On Tractable Φ-Equilibria in Non-Concave Games

Yang Cai
Yale University
yang.cai@yale.edu

Constantinos Daskalakis
MIT CSAIL
costis@csail.mit.edu

Haipeng Luo
University of Southern California
haipengl@usc.edu

Chen-Yu Wei
University of Virginia
chenyu.wei@virginia.edu

Weiqiang Zheng
Yale University
weiqiang.zheng@yale.edu

Abstract

While Online Gradient Descent and other no-regret learning procedures are known to efficiently converge to a coarse correlated equilibrium in games where each agent's utility is concave in their own strategy, this is not the case when utilities are non-concave – a common scenario in machine learning applications involving strategies parameterized by deep neural networks, or when agents' utilities are computed by neural networks, or both. Non-concave games introduce significant game-theoretic and optimization challenges: (i) Nash equilibria may not exist; (ii) local Nash equilibria, though they exist, are intractable; and (iii) mixed Nash, correlated, and coarse correlated equilibria generally have infinite support and are intractable. To sidestep these challenges, we revisit the classical solution concept of Φ-equilibria introduced by Greenwald and Jafari [GJ03], which is guaranteed to exist for an arbitrary set of strategy modifications Φ even in non-concave games [SL07]. However, the tractability of Φ-equilibria in such games remains elusive. In this paper, we initiate the study of tractable Φ-equilibria in non-concave games and examine several natural families of strategy modifications. We show that when Φ is finite, there exists an efficient uncoupled learning algorithm that approximates the corresponding Φ-equilibria. Additionally, we explore cases where Φ is infinite but consists of local modifications, showing that Online Gradient Descent can efficiently approximate Φ-equilibria in non-trivial regimes.

1 Introduction

Von Neumann's celebrated minimax theorem establishes the existence of Nash equilibrium in all two-player zero-sum games where the players' utilities are continuous as well as *concave* in their own strategy [Neu28].[1] This assumption that players' utilities are concave, or quasi-concave, in their own strategies has been a cornerstone for the development of equilibrium theory in Economics, Game Theory, and a host of other theoretical and applied fields that make use of equilibrium concepts. In particular, (quasi-)concavity is key for showing the existence of many types of equilibrium, from generalizations of min-max equilibrium [Fan53; Sio58] to competitive equilibrium in exchange economies [AD54; McK54], mixed Nash equilibrium in finite normal-form games [Nas50], and, more generally, Nash equilibrium in (quasi-)concave games [Deb52; Ros65].

---

[1]Throughout this paper, we model games using the standard convention in Game Theory that each player has a utility function that they want to maximize. This is, of course, equivalent to modeling the players as loss minimizers, a convention more common in learning. When we say that a player's utility is concave (respectively non-concave) in their strategy, this is the same as saying that the player's loss is convex (respectively non-convex) in their strategy.

Table 1: A comparison between different solution concepts in multi-player non-concave games. We include definitions of Nash equilibrium, mixed Nash equilibrium, (coarse) correlated equilibrium, strict local Nash equilibrium, and second-order local Nash equilibrium in Appendix B. We also give a detailed discussion on the existence and complexity of these solution concepts in Appendix B.

Solution Concept	Incentive Guarantee	Existence	Complexity
Nash equilibrium		✗	
Mixed Nash equilibrium	Global stability	✓	NP-hard
(Coarse) Correlated equilibrium		✓	[MK87; AZ22]
Strict local Nash equilibrium	Local stability	✗	
Second-order local Nash equilibrium	Second-order stability	✗	
Local Nash equilibrium	First-order stability	✓	PPAD-hard [DSZ21]
Φ-equilibrium (finite $|\Phi|$)	Stability against finite deviations	✓	Efficient ε-approximation for any $\varepsilon > 0$ (Theorem 2)
$\mathrm{Conv}(\Phi(\delta))$-equilibrium (finite $|\Phi(\delta)|$)		✓	Efficient ε-approximation for $\varepsilon = \Omega(\delta^2)$ (Theorem 4)
$\Phi_{\mathrm{proj}}(\delta)$-equilibrium	First-order stability when $\varepsilon = \Omega(\delta^2)$	✓	Efficient ε-approximation via GD/OG for $\varepsilon = \Omega(\delta^2)$ (Theorem 3,9)
$\Phi_{\mathrm{Int}}(\delta)$-equilibrium		✓	Efficient ε-approximation via no-regret learning for $\varepsilon = \Omega(\delta^2)$ (Theorem 5)

Not only are equilibria guaranteed to exist in concave games, but it is also well-established—thanks to a long line of work at the interface of game theory, learning and optimization whose origins can be traced to Dantzig's work on linear programming [Geo63], Brown and Robinson's work on fictitious play [Bro51; Rob51], Blackwell's approachability theorem [Bla56] and Hannan's consistency theory [Han57]—that several solution concepts are efficiently computable both centrally and via decentralized learning dynamics. For instance, it is well-known that the learning dynamics produced when the players of a game iteratively update their strategies using no-regret learning algorithms, such as online gradient descent, is guaranteed to converge to Nash equilibrium in two-player zero-sum concave games, and to coarse correlated equilibrium in multi-player general-sum concave games [CL06]. The existence of such simple decentralized dynamics further justifies using these solution concepts to predict the outcome of real-life multi-agent interactions where agents deploy strategies, obtain feedback, and use that feedback to update their strategies.

While (quasi-)concave utilities have been instrumental in the development of equilibrium theory, as described above, they are also too restrictive an assumption. Several modern applications and outstanding challenges in Machine Learning, from training Generative Adversarial Networks (GANs) to Multi-Agent Reinforcement Learning (MARL) as well as generic multi-agent Deep Learning settings where the agents' strategies are parameterized by deep neural networks or their utilities are computed by deep neural networks, or both, give rise to games where the agents' utilities are *non-concave* in their own strategies. We call these games *non-concave*, following [Das22].

Unfortunately, classical equilibrium theory quickly hits a wall in non-concave games. First, Nash equilibria are no longer guaranteed to exist. Second, while mixed Nash, correlated and coarse correlated equilibria do exist—under convexity and compactness of the strategy sets [Gli52], which we have been assuming all along in our discussion so far, they have infinite support, in general [Kar14]. Finally, they are computationally intractable; so, a fortiori, they are also intractable to attain via decentralized learning dynamics.

In view of the importance of non-concave games in emerging ML applications and the afore-described state-of-affairs, our investigation is motivated by the following broad and largely open question:

Question from [Das22]: *Is there a theory of non-concave games? What solution concepts are meaningful, universal, and tractable?*

1.1 Contributions

We study Daskalakis' question through the lens of the classical solution concept of Φ-equilibria introduced by Greenwald and Jafari [GJ03]. This concept is guaranteed to exist for virtually any set of strategy modifications Φ, even in non-concave games, as demonstrated by Stoltz and Lugosi [SL07].[2] However, the tractability of Φ-equilibria in such games remains elusive. In this paper, we initiate the study of tractable Φ-equilibria in non-concave games and examine several natural families of strategy modifications.

---

[2] Stoltz and Lugosi [SL07] only require the elements of Φ to be measurable functions.

Figure 1: The relationship between different solution concepts in non-concave games. An arrow from one solution concept to another means the former is contained in the latter. The dashed arrow from $\mathrm{Conv}(\Phi(\delta))$-equilibria to Φ_{Finite}-equilibria means the former is contained in the latter when $\Phi(\delta) = \Phi_{\mathrm{Finite}}$.
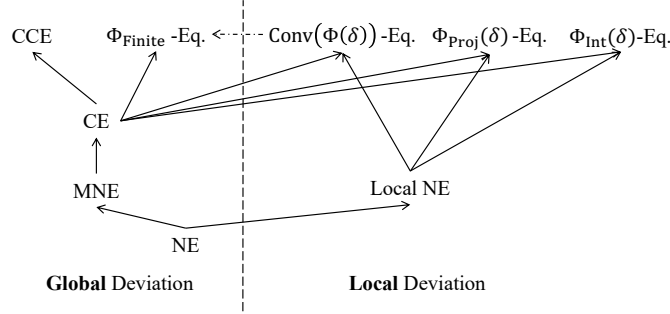
Φ-Equilibrium. The concept of Φ-equilibrium generalizes (coarse) correlated equilibrium. A Φ-equilibrium is a joint distribution over $\Pi_{i=1}^{n}\mathcal{X}_i$, the Cartesian product of all players' strategy sets, and is defined in terms of a set, $\Phi^{\mathcal{X}_i}$, of *strategy modifications*, for each player i. The set $\Phi^{\mathcal{X}_i}$ contains functions mapping $\mathcal{X}_i$ to itself. A joint distribution over strategy profiles qualifies as a $\Phi = \Pi_{i=1}^{n}\Phi^{\mathcal{X}_i}$-equilibrium if no player i can increase their expected utility by using any strategy modification function, $\phi_i \in \Phi^{\mathcal{X}_i}$, on the strategy sampled from the joint distribution. The larger the set Φ, the stronger the incentive guarantee offered by the Φ-equilibrium. For example, if $\Phi^{\mathcal{X}_i}$ contains all constant functions, the corresponding Φ-equilibrium coincides with the notion of coarse correlated equilibrium. Throughout the paper, we also consider ε-approximate Φ-equilibria, where no player can gain more than ε by deviating using any function from $\Phi^{\mathcal{X}_i}$. We study several families of Φ and illustrate their relationships in Figure 1.

Finite Set of Global Deviations. The first case we consider is when each player i's set of strategy modifications, $\Phi^{\mathcal{X}_i}$, contains a finite number of arbitrary functions mapping $\mathcal{X}_i$ to itself. As shown in [GJ03], if there exists an online learning algorithm where each player i is guaranteed to have sublinear $\Phi^{\mathcal{X}_i}$-regret, the empirical distribution of joint strategies played converges to a $\Phi = \Pi_{i=1}^{n}\Phi^{\mathcal{X}_i}$-equilibrium. Gordon, Greenwald, and Marks [GGM08] consider Φ-regret minimization but for concave reward functions, and their results, therefore, do not apply to non-concave games. Stoltz and Lugosi [SL07] provide an algorithm that achieves no $\Phi^{\mathcal{X}_i}$-regret in non-concave games; however, their algorithm requires a fixed-point computation per step, making it computationally inefficient.[3] Our first contribution is to provide an efficient randomized algorithm that achieves no $\Phi^{\mathcal{X}_i}$-regret for each player i with high probability.

> **Contribution 1:** Let $\mathcal{X}$ be a strategy set (not necessarily compact or convex), and Ψ an arbitrary finite set of strategy modification functions for $\mathcal{X}$. We design a randomized online learning algorithm that achieves $O\left(\sqrt{T \log|\Psi|}\right)$ Ψ-regret, with high probability, for *arbitrary* bounded reward functions on $\mathcal{X}$ (Theorem 2). The algorithm operates in time $\sqrt{T}|\Psi|$ per iteration. If every player in a *non-concave* game adopts this algorithm, the empirical distribution of strategy profiles played forms an ε-approximate $\Phi = \Pi_{i=1}^{n}\Phi^{\mathcal{X}_i}$-equilibrium, with high probability, for any $\varepsilon > 0$, after $\mathrm{poly}\left(\frac{1}{\varepsilon}, \log\left(\max_i |\Phi^{\mathcal{X}_i}|\right), \log n\right)$ iterations.

If players have infinitely many global strategy modifications, we can extend Algorithm 1 by discretizing the set of strategy modifications under mild assumptions, such as the modifications being Lipschitz (Corollary 1). The empirical distribution of the strategy profiles still converges to the corresponding Φ-equilibrium, but at a much slower rate of $O(T^{-\frac{1}{d+2}})$, where d is the dimension of the set of strategies. Additionally, the algorithm requires exponential time in the dimension per iteration, making it inefficient. This inefficiency is unavoidable, as the problem remains intractable even when Φ contains only constant functions.

---

[3]The existence of the fixed point is guaranteed by the Schauder-Cauty fixed-point theorem [Cau01], a generalization of the Brouwer fixed-point theorem. Hence, it's unlikely such fixed points are tractable.

To address the limitations associated with infinitely large global strategy modifications, a natural approach is to focus on local deviations instead. The corresponding Φ-equilibrium will guarantee local stability. The study of local equilibrium concepts in non-concave games has received significant attention in recent years—see e.g., [RBS16; HSZ17; DP18; JNJ20; DSZ21]. However, these solution concepts either are not guaranteed to exist, are restricted to sequential two-player zero-sum games [MV21], only establish local convergence guarantees for learning dynamics—see e.g., [DP18; WZB20; FCR20], only establish asymptotic convergence guarantees—see e.g., [Das+23], or involve non-standard solution concepts where local stability is not with respect to a distribution over strategy profiles [HSZ17].

We study the tractability of Φ-equilibrium with infinitely large Φ sets that consist solely of local strategy modifications. These local solution concepts are guaranteed to exist in general multi-player non-concave games. Specifically, we focus on the following three families of natural deviations.

- **Projection based Local Deviations:** Each player i's set of strategy modifications, denoted by $\Phi_{\text{Proj}}^{\mathcal{X}_i}(\delta)$, contains all deviations that attempt a small step from their input in a fixed direction and project if necessary, namely are of the form $\phi_v(x) = \Pi_{\mathcal{X}_i}[x - v]$, where $\|v\| \leq \delta$ and $\Pi_{\mathcal{X}_i}$ stands for the ℓ_2-projection onto $\mathcal{X}_i$.

- **Convex Combination of Finitely Many Local Deviations:** Each player i's set of strategy modifications, denoted by $\text{Conv}(\Phi^{\mathcal{X}_i}(\delta))$, contains all deviations that can be represented as a convex combination of a finite set of δ-local strategy modifications, i.e., $\|\phi(x) - x\| \leq \delta$ for all $\phi \in \Phi^{\mathcal{X}_i}(\delta)$.

- **Interpolation based Local Deviations:** each player i's set of local strategy modifications, denoted by $\Phi_{\text{Int}}^{\mathcal{X}_i}(\delta)$, that contains all deviations that *interpolate* between the input strategy and another strategy in $\mathcal{X}_i$. Formally, each element $\phi_{\lambda,x'}(x)$ of $\Phi_{\text{Int}}^{\mathcal{X}_i}(\delta)$ can be represented as $(1 - \lambda)x + \lambda x'$ for some $x' \in \mathcal{X}_i$ and $\lambda \leq \delta/D_{\mathcal{X}_i}$ ($D_{\mathcal{X}_i}$ is the diameter of $\mathcal{X}_i$).

For our three families of local strategy modifications, we explore the tractability of Φ-equilibrium within a regime we term the *first-order stationary regime*, where $\varepsilon = \Omega(\delta^2)$[4], with δ representing the maximum deviation allowed for a player. An ε-approximate Φ-equilibrium in this regime ensures first-order stability. This regime is particularly interesting for two reasons: (i) Daskalakis, Skoulakis, and Zampetakis [DSZ21] have demonstrated that computing an ε-approximate δ-local Nash equilibrium in this regime is intractable.[5] This poses an intriguing question: can correlating the players' strategies, as in a Φ-equilibrium, potentially make the problem tractable? (ii) Extending our algorithm, initially designed for finite sets of strategy modifications, to these three sets of local deviations results in inefficiency; specifically, the running time becomes exponential in one of the problem's natural parameters. Designing efficient algorithms for this regime thus presents challenges. Despite these, we show the following:

> **Contribution 2:** For any $\delta > 0$, for each of the three families of infinite δ-local strategy modifications mentioned above, there exists an efficient uncoupled learning algorithm that converges to an ε-approximate Φ-equilibrium of the non-concave game in the first-order stationary regime, i.e., $\varepsilon = \Omega(\delta^2)$.

We present our results for the projection-based local deviation in Theorem 3 and Theorem 9. Our result for the convex combination of local deviations can be found in Theorem 4. Theorem 5 contains our result for the interpolation-based local deviations. Similar to the finite case, our algorithms build on the connection between Φ-regret minimization and Φ-equilibrium. Given that our strategy modifications are non-standard, it is a priori unclear how to minimize the corresponding Φ-regret. For instance, to our knowledge, no algorithm is known to minimize $\Phi_{\text{Proj}}^{\mathcal{X}}(\delta)$-regret even when the reward functions are concave, and provably $\Phi_{\text{Proj}}^{\mathcal{X}}(\delta)$-regret is incomparable to external regret (Examples 3 and 4). However, via a novel analysis, we show that Online Gradient Descent (GD) and Optimistic Gradient (OG) achieve a near-optimal $\Phi_{\text{Proj}}^{\mathcal{X}}(\delta)$-regret guarantee (Theorem 3 and Theorem 8). Our results provide efficient uncoupled algorithms to compute ε-approximate $\Phi(\delta)$-equilibria in the first-order stationary regime $\varepsilon = \Omega(\delta^2)$.

Further related work is discussed in Appendix A.

---

[4]The regime $\varepsilon = \Omega(\delta)$ is trivial when the utility is Lipschitz.

[5]A strategy profile is considered an ε-approximate δ-local Nash equilibrium if no player can gain more than ε by deviating within a δ distance.

2 Preliminaries

A ball of radius $r > 0$ centered at $x \in \mathbb{R}^d$ is denoted by $B_d(x, r) := \{x' \in \mathbb{R}^d : \|x - x'\| \leq r\}$. We use $\| \cdot \|$ for ℓ_2 norm throughout. We also write $B_d(\delta)$ for a ball centered at the origin with radius δ. For $a \in \mathbb{R}$, we use $[a]^+$ to denote $\max\{0, a\}$. We denote $D_{\mathcal{X}}$ as the diameter of a set $\mathcal{X}$.

Continuous / Smooth Games. An n-player *continuous game* has a set of n players $[n] := \{1, 2, \ldots, n\}$. Each player $i \in [n]$ has a nonempty convex and compact strategy set $\mathcal{X}_i \subseteq \mathbb{R}^{d_i}$. For a joint strategy profile $x = (x_i, x_{-i}) \in \prod_{j=1}^{n} \mathcal{X}_j$, the reward of player i is determined by a utility function $u_i : \prod_{j=1}^{n} \mathcal{X}_j \to [0, 1]$. We denote by $d = \sum_{i=1}^{n} d_i$ the dimensionality of the game and assume $\max_{i \in [n]} \{D_{\mathcal{X}_i}\} \leq D$. A *smooth game* is a continuous game whose utility functions further satisfy the following assumption.

Assumption 1 (Smooth Games). *The utility function $u_i(x_i, x_{-i})$ for any player $i \in [n]$ is differentiable and satisfies*

1. *(G-Lipschitzness) $\|\nabla_{x_i} u_i(x)\| \leq G$ for all i and $x \in \prod_{j=1}^{n} \mathcal{X}_j$;*

2. *(L-smoothness) there exists $L_i > 0$ such that $\|\nabla_{x_i} u_i(x_i, x_{-i}) - \nabla_{x_i} u_i(x_i', x_{-i})\| \leq L_i \|x_i - x_i'\|$ for all $x_i, x_i' \in \mathcal{X}_i$ and $x_{-i} \in \Pi_{j \neq i} \mathcal{X}_j$. We denote $L = \max_i L_i$ as the smoothness of the game.*

Crucially, we make no assumption on the concavity of $u_i(x_i, x_{-i})$.

Φ-equilibrium and Φ-regret. Below we formally introduce the concept of Φ-equilibrium and its relationship with online learning and Φ-regret minimization.

Definition 1 (Φ-equilibrium [GJ03; SL07]). *In a continuous game, a distribution σ over joint strategy profiles $\Pi_{i=1}^{n} \mathcal{X}_i$ is an ε-approximate Φ-equilibrium for some $\varepsilon \geq 0$ and a profile of strategy modification sets $\Phi = \Pi_{i=1}^{n} \Phi_i$ if and only if for all player $i \in [n]$, $\max_{\phi \in \Phi_i} \mathbb{E}_{x \sim \sigma} [u_i(\phi(x_i), x_{-i})] \leq \mathbb{E}_{x \sim \sigma} [u_i(x)] + \varepsilon$. When $\varepsilon = 0$, we call σ a Φ-equilibrium.*

We consider the standard online learning setting: at each day $t \in [T]$, the learner chooses an action x^t from a nonempty convex compact set $\mathcal{X} \subseteq \mathbb{R}^m$ and the adversary chooses a possibly non-convex loss function $f^t : \mathcal{X} \to \mathbb{R}$, then the learner suffers a loss $f^t(x^t)$ and receives feedback. In this paper, we focus on two feedback models: (1) the player receives an oracle for $f^t(\cdot)$; (2) the player receives only the gradient $\nabla f^t(x^t)$. The classic goal of an online learning algorithm is to minimize the *external regret* defined as $\mathrm{Reg}^T := \max_{x \in \mathcal{X}} \sum_{t=1}^{T} (f^t(x^t) - f^t(x))$. An algorithm is called *no-regret* if its external regret is sublinear in T. The notion of Φ-regret generalizes external regret by allowing more general strategy modifications.

Definition 2 (Φ-regret). *Let Φ be a set of strategy modification functions $\{\phi : \mathcal{X} \to \mathcal{X}\}$. For $T \geq 1$, the Φ-regret of an online learning algorithm is $\mathrm{Reg}_{\Phi}^T := \max_{\phi \in \Phi} \sum_{t=1}^{T} (f^t(x^t) - f^t(\phi(x^t)))$. An algorithm is called* no Φ-regret *if its Φ-regret is sublinear in T.*

Many classic notions of regret can be interpreted as Φ-regret. For example, the external regret is Φ_{ext}-regret where Φ_{ext} contains all constant strategy modifications $\phi_{x^*}(x) = x^*$ for all $x^* \in \mathcal{X}$. The *swap regret* on simplex Δ^m is Φ_{swap}-regret where Φ_{swap} contains all linear transformations $\phi : \Delta^m \to \Delta^m$. A fundamental result for learning in games is that no-Φ-regret learning dynamics in games converge to an approximate Φ-equilibrium [GJ03].

Theorem 1 ([GJ03]). *If each player i's Φ_i-regret is upper bounded by $\mathrm{Reg}_{\Phi_i}^T$, then their empirical distribution of strategy profiles played is an $(\max_{i \in [n]} \mathrm{Reg}_{\Phi_i}^T / T)$-approximate Φ-equilibrium.*

3 Tractable Φ-Equilibrium for Finite Φ via Sampling

In this section, we revisit the problem of computing and learning an Φ-equilibrium in non-concave games when each player's set of strategy modifications $\Phi^{\mathcal{X}_i}$ is finite.

The pioneering work of Stoltz and Lugosi [SL07] gives a no-Φ-regret algorithm for this case where each player chooses a distribution over strategies in each round. This result also implies convergence to Φ-equilibrium. However, the algorithm by Stoltz and Lugosi [SL07] is not computationally efficient. In each iteration, their algorithm requires computing a distribution that is stationary under a

transformation that can be represented as a mixture of the modifications in Φ. The existence of such a stationary distribution is guaranteed by the Schauder-Cauty fixed-point theorem [Cau01], but the distribution might require exponential support and be intractable to find.

Our main result in this section is an efficient Φ-regret minimization algorithm (Algorithm 1) that circumvents the step of the exact computation of a stationary distribution. Consequently, our algorithm also ensures efficient convergence to a Φ-equilibrium when adopted by all players.

Algorithm 1: Φ-regret minimization for non-concave reward via sampling

---

Input: $x_{\text{root}} \in \mathcal{X}$, $h \geq 2$, an external regret minimization algorithm $\mathfrak{R}_\Phi$ over Φ
Output: A Φ-regret minimization algorithm for $\mathcal{X}$

1 **function** NEXTSTRATEGY()
2      $p^t \leftarrow \mathfrak{R}_\Phi$.NEXTSTRATEGY(). Note that p_t is a distribution over Φ.
3   **return** $x^t \leftarrow$ SAMPLESTRATEGY(x_{root}, h, p^t).
4 **function** OBSERVEREWARD($u^t(\cdot)$)
5   Set $u^t_\Phi(\phi) = u^t(\phi(x^t))$ for all $\phi \in \Phi$.
6      $\mathfrak{R}_\Phi$.OBSERVEREWARD($u^t_\Phi(\cdot)$).

Algorithm 2: SAMPLESTRATEGY

Input: $x_{\text{root}} \in \mathcal{X}$, $h \geq 2$, $p^t \in \Delta(\Phi)$
Output: $x \in \mathcal{X}$

1 $x_1 \leftarrow x_{\text{root}}$.
2 **for** $2 \leq k \leq h$ **do**
3      $\phi \leftarrow$ sample form Φ according to p^t.
4      $x_k = \phi(x_{k-1})$.
5 **return** x from $\{x_1, \ldots, x_h\}$ uniformly at random.

Theorem 2. *Let $\mathcal{X}$ be a strategy set (not necessarily compact or convex), Φ be an arbitrary finite set of strategy modifications over $\mathcal{X}$, and $u^1(\cdot), \ldots, u^T(\cdot)$ be an arbitrary sequence of possibly non-concave reward functions from $\mathcal{X}$ to $[0, 1]$. If we instantiate Algorithm 1 with $\mathfrak{R}_\Phi$ being the Hedge algorithm over Φ and $h = \sqrt{T}$, the algorithm guarantees that, with probability at least $1 - \beta$, it produces a sequence of strategies $x^1, \ldots, x^T$ with Φ-regret at most $\max_{\phi \in \Phi} \sum_{t=1}^{T} u^t(\phi(x^t)) - \sum_{t=1}^{T} u^t(x^t) \leq 8\sqrt{T(\log|\Phi| + \log(1/\beta))}$. Moreover, the algorithm runs in time $O(\sqrt{T}|\Phi|)$ per iteration.*

If all players in a non-concave continuous game employ Algorithm 1, then with probability at least $1 - \beta$, for any $\varepsilon > 0$, the empirical distribution of strategy profiles played forms an ε-approximate $\Phi = \Pi_{i=1}^{n} \Phi^{\mathcal{X}_i}$-equilibrium, after $\text{poly}\left(\frac{1}{\varepsilon}, \log\left(\max_i |\Phi^{\mathcal{X}_i}|\right), \log\frac{n}{\beta}\right)$ iterations.

High-level ideas. We adopt the framework in [SL07]. The framework contains two steps in each iteration t: (1) the learner runs a no-external-regret algorithm over Φ which outputs $p^t \in \Delta(\Phi)$ in each iteration t; (2) the learner chooses a stationary distribution $\mu^t = \sum_{\phi \in \Phi} p^t \phi(\mu^t)$, where we slightly abuse notation to use $\phi(\mu^t)$ to denote the image measure of μ by ϕ. However, how to compute the stationary distribution μ^t efficiently is unclear. We essentially provide a computationally efficient way to carry out step (2) without computing this stationary distribution.

- We first construct an ε-approximate stationary distribution by recursively applying strategy modifications from Φ. The constructed distribution can be viewed as a tree. Our construction is inspired by the recent work of Zhang, Anagnostides, Farina, and Sandholm [Zha+24] for concave games. The main difference here is that for non-concave games, the distribution needs to be approximately stationary with respect to a *mixture* of strategy modifications rather than a single one as in concave games. Consequently, this leads to an approximate stationary distribution with prohibitively high support size $(|\Phi|)^{\sqrt{T}}$, as opposed to $\sqrt{T}$ in [Zha+24] for concave games.

- Despite the exponentially large support size of the distribution, we utilize its tree structure to design a simple and efficient sampling procedure that runs in time $\sqrt{T}$. Equipped with such a sampling

6

procedure, we provide an efficient randomized algorithm that generates a sequence of strategies so that, with high probability, the Φ-regret for this sequence of strategies is at most $O(\sqrt{T \log |\Phi|})$.

We defer the full proof of Theorem 2 to Section 3.1. An extension of Theorem 2 to infinite Φ holds when the rewards $\{u^t\}_{t \in [T]}$ are G-Lipschitz and Φ admits an α-cover with size $N(\alpha)$. In particular, when Φ is the set of all M-Lipschitz functions over $[0,1]^d$, Φ admits an α-cover with $\log N(\alpha)$ of the order $(1/\alpha)^d$ [SL07]. In this case, we have

Corollary 1. *There is a randomized algorithm such that, with probability at least $1 - \beta$, the Φ-regret is bounded by $c \cdot T^{\frac{d+1}{d+2}} \cdot \log(1/\beta)$, where c only depends on G and M. The algorithm runs in time* $\mathrm{poly}(T, N(T^{-1/(d+2)}))$.

3.1 Proof of Theorem 2

For a distribution $\mu \in \Delta(\mathcal{X})$ over strategy space $\mathcal{X}$, we slightly abuse notation and define its expected utility as $u^t(\mu) := \mathbb{E}_{x \sim \mu}[u^t(x)] \in [0,1]$. We define $\phi(\mu)$ the image of μ under transformation ϕ. In each iteration t, the learner chooses their strategy $x^t \in \mathcal{X}$ according to the distribution μ^t. For a sequence of strategies $\{x^t\}_{t \in [T]}$, the Φ-regret is $\mathrm{Reg}_\Phi^T := \max_{\phi \in \Phi} \left\{ \sum_{t=1}^{T} \left(u^t(\phi(x^t)) - u^t(x^t) \right) \right\}$. Algorithm 1 uses an external regret minimization algorithm R_Φ over Φ which outputs a distribution $p^t \in \Delta(\Phi)$. We can then decompose the Φ-regret into two parts.

$$\mathrm{Reg}_\Phi^T = \max_{\phi \in \Phi} \underbrace{\left\{ \sum_{t=1}^{T} u^t(\phi(x^t)) - \mathbb{E}_{\phi' \sim p^t}\left[u^t(\phi'(x^t))\right] \right\}}_{\text{I: external regret over } \Phi} + \underbrace{\sum_{t=1}^{T} \mathbb{E}_{\phi' \sim p^t}\left[u^t(\phi'(x^t))\right] - u^t(x^t)}_{\text{II: approximation error of stationary distribution}}.$$

I: Bounding the external regret over Φ. The external regret over Φ can be bounded directly. This is equivalent to an online expert problem: in each iteration t, the external regret minimizer $\mathfrak{R}_\Phi$ chooses $p^t \in \Delta(\Phi)$ and the adversary then determines the utility of each expert $\phi \in \Phi$ as $u^t(\phi(x^t))$. We choose the external regret minimizer $\mathfrak{R}_\Phi$ to be the Hedge algorithm [FS99]. Then we have $\max_{\phi \in \Phi} \left\{ \sum_{t=1}^{T} u^t(\phi(x^t)) - \mathbb{E}_{\phi' \sim p^t}[u^t(\phi'(x^t))] \right\} \leq 2\sqrt{T \log |\Phi|}$ (Theorem 6), where we use the fact that the utility function u^t is bounded in $[0,1]$.

II: Bounding error due to sampling from an approximate stationary distribution. We first define a distribution μ^t over $\mathcal{X}$ using a complete $|\Phi|$-ary tree with depth h. The root of this tree is an arbitrary strategy $x_{\text{root}} \in \mathcal{X}$. Each internal node x has exactly $|\Phi|$ children, denoted as $\{\phi(x)\}_{\phi \in \Phi}$. The distribution μ^t is supported on the nodes of this tree. Next, we define the probability for each node under the distribution μ^t. The root node x_{root} receives probability $\frac{1}{h}$. The probability of other nodes is defined in a recursive manner. For every node $x = \phi(x_p)$ where x_p is its parent, x receives probability $\mathrm{Pr}_{\mu^t}[x] = \mathrm{Pr}_{\mu^t}[x_p] \cdot p^t(\phi)$. It is then clear that the total probability of the children of a node x_p is exactly $\mathrm{Pr}_{\mu^t}[x \text{ is } x_p\text{'s child}] = \sum_\phi \mathrm{Pr}[x_p] \cdot p^t(\phi) = \mathrm{Pr}[x_p]$. Denote the set of nodes in depth k as N_k. We have $\mathrm{Pr}_{\mu^t}[x \in N_k] = \frac{1}{h}$ for every depth $1 \leq k \leq h$. Thus the distribution μ^t supports on $\frac{|\Phi|^h - 1}{|\Phi| - 1}$ points and is well-defined. By the construction above, we know x^t output by Algorithm 2 is a sample from μ^t.

Now we show that the approximation error of μ^t is bounded by $O(\frac{1}{h})$. We can evaluate the approximation error of μ^t:

$$\mathbb{E}_{\phi \sim p^t}\left[u^t(\phi(\mu^t))\right] - u^t(\mu^t) = \mathbb{E}_{\phi \sim p^t}\left[\sum_{k=1}^{h} \sum_{x \in N_k} \mathrm{Pr}_{\mu^t}[x] u^t(\phi(x))\right] - \left[\sum_{k=1}^{h} \sum_{x \in N_k} \mathrm{Pr}_{\mu^t}[x] u^t(x)\right]$$

$$= \sum_{k=1}^{h} \sum_{x \in N_k} \left(\mathbb{E}_{\phi \sim p^t}\left[\mathrm{Pr}_{\mu^t}[x] u^t(\phi(x))\right] - \mathrm{Pr}_{\mu^t}[x] u^t(x) \right).$$

7

We recall that for a node $x = \phi(x_p)$ with x_p being its parent, we have $\Pr_{\mu^t}[x] = \Pr_{\mu^t}[x_p] \cdot p^t(\phi)$. Thus for any $1 \le k \le h-1$, we have

$$\sum_{x \in N_k} \left(\mathbb{E}_{\phi \sim p^t} \left[\Pr_{\mu^t}[x] u^t(\phi(x)) \right] - \Pr_{\mu^t}[x] u^t(x) \right) = \sum_{x \in N_k} \left(\sum_{\phi \in \Phi} p^t(\phi) \Pr_{\mu^t}[x] u^t(\phi(x)) - \Pr_{\mu^t}[x] u^t(x) \right)$$

$$= \sum_{x \in N_{k+1}} \Pr_{\mu^t}[x] u^t(x) - \sum_{x \in N_k} \Pr_{\mu^t}[x] u^t(x).$$

Using the above equality, we get

$$\mathbb{E}_{\phi \sim p^t} \left[u^t(\phi(\mu^t)) \right] - u^t(\mu^t)$$

$$= \sum_{k=1}^{h-1} \sum_{x \in N_{k+1}} \Pr_{\mu^t}[x] u^t(x) + \sum_{x \in N_h} \sum_{\phi \in \Phi} p^t(\phi) \Pr_{\mu^t}[x] u^t(\phi(x)) - \sum_{k=2}^{h} \sum_{x \in N_k} \Pr_{\mu^t}[x] u^t(x) - \Pr_{\mu^t}[x_{\mathrm{root}}] u^t(x_{\mathrm{root}})$$

$$= \sum_{x \in N_h} \sum_{\phi \in \Phi} p^t(\phi) \Pr_{\mu^t}[x] u^t(\phi(x)) - \Pr_{\mu^t}[x_{\mathrm{root}}] u^t(x_{\mathrm{root}}) \le \frac{1}{h},$$

where in the last inequality we use the fact that $\sum_{x \in N_k} \Pr_{\mu^t}[x] = \frac{1}{h}$ for all $1 \le k \le h$ and the utility function u^t is bounded in $[0,1]$. Therefore, for $x^t \sim \mu^t$, the sequence of random variables $\{\sum_{t=1}^{\tau} (\mathbb{E}_{\phi \sim p^t}[u^t(\phi(x^t))] - u^t(x^t) - \frac{1}{h})\}_{\tau \ge 1}$ is a super-martingale. Thanks to the boundedness of the utility function, we can apply the Hoeffding-Azuma Inequality and get for any $\varepsilon > 0$.

$$\Pr \left[\sum_{t=1}^{T} \left(\mathbb{E}_{\phi \sim p^t} \left[u^t(\phi(x^t)) \right] - u^t(x^t) - \frac{1}{h} \right) \ge \varepsilon \right] \le \exp \left(-\frac{\varepsilon^2}{8T} \right). \tag{1}$$

Combining the bounds for I and II with $\varepsilon = \sqrt{8T \log(1/\beta)}$ and $h = \sqrt{T}$, we have, with probability $1 - \beta$, that $\mathrm{Reg}_\Phi^T \le 2\sqrt{T \log |\Phi|} + \frac{T}{h} + \sqrt{8T \log(1/\beta)} \le 8\sqrt{T(\log |\Phi| + \log(1/\beta))}$.

Convergence to Φ-equilibrium. If all players in a non-concave continuous game employ Algorithm 1, then we know for each player i, with probability $1 - \frac{\beta}{n}$, its $\Phi^{\mathcal{X}_i}$-regret is upper bounded by $8\sqrt{T(\log |\Phi^{\mathcal{X}_i}| + \log(n/\beta))}$. By a union bound over all n players, we get with probability $1 - \beta$, every player i's $\Phi^{\mathcal{X}_i}$-regret is upper bounded by $8\sqrt{T(\log |\Phi^{\mathcal{X}_i}| + \log(n/\beta))}$. Now by Theorem 1, we know the empirical distribution of strategy profiles played forms an ε-approximate $\Phi = \Pi_{i=1}^{n} \Phi^{\mathcal{X}_i}$-equilibrium, as long as $T \ge 64(\log(\max_i |\Phi^{\mathcal{X}_i}|) + \log(n/\beta))/\varepsilon^2$ iterations. $\square$

4 Approximate Φ-Equilibria under Infinite Local Strategy Modifications

This section studies Φ-equilibrium when $|\Phi|$ is infinite. It is, in general, computationally hard to compute a Φ-equilibrium even if Φ contains all constant deviations. Instead, we focus on Φ that consists solely of local strategy modifications. We introduce several natural classes of local strategy modifications and provide efficient online learning algorithms that converge to ε-approximate Φ-equilibrium in the first-order stationary regime where $\varepsilon = \Omega(\delta^2 L)$. These approximate Φ-equilibria guarantee first-order stability.

Definition 3 (δ-local strategy modification). *For each agent i, we call a set of strategy modifications $\Phi^{\mathcal{X}_i}$ δ-local if for all $x \in \mathcal{X}_i$ and $\phi_i \in \Phi^{\mathcal{X}_i}$, $\|\phi_i(x) - x\| \le \delta$. We use notation $\Phi^{\mathcal{X}_i}(\delta)$ to denote a δ-local strategy modification set for agent i. We also use $\Phi(\delta) = \Pi_{i=1}^{n} \Phi^{\mathcal{X}_i}(\delta)$ to denote a profile of δ-local strategy modification sets.*

Below we present a useful reduction from computing an ε-approximate $\Phi(\delta)$-equilibrium in *non-concave* smooth games to $\Phi^{\mathcal{X}_i}(\delta)$-regret minimization against *convex* losses for any $\varepsilon \ge \frac{\delta^2 L}{2}$. The key observation here is that the L-smoothness of the utility function permits, within a δ-neighborhood, a $\frac{\delta^2 L}{2}$-approximation using a linear function. We defer the proof to Appendix D.

Lemma 1 (No $\Phi(\delta)$-Regret for Convex Losses to Approximate $\Phi(\delta)$-Equilibrium in Non-Concave Games)**. *For any $T \ge 1$ and $\delta > 0$, let $\mathcal{A}$ be an algorithm that guarantees to achieve no more than $\mathrm{Reg}_{\Phi^{\mathcal{X}_i}(\delta)}^T$ $\Phi^{\mathcal{X}_i}(\delta)$-regret for convex loss functions for each agent $i \in [n]$. Then*

1. *The $\Phi^{\mathcal{X}_i}(\delta)$-regret of $\mathcal{A}$ for* non-convex *and L-smooth losses is at most* $\mathrm{Reg}^T_{\Phi^{\mathcal{X}_i}(\delta)} + \frac{\delta^2 LT}{2}$, $\forall i$.

2. *When every agent employs $\mathcal{A}$ in a non-concave L-smooth game, their empirical distribution of the joint strategies played converges to a $\left(\max_{i \in [n]}\{\mathrm{Reg}^T_{\Phi^{\mathcal{X}_i}(\delta)}\}/T + \frac{\delta^2 L}{2}\right)$-approximate $\Phi(\delta)$-equilibrium.*

4.1 Projection-Based Local Strategy Modifications

In this section, we study a set of local strategy modifications based on projection. Specifically, the set $\Phi^{\mathcal{X}}_{\mathrm{Proj}}(\delta)$ encompasses all deviations that essentially add a fixed displacement vector v to the input strategy and project back to the feasible set: $\Phi^{\mathcal{X}}_{\mathrm{Proj}}(\delta) := \{\phi_{\mathrm{Proj},v}(x) = \Pi_{\mathcal{X}}[x - v] : v \in B_d(\delta)\}$. It is clear that $\|\phi_v(x) - x\| \le \|v\| \le \delta$. The induced $\Phi^{\mathcal{X}}_{\mathrm{Proj}}(\delta)$-regret is $\mathrm{Reg}^T_{\mathrm{Proj},\delta} := \max_{v \in B_d(\delta)} \sum_{t=1}^T (f^t(x^t) - f^t(\Pi_{\mathcal{X}}[x^t - v]))$. We also define $\Phi_{\mathrm{Proj}}(\delta) = \Pi_{i=1}^n \Phi^{\mathcal{X}_i}_{\mathrm{Proj}}(\delta)$

By Lemma 1, to achieve convergence to an approximate $\Phi_{\mathrm{Proj}}(\delta)$-equilibrium, it suffices to minimize $\Phi^{\mathcal{X}}_{\mathrm{Proj}}(\delta)$-regret against convex losses. However, to our knowledge, there does not exist an algorithm that minimize $\Phi^{\mathcal{X}}_{\mathrm{Proj}}(\delta)$-regret even in the convex case. In fact, external regret and $\Phi^{\mathcal{X}}_{\mathrm{Proj}}(\delta)$-regret are provably incomparable: a sequence of actions may suffer high Reg^T but low $\mathrm{Reg}^T_{\mathrm{Proj},\delta}$ (Example 3) or vise versa (Example 4). At a high level, the external regret competes against a fixed action, whereas $\Phi^{\mathcal{X}}_{\mathrm{Proj}}(\delta)$-regret is more akin to the notion of *dynamic regret*, competing with a sequence of varying actions. Despite this, surprisingly, we show that classical algorithms like Online Gradient Descent (GD) and Optimistic Gradient (OG), known for minimizing external regret, also attain near-optimal $\Phi^{\mathcal{X}}_{\mathrm{Proj}}(\delta)$-regret. We defer the examples and missing proofs to Appendix E.

$\Phi^{\mathcal{X}}_{\mathrm{Proj}}(\delta)$**-Regret Minimization in the Adversarial Setting.** We show that (GD) enjoys an $O(G\sqrt{\delta D_{\mathcal{X}} T})$ $\Phi^{\mathcal{X}}_{\mathrm{Proj}}(\delta)$-regret despite the difference between the external regret and $\Phi^{\mathcal{X}}_{\mathrm{Proj}}(\delta)$-regret. First, let us recall the update rule of GD: given initial point $x^1 \in \mathcal{X}$ and step size $\eta > 0$, GD updates in each iteration t:

$$x^{t+1} = \Pi_{\mathcal{X}}[x - \eta \nabla f^t(x^t)]. \tag{GD}$$

The key step in our analysis for GD is simple but novel and general (See Appendix E.2). We extend the analysis to the Optimistic Gradient (OG) algorithm in Appendix E.4.

Theorem 3. *Let $\delta > 0$ and $T \in \mathbb{N}$. For any convex and G-Lipschitz loss functions $\{f^t : \mathcal{X} \to \mathbb{R}\}_{t \in [T]}$, the $\Phi^{\mathcal{X}}_{\mathrm{Proj}}(\delta)$-regret of (GD) with step size $\eta > 0$ is $\mathrm{Reg}^T_{\mathrm{Proj},\delta} \le \frac{\delta^2}{2\eta} + \frac{\eta}{2} G^2 T + \frac{\delta D_{\mathcal{X}}}{\eta}$. We can choose η optimally as $\frac{\sqrt{\delta(\delta + D_{\mathcal{X}})}}{G\sqrt{T}}$ and attain $\mathrm{Reg}^T_{\mathrm{Proj},\delta} \le 2G\sqrt{\delta(\delta + D_{\mathcal{X}})T}$. For any $\delta > 0$ and any $\varepsilon > 0$, when all player employ GD in a smooth game, their empirical distribution of played strategy profiles converges to an $(\varepsilon + \frac{\delta^2 L}{2})$-approximate $\Phi_{\mathrm{Proj}}(\delta)$-equilibrium in $O(1/\varepsilon^2)$ iterations.*

Remark 1. *The $\Phi^{\mathcal{X}}_{\mathrm{Proj}}(\delta)$-regret can also be viewed as the* dynamic regret *[Zin03] with changing comparators $\{p^t := \Pi_{\mathcal{X}}[x - v]\}$. However, our analysis does not follow from standard $O(\frac{(1+P_T)}{\eta} + \eta T)$ dynamic regret bound of GD [Zin03] since P_T, defined as $\sum_{t=2}^T \|p^t - p^{t-1}\|$, can be $\Omega(\eta T)$.*

Lower bounds for $\Phi^{\mathcal{X}}_{\mathrm{Proj}}(\delta)$-regret. We complement our upper bound with two lower bounds for $\Phi^{\mathcal{X}}_{\mathrm{Proj}}(\delta)$-regret minimization. The first one is an $\Omega(\delta G\sqrt{T})$ lower bound for any online learning algorithms against linear loss functions (Theorem 7). The second one is an $\Omega(\delta^2 LT)$ lower bound for any algorithm that satisfies the *linear span* assumption (Proposition 1), which holds for GD and OG against L-smooth non-convex losses. Combining with Lemma 1, this lower bound suggests that GD attains nearly optimal $\Phi^{\mathcal{X}}_{\mathrm{Proj}}(\delta)$-regret, even in the non-convex setting, among a natural family of gradient-based algorithms. We defer the theorem statements and detailed discussion to Appendix E.3.

Improved $\Phi^{\mathcal{X}}_{\mathrm{Proj}}(\delta)$-Regret in the Game Setting. We complement the $\Omega(\sqrt{T})$ lower bound in the *adversarial* setting by considering the *game* setting where players interact with each other using the same algorithm setting, which has been extensively studied for concave games [Syr+15; CP20; DFG21; Ana+22a; Ana+22b; Far+22a]. We prove an improved $O(T^{\frac{1}{4}})$ individual $\Phi^{\mathcal{X}}_{\mathrm{Proj}}(\delta)$-regret bound for OG (Theorem 9). We defer the details to Appendix E.4

4.2 Convex Combination of Finite Local Strategy Modifications

This section considers $\mathrm{Conv}(\Phi)$ where Φ is a finite set of local strategy modifications. The set of infinite strategy modifications $\mathrm{Conv}(\Phi)$ is defined as $\mathrm{Conv}(\Phi) = \{\phi_p(x) = \sum_{\phi \in \Phi} p(\phi)\phi(x) : p \in \Delta(\Phi)\}$. Our main result is an efficient algorithm (Algorithm 3) that guarantees convergence to an ε-approximate $\mathrm{Conv}(\Phi)$-equilibrium in a smooth game satisfying Assumption 1 for any $\varepsilon > \delta^2 L$. Due to space constraints, we defer Algorithm 3 and the proof to Appendix F.

Theorem 4. *Let $\mathcal{X}$ be a convex and compact set, Φ be an arbitrary finite set of δ-local strategy modification functions for $\mathcal{X}$, and $u^1(\cdot), \ldots, u^T(\cdot)$ be a sequence of G-Lipschitz and L-smooth but possibly non-concave reward functions from $\mathcal{X}$ to $[0,1]$. If we instantiate Algorithm 3 with $\mathfrak{R}_\Phi$ being the Hedge algorithm over $\Delta(\Phi)$ and $K = \sqrt{T}$, the algorithm guarantees that, with probability at least $1 - \beta$, it produces a sequence of strategies $x^1, \ldots, x^T$ with $\mathrm{Conv}(\Phi)$-regret at most $8\sqrt{T}(G\delta\sqrt{\log|\Phi|} + \sqrt{\log(1/\beta)}) + \delta^2 LT$. The algorithm runs in time $\sqrt{T}|\Phi|$ per iteration.*

If all players in a non-concave smooth game employ Algorithm 3, then with probability $1 - \beta$, for any $\varepsilon > 0$, the empirical distribution of strategy profiles played forms an $(\varepsilon + \delta^2 L)$-approximate $\Phi = \Pi_{i=1}^n \Phi^{\mathcal{X}_i}$-equilibrium,, after $\mathrm{poly}\left(\frac{1}{\varepsilon}, G, \log\left(\max_i |\Phi^{\mathcal{X}_i}|\right), \log\frac{n}{\beta}\right)$ iterations.

4.3 Interpolation-Based Local Strategy Modifications

We introduce a natural set of local strategy modifications and the corresponding local equilibrium notion. Given any set of (possibly non-local) strategy modifications $\Psi = \{\psi : \mathcal{X} \to \mathcal{X}\}$, we define a set of *local* strategy modifications as follows: for $\delta \leq D_{\mathcal{X}}$ and $\lambda \in [0,1]$, each strategy modification $\phi_{\lambda,\psi}$ interpolates the input strategy x with the modified strategy $\psi(x)$: formally,

$$\Phi_{\mathrm{Int},\Psi}^{\mathcal{X}}(\delta) := \{\phi_{\lambda,\psi}(x) := (1-\lambda)x + \lambda\psi(x) : \psi \in \Psi, \lambda \leq \delta/D_{\mathcal{X}}\}.$$

Note that for any $\psi \in \Psi$ and $\lambda \leq \frac{\delta}{D_{\mathcal{X}}}$, we have $\|\phi_{\lambda,\psi}(x) - x\| = \lambda\|x - \psi(x)\| \leq \delta$, respecting the locality constraint. The induced $\Phi_{\mathrm{Int},\Psi}^{\mathcal{X}}(\delta)$-regret can be written as $\mathrm{Reg}_{\mathrm{Int},\Psi,\delta}^T :=$ $\max_{\psi \in \Psi, \lambda \leq \frac{\delta}{D_{\mathcal{X}}}} \sum_{t=1}^T (f^t(x^t) - f^t((1-\lambda)x^t + \lambda\psi(x^t)))$. To guarantee convergence to the corresponding Φ-equilibrium, it suffices to minimize $\Phi_{\mathrm{Int},\Psi}^{\mathcal{X}}(\delta)$-regret against convex losses, which we show further reduces to Ψ-regret minimization against convex losses (Theorem 10 in Appendix G).

CCE-like Instantiation. In the special case where Ψ contains only *constant* strategy modifications (i.e., $\psi(x) = x^*$ for all x), we get a coarse correlated equilibrium (CCE)-like instantiation of local equilibrium, which limits the gain by interpolating with any *fixed* strategy. We denote the resulting set of local strategy modification simply as $\Phi_{\mathrm{Int}}^{\mathcal{X}}(\delta)$. We can apply any no-external regret algorithm for efficient $\Phi_{\mathrm{Int}}^{\mathcal{X}}(\delta)$-regret minimization and computation of ε-approximate $\Phi_{\mathrm{Int}}(\delta)$-equilibrium in the first-order stationary regime as summarized in Theorem 5. We also discuss faster convergence rates in the game setting in Appendix G.

Theorem 5. *For the Online Gradient Descent algorithm (GD) [Zin03] with step size $\eta = \frac{D_{\mathcal{X}}}{G\sqrt{T}}$, its $\Phi_{\mathrm{Int}}^{\mathcal{X}}(\delta)$-regret is at most $2\delta G\sqrt{T}$. Furthermore, for any $\delta > 0$ and any $\varepsilon > \frac{\delta^2 L}{2}$, when all players employ the GD algorithm in a smooth game, their empirical distribution of played strategy profiles converges to an $(\varepsilon + \frac{\delta^2 L}{2})$-approximate $\Phi_{\mathrm{Int}}(\delta)$-equilibrium in $O(1/\varepsilon^2)$ iterations.*

5 Discussion and Future Directions

Lower Bound in the Global Regime When δ equals the diameter of our strategy set, it is NP-hard to compute an ε-approximate $\Phi(\delta)$-equilibrium (for $\Phi(\delta) = \Phi_{\mathrm{Proj}}(\delta), \Phi_{\mathrm{Int}}(\delta)$), even when $\varepsilon = \Theta(1)$ and $G, L = O(\mathrm{poly}(d))$. Moreover, given black-box access to value and gradient queries, finding such equilibria requires exponentially many queries in at least one of the parameters $d, G, L, 1/\varepsilon$. These results are presented as Theorem 12 and Theorem 13 in Appendix I.

More Efficient Φ-Equilibria We include the discussion of another natural class of local strategy modifications that is based on beam search, where GD suffers linear regret to Appendix H. This result shows that even for simple local strategy modification sets $\Phi(\delta)$, the landscape of efficient local $\Phi(\delta)$-regret minimization is already quite rich and many questions remain open. A fruitful future direction is to identify more classes of Φ that admit efficient regret minimization.

References

[AD54]   Kenneth J Arrow and Gerard Debreu. "Existence of an equilibrium for a competitive economy". In: *Econometrica* (1954), pp. 265–290.

[AFS23]   Ioannis Anagnostides, Gabriele Farina, and Tuomas Sandholm. "Near-Optimal Phi-Regret Learning in Extensive-Form Games". In: *International Conference on Machine Learning (ICML)*. 2023.

[AGH19]   Naman Agarwal, Alon Gonen, and Elad Hazan. "Learning in non-convex games with an optimization oracle". In: *Conference on Learning Theory*. PMLR. 2019, pp. 18–29.

[AHK12]   Sanjeev Arora, Elad Hazan, and Satyen Kale. "The multiplicative weights update method: a meta-algorithm and applications". In: *Theory of computing* 8.1 (2012), pp. 121–164.

[ALW21]   Jacob Abernethy, Kevin A Lai, and Andre Wibisono. "Last-iterate convergence rates for min-max optimization: Convergence of hamiltonian gradient descent and consensus optimization". In: *Algorithmic Learning Theory*. PMLR. 2021, pp. 3–47.

[Ana+22a]   Ioannis Anagnostides, Constantinos Daskalakis, Gabriele Farina, Maxwell Fishelson, Noah Golowich, and Tuomas Sandholm. "Near-optimal no-regret learning for correlated equilibria in multi-player general-sum games". In: *Proceedings of the 54th Annual ACM SIGACT Symposium on Theory of Computing (STOC)*. 2022.

[Ana+22b]   Ioannis Anagnostides, Gabriele Farina, Christian Kroer, Chung-Wei Lee, Haipeng Luo, and Tuomas Sandholm. "Uncoupled Learning Dynamics with $O(\log T)$ Swap Regret in Multiplayer Games". In: *Advances in Neural Information Processing Systems (NeurIPS)*. 2022.

[Aue+02]   Peter Auer, Nicolo Cesa-Bianchi, Yoav Freund, and Robert E Schapire. "The non-stochastic multiarmed bandit problem". In: *SIAM journal on computing* 32.1 (2002), pp. 48–77.

[AZ22]   Amir Ali Ahmadi and Jeffrey Zhang. "On the complexity of finding a local minimizer of a quadratic function over a polytope". In: *Mathematical Programming* 195.1 (2022), pp. 783–792.

[AZF19]   Sergul Aydore, Tianhao Zhu, and Dean P Foster. "Dynamic local regret for non-convex online forecasting". In: *Advances in neural information processing systems* 32 (2019).

[Bai+22]   Yu Bai, Chi Jin, Song Mei, Ziang Song, and Tiancheng Yu. "Efficient Phi-Regret Minimization in Extensive-Form Games via Online Mirror Descent". In: *Advances in Neural Information Processing Systems* 35 (2022), pp. 22313–22325.

[Ber+23]   Martino Bernasconi, Matteo Castiglioni, Alberto Marchesi, Francesco Trovò, and Nicola Gatti. "Constrained Phi-Equilibria". In: *International Conference on Machine Learning*. 2023.

[Bla56]   David Blackwell. "An analog of the minimax theorem for vector payoffs." In: *Pacific Journal of Mathematics* 6.1 (Jan. 1956). Publisher: Pacific Journal of Mathematics, A Non-profit Corporation, pp. 1–8.

[Bro51]   George W Brown. "Iterative solution of games by fictitious play". In: *Act. Anal. Prod Allocation* 13.1 (1951), p. 374.

[Bub+15]   Sébastien Bubeck et al. "Convex optimization: Algorithms and complexity". In: *Foundations and Trends® in Machine Learning* 8.3-4 (2015), pp. 231–357.

[Cau01]   Robert Cauty. "Solution du problème de point fixe de Schauder". en. In: *Fundamenta Mathematicae* 170 (2001). Publisher: Instytut Matematyczny Polskiej Akademii Nauk, pp. 231–246.

[CDT09]   Xi Chen, Xiaotie Deng, and Shang-Hua Teng. "Settling the complexity of computing two-player Nash equilibria". In: *Journal of the ACM (JACM)* 56.3 (2009), pp. 1–57.

[CL06]   Nicolo Cesa-Bianchi and Gábor Lugosi. *Prediction, learning, and games*. Cambridge university press, 2006.

[CP20]   Xi Chen and Binghui Peng. "Hedging in games: Faster convergence of external and swap regrets". In: *Advances in Neural Information Processing Systems (NeurIPS)* 33 (2020), pp. 18990–18999.

[CZ23]   Yang Cai and Weiqiang Zheng. "Accelerated Single-Call Methods for Constrained Min-Max Optimization". In: *International Conference on Learning Representations (ICLR)* (2023).

[Dag+24]   Yuval Dagan, Constantinos Daskalakis, Maxwell Fishelson, and Noah Golowich. "From External to Swap Regret 2.0: An Efficient Reduction for Large Action Spaces". In: *Proceedings of the 56th Annual ACM Symposium on Theory of Computing*. 2024, pp. 1216–1222.

[Das+23]   Constantinos Daskalakis, Noah Golowich, Stratis Skoulakis, and Emmanouil Zampetakis. "STay-ON-the-Ridge: Guaranteed Convergence to Local Minimax Equilibrium in Nonconvex-Nonconcave Games". In: *The Thirty Sixth Annual Conference on Learning Theory*. PMLR. 2023, pp. 5146–5198.

[Das22]   Constantinos Daskalakis. "Non-Concave Games: A Challenge for Game Theory's Next 100 Years". In: *Cowles Preprints* (2022).

[DDJ21]   Jelena Diakonikolas, Constantinos Daskalakis, and Michael Jordan. "Efficient methods for structured nonconvex-nonconcave min-max optimization". In: *International Conference on Artificial Intelligence and Statistics* (2021).

[Deb52]   Gerard Debreu. "A social equilibrium existence theorem". In: *Proceedings of the National Academy of Sciences* 38.10 (1952), pp. 886–893.

[DFG21]   Constantinos Daskalakis, Maxwell Fishelson, and Noah Golowich. "Near-optimal no-regret learning in general games". In: *Advances in Neural Information Processing Systems (NeurIPS)* (2021).

[DGP09]   Constantinos Daskalakis, Paul W Goldberg, and Christos H Papadimitriou. "The complexity of computing a Nash equilibrium". In: *Communications of the ACM* 52.2 (2009), pp. 89–97.

[DP18]   Constantinos Daskalakis and Ioannis Panageas. "The limit points of (optimistic) gradient descent in min-max optimization". In: *the 32nd Annual Conference on Neural Information Processing Systems (NeurIPS)*. 2018.

[DSZ21]   Constantinos Daskalakis, Stratis Skoulakis, and Manolis Zampetakis. "The complexity of constrained min-max optimization". In: *Proceedings of the 53rd Annual ACM SIGACT Symposium on Theory of Computing (STOC)*. 2021.

[Fan53]   Ky Fan. "Minimax theorems". In: *Proceedings of the National Academy of Sciences* 39.1 (1953), pp. 42–47.

[Far+22a] Gabriele Farina, Ioannis Anagnostides, Haipeng Luo, Chung-Wei Lee, Christian Kroer, and Tuomas Sandholm. "Near-optimal no-regret learning dynamics for general convex games". In: *Advances in Neural Information Processing Systems* 35 (2022), pp. 39076–39089.

[Far+22b] Gabriele Farina, Andrea Celli, Alberto Marchesi, and Nicola Gatti. "Simple Uncoupled No-regret Learning Dynamics for Extensive-form Correlated Equilibrium". In: *Journal of the ACM* 69.6 (2022).

[FCR20]   Tanner Fiez, Benjamin Chasnov, and Lillian Ratliff. "Implicit learning dynamics in stackelberg games: Equilibria characterization, convergence analysis, and empirical study". In: *International Conference on Machine Learning*. PMLR. 2020, pp. 3133–3144.

[FR21]   Tanner Fiez and Lillian J Ratliff. "Local convergence analysis of gradient descent ascent with finite timescale separation". In: *Proceedings of the International Conference on Learning Representation*. 2021.

[FS99]   Yoav Freund and Robert E. Schapire. "Adaptive Game Playing Using Multiplicative Weights". In: *Games and Economic Behavior* 29 (1999), pp. 79–103.

[Geo63]   George B. Dantzig. *Linear Programming and Extensions*. Princeton University Press, 1963.

[GGM08]   Geoffrey J Gordon, Amy Greenwald, and Casey Marks. "No-regret learning in convex games". In: *Proceedings of the 25th international conference on Machine learning*. 2008, pp. 360–367.

[GJ03]   Amy Greenwald and Amir Jafari. "A general class of no-regret learning algorithms and game-theoretic equilibria". In: *Learning Theory and Kernel Machines: 16th Annual Conference on Learning Theory and 7th Kernel Workshop, COLT/Kernel 2003, Washington, DC, USA, August 24-27, 2003. Proceedings*. Springer. 2003, pp. 2–12.

[Gli52]   Irving L Glicksberg. "A further generalization of the Kakutani fixed theorem, with application to Nash equilibrium points". In: *Proceedings of the American Mathematical Society* 3.1 (1952), pp. 170–174.

[GLZ18]   Xiand Gao, Xiaobo Li, and Shuzhong Zhang. "Online learning with non-convex losses and non-stationary regret". In: *International Conference on Artificial Intelligence and Statistics*. PMLR. 2018, pp. 235–243.

[GZL23]   Ziwei Guan, Yi Zhou, and Yingbin Liang. "Online Nonconvex Optimization with Limited Instantaneous Oracle Feedback". In: *The Thirty Sixth Annual Conference on Learning Theory*. PMLR. 2023, pp. 3328–3355.

[Han57]   James Hannan. "Approximation to Bayes risk in repeated play". In: *Contributions to the Theory of Games* 3 (1957), pp. 97–139.

[Hél+20]   Amélie Héliou, Matthieu Martin, Panayotis Mertikopoulos, and Thibaud Rahier. "Online non-convex optimization with imperfect feedback". In: *Advances in Neural Information Processing Systems* 33 (2020), pp. 17224–17235.

[HMC21a]   Nadav Hallak, Panayotis Mertikopoulos, and Volkan Cevher. "Regret minimization in stochastic non-convex learning via a proximal-gradient approach". In: *International Conference on Machine Learning*. PMLR. 2021, pp. 4008–4017.

[HMC21b]   Ya-Ping Hsieh, Panayotis Mertikopoulos, and Volkan Cevher. "The limits of min-max optimization algorithms: Convergence to spurious non-critical sets". In: *International Conference on Machine Learning*. PMLR. 2021, pp. 4337–4348.

[HSZ17]   Elad Hazan, Karan Singh, and Cyril Zhang. "Efficient regret minimization in non-convex games". In: *International Conference on Machine Learning*. PMLR. 2017, pp. 1433–1441.

[JNJ20]   Chi Jin, Praneeth Netrapalli, and Michael Jordan. "What is local optimality in nonconvex-nonconcave minimax optimization?" In: *International conference on machine learning (ICML)*. PMLR. 2020, pp. 4880–4889.

[Kar14]   Samuel Karlin. *Mathematical Methods and Theory in Games, Programming, and Economics: Volume 2: The Theory of Infinite Games*. Elsevier, 2014.

[Kar59]   Samuel Karlin. *Mathematical methods and theory in games, programming, and economics: Volume II: the theory of infinite games*. Vol. 2. Addision-Wesley, 1959.

[Kri+15]   Walid Krichene, Maximilian Balandat, Claire Tomlin, and Alexandre Bayen. "The hedge algorithm on a continuum". In: *International Conference on Machine Learning*. PMLR. 2015, pp. 824–832.

[McK54]   Lionel McKenzie. "On equilibrium in Graham's model of world trade and other competitive systems". In: *Econometrica* (1954), pp. 147–161.

[MK87]   Katta G. Murty and Santosh N. Kabadi. "Some NP-complete problems in quadratic and nonlinear programming". en. In: *Mathematical Programming* 39.2 (June 1987), pp. 117–129.

[MM10]   Odalric-Ambrym Maillard and Rémi Munos. "Online learning in adversarial lipschitz environments". In: *Joint european conference on machine learning and knowledge discovery in databases*. Springer. 2010, pp. 305–320.

[Mor+21a] 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: *Proceedings of the AAAI Conference on Artificial Intelligence*. Vol. 35. 6. 2021, pp. 5584–5594.

[Mor+21b] 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*. PMLR. 2021, pp. 7818–7828.

[MRS20] Eric Mazumdar, Lillian J Ratliff, and S Shankar Sastry. "On gradient-based learning in continuous games". In: *SIAM Journal on Mathematics of Data Science* 2.1 (2020), pp. 103–131.

[MV21] Oren Mangoubi and Nisheeth K Vishnoi. "Greedy adversarial equilibrium: an efficient alternative to nonconvex-nonconcave min-max optimization". In: *Proceedings of the 53rd Annual ACM SIGACT Symposium on Theory of Computing*. 2021, pp. 896–909.

[MZ19] Panayotis Mertikopoulos and Zhengyuan Zhou. "Learning in games with continuous action sets and unknown payoff functions". In: *Mathematical Programming* 173 (2019), pp. 465–507.

[Nas50] John F Nash Jr. "Equilibrium points in n-person games". In: *Proceedings of the national academy of sciences* 36.1 (1950), pp. 48–49.

[Neu28] J v. Neumann. "Zur theorie der gesellschaftsspiele". In: *Mathematische annalen* 100.1 (1928), pp. 295–320.

[Pet+22] Thomas Pethick, Puya Latafat, Panagiotis Patrinos, Olivier Fercoq, and Volkan Cevherå. "Escaping limit cycles: Global convergence for constrained nonconvex-nonconcave minimax problems". In: *International Conference on Learning Representations (ICLR)*. 2022.

[Pil+22] Georgios Piliouras, Mark Rowland, Shayegan Omidshafiei, Romuald Elie, Daniel Hennes, Jerome Connor, and Karl Tuyls. "Evolutionary dynamics and phi-regret minimization in games". In: *Journal of Artificial Intelligence Research* 74 (2022), pp. 1125–1158.

[PR24] Binghui Peng and Aviad Rubinstein. "Fast swap regret minimization and applications to approximate correlated equilibria". In: *Proceedings of the 56th Annual ACM Symposium on Theory of Computing*. 2024, pp. 1223–1234.

[RBS16] Lillian J Ratliff, Samuel A Burden, and S Shankar Sastry. "On the characterization of local Nash equilibria in continuous games". In: *IEEE transactions on automatic control* 61.8 (2016), pp. 2301–2307.

[Rob51] Julia Robinson. "An iterative method of solving a game". In: *Annals of mathematics* (1951), pp. 296–301.

[Ros65] J Ben Rosen. "Existence and uniqueness of equilibrium points for concave n-person games". In: *Econometrica* (1965), pp. 520–534.

[RS13] Sasha Rakhlin and Karthik Sridharan. "Optimization, learning, and games with predictable sequences". In: *Advances in Neural Information Processing Systems* (2013).

[RST11] Alexander Rakhlin, Karthik Sridharan, and Ambuj Tewari. "Online learning: Beyond regret". In: *Proceedings of the 24th Annual Conference on Learning Theory*. JMLR Workshop and Conference Proceedings. 2011, pp. 559–594.

[Sha24] Dravyansh Sharma. "No Internal Regret with Non-convex Loss Functions". In: *Proceedings of the AAAI Conference on Artificial Intelligence*. Vol. 38. 13. 2024, pp. 14919–14927.

[Sio58] Maurice Sion. "On general minimax theorems." In: *Pacific J. Math.* 8.4 (1958), pp. 171–176.

[SL07] Gilles Stoltz and Gábor Lugosi. "Learning correlated equilibria in games with compact sets of strategies". In: *Games and Economic Behavior* 59.1 (2007), pp. 187–208.

[SMB22] Ziang Song, Song Mei, and Yu Bai. "Sample-efficient learning of correlated equilibria in extensive-form games". In: *Advances in Neural Information Processing Systems* 35 (2022), pp. 4099–4110.

[SN20] Arun Sai Suggala and Praneeth Netrapalli. "Online non-convex learning: Following the perturbed leader is optimal". In: *Algorithmic Learning Theory*. PMLR. 2020, pp. 845–861.

[Syr+15] Vasilis Syrgkanis, Alekh Agarwal, Haipeng Luo, and Robert E Schapire. "Fast convergence of regularized learning in games". In: *Advances in Neural Information Processing Systems (NeurIPS)* (2015).

[VF08] Bernhard Von Stengel and Françoise Forges. "Extensive-form correlated equilibrium: Definition and computational complexity". In: *Mathematics of Operations Research* 33.4 (2008), pp. 1002–1022.

[WZB20]   Yuanhao Wang, Guodong Zhang, and Jimmy Ba. "On Solving Minimax Optimization Locally: A Follow-the-Ridge Approach". In: *International Conference on Learning Representations (ICLR)*. 2020.

[Zha+24]   Brian Hu Zhang, Ioannis Anagnostides, Gabriele Farina, and Tuomas Sandholm. "Efficient Φ-Regret Minimization with Low-Degree Swap Deviations in Extensive-Form Games". In: *Advances in Neural Information Processing Systems*. 2024.

[Zin03]   Martin Zinkevich. "Online convex programming and generalized infinitesimal gradient ascent". In: *Proceedings of the 20th international conference on machine learning (ICML)*. 2003.

Contents

A   Related Work

Non-Concave Games.   An important special case of multi-player games are two-player zero-sum games, which are defined in terms of some function $f : \mathcal{X} \times \mathcal{Y} \to \mathbb{R}$ that one of the two players say the one choosing $x \in \mathcal{X}$, wants to minimize, while the other player, the one choosing $y \in \mathcal{Y}$, wants to maximize. Finding Nash equilibrium in such games is tractable in the *convex-concave* setting, i.e. when $f(x, y)$ is convex with respect to the minimizing player's strategy, x, and concave with respect to the maximizing player's strategy, y, but it is computationally intractable in the general *nonconvex-nonconcave* setting. Namely, a Nash equilibrium may not exist, and it is NP-hard to determine if one exists and, if so, find it. Moreover, in this case, stable limit points of gradient-based dynamics are not necessarily Nash equilibria, not even local Nash equilibria [DP18; MRS20]. Moreover, there are examples including the "Polar Game" [Pet+22] and the "Forsaken Matching Pennies" [HMC21b] showing that for GD / OG and many other no-regret learning algorithms in nonconvex-nonconcave min-max optimization, the last-iterate does not converge and even the average-iterate fails to be a stationary point. We emphasize that the convergence guarantees we provide for GD / OG in Section 4.1 and Section 4.3 holds for the empirical distribution of play, not the average-iterate or the last-iterate.

A line of work focuses on computing Nash equilibrium under additional structure in the game. This encompasses settings where the game satisfies the (weak) Minty variational inequality [MZ19; DDJ21; Pet+22; CZ23], or is sufficiently close to being bilinear [ALW21]. However, the study of universal solution concepts in the nonconvex-nonconcave setting is sparse. Daskalakis, Skoulakis, and Zampetakis [DSZ21] proved the existence and computational hardness of local Nash equilibrium. In a more recent work, [Das+23] proposes second-order algorithms with asymptotic convergence to local Nash equilibrium. Several works study sequential two-player zero-sum games with additional assumptions about the player who goes second. They propose equilibrium concepts such as *local minimax points* [JNJ20], *differentiable Stackelberg equilibrium* [FCR20], and *greedy adversarial equilibrium* [MV21]. Notably, local minimax points are stable limit points of Gradient-Descent-Ascent (GDA) dynamics [JNJ20; WZB20; FR21] while greedy adversarial equilibrium can be computed efficiently using second-order algorithms in the unconstrained setting [MV21]. In contrast to these studies, we focus on the more general case of multi-player non-concave games.

Local Equilibrium.   To address the limitations associated with classical, global equilibrium concepts, a natural approach is to focus on developing equilibrium concepts that guarantee local stability instead. One definition of interest is the strict local Nash equilibrium, wherein each player's strategy corresponds to a local maximizer of their utility function, given the other players' strategies. Unfortunately, a strict local Nash equilibrium may not always exist, as demonstrated in Example 1. Furthermore, a weaker notion—the second-order local Nash equilibrium, where each player has no incentive to deviate based on the second-order Taylor expansion estimate of their utility, is also not guaranteed to exist as illustrated in Example 1. What's more, it is NP-hard to check whether a given strategy profile is a strict local Nash equilibrium or a second-order local Nash equilibrium, as implied by the result of Murty and Kabadi [MK87] and Ahmadi and Zhang [AZ22].[6] Finally, one can consider *local Nash equilibrium*, a first-order stationary solution, which is guaranteed to exist [DSZ21]. Unlike non-convex optimization, where targeting first-order local optima sidesteps the intractability of global optima, this first-order local Nash equilibrium has been recently shown to be intractable, even in

---

[6]Murty and Kabadi [MK87] shows that checking whether a point is a local maximizer of a multi-variate quadratic function is NP-hard, and Ahmadi and Zhang [AZ22] shows that computing a local maximize of a multi-variate quadratic function over a polytope is NP-hard.

two-player zero-sum non-concave games with joint feasibility constraints [DSZ21].[7] See Table 1 for a summary of solution concepts in non-concave games.

Online Learning with Non-Convex Losses. A line of work has studied online learning against non-convex losses. To circumvent the computational intractability of this problem, various approaches have been pursued: some works assume a restricted set of non-convex loss functions [GLZ18], while others assume access to a sampling oracle [MM10; Kri+15] or access to an offline optimization oracle [AGH19; SN20; Hél+20] or a weaker notion of regret [HSZ17; AZF19; HMC21a; GZL23]. The work most closely related to ours is [HSZ17]. The authors propose a notion of w-*smoothed local regret* against non-convex losses, and they also define a local equilibrium concept for non-concave games. They use the idea of *smoothing* to average the loss functions in the previous w iterations and design algorithms with optimal w-smoothed local regret. The concept of regret they introduce suggests a local equilibrium concept. However, their local equilibrium concept is non-standard in that its local stability is not with respect to a distribution over strategy profiles sampled by this equilibrium concept. Moreover, the path to attaining this local equilibrium through decentralized learning dynamics remains unclear. The algorithms provided in [HSZ17; GZL23] require that every agent i experiences (over several rounds) the average utility function of the previous w iterates, denoted as $F_{i,w}^t := \frac{1}{w} \sum_{\ell=0}^{w-1} u_i^{t-\ell}(\cdot, x_{-i}^{t-\ell})$. Implementing this imposes a significant coordination burden on the agents. In contrast, we focus on a natural concept of $\Phi(\delta)$-equilibrium, which is incomparable to that of [HSZ17], and we also show that efficient convergence to this concept is achieved via decentralized gradient-based learning dynamics.

Φ-regret and Φ-equilibrium. The concept of Φ-regret and the associated Φ-equilibrium is introduced by Greenwald and Jafari [GJ03] and has been broadly investigated in the context of concave games [GJ03; SL07; GGM08; RST11; Pil+22; Ber+23; Dag+24; PR24] and extensive-form games [VF08; Mor+21a; Mor+21b; Far+22b; Bai+22; SMB22; AFS23; Zha+24]. The work of [Sha24] studies internal regret minimization against non-convex losses. To our knowledge, no efficient algorithm exists for the classes of Φ-equilibria we consider for non-concave games. Specifically, all of these algorithms, when applied to compute a $(\varepsilon, \Phi(\delta))$-equilibrium for a general δ-local strategy modification set $\Phi(\delta)$ (using Lemma 1), require running time exponential in either $1/\varepsilon$ or the dimension d. In contrast, we show that for several natural choices of $\Phi(\delta)$, ε-approximate $\Phi(\delta)$-equilibrium can be computed efficiently, i.e. polynomial in $1/\varepsilon$ and d, using simple algorithms.

B Additional Preliminaries: Solution Concepts in Non-Concave Games

We present definitions of several solution concepts in the literature as well as the existence and computational complexity of each solution concept.

Definition 4 (Nash Equilibrium). *In a continuous game, a strategy profile $x \in \prod_{j=1}^n \mathcal{X}_j$ is a* Nash equilibrium (NE) *if and only if for every player $i \in [n]$,*

$$u_i(x_i', x_{-i}) \leq u_i(x), \forall x_i' \in \mathcal{X}_i$$

Definition 5 (Mixed Nash Equilibrium). *In a continuous game, a mixed strategy profile $p \in \prod_{j=1}^n \Delta(\mathcal{X}_j)$ (here we denote $\Delta(\mathcal{X}_i)$ as the set of probability measures over $\mathcal{X}_i$) is a* mixed Nash equilibrium (MNE) *if and only if for every player $i \in [n]$,*

$$u_i(p_i', p_{-i}) \leq u_i(p), \forall p_i' \in \Delta(\mathcal{X}_i)$$

Definition 6 ((Coarse) Correlated Equilibrium). *In a continuous game, a distribution σ over joint strategy profiles $\Pi_{i=1}^n \mathcal{X}_i$ is a* correlated equilibrium (CE) *if and only if for all player $i \in [n]$,*

$$\max_{\phi_i : \mathcal{X}_i \to \mathcal{X}_i} \mathbb{E}_{x \sim \sigma}[u_i(\phi_i(x_i), x_{-i})] \leq \mathbb{E}_{x \sim \sigma}[u_i(x)].$$

Similarly, a distribution σ over joint strategy profiles $\Pi_{i=1}^n \mathcal{X}_i$ is a coarse correlated equilibrium (CCE) *if and only if for all player $i \in [n]$,*

$$\max_{x_i' \in \mathcal{X}_i} \mathbb{E}_{x \sim \sigma}[u_i(x_i', x_{-i})] \leq \mathbb{E}_{x \sim \sigma}[u_i(x)].$$

---

[7] In general sum games, it is not hard to see that the intractability results [DGP09; CDT09] for computing *global* Nash equilibria in bimatrix games imply intractability for computing *local* Nash equilibria.

Definition 7 (Strict Local Nash Equilibrium). *In a continuous game, a strategy profile $x \in \prod_{j=1}^{n} \mathcal{X}_j$ is a* strict local Nash equilibrium *if and only if for every player $i \in [n]$, there exists $\delta > 0$ such that*

$$u_i(x_i', x_{-i}) \leq u_i(x), \forall x_i' \in B_{d_i}(x_i, \delta) \cap \mathcal{X}_i.$$

Definition 8 (Second-order Local Nash Equilibrium). *Consider a continuous game where each utility function $u_i(x_i, x_{-i})$ is twice-differentiable with respect to x_i for any fixed x_{-i}. A strategy profile $x \in \prod_{j=1}^{n} \mathcal{X}_j$ is a* second-order local Nash equilibrium *if and only if for every player $i \in [n]$, x_i maximizes the second-order Taylor expansion of its utility functions at x_i, or formally,*

$$\langle \nabla_{x_i} u_i(x), x_i' - x_i \rangle + (x_i' - x_i)^\top \nabla_{x_i}^2 u_i(x)(x_i' - x_i) \leq 0, \forall x_i' \in \mathcal{X}_i.$$

Existence Mixed Nash equilibria exist in continuous games, thus smooth games [Deb52; Gli52; Fan53]. By definition, an MNE is also a CE and a CCE. This also proves the existence of CE and CCE. In contrast, strict local Nash equilibria, second-order Nash equilibria, or (pure) Nash equilibria may not exist in a smooth non-concave game, as we show in the following example.

Example 1. *Consider a two-player zero-sum non-concave game: the action sets are $\mathcal{X}_1 = \mathcal{X}_2 = [-1, 1]$ and the utility functions are $u_1(x_1, x_2) = -u_2(x_1, x_2) = (x_1 - x_2)^2$. Let $x = (x_1, x_2) \in \mathcal{X}_1 \times \mathcal{X}_2$ be any strategy profile: if $x_1 = x_2$, then player 1 is not at a local maximizer; if $x_1 \neq x_2$, then player 2 is not at a local maximizer. Thus x is not a strict local Nash equilibrium. Since the utility function is quadratic, we conclude that the game also has no second-order local Nash equilibrium.*

Computational Complexity Consider a single-player smooth non-concave game with a quadratic utility function $f : \mathcal{X} \rightarrow \mathbb{R}$. The problem of finding a *local* maximizer of f can be reduced to the problem of computing a NE, a MNE, a CE, a CCE, a strict local Nash equilibrium, or a second-order local Nash equilibrium. Since computing a local maximizer or checking if a given point is a local maximizer is NP-hard [MK87], we know that the computational complexities of NE, MNE, CE, CCE, strict local Nash equilibria, and second-order local Nash equilibria are all NP-hard.

Representation Complexity Karlin [Kar59] present a two-player zero-sum non-concave game whose unique MNE has infinite support. Since in a two-player zero-sum game, the marginal distribution of a CE or a CCE is an MNE, it also implies that the representation complexity of any CE or CCE is infinite. We present the example in Karlin [Kar59] here for completeness and also prove that the game is Lipschitz and smooth.

Example 2 ([Kar59, Chapter 7.1, Example 3]). *We consider a two-player zero-sum game with action sets $\mathcal{X}_1 = \mathcal{X}_2 = [0, 1]$. Let p and q be two distributions over $[0, 1]$. The only requirement for p and q is that their cumulative distribution functions are not finite-step functions. For example, we can take $p = q$ to be the uniform distribution.*

Let μ_n and ν_n denote the n-th moments of p and q, respectively. Define the utility function

$$u(x, y) = u_1(x, y) = -u_2(x, y) = \sum_{n=0}^{\infty} \frac{1}{2^n}(x^n - \mu_n)(y^n - \nu_n), \quad 0 \leq x, y \leq 1.$$

Claim 1. *The game in Example 2 is 2-Lipschitz and 6-smooth, and (p, q) is its unique (mixed) Nash equilibrium.*

Proof. Fix any $y \in [0, 1]$, since $|\frac{1}{2^n}(y^n - \nu_n)nx^{n-1}| \leq \frac{n}{2^n}$, the series of $\nabla_x u(x, y)$ is uniformly convergent. We have $|\nabla_x u(x, y)| \leq \sum_{n=0}^{\infty} \frac{n}{2^n} \leq 2, \quad y \in [0, 1]$. Similarly, we have $|\nabla_x^2 u(x, y)| \leq \sum_{n=0}^{\infty} \frac{n^2}{2^n} \leq 6$ for all $y \in [0, 1]$. By symmetry, we also have $|\nabla_y(x, y)| \leq 2$ and $|\nabla_y^2(x, y)| \leq 6$ for all $x, y \in [0, 1]$. Thus, the game is 2-Lispchitz and 6-smooth.

Since $|\frac{1}{2^n}(x^n - \mu_n)(y^n - \nu_n)| \leq \frac{1}{2^n}$, the series of $u(x, y)$ is absolutely and uniformly convergent. We have

$$\int_0^1 u(x, y) \mathrm{d}F_p(x) = \sum_{n=0}^{\infty} \frac{1}{2^n}(y^n - \nu_n) \int_0^1 (x^n - \mu_n) \mathrm{d}F_p(x) \equiv 0,$$

$$\int_0^1 u(x, y) F_q(y) = \sum_{n=0}^{\infty} \frac{1}{2^n}(x^n - \mu_n) \int_0^1 (y^n - \nu_n) \mathrm{d}F_q(y) \equiv 0.$$

In particular, (p, q) is a mixed Nash equilibrium, and the value of the game is 0. Suppose (p', q') is also a mixed Nash equilibrium. Then (p, q') is a mixed Nash equilibrium. Note that p supports on every point in $[0, 1]$. As a consequence, we have

$$0 \equiv \int_0^1 u(x, y) \mathrm{d}F_{q'}(y) = \sum_{n=0}^{\infty} \frac{1}{2^n} (x^n - \mu_n)(\nu_n' - \nu_n)$$

for all $x \in [0, 1]$, where ν_n' is the n-th moment of q'. Since the series vanished identically, the coefficients of each power of x must vanish. Thus $\nu_n' = \nu_n$ and $q' = q$. Similarly, we have $p' = p$, and the mixed Nash equilibrium is unique. □

C   Rerget Bound for Hedge

Theorem 6 ([AHK12]). *In an N-expert problem, assume all the rewards are bounded, i.e., $u^t \in [-M, M]^N$, then the Hedge algorithm with step size $\eta = \min\{\frac{1}{M}, \frac{\sqrt{\log N}}{M\sqrt{T}}\}$ has regret*

$$\max_{p \in \Delta(N)} \sum_{t=1}^T \langle u^t, p \rangle - \sum_{t=1}^T \langle u^t, p^t \rangle \leq 2M\sqrt{T \log N}.$$

D   Proof of Lemma 1

Let $\{f^t\}_{t \in [T]}$ be a sequence of non-convex L-smooth loss functions satisfying Assumption 1. Let $\{x^t\}_{t \in [T]}$ be the iterates produced by $\mathcal{A}$ against $\{f^t\}_{t \in [T]}$. Then $\{x^t\}_{t \in [T]}$ is also the iterates produced by $\mathcal{A}$ against a sequence of linear loss functions $\{\langle \nabla f^t(x^t), \cdot \rangle\}$. For the latter, we know

$$\max_{\phi \in \Phi^{\mathcal{X}}(\delta)} \sum_{t=1}^T \langle \nabla f^t(x^t), x^t - \phi(x^t) \rangle \leq \mathrm{Reg}_{\Phi^{\mathcal{X}}(\delta)}^T.$$

Then using L-smoothness of $\{f^t\}$ and the fact that $\|\phi(x) - x\| \leq \delta$ for all $\phi \in \Phi(\delta)$, we get

$$\max_{\phi \in \Phi^{\mathcal{X}}(\delta)} \sum_{t=1}^T f^t(x^t) - f^t(\phi(x^t)) \leq \max_{\phi \in \Phi^{\mathcal{X}}(\delta)} \sum_{t=1}^T \left(\langle \nabla f^t(x^t), x^t - \phi(x^t) \rangle + \frac{L}{2} \|x^t - \phi(x^t)\|^2 \right)$$

$$\leq \mathrm{Reg}_{\Phi^{\mathcal{X}}(\delta)}^T + \frac{\delta^2 LT}{2}.$$

This completes the proof of the first part.

Let each player $i \in [n]$ employ algorithm $\mathcal{A}$ in a smooth game independently and produces iterates $\{x^t\}$. The averaged joint strategy profile σ^T that chooses x^t uniformly at random from $t \in [T]$ satisfies for any player $i \in [n]$,

$$\max_{\phi \in \Phi^{\mathcal{X}_i}(\delta)} \mathbb{E}_{x \sim \sigma}[u_i(\phi(x_i), x_{-i})] - \mathbb{E}_{x \sim \sigma}[u_i(x)]$$

$$= \max_{\phi \in \Phi^{\mathcal{X}_i}(\delta)} \frac{1}{T} \sum_{t=1}^T \left(u_i(\phi(x_i^t), x_{-i}^t) - u_i(x^t) \right)$$

$$\leq \frac{\mathrm{Reg}_{\Phi^{\mathcal{X}_i}(\delta)}^T}{T} + \frac{\delta^2 L}{2}.$$

Thus σ^T is a $(\max_{i \in [n]}\{\mathrm{Reg}_{\Phi^{\mathcal{X}_i}(\delta)}^T\} \cdot T^{-1} + \frac{\delta^2 L}{2})$-approximate $\Phi(\delta)$-equilibrium. This completes the proof of the second part.

E   Missing Details in Section 4.1

E.1   Differences between External Regret and $\Phi_{\mathrm{Proj}}^{\mathcal{X}}$-regret

In the following two examples, we show that $\Phi_{\mathrm{Proj}}^{\mathcal{X}}(\delta)$-regret is incomparable with external regret for convex loss functions . A sequence of actions may suffer high Reg^T but low $\mathrm{Reg}_{\mathrm{Proj},\delta}^T$ (Example 3), and vise versa (Example 4).

Example 3. *Let $f^1(x) = f^2(x) = |x|$ for $x \in \mathcal{X} = [-1,1]$. Then the $\Phi^{\mathcal{X}}_{\text{Proj}}(\delta)$-regret of the sequence $\{x^1 = \frac{1}{2}, x^2 = -\frac{1}{2}\}$ for any $\delta \in (0, \frac{1}{2})$ is 0. However, the external regret of the same sequence is 1. By repeating the construction for $\frac{T}{2}$ times, we conclude that there exists a sequence of actions with $\text{Reg}^T_{\text{Proj},\delta} = 0$ and $\text{Reg}^T = \frac{T}{2}$ for all $T \geq 2$.*

Example 4. *Let $f^1(x) = -2x$ and $f^2(x) = x$ for $x \in \mathcal{X} = [-1,1]$. Then the $\Phi^{\mathcal{X}}_{\text{Proj}}(\delta)$-regret of the sequence $\{x^1 = \frac{1}{2}, x^2 = 0\}$ for any $\delta \in (0, \frac{1}{2})$ is δ. However, the external regret of the same sequence is 0. By repeating the construction for $\frac{T}{2}$ times, we conclude that there exists a sequence of actions with $\text{Reg}^T_{\text{Proj},\delta} = \frac{\delta T}{2}$ and $\text{Reg}^T = 0$ for all $T \geq 2$.*

At a high level, the external regret competes against a fixed action, whereas $\Phi^{\mathcal{X}}_{\text{Proj}}(\delta)$-regret is more akin to the notion of *dynamic regret*, competing with a sequence of varying actions. When the environment is stationary, i.e., $f^t = f$ (Example 3), a sequence of actions that are far away from the global minimum must suffer high regret, but may produce low $\Phi^{\mathcal{X}}_{\text{Proj}}(\delta)$-regret since the change to the cumulative loss caused by a fixed-direction deviation could be neutralized across different actions in the sequence. In contrast, in a non-stationary (dynamic) environment (Example 4), every fixed action performs poorly, and a sequence of actions could suffer low regret against a fixed action but the $\Phi^{\mathcal{X}}_{\text{Proj}}(\delta)$-regret that competes with a fixed-direction deviation could be large. Nevertheless, despite these differences between the two notions of regret as shown above, they are *compatible* for convex loss functions: our main results in this section provide algorithms that minimize external regret and $\Phi^{\mathcal{X}}_{\text{Proj}}(\delta)$-regret simultaneously.

E.2 Proof of Theorem 3

Proof. Let us denote $v \in B_d(\delta)$ a fixed deviation and define $p^t = \Pi_{\mathcal{X}}[x^t - v]$. By standard analysis of GD [Zin03] (see also the proof of [Bub+15, Theorem 3.2]), we have

$$\sum_{t=1}^{T} \left(f^t(x^t) - f^t(p^t) \right) \leq \sum_{t=1}^{T} \frac{1}{2\eta} \left(\|x^t - p^t\|^2 - \|x^{t+1} - p^t\|^2 + \eta^2 \|\nabla f^t(x^t)\|^2 \right)$$

$$\leq \sum_{t=1}^{T-1} \frac{1}{2\eta} \left(\|x^{t+1} - p^{t+1}\|^2 - \|x^{t+1} - p^t\|^2 \right) + \frac{\delta^2}{2\eta} + \frac{\eta}{2} G^2 T,$$

where the last step uses $\|x^1 - p^1\| \leq \delta$ and $\|\nabla f^t(x^t)\| \leq G$. Here the terms $\|x^{t+1} - p^{t+1}\|^2 - \|x^{t+1} - p^t\|^2$ do not telescope, and we further relax them in the following key step.

Key Step: We relax the first term as:

$$\|x^{t+1} - p^{t+1}\|^2 - \|x^{t+1} - p^t\|^2 = \langle p^t - p^{t+1}, 2x^{t+1} - p^t - p^{t+1} \rangle$$

$$= \langle p^t - p^{t+1}, 2x^{t+1} - 2p^{t+1} \rangle - \|p^t - p^{t+1}\|^2$$

$$= 2\langle p^t - p^{t+1}, v \rangle + 2\langle p^t - p^{t+1}, x^{t+1} - v - p^{t+1} \rangle - \|p^t - p^{t+1}\|^2$$

$$\leq 2\langle p^t - p^{t+1}, v \rangle - \|p^t - p^{t+1}\|^2,$$

where in the last inequality we use the fact that p^{t+1} is the projection of $x^{t+1} - v$ onto $\mathcal{X}$ and p^t is in $\mathcal{X}$. Now we get a telescoping term $2\langle p^t - p^{t+1}, u \rangle$ and a negative term $-\|p^t - p^{t+1}\|^2$. The negative term is useful for improving the regret analysis in the game setting, but we ignore it for now. Combining the two inequalities above, we have

$$\sum_{t=1}^{T} \left(f^t(x^t) - f^t(p^t) \right) \leq \frac{\delta^2}{2\eta} + \frac{\eta}{2} G^2 T + \frac{1}{\eta} \sum_{t=1}^{T-1} \langle p^t - p^{t+1}, v \rangle$$

$$= \frac{\delta^2}{2\eta} + \frac{\eta}{2} G^2 T + \frac{1}{\eta} \langle p^1 - p^T, v \rangle \leq \frac{\delta^2}{2\eta} + \frac{\eta}{2} G^2 T + \frac{\delta D_{\mathcal{X}}}{\eta}.$$

Since the above holds for any v with $\|v\| \leq \delta$, it also upper bounds $\text{Reg}^T_{\text{Proj},\delta}$. $\qquad\square$

E.3 Lower bounds for $\Phi^{\mathcal{X}}_{\mathrm{Proj}}$-Regret

Theorem 7 (Lower bound for $\Phi^{\mathcal{X}}_{\mathrm{Proj}}(\delta)$-regret against convex losses). *For any $T \geq 1$, $D_{\mathcal{X}} > 0$, $0 < \delta \leq D_{\mathcal{X}}$, and $G \geq 0$, there exists a distribution $\mathcal{D}$ on G-Lipschitz linear loss functions $f^1, \ldots, f^T$ over $\mathcal{X} = [-D_{\mathcal{X}}, D_{\mathcal{X}}]$ such that for any online algorithm, its $\Phi^{\mathcal{X}}_{\mathrm{Proj}}(\delta)$-regret on the loss sequence satisfies $\mathbb{E}_{\mathcal{D}}[\mathrm{Reg}^T_{\mathrm{Proj},\delta}] = \Omega(\delta G \sqrt{T})$.*

Remark 2. *A keen reader may notice that the $\Omega(G\delta\sqrt{T})$ lower bound in Theorem 7 does not match the $O(G\sqrt{\delta D_{\mathcal{X}} T})$ upper bound in Theorem 3, especially when $D_{\mathcal{X}} \gg \delta$. A natural question is:* which of them is tight? *We conjecture that the lower bound is tight. In fact, for the special case where the feasible set $\mathcal{X}$ is a* box, *we obtain a $D_{\mathcal{X}}$-independent bound $O(d^{\frac{1}{4}} G\delta\sqrt{T})$ using a modified version of GD, which is tight when $d = 1$. See Appendix J for a detailed discussion.*

This lower bound suggests that GD achieves near-optimal $\Phi^{\mathcal{X}}_{\mathrm{Proj}}(\delta)$-regret for convex losses. For L-smooth *non-convex* loss functions, we provide another $\Omega(\delta^2 LT)$ lower bound for algorithms that satisfy the linear span assumption. The *linear span* assumption states that the algorithm produces $x^{t+1} \in \{\Pi_{\mathcal{X}}[\sum_{i \in [t]} a_i \cdot x^i + b_i \cdot \nabla f^i(x^i)] : a_i, b_i \in \mathbb{R}, \forall i \in [t]\}$ as essentially the linear combination of the previous iterates and their gradients. Many online algorithms such as online gradient descent and optimistic gradient satisfy the linear span assumption. Combining with Lemma 1, this lower bound suggests that GD attains nearly optimal $\Phi^{\mathcal{X}}_{\mathrm{Proj}}(\delta)$-regret, even in the non-convex setting, among a natural family of gradient-based algorithms.

Proposition 1 (Lower bound for $\Phi^{\mathcal{X}}_{\mathrm{Proj}}(\delta)$-regret against non-convex losses). *For any $T \geq 1$, $\delta \in (0, 1)$, and $L \geq 0$, there exists a sequence of L-Lipschitz and L-smooth non-convex loss functions $f^1, \ldots, f^T$ on $\mathcal{X} = [-1, 1]$ such that for any algorithm that satisfies the linear span assumption, its $\Phi^{\mathcal{X}}_{\mathrm{Proj}}(\delta)$-regret on the loss sequence is $\mathrm{Reg}^T_{\mathrm{Proj},\delta} \geq \frac{\delta^2 LT}{2}$.*

E.3.1 Proof of Theorem 7

Our proof technique comes from the standard one used in multi-armed bandits [Aue+02, Theorem 5.1]. Suppose that $f^t(x) = g^t x$. We construct two possible environments. In the first environment, $g^t = G$ with probability $\frac{1+\varepsilon}{2}$ and $g^t = -G$ with probability $\frac{1-\varepsilon}{2}$; in the second environment, $g^t = G$ with probability $\frac{1-\varepsilon}{2}$ and $g^t = -G$ with probability $\frac{1+\varepsilon}{2}$. We use $\mathbb{E}_i$ and $\mathbb{P}_i$ to denote the expectation and probability measure under environment i, respectively, for $i = 1, 2$. Suppose that the true environment is uniformly chosen from one of these two environments. Below, we show that the expected regret of the learner is at least $\Omega(\delta G \sqrt{T})$.

Define $N_+ = \sum_{t=1}^T \mathbb{I}\{x^t \geq 0\}$ be the number of times x^t is non-negative, and define $f^{1:T} = (f^1, \ldots, f^T)$. Then we have

$$|\mathbb{E}_1[N_+] - \mathbb{E}_2[N_+]| = \left| \sum_{f^{1:T}} \left(\mathbb{P}_1(f^{1:T})\mathbb{E}\left[N_+ \mid f^{1:T}\right] - \mathbb{P}_2(f^{1:T})\mathbb{E}\left[N_+ \mid f^{1:T}\right] \right) \right|$$

(enumerate all possible sequences of $f^{1:T}$)

$$\leq T \sum_{f^{1:T}} \left| \mathbb{P}_1(f^{1:T}) - \mathbb{P}_2(f^{1:T}) \right|$$

$$= T\|\mathbb{P}_1 - \mathbb{P}_2\|_{\mathrm{TV}}$$

$$\leq T\sqrt{(2\ln 2)\mathrm{KL}(\mathbb{P}_1, \mathbb{P}_2)} \qquad \text{(Pinsker's inequality)}$$

$$= T\sqrt{(2\ln 2)T \cdot \mathrm{KL}\left(\mathrm{Bernoulli}\left(\frac{1+\varepsilon}{2}\right), \mathrm{Bernoulli}\left(\frac{1-\varepsilon}{2}\right)\right)}$$

$$= T\sqrt{(2\ln 2)T\varepsilon \ln \frac{1+\varepsilon}{1-\varepsilon}} \leq T\sqrt{(4\ln 2)T\varepsilon^2}. \qquad (2)$$

In the first environment, we consider the regret with respect to $v = \delta$. Then we have

$$
\mathbb{E}_1\left[\operatorname{Reg}_{\mathrm{Proj},\delta}^T\right] \geq \mathbb{E}_1\left[\sum_{t=1}^{T} f^t(x^t) - f^t(\Pi_{\mathcal{X}}[x^t - \delta])\right] = \mathbb{E}_1\left[\sum_{t=1}^{T} g^t(x^t - \Pi_{\mathcal{X}}[x^t - \delta])\right]
$$

$$
= \mathbb{E}_1\left[\sum_{t=1}^{T} \varepsilon G(x^t - \Pi_{\mathcal{X}}[x^t - \delta])\right] \geq \varepsilon\delta G \mathbb{E}_1\left[\sum_{t=1}^{T} \mathbb{I}\{x^t \geq 0\}\right] = \varepsilon\delta G \mathbb{E}_1[N_+],
$$

where in the last inequality we use the fact that if $x^t \geq 0$ then $x^t - \Pi_{\mathcal{X}}[x^t - \delta] = x^t - (x^t - \delta) = \delta$ because $D \geq \delta$. In the second environment, we consider the regret with respect to $v = -\delta$. Then similarly, we have

$$
\mathbb{E}_2\left[\operatorname{Reg}_{\mathrm{Proj},\delta}^T\right] \geq \mathbb{E}_2\left[\sum_{t=1}^{T} f^t(x^t) - f^t(\Pi_{\mathcal{X}}[x^t + \delta])\right] = \mathbb{E}_2\left[\sum_{t=1}^{T} g^t(x^t - \Pi_{\mathcal{X}}[x^t + \delta])\right]
$$

$$
= \mathbb{E}_2\left[\sum_{t=1}^{T} -\varepsilon G(x^t - \Pi_{\mathcal{X}}[x^t + \delta])\right] \geq \varepsilon\delta G \mathbb{E}_2\left[\sum_{t=1}^{T} \mathbb{I}\{x^t < 0\}\right] = \varepsilon\delta G\left(T - \mathbb{E}_2[N_+]\right).
$$

Summing up the two inequalities, we get

$$
\frac{1}{2}\left(\mathbb{E}_1\left[\operatorname{Reg}_{\mathrm{Proj},\delta}^T\right] + \mathbb{E}_2\left[\operatorname{Reg}_{\mathrm{Proj},\delta}^T\right]\right) \geq \frac{1}{2}\left(\varepsilon\delta GT + \varepsilon\delta G(\mathbb{E}_1[N_+] - \mathbb{E}_2[N_+])\right)
$$

$$
\geq \frac{1}{2}\left(\varepsilon\delta GT - \varepsilon\delta GT\varepsilon\sqrt{(4\ln 2)T}\right). \qquad \text{(by (2))}
$$

Choosing $\varepsilon = \frac{1}{\sqrt{(16\ln 2)T}}$, we can lower bound the last expression by $\Omega(\delta G\sqrt{T})$. The theorem is proven by noticing that $\frac{1}{2}\left(\mathbb{E}_1\left[\operatorname{Reg}_{\mathrm{Proj},\delta}^T\right] + \mathbb{E}_2\left[\operatorname{Reg}_{\mathrm{Proj},\delta}^T\right]\right)$ is the expected regret of the learner.

E.3.2 Proof of Proposition 1

Proof. Consider $f : [-1, 1] \to \mathbb{R}$ such that $f(x) = -\frac{L}{2}x^2$ and let $f^t = f$ for all $t \in [T]$. Then any first-order methods that satisfy the linear span assumption with initial point $x^1 = 0$ will produce $x^t = 0$ for all $t \in [T]$. The $\Phi_{\mathrm{Proj}}^{\mathcal{X}}(\delta)$-regret is thus $\sum_{t=1}^{T}(f(0) - f(\delta)) = \frac{\delta^2 LT}{2}$. $\qquad\square$

E.4 Improved $\Phi_{\mathrm{Proj}}^{\mathcal{X}}(\delta)$-Regret in the Game Setting and Proof of Theorem 9

Any online algorithm suffers an $\Omega(\sqrt{T})$ $\Phi_{\mathrm{Proj}}^{\mathcal{X}}(\delta)$-regret even against linear loss functions by Theorem 7. This lower bound, however, holds only in the *adversarial* setting. In this section, we show an improved $O(T^{\frac{1}{4}})$ individual $\Phi_{\mathrm{Proj}}^{\mathcal{X}}(\delta)$-regret bound under a slightly stronger smoothness assumption (Assumption 2) in the *game* setting, where players interact with each other using the same algorithm, previous results show improved external regret [Syr+15; CP20; DFG21; Ana+22a; Ana+22b; Far+22a]. This assumption is naturally satisfied by finite normal-form games and is also made for results about concave games [Far+22a].

Assumption 2. *For any player $i \in [n]$, the utility $u_i(x)$ satisfies $\|\nabla_{x_i} u_i(x) - \nabla_{x_i} u_i(x')\| \leq L\|x - x'\|$ for all $x, x' \in \mathcal{X}$.*

We study the Optimistic Gradient (OG) algorithm [RS13], an optimistic variant of GD that has been shown to have improved individual *external* regret guarantee in the game setting [Syr+15]. The OG algorithm initializes $w^0 \in \mathcal{X}$ arbitrarily and $g^0 = 0$. In each step $t \geq 1$, the algorithm plays x^t, receives feedback $g^t := \nabla f^t(x^t)$, and updates w^t, as follows:

$$
x^t = \Pi_{\mathcal{X}}\left[w^{t-1} - \eta g^{t-1}\right], \quad w^t = \Pi_{\mathcal{X}}\left[w^{t-1} - \eta g^t\right]. \tag{OG}
$$

We show that OG has $O(\sqrt{T})$ $\Phi_{\mathrm{Proj}}^{\mathcal{X}}(\delta)$-regret in the adversarial setting and fast $O(T^{1/4})$ $\Phi_{\mathrm{Proj}}^{\mathcal{X}}(\delta)$-regret and convergence to approximate $\Phi_{\mathrm{Proj}}^{\mathcal{X}}(\delta)$-equilibrium in games.

Theorem 8 (Adversarial Regret Bound for OG). *Let $\delta > 0$ and $T \in \mathbb{N}$. For convex and G-Lipschitz loss functions $\{f^t : \mathcal{X} \to \mathbb{R}\}_{t \in [T]}$, the $\Phi_{\mathrm{Proj}}^{\mathcal{X}}(\delta)$-regret of (OG) with step size $\eta > 0$ is*

$$
\operatorname{Reg}_{\mathrm{Proj},\delta}^T \leq \frac{\delta D_{\mathcal{X}}}{\eta} + \eta \sum_{t=1}^{T}\left\|g^t - g^{t-1}\right\|^2.
$$

Choosing step size $\eta = \frac{\sqrt{\delta D_{\mathcal{X}}}}{2G\sqrt{T}}$, *we have* $\mathrm{Reg}^T_{\mathrm{Proj},\delta} \leq 4G\sqrt{\delta D_{\mathcal{X}} T}$.

Theorem 9 (Improved Individual $\Phi^{\mathcal{X}}_{\mathrm{Proj}}(\delta)$-Regret of OG in the Game Setting)**.** *In a G-Lipschitz L-smooth (in the sense of Assumption 2) game, when all players employ OG with step size $\eta > 0$, then for each player i, $\delta > 0$, and $T \geq 1$, their individual $\Phi^{\mathcal{X}_i}_{\mathrm{Proj}}(\delta)$-regret denoted as $\mathrm{Reg}^{T,i}_{\mathrm{Proj},\delta}$ is $\mathrm{Reg}^{T,i}_{\mathrm{Proj},\delta} \leq \frac{\delta D}{\eta} + \eta G^2 + 3nL^2G^2\eta^3 T$. Choosing $\eta = \min\{(\delta D/(nL^2G^2T))^{\frac{1}{4}}, (\delta D)^{\frac{1}{2}}/G\}$, we have $\mathrm{Reg}^{T,i}_{\mathrm{Proj},\delta} \leq 4(\delta D)^{\frac{3}{4}}(nL^2G^2T)^{\frac{1}{4}} + 2\sqrt{\delta D}G$. Furthermore, for any $\delta > 0$ and any $\varepsilon > 0$, their empirical distribution of played strategy profiles converges to an $(\varepsilon + \frac{\delta^2 L}{2})$-approximate $\Phi_{\mathrm{Proj}}(\delta)$-equilibrium in $O(1/\varepsilon^{\frac{4}{3}})$ iterations.*

E.4.1 Proof of Theorem 8

Proof. Fix any deviation v that is bounded by δ. Let us define $p^0 = w^0$ and $p^t = \Pi_{\mathcal{X}}[x^t - v]$. Following standard analysis of OG [RS13], we have

$$
\sum_{t=1}^T f^t(x^t) - f^t(p^t) \leq \sum_{t=1}^T \langle \nabla f^t(x^t), x^t - p^t \rangle
$$

$$
\leq \sum_{t=1}^T \frac{1}{2\eta}\left(\left\|w^{t-1} - p^t\right\|^2 - \left\|w^t - p^t\right\|^2\right) + \eta\left\|g^t - g^{t-1}\right\|^2 - \frac{1}{2\eta}\left(\left\|x^t - w^t\right\|^2 + \left\|x^t - w^{t-1}\right\|^2\right)
$$

$$
\leq \sum_{t=1}^T \left(\frac{1}{2\eta}\left\|w^{t-1} - p^t\right\|^2 - \frac{1}{2\eta}\left\|w^{t-1} - p^{t-1}\right\|^2 + \eta\left\|g^t - g^{t-1}\right\|^2 - \frac{1}{2\eta}\left\|x^t - w^{t-1}\right\|^2\right) \tag{3}
$$

Now we apply a similar analysis from Theorem 3 to upper bound the term $\left\|w^{t-1} - p^t\right\|^2 - \left\|w^{t-1} - p^{t-1}\right\|^2$:

$$
\begin{aligned}
&\left\|w^{t-1} - p^t\right\|^2 - \left\|w^{t-1} - p^{t-1}\right\|^2 \\
&= \langle p^{t-1} - p^t, 2w^{t-1} - p^{t-1} - p^t \rangle \\
&= \langle p^{t-1} - p^t, 2w^{t-1} - 2p^t \rangle - \left\|p^t - p^{t-1}\right\|^2 \\
&= 2\langle p^{t-1} - p^t, v \rangle + 2\langle p^{t-1} - p^t, w^{t-1} - v - p^t \rangle - \left\|p^t - p^{t-1}\right\|^2 \\
&= 2\langle p^{t-1} - p^t, v \rangle + 2\langle p^{t-1} - p^t, x^t - v - p^t \rangle + 2\langle p^{t-1} - p^t, w^{t-1} - x^t \rangle - \left\|p^t - p^{t-1}\right\|^2 \\
&\leq 2\langle p^{t-1} - p^t, v \rangle + \left\|x^t - w^{t-1}\right\|^2,
\end{aligned}
$$

where in the last-inequality we use $\langle p^{t-1} - p^t, x^t - u - p^t \rangle \leq 0$ since $p^t = \Pi_{\mathcal{X}}[x^t - v]$ and $\mathcal{X}$ is a compact convex set; we also use $2\langle a, b \rangle - b^2 \leq a^2$. In the analysis above, unlike the analysis of GD where we drop the negative term $-\left\|p^t - p^{t-1}\right\|^2$, we use $-\left\|p^t - p^{t-1}\right\|^2$ to get a term $\left\|x^t - w^{t-1}\right\|^2$ which can be canceled by the last term in (3).

Now we combine the above two inequalities. Since the term $\left\|x^t - w^{t-1}\right\|^2$ cancels out and $2\langle p^{t-1} - p^t, v \rangle$ telescopes, we get

$$
\sum_{t=1}^T f^t(x^t) - f^t(p^t) \leq \frac{\langle p^0 - p^T, u \rangle}{\eta} + \sum_{t=1}^T \eta\left\|g^t - g^{t-1}\right\|^2 \leq \frac{\delta D_{\mathcal{X}}}{\eta} + \eta\sum_{t=1}^T \left\|g^t - g^{t-1}\right\|^2.
$$

$\square$

E.4.2 Proof of Theorem 9

In the analysis of Theorem 8 for the adversarial setting, the term $\left\|g^t - g^{t-1}\right\|^2$ can be as large as $4G^2$. In the game setting where every player i employs OG, g_i^t, i.e., $-\nabla_{x_i} u_i(x)$, depends on other players' action x_{-i}^t. Note that the change of the players' actions $\left\|x^t - x^{t-1}\right\|^2$ is only $O(\eta^2)$. Such stability of the updates leads to an improved upper bound on $\left\|g_i^t - g_i^{t-1}\right\|^2$ and hence also an improved $O(T^{\frac{1}{4}})$ $\Phi^{\mathcal{X}}_{\mathrm{Proj}}(\delta)$-regret for the player.

Proof. Let us fix any player $i \in [n]$ in the smooth game. In every step t, player i's loss function $f^t : \mathcal{X}_i \to \mathbb{R}$ is $\langle -\nabla_{x_i} u_i(x^t), \cdot \rangle$ determined by their utility function u_i and all players' actions x^t. Therefore, their gradient feedback is $g^t = -\nabla_{x_i} u_i(x^t)$. For all $t \geq 2$, we have

$$
\begin{aligned}
\left\| g^t - g^{t-1} \right\|^2 &= \left\| \nabla u_i(x^t) - \nabla u_i(x^{t-1}) \right\|^2 \\
&\leq L^2 \left\| x^t - x^{t-1} \right\|^2 \\
&= L^2 \sum_{i=1}^n \left\| x_i^t - x_i^{t-1} \right\|^2 \\
&\leq 3L^2 \sum_{i=1}^n \left(\left\| x_i^t - w_i^t \right\|^2 + \left\| w_i^t - w_i^{t-1} \right\|^2 + \left\| w_i^{t-1} - x_i^{t-1} \right\|^2 \right) \\
&\leq 3nL^2\eta^2 G^2,
\end{aligned}
$$

where we use L-smoothness of the utility function u_i in the first inequality; we use the update rule of OG and the fact that gradients are bounded by G in the last inequality.

Applying the above inequality to the regret bound obtained in Theorem 8, the individual $\Phi_{\mathrm{Proj}}^{\mathcal{X}}(\delta)$-regret of player i is upper bounded by

$$
\mathrm{Reg}_{\mathrm{Proj},\delta}^{T,i} \leq \frac{\delta D}{\eta} + \eta G^2 + 3nL^2 G^2 \eta^3 T.
$$

Choosing $\eta = \min\{(\delta D/(nL^2 G^2 T))^{\frac{1}{4}}, (\delta D)^{\frac{1}{2}}/G\}$, we have $\mathrm{Reg}_{\mathrm{Proj},\delta}^{T,i} \leq 4(\delta D)^{\frac{3}{4}}(nL^2 G^2 T)^{\frac{1}{4}} + 2\sqrt{\delta D} G$. Using Lemma 1, we have the empirical distribution of played strategy profiles converge to an $(\varepsilon + \frac{\delta^2 L}{2})$-approximate $\Phi_{\mathrm{Proj}}(\delta))$-equilibrium in $O(1/\varepsilon^{\frac{4}{3}})$ iterations. $\qquad\square$

F Missing Details in Section 4.2

Algorithm 3: $\mathrm{Conv}(\Phi)$-regret minimization for Lipschitz smooth non-concave rewards

---
Input: $x_1 \in \mathcal{X}$, $K \geq 2$, a no-external-regret algorithm $\mathfrak{R}_\Phi$ against linear reward over $\Delta(\Phi)$
Output: A $\mathrm{Conv}(\Phi)$-regret minimization algorithm over $\mathcal{X}$

1 **function** NEXTSTRATEGY()
2 $\quad$ $p^t \leftarrow \mathfrak{R}_\Phi$.NEXTSTRATEGY(). Note that p_t is a distribution over Φ.
3 $\quad$ $x_k \leftarrow \phi_{p^t}(x_{k-1})$, for all $2 \leq k \leq K$
4 $\quad$ **return** $x^t \leftarrow$ uniformly at random from $\{x_1, \dots, x_K\}$.
5 **function** OBSERVEREWARD($\nabla_x u^t(x^t)$)
6 $\quad$ $u_\Phi^t(\cdot) \leftarrow$ a linear reward over $\Delta(\Phi)$ with $u_\Phi^t(\phi) = \langle \nabla_x u^t(x^t), \phi(x^t) - x^t \rangle$ for all $\phi \in \Phi$.
7 $\quad$ $\mathfrak{R}_\Phi$.OBSERVEREWARD($u_\Phi^t(\cdot)$).

F.1 Proof of Theorem 4

Proof Sketch. We adopt the framework in [SL07; GGM08] (as described in Section 3) with two main modifications. First, we utilize the L-smoothness of the utilities to transform the problem of external regret over $\Delta(\Phi)$ against non-concave rewards into a linear optimization problem. Second, we use the technique of "fixed point in expectation" [Zha+24] to circumvent the intractable problem of finding a fixed point.

Proof. For a sequence of strategies $\{x^t\}_{t \in [T]}$, its $\mathrm{Conv}(\Phi)$-regret is

$$
\begin{aligned}
\mathrm{Reg}_{\mathrm{Conv}(\Phi)}^T &= \max_{\phi \in \mathrm{Conv}(\Phi)} \left\{ \sum_{t=1}^T \left(u^t(\phi(x^t)) - u^t(x^t) \right) \right\} \\
&= \underbrace{\max_{p \in \Delta(\Phi)} \left\{ \sum_{t=1}^T u^t(\phi_p(x^t)) - u^t(\phi_{p^t}(x^t)) \right\}}_{\text{I: external regret over } \Delta(\Phi)} + \underbrace{\sum_{t=1}^T u^t(\phi_{p^t}(x^t)) - u^t(x^t)}_{\text{II: approximation error of fixed point}}.
\end{aligned}
$$

Bounding External Regret over $\Delta(\Phi)$ We can define a new reward function $f^t(p) := u^t(\phi_p(x^t))$ over $p \in \Delta(\Phi)$. Since u^t is non-concave, the reward f^t is also non-concave and it is computational intractable to minimize external regret. We use locality to avoid computational barrier. Here we use the fact that $\Phi = \Phi(\delta)$ contains only δ-local strategy modifications. Then by L-smoothness of u^t, we know for any $p \in \Delta(\Phi)$

$$\left| u^t(\phi_p(x^t)) - u^t(x^t) - \langle \nabla u^t(x^t), \phi_p(x^t) - x^t \rangle \right| \leq \frac{L}{2} \left\| \phi_p(x^t) - x^t \right\|^2 \leq \frac{\delta^2 L}{2}.$$

Thus we can approximate the non-concave optimization problem by a linear optimization problem over $\Delta(\Phi)$ with only second-order error $\frac{\delta^2 L}{2}$. Here we use the notation $a = b \pm c$ to mean $b - c \leq a \leq b + c$.

$$u^t(\phi_p(x^t)) - u^t(x^t) = \langle \nabla u^t(x^t), \phi_p(x^t) - x^t \rangle \pm \frac{\delta^2 L}{2}$$

$$= \left\langle \nabla u^t(x^t), \sum_{\phi \in \Phi} p(\phi)\phi(x^t) - x^t \right\rangle \pm \frac{\delta^2 L}{2}$$

$$= \sum_{\phi \in \Phi} p(\phi) \langle \nabla u^t(x^t), \phi(x^t) - x^t \rangle \pm \frac{\delta^2 L}{2}.$$

We can then instantiate the external regret $\mathfrak{R}_\Phi$ as the Hedge algorithm over reward $f^t(p) = \sum_{\phi \in \Phi} p(\phi) \langle \nabla u^t(x^t), \phi(x^t) - x^t \rangle$ and get

$$\max_{p \in \Delta(\Phi)} \left\{ \sum_{t=1}^{T} u^t(\phi_p(x^t)) - u^t(\phi_{p^t}(x^t)) \right\}$$

$$\leq \max_{p \in \Delta(\Phi)} \left\{ \sum_{t=1}^{T} \sum_{\phi \in \Phi} (p(\phi) - p^t(\phi)) \langle \nabla u^t(x^t), \phi(x^t) - x^t \rangle \right\} + \delta^2 LT$$

$$\leq 2G\delta\sqrt{T \log |\Phi|} + \delta^2 LT,$$

where we use the fact that $\langle \nabla u^t(x^t), \phi(x^t) - x^t \rangle \leq \|\nabla u^t(x^t)\| \cdot \|\phi(x^t) - x^t\| \leq G\delta$.

Bounding error due to sampling from a fixed point in expectation   We choose x_1 as an arbitrary point in $\mathcal{X}$. Then we recursively apply ϕ_{p^t} to get

$$x_k = \phi_{p^t}(x_{k-1}) = \sum_{\phi \in \Phi} p^t(\phi)\phi(x_{k-1}), \forall 2 \leq k \leq K.$$

We denote $\mu^t = \text{Uniform}\{x_k : 1 \leq k \leq K\}$. Then the strategy $x^t \sim \mu^t$ is sampled from μ^t. We have that μ^t is an approximate fixed-point in expectation / stationary distribution in the sense that

$$\mathbb{E}_{\mu^t}\left[u^t(\phi_{p^t}(x^t)) - u^t(x^t) \right] = \frac{1}{K} \sum_{k=1}^{K} u^t(\phi_{p^t}(x_k) - u^t(x_k))$$

$$= \frac{1}{K}\left(u^t(\phi_{p^t}(x_K)) - u^t(x_1) \right)$$

$$\leq \frac{1}{K}.$$

Thanks to the boundedness of u^t, we can use Hoeffding-Azuma's inequality to conclude that

$$\Pr\left[\sum_{t=1}^{T}\left(u^t(\phi_{p^t}(x^t)) - u^t(x^t) - \frac{1}{K} \right) \geq \epsilon \right] \leq \exp\left(-\frac{\epsilon^2}{8T} \right). \tag{4}$$

for any $\varepsilon > 0$. Combining the above with $\varepsilon = \sqrt{8T \log(1/\beta)}$ and $K = \sqrt{T}$, we get with probability at least $1 - \beta$,

$$\text{Reg}^T_{\text{Conv}(\Phi)} \leq 2G\delta\sqrt{T \log |\Phi|} + \delta^2 LT + \sqrt{T} + \sqrt{8T \log(1/\beta)}$$

$$\leq 8\sqrt{T}\left(G\delta\sqrt{\log |\Phi|} + \sqrt{\log(1/\beta)} \right) + \delta^2 LT.$$

Convergence to Φ-equilibrium   If all players in a non-concave continuous game employ Algorithm 1, then we know for each player i, with probability $1 - \frac{\beta}{n}$, its $\Phi^{\mathcal{X}_i}$-regret is upper bounded by

$$8\sqrt{T}\left(G\delta\sqrt{\log|\Phi^{\mathcal{X}_i}|} + \sqrt{\log(n/\beta)}\right) + \delta^2 LT.$$

By a union bound over all n players, we get with probability $1 - \beta$, every player i's $\Phi^{\mathcal{X}_i}$-regret is upper bounded by $8\sqrt{T}(G\delta\sqrt{\log|\Phi^{+\mathcal{X}_i}|} + \sqrt{\log(n/\beta)}) + \delta^2 LT$. Now by Theorem 1, we know the empirical distribution of strategy profiles played forms an $(\varepsilon + \delta^2 L)$-approximate $\Phi = \Pi_{i=1}^n \Phi^{\mathcal{X}_i}$-equilibrium, as long as $T \geq \frac{128(G^2\delta^2 \log|\Phi^{\mathcal{X}_i}| + \log(n/\beta))}{\varepsilon^2}$ iterations. $\qquad\square$

G   Missing details in Section 4.3

We introduce a natural set of local strategy modifications and the corresponding local equilibrium notion. Given any set of (possibly non-local) strategy modifications $\Psi = \{\psi : \mathcal{X} \to \mathcal{X}\}$, we define a set of *local* strategy modifications as follows: for $\delta \leq D_{\mathcal{X}}$ and $\lambda \in [0, 1]$, each strategy modification $\phi_{\lambda,\psi}$ interpolates the input strategy x with the modified strategy $\psi(x)$: formally,

$$\Phi_{\text{Int},\Psi}^{\mathcal{X}}(\delta) := \{\phi_{\lambda,\psi}(x) := (1 - \lambda)x + \lambda\psi(x) : \psi \in \Psi, \lambda \leq \delta/D_{\mathcal{X}}\}.$$

Note that for any $\psi \in \Psi$ and $\lambda \leq \frac{\delta}{D_{\mathcal{X}}}$, we have $\|\phi_{\lambda,\psi}(x) - x\| = \lambda\|x - \psi(x)\| \leq \delta$, respecting the locality constraint. The induced $\Phi_{\text{Int},\Psi}^{\mathcal{X}}(\delta)$-regret can be written as $\text{Reg}_{\text{Int},\Psi,\delta}^T :=$ $\max_{\psi \in \Psi, \lambda \leq \frac{\delta}{D_{\mathcal{X}}}} \sum_{t=1}^T (f^t(x^t) - f^t((1 - \lambda)x^t + \lambda\psi(x^t)))$. We now define the corresponding $\Phi_{\text{Int},\Psi}(\delta)$-equilibrium.

Definition 9. *Define $\Phi_{\text{Int},\Psi}(\delta) = \Pi_{j=1}^n \Phi_{\text{Int},\Psi_j}^{\mathcal{X}_j}(\delta)$. In a continuous game, a distribution σ over strategy profiles is an (ε-approximate $\Phi_{\text{Int},\Psi}(\delta)$-equilibrium if and only if for all player $i \in [n]$,*

$$\max_{\psi \in \Psi_i, \lambda \leq \delta/D_{\mathcal{X}_i}} \mathbb{E}_{x \sim \sigma}[u_i((1 - \lambda)x_i + \lambda\psi(x_i), x_{-i})] \leq \mathbb{E}_{x \sim \sigma}[u_i(x)] + \varepsilon.$$

Intuitively speaking, when a correlation device recommends strategies to players according to an ε-approximate $\Phi_{\text{Int},\Psi}(\delta)$-equilibrium, no player can increase their utility by more than ε through a local deviation by interpolating with a (possibly global) strategy modification $\psi \in \Psi$. The richness of Ψ determines the incentive guarantee provided by an ε-approximate $\Phi_{\text{Int},\Psi}(\delta)$-equilibriumas well as its computational complexity. When we choose Ψ to be the set of all possible strategy modifications, the corresponding notion of local equilibrium—limiting the gain of a player by interpolating with any strategy—resembles that of a *correlated equilibrium*.

Computation of ε-approximate $\Phi_{\text{Int},\Psi}(\delta)$-Equilibrium.   By Lemma 1, we know computing an ε-approximate $\Phi_{\text{Int},\Psi}(\delta)$-equilibrium reduces to minimizing $\Phi_{\text{Int},\Psi}^{\mathcal{X}}(\delta)$-regret against convex loss functions. We show that minimizing $\Phi_{\text{Int},\Psi}^{\mathcal{X}}(\delta)$-regret against convex loss functions further reduces to Ψ-regret minimization against linear loss functions.
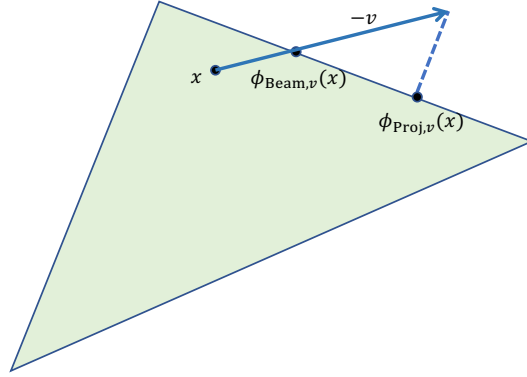
Theorem 10. *Let $\mathcal{A}$ be an algorithm with Ψ-regret $\text{Reg}_{\Psi}^T(G, D_{\mathcal{X}})$ for linear and G-Lipschitz loss functions over $\mathcal{X}$. Then, for any $\delta > 0$, the $\Phi_{\text{Int},\Psi}^{\mathcal{X}}(\delta)$-regret of $\mathcal{A}$ for convex and G-Lipschitz loss functions over $\mathcal{X}$ is at most $\frac{\delta}{D_{\mathcal{X}}} \cdot [\text{Reg}_{\Psi}^T(G, D_{\mathcal{X}})]^+$.*

Proof. By definition and convexity of f^t, we get

$$\max_{\phi \in \Phi_{\text{Int},\Psi}^{\mathcal{X}}(\delta)} \sum_{t=1}^T f^t(x^t) - f^t(\phi(x^t)) = \max_{\psi \in \Psi, \lambda \leq \frac{\delta}{D_{\mathcal{X}}}} \sum_{t=1}^T f^t(x^t) - f^t((1 - \lambda)x^t + \lambda\psi(x^t))$$

$$\leq \frac{\delta}{D_{\mathcal{X}}}\left[\max_{\psi \in \Psi} \sum_{t=1}^T \langle\nabla f^t(x^t), x^t - \psi(x^t)\rangle\right]^+.$$

$\qquad\square$

Figure 2: Illustration of $\phi_{\mathrm{Proj},v}(x)$ and $\phi_{\mathrm{Beam},v}(x)$

Note that when f^t is linear, the reduction is without loss. Thus, any worst-case $\Omega(r(T))$-lower bound for Ψ-regret implies a $\Omega(\frac{\delta}{D_{\mathcal{X}}} \cdot r(T))$ lower bound for $\Phi_{\mathrm{Int},\Psi}(\delta)$-regret. Moreover, for any set Ψ that admits efficient Ψ-regret minimization algorithms such as swap transformations over the simplex and more generally any set such that (i) all modifications in the set can be represented as linear transformations in some finite-dimensional space and (ii) fixed point computation can be carried out efficiently for any linear transformations [GGM08], we also get an efficient algorithm for computing an ε-approximate $\Phi_{\mathrm{Int},\Psi}(\delta)$-equilibrium in the first-order stationary regime.

CCE-like Instantiation In the special case where Ψ contains only *constant* strategy modifications (i.e. $\psi(x) = x^*$ for all x), we get a coarse correlated equilibrium (CCE)-like instantiation of local equilibrium, which limits the gain by interpolating with any *fixed* strategy. We denote the resulting set of local strategy modification simply as $\Phi_{\mathrm{Int}}^{\mathcal{X}}$. We can apply any no-external regret algorithm for efficient $\Phi_{\mathrm{Int}}^{\mathcal{X}}$-regret minimization and computation of ε-approximate $\Phi_{\mathrm{Int}}(\delta)$-equilibrium in the first-order stationary regime as summarized in Theorem 5.

The above $\Phi_{\mathrm{Int}}^{\mathcal{X}}(\delta)$-regret bound of $O(\sqrt{T})$ is derived for the adversarial setting. In the game setting, where each player employs the same algorithm, players may have substantially lower external regret [Syr+15; CP20; DFG21; Ana+22a; Ana+22b; Far+22a] but we need a slightly stronger smoothness assumption than Assumption 1. This assumption is naturally satisfied by finite normal-form games and is also made for results about concave games [Far+22a]. Using Assumption 2 and Lemma 1, the no-regret learning dynamics of [Far+22a] that guarantees $O(\log T)$ individual external regret in concave games can be applied to smooth non-concave games so that the individual $\Phi_{\mathrm{Int}}^{\mathcal{X}}(\delta)$-regret of each player is at most $O(\log T) + \frac{\delta^2 L T}{2}$. This gives an algorithm with faster $\tilde{O}(1/\varepsilon)$ convergence to an $(\varepsilon + \frac{\delta^2 L}{2})$-approximate $\Phi_{\mathrm{Int}}(\delta)$-equilibrium than GD.

H Beam-Search Local Strategy Modifications and Local Equilibria

In Section 4.1 and Section 4.3, we have shown that GD achieves near-optimal performance for both $\Phi_{\mathrm{Int}}^{\mathcal{X}}(\delta)$-regret and $\Phi_{\mathrm{Proj}}^{\mathcal{X}}(\delta)$-regret. In this section, we introduce another natural set of local strategy modifications, $\Phi_{\mathrm{Beam}}^{\mathcal{X}}(\delta)$, which is similar to $\Phi_{\mathrm{Proj}}^{\mathcal{X}}(\delta)$. Specifically, the set $\Phi_{\mathrm{Beam}}^{\mathcal{X}}(\delta)$ contains deviations that try to move as far as possible in a fixed direction (see Figure 2 for an illustration of the difference between $\phi_{\mathrm{Beam},v}(x)$ and $\phi_{\mathrm{Proj},v}(x)$):

$$\Phi_{\mathrm{Beam}}^{\mathcal{X}}(\delta) := \{\phi_{\mathrm{Beam},v}(x) = x - \lambda^* v : v \in B_d(\delta), \lambda^* = \max\{\lambda : x - \lambda v \in \mathcal{X}, \lambda \in [0,1]\}\}.$$

It is clear that $\|\phi_{\mathrm{Beam},v}(x) - x\| \le \|v\| \le \delta$. We can similarly derive the notion of $\Phi_{\mathrm{Beam}}^{\mathcal{X}}$-regret and $(\varepsilon, \Phi_{\mathrm{Beam}}(\delta))$-equilibrium. Surprisingly, we show that GD suffers linear $\Phi_{\mathrm{Beam}}^{\mathcal{X}}(\delta)$-regret (proof deferred to Appendix H.1).

Theorem 11. *For any $\delta, \eta < \frac{1}{2}$ and $T \ge 1$, there exists a sequence of linear loss functions $\{f^t : \mathcal{X} \subseteq [0,1]^2 \to \mathbb{R}\}_{t \in [T]}$ such that GD with step size η suffers $\Omega(\delta T)$ $\Phi_{\mathrm{Beam}}^{\mathcal{X}}(\delta)$-regret.*

H.1 Proof of Theorem 11

Let $\mathcal{X} \subset \mathbb{R}^2$ be a triangle region with vertices $A = (0,0)$, $B = (1,1)$, $C = (\delta,0)$. Consider $v = (-\delta, 0)$. The initial point is $x_1 = (0,0)$.

The adversary will choose ℓ_t adaptively so that x_t remains on the boundary of $\mathcal{X}$ and cycles clockwise (i.e., $A \to \cdots \to B \to \cdots \to C \to \cdots \to A \to \cdots$). To achieve this, the adversary will repeat the following three phases:

1. Keep choosing $\ell_t = u_{\overrightarrow{BA}}$ ($u_{\overrightarrow{BA}}$ denotes the unit vector in the direction of $\overrightarrow{BA}$) until x_{t+1} reaches B.
2. Keep choosing $\ell_t = u_{\overrightarrow{CB}}$ until x_{t+1} reaches C.
3. Keep choosing $\ell_t = u_{\overrightarrow{AC}}$ until x_{t+1} reaches A.

In Phase 1, $x_t \in \overline{AB}$. By the choice of $v = (-\delta, 0)$, we have $x_t - \phi_v(x_t) = (-\delta(1 - x_{t,1}), 0)$, and the instantaneous regret is $\frac{\delta(1 - x_{t,1})}{\sqrt{2}} \geq 0$.

In Phase 2, $x_t \in \overline{BC}$. By the choice of $v = (-\delta, 0)$, we have $x_t - \phi_v(x_t) = (0,0)$, and the instantaneous regret is 0.

In Phase 3, $x_t \in \overline{CA}$. By the choice of $v = (-\delta, 0)$, we have $x_t - \phi_v(x_t) = (-\delta + x_{t,1}, 0)$, and the instantaneous regret is $-\delta + x_{t,1} \leq 0$.

In each cycle, the number of rounds in Phase 1 is of order $\Theta(\frac{\sqrt{2}}{\eta})$, the number of rounds in Phase 2 is between $O(\frac{1}{\eta})$ and $O(\frac{\sqrt{2}}{\eta})$, the number of rounds in Phase 3 is of order $\Theta(\frac{\delta}{\eta})$.

Therefore, the cumulative regret in each cycle is roughly

$$\frac{\sqrt{2}}{\eta} \times \frac{0.5\delta}{\sqrt{2}} + 0 + \frac{\delta}{\eta}(-0.5\delta) = \frac{0.5\delta - 0.5\delta^2}{\eta}.$$

On the other hand, the number of cycles is no less than $\frac{T}{\frac{\sqrt{2}}{\eta} + \frac{\sqrt{2}}{\eta} + \frac{\delta}{\eta}} = \Theta(\eta T)$. Overall, the cumulative regret is at least $\frac{0.5\delta - 0.5\delta^2}{\eta} \times \Theta(\eta T) = \Theta(\delta T)$ as long as $\delta < 0.5$.

I Hardness in the Global Regime

In the first-order stationary regime $\delta \leq \sqrt{2\varepsilon/L}$, (ε, δ)-local Nash equilibrium is intractable, and we have shown polynomial-time algorithms for computing the weaker notions of ε-approximate $\Phi_{\mathrm{Int}}(\delta))$-equilibrium and ε-approximate $\Phi_{\mathrm{Proj}}(\delta))$-equilibrium. A natural question is whether correlation enables efficient computation of ε-approximate $\Phi(\delta))$-equilibrium when δ is in the global regime, i.e., $\delta = \Omega(\sqrt{d})$. In this section, we prove both computational hardness and a query complexity lower bound for both notions in the global regime

To prove the lower bound results, we only require a single-player game. The problem of computing an ε-approximate $\Phi(\delta)$-equilibrium becomes: given scalars $\varepsilon, \delta, G, L > 0$ and a polynomial-time Turing machine $\mathcal{C}_f$ evaluating a G-Lipschitz and L-smooth function $f : [0,1]^d \to [0,1]$ and its gradient $\nabla f : [0,1]^d \to \mathbb{R}^d$, we are asked to output a distribution σ that is an ε-approximate $\Phi(\delta)$-equilibrium or $\perp$ if such equilibrium does not exist.

Hardness of finding ε-approximate $\Phi_{\mathrm{Int}}^{\mathcal{X}}(\delta)$-equilibria in the global regime   When $\delta = \sqrt{d}$, which equals to the diameter D of $[0,1]^d$, then the problem of finding an ε-approximate $\Phi_{\mathrm{Int}}^{\mathcal{X}}(\delta)$-equilibrium is equivalent to finding a (ε, δ)-local minimum of f: assume σ is an ε-approximate $\Phi_{\mathrm{Int}}^{\mathcal{X}}(\delta)$-equilibrium of f, then there exists $x \in [0,1]^d$ in the support of σ such that

$$f(x) - \min_{x^* \in [0,1]^d \cap B_d(x^*, \delta)} f(x^*) \leq \varepsilon.$$

Then hardness of finding an ε-approximate $\Phi_{\mathrm{Int}}^{\mathcal{X}}(\delta)$-equilibrium follows from hardness of finding a (ε, δ)-local minimum of f [DSZ21]. The following Theorem is a corollary of Theorem 10.3 and 10.4 in [DSZ21].

Theorem 12 (Hardness of finding ε-approximate $\Phi_{\mathrm{Int}}^{\mathcal{X}}(\delta)$-equilibria in the global regime). *In the worst case, the following two holds.*

- *Computing an ε-approximate $\Phi_{\mathrm{Int}}^{\mathcal{X}}(\delta)$-equilibrium for a game on $\mathcal{X} = [0,1]^d$ with $G = \sqrt{d}$, $L = d$, $\varepsilon \le \frac{1}{24}$, $\delta = \sqrt{d}$ is NP-hard.*

- *$\Omega(2^d/d)$ value/gradient queries are needed to determine an ε-approximate $\Phi_{\mathrm{Int}}^{\mathcal{X}}(\delta)$-equilibrium for a game on $\mathcal{X} = [0,1]^d$ with $G = \Theta(d^{15})$, $L = \Theta(d^{22})$, $\varepsilon < 1$, $\delta = \sqrt{d}$.*

Hardness of finding ε-approximate $\Phi_{\mathrm{Proj}}^{\mathcal{X}}(\delta)$-equilibria in the global regime

Theorem 13 (Hardness of of finding ε-approximate $\Phi_{\mathrm{Proj}}^{\mathcal{X}}(\delta)$-equilibria in the global regime). *In the worst case, the following two holds.*

- *Computing an ε-approximate $\Phi_{\mathrm{Proj}}^{\mathcal{X}}(\delta)$-equilibrium for a game on $\mathcal{X} = [0,1]^d$ with $G = \Theta(d^{15})$, $L = \Theta(d^{22})$, $\varepsilon < 1$, $\delta = \sqrt{d}$ is NP-hard.*

- *$\Omega(2^d/d)$ value/gradient queries are needed to determine an ε-approximate $\Phi_{\mathrm{Proj}}^{\mathcal{X}}(\delta)$-equilibrium for a game on $\mathcal{X} = [0,1]^d$ with $G = \Theta(d^{15})$, $L = \Theta(d^{22})$, $\varepsilon < 1$, $\delta = \sqrt{d}$.*

The hardness of computing ε-approximate $\Phi_{\mathrm{Proj}}^{\mathcal{X}}(\delta)$-equilibrium also implies a lower bound on $\Phi_{\mathrm{Proj}}^{\mathcal{X}}(\delta)$-regret in the global regime.

Corollary 2 (Lower bound of $\Phi_{\mathrm{Proj}}^{\mathcal{X}}(\delta)$-regret against non-convex functions). *In the worst case, the $\Phi_{\mathrm{Proj}}^{\mathcal{X}}(\delta)$-regret of any online algorithm is at least $\Omega(2^d/d, T)$ even for loss functions $f : [0,1]^d \to [0,1]$ with $G, L = \mathrm{poly}(d)$ and $\delta = \sqrt{d}$.*

The proofs of Theorem 13 and Corollary 2 can be found in the next two sections.

I.1 Proof of Theorem 13

We will reduce the problem of finding an ε-approximate $\Phi_{\mathrm{Proj}}^{\mathcal{X}}(\delta)$-equilibrium in smooth games to finding a satisfying assignment of a boolean function, which is NP-complete.

Fact 1. *Given only* black-box access *to a boolean formula $\phi : \{0,1\}^d \to \{0,1\}$, at least $\Omega(2^d)$ queries are needed in order to determine whether ϕ admits a satisfying assignment x^* such that $\phi(x^*) = 1$. The term* black-box access *refers to the fact that the clauses of the formula are not given, and the only way to determine whether a specific boolean assignment is satisfying is by querying the specific binary string. Moreover, the problem of finding a satisfying assignment of a general boolean function is NP-hard.*

We revisit the construction of the hard instance in the proof of [DSZ21, Theorem 10.4] and use its specific structures. Given black-box access to a boolean formula ϕ as described in Fact 1, following [DSZ21], we construct the function $f_\phi(x) : [0,1]^d \to [0,1]$ as follows:

1. for each corner $v \in V = \{0,1\}^d$ of the $[0,1]^d$ hypercube, we set $f_\phi(x) = 1 - \phi(x)$.

2. for the rest of the points $x \in [0,1]^d/V$, we set $f_\phi(x) = \sum_{v \in V} P_v(x) \cdot f_\phi(v)$ where $P_v(x)$ are non-negative coefficients defined in [DSZ21, Definition 8.9].

The function f_ϕ satisfies the following properties:

1. if ϕ is not satisfiable, then $f_\phi(x) = 1$ for all $x \in [0,1]^d$ since $f_\phi(v) = 1$ for all $v \in V$; if ϕ has a satisfying assignment v^*, then $f_\phi(v^*) = 0$.

2. f_ϕ is $\Theta(d^{12})$-Lipschitz and $\Theta(d^{25})$-smooth.

3. for any point $x \in [0,1]^d$, the set $V(x) := \{v \in V : P_v(x) \ne 0\}$ has cardinality at most $d + 1$ while $\sum_{v \in V} P_v(x) = 1$; any value / gradient query of f_ϕ can be simulated by $d + 1$ queries on ϕ.

29

In the case there exists a satisfying argument v^*, then $f_\phi(v^*) = 0$. Define the deviation e so that $e[i] = 1$ if $v^*[i] = 0$ and $e[i] = -1$ if $v^*[i] = 1$. It is clear that $\|e\| = \sqrt{d} = \delta$. By properties of projection on $[0,1]^d$, for any $x \in [0,1]^d$, we have $\Pi_{[0,1]^n}[x - v] = v^*$. Then any ε-approximate $\Phi^{\mathcal{X}}_{\text{Proj}}(\delta)$-equilibrium σ must include some $x^* \in \mathcal{X}$ with $f_\phi(x^*) < 1$ in the support, since $\varepsilon < 1$. In case there exists an algorithm $\mathcal{A}$ that computes an ε-approximate $\Phi^{\mathcal{X}}_{\text{Proj}}(\delta)$-equilibrium, $\mathcal{A}$ must have queried some x^* with $f_\phi(x^*) < 1$. Since $f_\phi(x^*) = \sum_{v \in V(x^*)} P_v(x^*) f_\phi(v) < 1$, there exists $\hat{v} \in V(x^*)$ such that $f_\phi(\hat{v}) = 0$. Since $|V(x^*)| \le d+1$, it takes addition $d+1$ queries to find $\hat{v}$ with $f_\phi(\hat{v}) = 0$. By Fact 1 and the fact that we can simulate every value/gradient query of f_ϕ by $d+1$ queries on ϕ, $\mathcal{A}$ makes at least $\Omega(2^d/d)$ value/gradient queries.

Suppose there exists an algorithm $\mathcal{B}$ that outputs an ε-approximate $\Phi^{\mathcal{X}}_{\text{Proj}}(\delta)$-equilibrium σ in time $T(\mathcal{B})$ for $\varepsilon < 1$ and $\delta = \sqrt{d}$. We construct another algorithm $\mathcal{C}$ for SAT that terminates in time $T(\mathcal{B}) \cdot \text{poly}(d)$. $\mathcal{C}$: (1) given a boolean formula ϕ, construct f_ϕ as described above; (2) run $\mathcal{B}$ and get output σ (3) check the support of σ to find $v \in \{0,1\}^d$ such that $f_\phi(v) = 0$; (3) if finds $v \in \{0,1\}^d$ such that $f_\phi(v) = 0$, then ϕ is satisfiable, otherwise ϕ is not satisfiable. Since we can evaluate f_ϕ and ∇f_ϕ in $\text{poly}(d)$ time and the support of σ is smaller than $T(\mathcal{B})$, the algorithm $\mathcal{C}$ terminates in time $O(T(\mathcal{B}) \cdot \text{poly}(d))$. The above gives a polynomial time reduction from SAT to finding an ε-approximate $\Phi^{\mathcal{X}}_{\text{Proj}}(\delta)$-equilibrium and proves the NP-hardness of the latter problem.

I.2 Proof of Corollary 2

Let $\phi : \{0,1\}^d \to \{0,1\}$ be a boolean formula and define $f_\phi : [0,1]^d \to [0,1]$ the same as that in Theorem 13. We know f_ϕ is $\Theta(\text{poly}(d))$-Lipschitz and $\Theta(\text{poly}(d))$-smooth. Now we let the adversary pick f_ϕ each time. For any $T \le O(2^d/d)$, in case there exists an online learning algorithm with $\text{Reg}^T_{\text{Proj},\delta} < \frac{T}{2}$, then $\sigma := \frac{1}{T} \sum_{t=1}^T 1_{x^t}$ is an $(\frac{1}{2}, \delta)$-equilibrium. Applying Theorem 13 and the fact that in this case, $\text{Reg}^T_{\text{Proj},\delta}$ is non-decreasing with respect to T concludes the proof.

J Removing the D dependence for $\Phi^{\mathcal{X}}_{\text{Proj}}$-regret

For the regime $\delta \le D_{\mathcal{X}}$ which we are more interested in, the lower bound in Theorem 7 is $\Omega(G\delta\sqrt{T})$ while the upper bound in Theorem 3 is $O(G\sqrt{\delta D_{\mathcal{X}} T})$. They are not tight especially when $D_{\mathcal{X}} \gg \delta$. A natural question is: *which of them is the tight bound?* We conjecture that the lower bound is tight. In fact, for the special case where the feasible set $\mathcal{X}$ is a *box*, we have a way to obtain a $D_{\mathcal{X}}$-independent bound $O(d^{\frac{1}{4}} G\delta\sqrt{T})$, which is tight when $d = 1$. Below, we first describe the improved strategy in 1-dimension. Then we show how to extend it to the d-dimensional box setting.

J.1 One-Dimensional Case

In one-dimension, we assume that $\mathcal{X} = [a, b]$ for some $b - a \ge 2\delta$ (if $b - a \le 2\delta$, then our original bound in Theorem 3 is already of order $G\delta\sqrt{T}$). We first investigate the case where $f^t(x)$ is a linear function, i.e., $f^t(x) = g^t x$ for some $g^t \in [-G, G]$. The key idea is that we will only select x^t from the two intervals $[a, a + \delta]$ and $[b - \delta, b]$, and never play $x^t \in (a + \delta, b - \delta)$. To achieve so, we concatenate these two intervals, and run an algorithm in this region whose diameter is only 2δ. The key property we would like to show is that the regret is preserved in this modified problem.

More precisely, given the original feasible set $\mathcal{X} = [a, b]$, we create a new feasible set $\mathcal{Y} = [-\delta, \delta]$ and apply our algorithm GD in this new feasible set. The loss function is kept as $f^t(x) = g^t x$. Whenever the algorithm for $\mathcal{Y}$ outputs $y^t \in [-\delta, 0]$, we play $x^t = y^t + a + \delta$ in $\mathcal{X}$; whenever it outputs $y^t \in (0, \delta]$, we play $x^t = y^t + b - \delta$. Below we show that the regret is the same in these two

problems. Notice that when $y^t \le 0$, we have for any $v \in [-\delta, \delta]$,

$$
\begin{aligned}
x^t - \Pi_{\mathcal{X}}[x^t - v] &= x^t - \max\left(\min\left(x^t - v, b\right), a\right) \\
&= x^t - \max\left(x^t - v, a\right) \\
&\qquad\qquad\qquad\qquad (x^t - v = y^t + a + \delta - v \le a + 2\delta \le b \text{ always holds}) \\
&= y^t + a + \delta - \max\left(y^t + a + \delta - v, a\right) \\
&= y^t - \max\left(y^t - v, -\delta\right) \\
&= y^t - \max\left(\min\left(y^t - v, \delta\right), -\delta\right) \qquad (y^t - v \le \delta \text{ always holds}) \\
&= y^t - \Pi_{\mathcal{Y}}[y^t - v]
\end{aligned}
$$

Similarly, when $y^t > 0$, we can follow the same calculation and prove $x^t - \Pi_{\mathcal{X}}[x^t - v] = y^t - \Pi_{\mathcal{Y}}[y^t - v]$. Thus, the regret in the two problems:

$$
g^t\left(x^t - \Pi_{\mathcal{X}}[x^t - v]\right) \quad \text{and} \quad g^t\left(y^t - \Pi_{\mathcal{Y}}[y^t - v]\right)
$$

are exactly the same for any v. Finally, observe that the diameter of $\mathcal{Y}$ is only of order $O(\delta)$. Thus, the upper bound in Theorem 3 would give us an upper bound of $O(G\sqrt{\delta \cdot \delta T}) = O(G\delta\sqrt{T})$.

For convex f^t, we run the algorithm above with $g^t = \nabla f^t(x^t)$. Then by convexity we have

$$
f^t(x^t) - f^t(\Pi_{\mathcal{X}}[x^t - v]) \le g^t(x^t - \Pi_{\mathcal{X}}[x^t - v]) = g^t(y^t - \Pi_{\mathcal{Y}}[y^t - v]),
$$

so the regret in the modified problem (which is $O(G\delta\sqrt{T})$) still serves as a regret upper bound for the original problem.

J.2 d-Dimensional Box Case

A d-dimensional box is of the form $\mathcal{X} = [a_1, b_1] \times [a_2, b_2] \times \cdots \times [a_d, b_d]$. The box case is easy to deal with because we can decompose the regret into individual components in each dimension. Namely, we have

$$
\begin{aligned}
f^t(x^t) - f^t(\Pi_{\mathcal{X}}[x^t - v]) &\le \nabla f^t(x^t)^\top \left(x^t - \Pi_{\mathcal{X}}[x^t - v]\right) \\
&= \sum_{i=1}^{d} g_i^t\left(x_i^t - \Pi_{\mathcal{X}_i}[x_i^t - v_i]\right)
\end{aligned}
$$

where we define $\mathcal{X}_i = [a_i, b_i]$, $g^t = \nabla f^t(x^t)$, and use subscript i to indicate the i-th component of a vector. The last equality above is guaranteed by the box structure. This decomposition allows as to view the problem as d independent 1-dimensional problems.

Now we follow the strategy described in Section J.1 to deal with individual dimensions (if $b_i - a_i < 2\delta$ then we do not modify $\mathcal{X}_i$; otherwise, we shrink $\mathcal{X}_i$ to be of length 2δ). Applying the analysis of Theorem 3 to each dimension, we get

$$
\begin{aligned}
&\sum_{i=1}^{d} g_i^t\left(x_i^t - \Pi_{\mathcal{X}_i}[x_i^t - v_i]\right) \\
&\le \sum_{i=1}^{d}\left(\frac{v_i^2}{2\eta} + \frac{\eta}{2}\sum_{t=1}^{T}(g_i^t)^2 + \frac{|v_i| \times 2\delta}{\eta}\right) \quad \text{(the diameter in each dimension is now bounded by } 2\delta) \\
&\le O\left(\frac{\delta \sum_{i=1}^{d}|v_i|}{\eta} + \eta G^2 T\right) \\
&\le O\left(\frac{\delta^2 \sqrt{d}}{\eta} + \eta G^2 T\right). \qquad\qquad \text{(by Cauchy-Schwarz, } \sum_i |v_i| \le \sqrt{d}\sqrt{\sum_i |v_i|^2} \le \delta\sqrt{d})
\end{aligned}
$$

Choosing the optimal $\eta = \frac{d^{\frac{1}{4}}\delta}{G\sqrt{T}}$, we get the regret upper bound of order $O\left(d^{\frac{1}{4}}G\delta\sqrt{T}\right)$.

NeurIPS Paper Checklist

1. **Claims**

   Question: Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope?

   Answer: [Yes]

   Justification: We state the problem setting and our main contributions in the abstract and introduction.

   Guidelines:
   - The answer NA means that the abstract and introduction do not include the claims made in the paper.
   - The abstract and/or introduction should clearly state the claims made, including the contributions made in the paper and important assumptions and limitations. A No or NA answer to this question will not be perceived well by the reviewers.
   - The claims made should match theoretical and experimental results, and reflect how much the results can be expected to generalize to other settings.
   - It is fine to include aspirational goals as motivation as long as it is clear that these goals are not attained by the paper.

2. **Limitations**

   Question: Does the paper discuss the limitations of the work performed by the authors?

   Answer: [Yes]

Justification: Our results provide efficient uncoupled algorithms to compute ε-approximate Φ-equilibria in the first-order stationary regime. We leave the computational complexity of finding ε-approximate Φ-equilibria beyond the first-order stationary regime as an open question.

Guidelines:

- The answer NA means that the paper has no limitation while the answer No means that the paper has limitations, but those are not discussed in the paper.
- The authors are encouraged to create a separate "Limitations" section in their paper.
- The paper should point out any strong assumptions and how robust the results are to violations of these assumptions (e.g., independence assumptions, noiseless settings, model well-specification, asymptotic approximations only holding locally). The authors should reflect on how these assumptions might be violated in practice and what the implications would be.
- The authors should reflect on the scope of the claims made, e.g., if the approach was only tested on a few datasets or with a few runs. In general, empirical results often depend on implicit assumptions, which should be articulated.
- The authors should reflect on the factors that influence the performance of the approach. For example, a facial recognition algorithm may perform poorly when image resolution is low or images are taken in low lighting. Or a speech-to-text system might not be used reliably to provide closed captions for online lectures because it fails to handle technical jargon.
- The authors should discuss the computational efficiency of the proposed algorithms and how they scale with dataset size.
- If applicable, the authors should discuss possible limitations of their approach to address problems of privacy and fairness.
- While the authors might fear that complete honesty about limitations might be used by reviewers as grounds for rejection, a worse outcome might be that reviewers discover limitations that aren't acknowledged in the paper. The authors should use their best judgment and recognize that individual actions in favor of transparency play an important role in developing norms that preserve the integrity of the community. Reviewers will be specifically instructed to not penalize honesty concerning limitations.

3. **Theory Assumptions and Proofs**

   Question: For each theoretical result, does the paper provide the full set of assumptions and a complete (and correct) proof?

   Answer: [Yes]

   Justification: We provide assumptions and proofs for our theoretical results in the main body and the appendix.

   Guidelines:

   - The answer NA means that the paper does not include theoretical results.
   - All the theorems, formulas, and proofs in the paper should be numbered and cross-referenced.
   - All assumptions should be clearly stated or referenced in the statement of any theorems.
   - The proofs can either appear in the main paper or the supplemental material, but if they appear in the supplemental material, the authors are encouraged to provide a short proof sketch to provide intuition.
   - Inversely, any informal proof provided in the core of the paper should be complemented by formal proofs provided in appendix or supplemental material.
   - Theorems and Lemmas that the proof relies upon should be properly referenced.

4. **Experimental Result Reproducibility**

   Question: Does the paper fully disclose all the information needed to reproduce the main experimental results of the paper to the extent that it affects the main claims and/or conclusions of the paper (regardless of whether the code and data are provided or not)?

   Answer: [NA]

Justification: This paper does not contain experimental results.

Guidelines:

- The answer NA means that the paper does not include experiments.
- If the paper includes experiments, a No answer to this question will not be perceived well by the reviewers: Making the paper reproducible is important, regardless of whether the code and data are provided or not.
- If the contribution is a dataset and/or model, the authors should describe the steps taken to make their results reproducible or verifiable.
- Depending on the contribution, reproducibility can be accomplished in various ways. For example, if the contribution is a novel architecture, describing the architecture fully might suffice, or if the contribution is a specific model and empirical evaluation, it may be necessary to either make it possible for others to replicate the model with the same dataset, or provide access to the model. In general. releasing code and data is often one good way to accomplish this, but reproducibility can also be provided via detailed instructions for how to replicate the results, access to a hosted model (e.g., in the case of a large language model), releasing of a model checkpoint, or other means that are appropriate to the research performed.
- While NeurIPS does not require releasing code, the conference does require all submissions to provide some reasonable avenue for reproducibility, which may depend on the nature of the contribution. For example
  (a) If the contribution is primarily a new algorithm, the paper should make it clear how to reproduce that algorithm.
  (b) If the contribution is primarily a new model architecture, the paper should describe the architecture clearly and fully.
  (c) If the contribution is a new model (e.g., a large language model), then there should either be a way to access this model for reproducing the results or a way to reproduce the model (e.g., with an open-source dataset or instructions for how to construct the dataset).
  (d) We recognize that reproducibility may be tricky in some cases, in which case authors are welcome to describe the particular way they provide for reproducibility. In the case of closed-source models, it may be that access to the model is limited in some way (e.g., to registered users), but it should be possible for other researchers to have some path to reproducing or verifying the results.

5. **Open access to data and code**

Question: Does the paper provide open access to the data and code, with sufficient instructions to faithfully reproduce the main experimental results, as described in supplemental material?

Answer: [NA]

Justification: This paper does not contain experimental results.

Guidelines:

- The answer NA means that paper does not include experiments requiring code.
- Please see the NeurIPS code and data submission guidelines (https://nips.cc/public/guides/CodeSubmissionPolicy) for more details.
- While we encourage the release of code and data, we understand that this might not be possible, so "No" is an acceptable answer. Papers cannot be rejected simply for not including code, unless this is central to the contribution (e.g., for a new open-source benchmark).
- The instructions should contain the exact command and environment needed to run to reproduce the results. See the NeurIPS code and data submission guidelines (https://nips.cc/public/guides/CodeSubmissionPolicy) for more details.
- The authors should provide instructions on data access and preparation, including how to access the raw data, preprocessed data, intermediate data, and generated data, etc.
- The authors should provide scripts to reproduce all experimental results for the new proposed method and baselines. If only a subset of experiments are reproducible, they should state which ones are omitted from the script and why.

- At submission time, to preserve anonymity, the authors should release anonymized versions (if applicable).
- Providing as much information as possible in supplemental material (appended to the paper) is recommended, but including URLs to data and code is permitted.

6. **Experimental Setting/Details**

Question: Does the paper specify all the training and test details (e.g., data splits, hyper-parameters, how they were chosen, type of optimizer, etc.) necessary to understand the results?

Answer: [NA]

Justification: This paper does not contain experimental results.

Guidelines:

- The answer NA means that the paper does not include experiments.
- The experimental setting should be presented in the core of the paper to a level of detail that is necessary to appreciate the results and make sense of them.
- The full details can be provided either with the code, in appendix, or as supplemental material.

7. **Experiment Statistical Significance**

Question: Does the paper report error bars suitably and correctly defined or other appropriate information about the statistical significance of the experiments?

Answer: [NA]

Justification: This paper does not contain experimental results.

Guidelines:

- The answer NA means that the paper does not include experiments.
- The authors should answer "Yes" if the results are accompanied by error bars, confidence intervals, or statistical significance tests, at least for the experiments that support the main claims of the paper.
- The factors of variability that the error bars are capturing should be clearly stated (for example, train/test split, initialization, random drawing of some parameter, or overall run with given experimental conditions).
- The method for calculating the error bars should be explained (closed form formula, call to a library function, bootstrap, etc.)
- The assumptions made should be given (e.g., Normally distributed errors).
- It should be clear whether the error bar is the standard deviation or the standard error of the mean.
- It is OK to report 1-sigma error bars, but one should state it. The authors should preferably report a 2-sigma error bar than state that they have a 96% CI, if the hypothesis of Normality of errors is not verified.
- For asymmetric distributions, the authors should be careful not to show in tables or figures symmetric error bars that would yield results that are out of range (e.g. negative error rates).
- If error bars are reported in tables or plots, The authors should explain in the text how they were calculated and reference the corresponding figures or tables in the text.

8. **Experiments Compute Resources**

Question: For each experiment, does the paper provide sufficient information on the computer resources (type of compute workers, memory, time of execution) needed to reproduce the experiments?

Answer: [NA]

Justification: This paper does not contain experimental results.

Guidelines:

- The answer NA means that the paper does not include experiments.

- The paper should indicate the type of compute workers CPU or GPU, internal cluster, or cloud provider, including relevant memory and storage.
- The paper should provide the amount of compute required for each of the individual experimental runs as well as estimate the total compute.
- The paper should disclose whether the full research project required more compute than the experiments reported in the paper (e.g., preliminary or failed experiments that didn't make it into the paper).

9. **Code Of Ethics**

Question: Does the research conducted in the paper conform, in every respect, with the NeurIPS Code of Ethics https://neurips.cc/public/EthicsGuidelines?

Answer: [Yes]

Justification: We have reviewed the NeurIPS Code of Ethics and the current paper conforms the NeurIPS Code of Ethics.

Guidelines:

- The answer NA means that the authors have not reviewed the NeurIPS Code of Ethics.
- If the authors answer No, they should explain the special circumstances that require a deviation from the Code of Ethics.
- The authors should make sure to preserve anonymity (e.g., if there is a special consideration due to laws or regulations in their jurisdiction).

10. **Broader Impacts**

Question: Does the paper discuss both potential positive societal impacts and negative societal impacts of the work performed?

Answer: [NA]

Justification: This is a purely theoretical paper and we do not see any immediate societal impact.

Guidelines:

- The answer NA means that there is no societal impact of the work performed.
- If the authors answer NA or No, they should explain why their work has no societal impact or why the paper does not address societal impact.
- Examples of negative societal impacts include potential malicious or unintended uses (e.g., disinformation, generating fake profiles, surveillance), fairness considerations (e.g., deployment of technologies that could make decisions that unfairly impact specific groups), privacy considerations, and security considerations.
- The conference expects that many papers will be foundational research and not tied to particular applications, let alone deployments. However, if there is a direct path to any negative applications, the authors should point it out. For example, it is legitimate to point out that an improvement in the quality of generative models could be used to generate deepfakes for disinformation. On the other hand, it is not needed to point out that a generic algorithm for optimizing neural networks could enable people to train models that generate Deepfakes faster.
- The authors should consider possible harms that could arise when the technology is being used as intended and functioning correctly, harms that could arise when the technology is being used as intended but gives incorrect results, and harms following from (intentional or unintentional) misuse of the technology.
- If there are negative societal impacts, the authors could also discuss possible mitigation strategies (e.g., gated release of models, providing defenses in addition to attacks, mechanisms for monitoring misuse, mechanisms to monitor how a system learns from feedback over time, improving the efficiency and accessibility of ML).

11. **Safeguards**

Question: Does the paper describe safeguards that have been put in place for responsible release of data or models that have a high risk for misuse (e.g., pretrained language models, image generators, or scraped datasets)?

Answer: [NA]

Justification: This is a purely theoretical paper

Guidelines:

- The answer NA means that the paper poses no such risks.
- Released models that have a high risk for misuse or dual-use should be released with necessary safeguards to allow for controlled use of the model, for example by requiring that users adhere to usage guidelines or restrictions to access the model or implementing safety filters.
- Datasets that have been scraped from the Internet could pose safety risks. The authors should describe how they avoided releasing unsafe images.
- We recognize that providing effective safeguards is challenging, and many papers do not require this, but we encourage authors to take this into account and make a best faith effort.

12. **Licenses for existing assets**

Question: Are the creators or original owners of assets (e.g., code, data, models), used in the paper, properly credited and are the license and terms of use explicitly mentioned and properly respected?

Answer: [NA]

Justification: This is a purely theoretical paper.

Guidelines:

- The answer NA means that the paper does not use existing assets.
- The authors should cite the original paper that produced the code package or dataset.
- The authors should state which version of the asset is used and, if possible, include a URL.
- The name of the license (e.g., CC-BY 4.0) should be included for each asset.
- For scraped data from a particular source (e.g., website), the copyright and terms of service of that source should be provided.
- If assets are released, the license, copyright information, and terms of use in the package should be provided. For popular datasets, paperswithcode.com/datasets has curated licenses for some datasets. Their licensing guide can help determine the license of a dataset.
- For existing datasets that are re-packaged, both the original license and the license of the derived asset (if it has changed) should be provided.
- If this information is not available online, the authors are encouraged to reach out to the asset's creators.

13. **New Assets**

Question: Are new assets introduced in the paper well documented and is the documentation provided alongside the assets?

Answer: [NA]

Justification: This is a purely theoretical paper.

Guidelines:

- The answer NA means that the paper does not release new assets.
- Researchers should communicate the details of the dataset/code/model as part of their submissions via structured templates. This includes details about training, license, limitations, etc.
- The paper should discuss whether and how consent was obtained from people whose asset is used.
- At submission time, remember to anonymize your assets (if applicable). You can either create an anonymized URL or include an anonymized zip file.

14. **Crowdsourcing and Research with Human Subjects**

Question: For crowdsourcing experiments and research with human subjects, does the paper include the full text of instructions given to participants and screenshots, if applicable, as well as details about compensation (if any)?

Answer: [NA]

Justification: This paper does not contain experiments.

Guidelines:

- The answer NA means that the paper does not involve crowdsourcing nor research with human subjects.
- Including this information in the supplemental material is fine, but if the main contribution of the paper involves human subjects, then as much detail as possible should be included in the main paper.
- According to the NeurIPS Code of Ethics, workers involved in data collection, curation, or other labor should be paid at least the minimum wage in the country of the data collector.

15. **Institutional Review Board (IRB) Approvals or Equivalent for Research with Human Subjects**

Question: Does the paper describe potential risks incurred by study participants, whether such risks were disclosed to the subjects, and whether Institutional Review Board (IRB) approvals (or an equivalent approval/review based on the requirements of your country or institution) were obtained?

Answer: [NA]

Justification: This paper does not contain experiments.

Guidelines:

- The answer NA means that the paper does not involve crowdsourcing nor research with human subjects.
- Depending on the country in which research is conducted, IRB approval (or equivalent) may be required for any human subjects research. If you obtained IRB approval, you should clearly state this in the paper.
- We recognize that the procedures for this may vary significantly between institutions and locations, and we expect authors to adhere to the NeurIPS Code of Ethics and the guidelines for their institution.
- For initial submissions, do not include any information that would break anonymity (if applicable), such as the institution conducting the review.