_i^\mu(1)\,|\mathbb{E}_\mu[u_i'(\boldsymbol{a}) \mid a_i = 1] - v_i| \\ &\ge \min\{p_i^\mu(0), p_i^\mu(1)\}\,|\mathbb{E}_\mu[u_i'(\boldsymbol{a}) \mid a_i = 1] - \mathbb{E}_\mu[u_i'(\boldsymbol{a}) \mid a_i = 0]| \\ &= \min\{p_i^\mu(0), p_i^\mu(1)\} \left|\frac{1}{2}\mathbb{E}_\mu[d_i(\boldsymbol{a}_{-i}) \mid a_i = 1] - \alpha\mathbb{E}_\mu[d_i(\boldsymbol{a}_{-i}) \mid a_i = 0]\right|,\end{aligned}$$

which proves the claim. □

This lemma, together with simple calculations, establishes the claimed equilibrium collapse, as we formalize below.

Lemma 5.5. *If μ is an (ϵ, δ)-constant-expectation correlated equilibrium of $\mathcal{G}'$ and $\boldsymbol{x}$ is its marginal strategy profile, then $\boldsymbol{x}$ is an η-Nash equilibrium of $\mathcal{G}$, where $\eta = \epsilon/\alpha + \delta/(\alpha(1 - 2\alpha))$.*

Proof. By Lemma 5.3, we have that for any player i,

$$u_i(1, \boldsymbol{x}_{-i}) - u_i(0, \boldsymbol{x}_{-i}) = \mathbb{E}_\mu[d_i(\boldsymbol{a}_{-i})].$$

If $p_i^\mu(0) = 1$, obedience gives

$$\max_{a_i'\in\{0,1\}} u_i(a_i', \boldsymbol{x}_{-i}) - u_i(\boldsymbol{x}) = \max\{u_i(1, \boldsymbol{x}_{-i}) - u_i(0, \boldsymbol{x}_{-i}), 0\} \le \frac{\delta}{1/2 - \alpha}.$$

Similarly, if $p_i^\mu(1) = 1$,

$$\max_{a_i'\in\{0,1\}} u_i(a_i', \boldsymbol{x}_{-i}) - u_i(\boldsymbol{x}) \quad = \max\{u_i(0, \boldsymbol{x}_{-i}) - u_i(1, \boldsymbol{x}_{-i}), 0\} \le \frac{\delta}{1/2 - \alpha}.$$

Since $2\alpha < 1$, these bounds are at most $\epsilon/\alpha + \delta/(\alpha(1-2\alpha))$. If both probabilities are positive,

$$u_i(1, \boldsymbol{x}_{-i}) - u_i(0, \boldsymbol{x}_{-i}) = p_i^\mu(0)\mathbb{E}_\mu[d_i(\boldsymbol{a}_{-i}) \mid a_i = 0] + p_i^\mu(1)\mathbb{E}_\mu[d_i(\boldsymbol{a}_{-i}) \mid a_i = 1].$$

If $u_i(1, \boldsymbol{x}_{-i}) \geq u_i(0, \boldsymbol{x}_{-i})$, then

$$\begin{aligned}
\max_{a_i' \in \{0,1\}} u_i(a_i', \boldsymbol{x}_{-i}) - u_i(\boldsymbol{x}) &= p_i^\mu(0)\big(u_i(1, \boldsymbol{x}_{-i}) - u_i(0, \boldsymbol{x}_{-i})\big) \\
&= 2p_i^\mu(0)p_i^\mu(1)\left(\frac{1}{2}\mathbb{E}_\mu[d_i(\boldsymbol{a}_{-i}) \mid a_i = 1] - \alpha\mathbb{E}_\mu[d_i(\boldsymbol{a}_{-i}) \mid a_i = 0]\right) \\
&\quad + p_i^\mu(0)\big(p_i^\mu(0) + 2\alpha p_i^\mu(1)\big)\mathbb{E}_\mu[d_i(\boldsymbol{a}_{-i}) \mid a_i = 0] \\
&\leq 2\epsilon + \frac{\delta}{1/2 - \alpha},
\end{aligned}$$

where we used the fact that $p_i^\mu(0)p_i^\mu(1) \leq \min\{p_i^\mu(0), p_i^\mu(1)\}$, $p_i^\mu(0) + 2\alpha p_i^\mu(1) \leq 1$, and we applied both bounds from Lemma 5.4. Similarly, if $u_i(1, \boldsymbol{x}_{-i}) \leq u_i(0, \boldsymbol{x}_{-i})$,

$$\begin{aligned}
\max_{a_i' \in \{0,1\}} u_i(a_i', \boldsymbol{x}_{-i}) - u_i(\boldsymbol{x}) &= p_i^\mu(1)\big(u_i(0, \boldsymbol{x}_{-i}) - u_i(1, \boldsymbol{x}_{-i})\big) \\
&= \frac{p_i^\mu(0)p_i^\mu(1)}{\alpha}\left(\frac{1}{2}\mathbb{E}_\mu[d_i(\boldsymbol{a}_{-i}) \mid a_i = 1] - \alpha\mathbb{E}_\mu[d_i(\boldsymbol{a}_{-i}) \mid a_i = 0]\right) \\
&\quad - p_i^\mu(1)\left(p_i^\mu(1) + \frac{p_i^\mu(0)}{2\alpha}\right)\mathbb{E}_\mu[d_i(\boldsymbol{a}_{-i}) \mid a_i = 1] \\
&\leq \frac{\epsilon}{\alpha} + \frac{\delta}{\alpha(1-2\alpha)},
\end{aligned}$$

where we used the fact that $p_i^\mu(1) + p_i^\mu(0)/(2\alpha) \leq 1/(2\alpha)$ and Lemma 5.4. Since $\alpha < 1/2$, the second upper bound also dominates the first, proving the claim. □

We put everything together below.

Theorem 5.6 (PPAD-hardness)**.** *For every pair of constants such that $0 < \epsilon < 1/24$ and $0 \leq \delta < (1-24\epsilon)^2/96$, computing an (ϵ, δ)-constant-expectation correlated equilibrium in a normalized polymatrix game of maximum degree three with two actions per player is* PPAD*-hard.*

Proof. We consider constants $0 < \epsilon < 1/24$ and $0 \leq \delta < (1-24\epsilon)^2/96$, and set $\alpha = 1/4 + 6\epsilon$ and $\eta = \epsilon/\alpha + \delta/(\alpha(1-2\alpha))$. Since $\alpha(1-2\alpha) > 0$, we have

$$\frac{\epsilon}{\alpha} + \frac{\delta}{\alpha(1-2\alpha)} < \frac{1}{12} \quad \iff \quad \delta < (\alpha/12 - \epsilon)(1-2\alpha) = \frac{(1-24\epsilon)^2}{96}.$$

Given an instance $\mathcal{G}$ of the problem in Theorem 5.1 for this $\eta < 1/12$, construct $\mathcal{G}'$ as above. The construction is polynomial. From any (ϵ, δ)-constant-expectation correlated equilibrium of $\mathcal{G}'$, we compute its marginals $\boldsymbol{x}$. By Lemma 5.5, $\boldsymbol{x}$ is an η-Nash equilibrium of $\mathcal{G}$, and the result follows from Theorem 5.1. □

Next, we provide the proofs pertaining to the two-player reduction, which starts from a rank-2 game.

Lemma 5.7. *$\mathcal{G}'$ has payoffs in $[0, 1]$, has rank at most two, and has the same set of CEs as $\mathcal{G}$. Moreover, μ is a δ-CE of $\mathcal{G}'$ if and only if it is a $(4K\delta)$-CE of $\mathcal{G}$.*

Proof. Normalization is immediate. The rank bound follows from the fact that

$$4K(u_1'(\boldsymbol{a}) + u_2'(\boldsymbol{a})) = (f_1(a_1) + K)(g_1(a_2) + 2) + (f_2(a_1) + 2)(g_2(a_2) + K).$$

Finally, each transformed payoff is a positive multiple of the original payoff plus a term depending only on the opponent's action. The additional term cancels from every obedience inequality, so the two games have exactly the same correlated equilibria. More generally, for every deviation map $\phi_i : \mathcal{A}_i \to \mathcal{A}_i$,

$$\mathbb{E}_\mu[u_i'(\phi_i(a_i), \boldsymbol{a}_{-i}) - u_i'(\boldsymbol{a})] = \frac{1}{4K}\mathbb{E}_\mu[u_i(\phi_i(a_i), \boldsymbol{a}_{-i}) - u_i(\boldsymbol{a})],$$

because the opponent-dependent offsets cancel. Thus, μ is a δ-CE of $\mathcal{G}'$ if and only if it is a $(4K\delta)$-CE of $\mathcal{G}$. □

Lemma 5.8. *If μ is an η-CE of $\mathcal{G}$ with marginal strategy profile $\boldsymbol{x}$, then*

$$\sum_{i=1}^{2}\left(\max_{a_i'\in\mathcal{A}_i} u_i(a_i',\boldsymbol{x}_{-i}) - u_i(\boldsymbol{x})\right) \le 2\eta + \sum_{\ell=1}^{2}\mathrm{Cov}_\mu(f_\ell(a_1), g_\ell(a_2)).$$

Proof. For each player i, we have $\mathbb{E}_\mu[u_i(\boldsymbol{a})] \ge \max_{a_i'\in\mathcal{A}_i} u_i(a_i',\boldsymbol{x}_{-i}) - \eta$ by applying η-CE obedience to constant deviation maps. Subtracting $u_i(\boldsymbol{x})$, summing over both players, and using (11) gives the upper bound

$$\begin{aligned}\sum_{i=1}^{2}\left(\max_{a_i'} u_i(a_i',\boldsymbol{x}_{-i}) - u_i(\boldsymbol{x})\right) &\le 2\eta + \sum_{i=1}^{2}\left(\mathbb{E}_\mu[u_i(\boldsymbol{a})] - u_i(\boldsymbol{x})\right)\\ &= 2\eta + \sum_{\ell=1}^{2}\mathrm{Cov}_\mu(f_\ell(a_1), g_\ell(a_2)).\end{aligned}$$

□

In light of the previous lemma, it suffices to bound the covariance terms. This is done in the following lemma, crucially relying on the constant expectation property.

Lemma 5.9. *If μ is an (ϵ,δ)-constant-expectation correlated equilibrium of $\mathcal{G}'$, then, for $\ell\in\{1,2\}$,*

$$|\mathrm{Cov}_\mu(f_\ell(a_1), g_\ell(a_2))| \le 8\epsilon + \frac{6}{K}.$$

Proof. We use the notation $u_i'(a_i) := \mathbb{E}_\mu[u_i'(\boldsymbol{a}) \mid a_i]$ for supported actions, and let $\bar{u}_i := \mathbb{E}_\mu[u_i'(\boldsymbol{a})]$. Since $\bar{u}_i = \sum_{a_i\in\mathrm{supp}(p_i^\mu)} p_i^\mu(a_i)u_i'(a_i)$, the averaged condition in Definition 3.1 gives

$$\sum_{a_i\in\mathrm{supp}(p_i^\mu)} p_i^\mu(a_i)|u_i'(a_i) - \bar{u}_i| \le \sum_{a_i\in\mathrm{supp}(p_i^\mu)} p_i^\mu(a_i)|u_i'(a_i) - v_i| + |v_i - \bar{u}_i| \le 2\epsilon. \tag{13}$$

Taking conditional and unconditional expectations in (12), and using $|u_1 + 2g_2| \le 3$, gives

$$|\mathbb{E}_\mu[g_1(a_2) \mid a_1] - \mathbb{E}_\mu[g_1(a_2)]| \le 4|u_1'(a_1) - \bar{u}_1| + \frac{6}{K}.$$

Averaging over a_1 and applying (13), we obtain

$$\sum_{a_1\in\mathrm{supp}(p_1^\mu)} p_1^\mu(a_1)\,|\mathbb{E}_\mu[g_1(a_2) \mid a_1] - \mathbb{E}_\mu[g_1(a_2)]| \le 8\epsilon + \frac{6}{K}.$$

Consequently, since $|f_1| \le 1$,

$$\begin{aligned}|\,\mathrm{Cov}_\mu(f_1(a_1), g_1(a_2))| &= |\mathbb{E}_\mu\left[f_1(a_1)\left(\mathbb{E}_\mu[g_1(a_2) \mid a_1] - \mathbb{E}_\mu[g_1(a_2)]\right)\right]|\\ &\le 8\epsilon + \frac{6}{K}.\end{aligned}$$

Applying the same argument to Player 2, using $|u_2 + 2f_1| \le 3$, bounds the expected absolute deviation of $\mathbb{E}_\mu[f_2(a_1) \mid a_2]$ from its unconditional expectation by $8\epsilon + 6/K$. Multiplying by $g_2(a_2)$ and using $|g_2| \le 1$ proves the second covariance bound. □

Theorem 5.10. *Computing an (ϵ,δ)-constant-expectation correlated equilibrium is* PPAD*-hard in bimatrix games when ϵ and δ are exponentially small.*

Proof. Given a game $\mathcal{G}$ as above, we choose $\eta = 2^{-p(L)} \in (0,1]$, where L is its bit complexity and p is a polynomial, sufficiently small that any rational η-Nash equilibrium can be converted in polynomial time into an exact Nash equilibrium; for example, this can be done through an LP (Etessami & Yannakakis, 2010). We set the parameters as $K := 48/\eta$, $\epsilon := \eta/32$, $\delta := \eta^2/1536 = \eta/(32K)$, and construct $\mathcal{G}'$ using (12). The construction is polynomial in L.

Let μ be any (ϵ, δ)-constant-expectation correlated equilibrium of $\mathcal{G}'$, and let $\boldsymbol{x}$ be its marginal strategy profile. By Lemma 5.7, μ is a $(4K\delta)$-CE of $\mathcal{G}$. Therefore, Lemmas 5.8 and 5.9 give

$$\sum_{i=1}^{2} \left(\max_{a_i'} u_i(a_i', \boldsymbol{x}_{-i}) - u_i(\boldsymbol{x}) \right) \leq 16\epsilon + 8K\delta + \frac{12}{K} = \eta,$$

where the equality follows by our parameter choice. This proves that $\boldsymbol{x}$ is an η-Nash equilibrium, from which we can recover an exact Nash equilibrium of $\mathcal{G}$. □

C HARDNESS FOR INDEPENDENT STATE RECOMMENDATIONS

In this section, we show that, in a sense, constant-expectation CEs are *necessary* to compute NFCEs in Markov games. Thus, applying our hardness results (Theorems 5.6 and 5.10), it will follow that computing an ϵ-NFCE with independent state recommendations is PPAD-hard for games with a constant number of players when ϵ is exponentially small, and for $n+1$ players when $\epsilon \leq 1/(3072n)$.

Theorem 5.11. *Let $\mathcal{G}$ be an n-player normal-form game with utilities bounded in $[0, 1]$, and $0 \leq \epsilon \leq 1/(16n)$. There is a Markov game $\mathcal{M}$ with $n+1$ players, horizon $H = 2$, $S = n+1$ states, and stage rewards bounded in $[0, 1/2]$, such that every ϵ-NFCE with independent state recommendations yields a $(16n\epsilon, 16\epsilon)$-constant-expectation CE of $\mathcal{G}$.*

Proof. Player $n+1$ will be called the *tester*. The Markov game is constructed as follows. At the first stage, a player $i \in [n]$ is selected uniformly at random. That player plays a game of matching pennies against the tester: if the pennies match, then i gets 2; else, the tester gets 2. If the player played tails, the game ends. Otherwise, at stage 2, the n players (excluding the tester) play $\mathcal{G}$ and gain the associated rewards. There are n states at stage 1 and one state at stage 2, for a total of $n+1$ states. All unspecified rewards are zero, and inactive players have a single dummy action.

Consider any ϵ-NFCE of $\mathcal{M}$ with independent state recommendations. (Throughout this proof, we use the ex-ante deviation inequalities of an ϵ-NFCE.) Let μ be the correlated profile recommended at $\mathcal{G}$. We claim that μ is a $(4n\epsilon, 4\epsilon)$-constant-expectation CE. Let $v_i := \mathbb{E}_{\boldsymbol{a}\sim\mu}[u_i(\boldsymbol{a})]$. Independence across states implies that deviations affecting only the first-stage action, and ignoring the second-stage recommendation, have continuation value v_i if $\mathcal{G}$ is reached. If Player i is selected, then Player i and the tester play a game strategically equivalent to the zero-sum game with row utility matrix

	H	T
H	$2 + v_i$	v_i
T	0	2

where the player is the row player and the tester is the column player. The unique Nash equilibrium has the player playing heads with probability $1/2$, the tester playing heads with probability $(2 - v_i)/4 \in [1/4, 1/2]$, and value $V_i := 1 + v_i/2$ for i. Let $\tilde{V}_i$ be the row player's expected payoff conditional on reaching this state. The state is reached with probability $1/n$, so a deviation confined to this state has conditional gain at most $n\epsilon$. Therefore, we have

$$|V_i - \tilde{V}_i| \leq n\epsilon.$$

The tester's deviation to H gives

$$\tilde{V}_i \leq (2 + v_i)q_i + n\epsilon,$$

where q_i is the probability that Player i plays heads. Therefore,

$$q_i \geq \frac{\tilde{V}_i - n\epsilon}{2 + v_i} \geq \frac{V_i - 2n\epsilon}{2V_i} \geq \frac{1}{2} - n\epsilon \geq \frac{1}{4},$$

where we use the inequalities $\tilde{V}_i \geq V_i - n\epsilon$, $V_i \geq 1$, and $n\epsilon \leq 1/4$. Therefore, $\mathcal{G}$ is reached with probability at least $1/4$, and so μ is a 4ϵ-CE.

It remains to show the constant-expectation property. Let p_i be the probability with which the tester is recommended to play heads against i. For every recommendation a_i in the support of p_i^μ, let

$$u_i(a_i) := \mathbb{E}_{\boldsymbol{a}\sim\mu}[u_i(\boldsymbol{a}) \mid a_i].$$

Consider a deviation that, only when i is selected, ignores its first-stage recommendation and chooses heads or tails using only a_i, while obeying at $\mathcal{G}$ if it is reached. By independence across states, heads has expected value $2p_i + u_i(a_i)$ and tails has expected value $2(1 - p_i)$. Therefore, choosing the most profitable action and averaging over a_i yields conditional utility

$$V_i^* = \sum_{a_i} p_i^\mu(a_i) \max\{2p_i + u_i(a_i),\ 2(1 - p_i)\} = V_i + \frac{1}{2} \sum_{a_i} p_i^\mu(a_i) |(2 - 4p_i) - u_i(a_i)|.$$

This deviation cannot increase utility conditional on the selected state by more than $n\epsilon$. Consequently, $V_i^* \le \tilde{V}_i + n\epsilon \le V_i + 2n\epsilon$, and

$$\sum_{a_i \in \mathrm{supp}(p_i^\mu)} p_i^\mu(a_i) |(2 - 4p_i) - u_i(a_i)| \le 4n\epsilon.$$

and therefore choosing $v_i' := 2 - 4p_i$, and dividing all rewards by 4 to satisfy the normalization assumption, completes the proof. We finally note that the construction preserves succinct payoff representations with polynomial overhead. □