Proceedings of Machine Learning Research vol 313:1–23, 2026 37th International Conference on Algorithmic Learning Theory

Efficient Opportunistic Approachability

Teodor Vanislavov Marinov TVMARINOV@GOOGLE.COM
Google Research

Mehryar Mohri MOHRI@GOOGLE.COM
Google Research

Princewill Okoroafor PCO9@CORNELL.EDU
Cornell University

Jon Schneider JSCHNEI@GOOGLE.COM
Google Research

Julian Zimmert ZIMMERT@GOOGLE.COM
Google Research

Editors: Matus Telgarsky and Jonathan Ullman

Abstract

We study the problem of *opportunistic approachability*: a generalization of Blackwell approachability where the learner would like to obtain stronger guarantees (i.e., approach a smaller set) when their adversary limits themselves to a subset of their possible action space. Bernstein and Shimkin (2015) introduced this problem in 2014 and presented an algorithm that guarantees sublinear approachability rates for opportunistic approachability. However, this algorithm requires the ability to produce calibrated online predictions of the adversary's actions, a problem whose standard implementations require time exponential in the ambient dimension and result in approachability rates that scale as $T^{-O(1/d)}$. In this paper, we present an efficient algorithm for opportunistic approachability that achieves a rate of $O(T^{-1/4})$ (and an inefficient one that achieves a rate of $O(T^{-1/3})$), bypassing the need for an online calibration subroutine. Moreover, in the case where the dimension of the adversary's action set is at most two, we show it is possible to obtain the optimal rate of $O(T^{-1/2})$.

Keywords: approachability, online learning, opportunistic

1. Introduction

Blackwell's approachability theorem has proven itself to be an incredibly valuable tool for tackling a wide range of different problems in machine learning, economics, and game theory involving repeated interactions with multi-objective constraints. At a high level, this theorem allows a learner repeatedly taking decisions in an adversarial environment to steer the average value of a vector-valued payoff function so that it "approaches" a certain set.

Slightly more formally, in the standard setting of Blackwell approachability, a learner and an adversary play a repeated game for T rounds. Every round $t \in \{1, 2, \ldots, T\}$, the learner selects an action p_t from a convex action set $\mathcal{P}$, and simultaneously the adversary selects an action ℓ_t from a convex action set $\mathcal{L}$. Upon doing so, the learner receives a vector-valued payoff $u(p_t, \ell_t) \in \mathbb{R}^d$, for some bilinear function $u(\cdot, \cdot)$. The goal of the learner is to steer the average value $\frac{1}{T}\sum_t u(p_t, \ell_t)$ of this payoff so that it converges to a convex set S. If this is possible, the set is *approachable*.

Echoing von Neumann's minimax theorem, Blackwell's approachability theorem states that a convex set S is approachable if and only if for every adversary action $\ell \in \mathcal{L}$, there exists a response

© 2026 T.V. Marinov, M. Mohri, P. Okoroafor, J. Schneider & J. Zimmert.

action $p^*(\ell) \in \mathcal{P}$ such that $u(p^*(\ell), \ell)$ lies in S. In particular, this theorem implies that for any response function $p^* : \mathcal{L} \to \mathcal{P}$, it is possible for the learner to guarantee that the time-average value of $u(p_t, \ell_t)$ converges to the convex hull $S = \text{conv}\{u(p^*(\ell), \ell) \mid \ell \in \mathcal{L}\}$.

While Blackwell's theorem provides a sharp characterization of when a set can be approached in the worst-case sense, this guarantee can be overly restrictive in practice. In many situations, the adversary's behavior is not fully adversarial but instead follows some structured, stochastic, or cooperative pattern. This motivates the notion of *opportunistic approachability* (Bernstein et al., 2014), which asks not whether the learner can enforce convergence to S against all possible adversary strategies, but whether the learner can exploit opportunities to approach S whenever the adversary's play makes this feasible. In particular, (one framing of) opportunistic approachability asks for a learning algorithm with the following "data-dependent" guarantee: if the adversary only plays actions in the convex set $Q \subseteq \mathcal{L}$, then the learning algorithm should guarantee that the time-average value of $u(x_t, y_t)$ converges to the restricted set $S(Q) = \text{conv}\{u(p^*(y), y) \mid y \in Q\}$. Importantly, the algorithm should accomplish this without any prior knowledge of the set Q.

In their paper originating this notion, Bernstein and Shimkin (2015) provide an algorithm for opportunistic approachability based on the technique of best responding to calibrated predictions of the adversaries actions y_t. While this algorithm does satisfy the requirements of opportunistic approachability, the *rates of convergence* it guarantees to the set $S(Q)$ are quite poor[1]. In particular, if $d_{\mathcal{L}} = \dim(\mathcal{L})$ is the dimension of the adversary's action set, then the algorithm of Bernstein and Shimkin (2015) only guarantees that the distance between $\frac{1}{T}\sum_t u(p_t, \ell_t)$ and $S(Q)$ is at most $T^{-O(1/d_{\mathcal{L}})}$ – in other words, it requires at least $T = \delta^{-\Omega(d_{\mathcal{L}})}$ rounds to guarantee δ-convergence to the set $S(Q)$. The algorithm is similarly computationally inefficient – obtaining a convergence rate of δ requires maintaining a set of weights, one for each point in an δ-discretization of the adversary's action space.

1.1. Our Results and Technical Overview

In this paper, we present several new algorithms for opportunistic approachability with *polynomial* rates of convergence – i.e., that guarantee δ-convergence to the set $S(Q)$ after at most $T = \text{poly}(1/\delta)$ rounds. Many of these algorithms are also computationally efficient, running in $\text{poly}(1/\delta)$ time per round.

We distinguish between two variants of opportunistic approachability. In *strict* opportunistic approachability (the variant defined above), we require approachability to the set $S(Q)$ only for convex sets Q which contain *every* adversary action ℓ_t. In (ε-)statistical opportunistic approachability, we require the stronger guarantee of approachability to the set $S(Q)$ for any convex set Q which contains at least a $(1-\varepsilon)$ fraction of all adversary actions ℓ_t (see Definition 8 for a precise definition).

We design algorithms for both settings of opportunistic approachability. Let $d_{\mathcal{P}} = \dim(\mathcal{P})$ equal the dimension of the learner's action set. For strict opportunistic approachability, we first present an algorithm **(Theorem 14)** achieving rate $O(d_{\mathcal{P}}^{1/2} T^{-1/4})$ that runs in polynomial time per iteration, followed by an algorithm **(Theorem 21)** with improved rate $O(d_{\mathcal{P}}^{1/3} T^{-1/3})$, at the cost of a worse

1. Bernstein et al. (2014) do not directly study the convergence rates or computational complexity of their algorithm, and use the existence of online calibrated predictions as a black box. However, the standard algorithms for constructing d-dimensional online calibrated predictions produce predictions with average calibration error $T^{-O(1/d)}$ and are computationally inefficient, and the opportunistic approachability algorithms for Bernstein et al. (2014) inherit similar rates. We discuss this further in the related work.

computational complexity (possibly exponential in $d_{\mathcal{P}}$). Similarly, for ε-statistical opportunistic approachability, we first present a simple algorithm (**Theorem 17**) achieving rate $O(d_{\mathcal{P}}^{1/2}T^{-1/4} + \varepsilon^{1/3}d_{\mathcal{P}}^{1/3})$, followed by a slightly more complex algorithm (**Theorem 23**) achieving the improved rate $O(d_{\mathcal{P}}^{1/3}T^{-1/3} + \varepsilon^{1/3}d_{\mathcal{P}}^{1/3})$.

In contrast to the strictly opportunistic setting (where the algorithm of Theorem 14 both has polynomial rates and is computationally efficient), all algorithms for the statistical opportunistic approachability generally require time exponential in the payoff dimension d. However, under some assumptions it is possible to implement the $O(T^{-1/4})$-rate algorithm efficiently (for example, if the response function p^* is Lipschitz) – see Section 3.2 for additional details.

Our algorithms sidestep the need for a high-dimensional online calibration subroutine as used by Bernstein et al. (2014). Instead, they employ a novel epoch-based framework that decomposes the learning task into two coupled online learners operating at different timescales. In the outer loop of this framework, the outer online learner outputs a *direction vector* in payoff space (trying to continually learn a direction in which the current time-averaged payoff is maximally separated from a current "target payoff" guaranteed to be in the set $S(Q)$). In the inner loop of this framework, the inner online learner then produces actions to play that minimize the loss in the direction vector output by the online learner.

Key to obtaining many of our results efficiently is the idea that we can relax the losses fed to both learners in this framework: in fact, all that is necessary is that we have 1. a *consistent target function* that takes an interval of losses played by the adversary and outputs an element $U \in \mathbb{R}^d$ guaranteed to belong to the final target set, and 2. a *valid loss rule* $F(p;\lambda) \in \mathbb{R}$ that upper bounds the largest possible payoff $\langle \lambda, u(p,\ell)\rangle$ in the direction λ obtainable by playing p against a worst-case loss ℓ. For both settings of opportunistic approachability, simple choices for these two functions lead to $T^{-1/4}$ rates, whereas more sophisticated choices (combined with a more careful path-length based analysis of regret of the inner learner) lead to $T^{-1/3}$ rates.

Finally, we consider the question of whether it is possible to achieve the optimal rate of $O(T^{-1/2})$, even in the strictly opportunistic setting. We provide some positive evidence that this might be the case by constructing efficient algorithms with this rate when either the dimension d of the payoff is 1 (Theorem 24), or when the dimension $d_{\mathcal{L}}$ of the adversary's action set is at most 2 (Theorem 26). Interestingly, the algorithm of Theorem 26 adopts a convex geometric approach from Leme and Schneider (2018) to prove that the myopic approach of assuming that the adversary will continue to play losses in the convex hull of the losses already played achieves this optimal rate (unfortunately, this approach breaks down for dimensions greater than 3, and does not appear to produce any rates better than the existing $O(T^{-1/3})$ rate).

We conclude with two open questions:

1. Does there exist an efficient algorithm that achieves a $O(T^{-1/2})$ rate for strictly opportunistic approachability in high dimensions ($d > 3$)? Alternatively, can we prove a lower bound of $\Omega(T^{-1/3})$?

2. Is statistically opportunistic approachability fundamentally harder than strictly opportunistic approachability? Does there exist a computationally efficient algorithm for statistically opportunistic approachability that achieves polynomial rates of convergence?

1.2. Related Work

High-Dimensional Calibration The original algorithms of Bernstein et al. (2014) for opportunistic approachability require a subroutine for calibrated predictions of the adversary's actions. Such a subroutine was first shown to exist by Foster and Vohra (1998), who used this to construct the first algorithms for minimizing swap regret (and hence converging to correlated equilibria) in games. This algorithm suffers from an exponential dependence on the dimension d of the adversary's action set (i.e., producing δ-calibrated predictions required at least $T = \delta^{-\Omega(d)}$ rounds). For a while, it was an open question whether this exponential dependence was necessary (Abernethy and Mannor, 2011); this was largely resolved by recent work showing that although it is possible to construct algorithms producing δ-calibrated predictions in $T = d^{\Omega(1/\delta)}$ rounds, there is no such algorithm with T polynomial in both d and $1/\delta$ (Peng, 2025; Fishelson et al., 2025).

Other Algorithms for Blackwell Approachability We briefly survey here other known algorithms for solving the Blackwell approachability problem and how they relate to the two-stage framework we introduce in this paper.

Perhaps the most common technique for constructing an online learning algorithm for the approachability problem is the reduction to online linear optimization popularized by Abernethy et al. (2011), also implicit in the original construction of Blackwell (1956) and several follow-up works (Mannor and Perchet, 2013; Kwon, 2016; Dann et al., 2023). This reduction relies on a sub-algorithm that produces a low-regret sequence H_t of half-spaces containing the target set S, and then playing actions p_t that guarantee that each individual payoff $u(p_t, \ell)$ lies in H_t (regardless of the choice of adversary action ℓ). In some ways this is analogous to the outer loop of our framework (which also chooses directions in utility space). One major difference, however, is that the outer loop of our procedure produces *directions* in utility space, whereas this sub-algorithm produces *half-spaces* (directions with achievable target values); this distinction is important, as we cannot tell whether a given half-space is valid until we have identified the final target set $S(Q)$ (the inner learner in our framework can perhaps be thought of as trying to learn the optimal half-space in a given direction).

Another method for solving approachability instances is provided by Bernstein and Shimkin (2015) in their study of *response-based approachability* (i.e., approachability instances defined in terms of a response function $p^* : \mathcal{L} \to \mathcal{P}$). The algorithm they propose avoids the need to do online linear optimization over the set of half-spaces containing the target set (a task that can be computationally challenging (Bernstein and Shimkin, 2015; Daskalakis et al., 2025)). Similar to the outer loop of our framework, their algorithm identifies an approachability direction by looking at the displacement between the current time-average payoff and some target payoff guaranteed to be in S. Like the Abernethy et al. (2011) reduction, this does not immediately generalize to opportunistic approachability due to the difficulty of guaranteeing a specific action p is optimal for a given (scalar) payoff function without knowledge of the adversary's set of actions Q.

Applications of (Opportunistic) Approachability As mentioned, Blackwell approachability is an incredibly powerful tool for the design of online algorithms. Among many other applications, in recent years this has allowed researchers to design learning algorithms with strong strategic robustness properties in general games (Arunachaleswaran et al., 2025), online prediction algorithms with strong downstream guarantees (Okoroafor et al., 2024, 2025), fair online learning algorithms

(Chzhen et al., 2021), and reinforcement learning algorithms satisfying multi-objective constraints (Miryoosefi et al., 2019).

Any application of Blackwell approachability can in theory benefit from the additional guarantees provided by opportunistic approachability. We highlight two specific applications here for which these additional guarantees are particularly meaningful.

- **Approximately approaching unapproachable sets**: It is not uncommon to have an online learning problem with a set of desired constraints that are not possible to guarantee simultaneously (i.e., a target set that is not in general approachable). This is the case for example in fair online learning: Chzhen et al. (2021) present three fairness desiderata that they prove are not simultaneously attainable. Mannor et al. (2014) provide examples of other approachability instances where this is the case.

  In such cases, one may wish to approach the target set as closely as possible. By setting the response function $p^*(\ell)$ to be the response to ℓ that minimizes the distance to the target set, opportunistic approachability guarantees that one does approach the target set if the adversary limits themselves to a "feasible" convex subset of strategies Q.

- **Stronger profile swap regret guarantees**: Arunachaleswaran et al. (2025) recently introduced the notion of *profile swap regret* in general linear games, an efficiently minimizable regret notion with the property that any no-profile-swap-regret player is "non-manipulable" (no opponent can extract more per-round utility from them than they could obtain against a rational opponent in a single-shot game). Interestingly, minimizing profile swap regret corresponds exactly to the response-based approachability problem where $u(p, \ell) = p \otimes \ell$ and $p^*(\ell)$ is the best response to the action ℓ. Running an opportunistic approachability algorithm therefore provides stronger non-manipulability guarantees in settings where the opponent's true action set is unknown.

2. Model and Preliminaries

2.1. Notation

We write $[T]$ to denote the set $\{1, 2, \ldots, T\}$, $\|\cdot\|$ to denote the ℓ_2 norm, and $B_d(r) \subset \mathbb{R}^d$ to denote the d-dimensional ℓ_2 ball of radius r. The support function for a set S at $\lambda \in \mathbb{R}^d$ is denoted by $\sigma_S(\lambda) = \sup_{s \in S} \langle s, \lambda \rangle$. We note that the distance of a point $x \in \mathbb{R}^d$ to S can be written as $\texttt{dist}(x, S) = \sup_{\lambda \in B_d(1)} \langle \lambda, x \rangle - \sigma_S(\lambda)$.

2.2. Blackwell Approachability and Opportunistic Approachability

Blackwell's approachability theorem (Blackwell, 1956) extends repeated two-player zero-sum games to settings involving vector-valued payoffs. In this model, at each timestep $t = 1, 2, \ldots, T$ of the game, the learner selects an action $p_t \in \mathcal{P} \subset \mathbb{R}^{d_\mathcal{P}}$, and the adversary selects an action $\ell_t \in \mathcal{L} \subset \mathbb{R}^{d_\mathcal{L}}$. The learner then receives a vector-valued payoff $u(p_t, \ell_t)$, for some bi-affine payoff function $u : \mathcal{P} \times \mathcal{L} \to \mathbb{R}^d$. We will assume the action sets $\mathcal{P}$ and $\mathcal{L}$ are compact and convex subsets, and that u is bounded so that $\|u(p, \ell)\| \leq 1$ for any $p \in \mathcal{P}$ and $\ell \in \mathcal{L}$. We will further assume that the action set $\mathcal{P}$ of the learner is in isotropic position (and in particular assume that $B_{d_\mathcal{P}}(1) \subseteq \mathcal{P} \subseteq B_{d_\mathcal{P}}(d_\mathcal{P})$, where $B_d(r)$ denotes the d-dimensional ball of radius r centered at the origin).

The learner's goal is to select actions $\{p_t\}_{t=1}^T$ (adaptively, choosing p_t based on the current history $(p_1, \ell_1, \ldots, p_{t-1}, \ell_{t-1})$) such that their average payoff vector converges to a given closed convex

target set $S \subseteq \mathbb{R}^d$. This convergence must be guaranteed regardless of the adversary's choices $\{\ell_t\}_{t=1}^T$. For a given closed convex set S, the learner aims to ensure that:

$$\texttt{dist}\left(\frac{1}{T}\sum_{t=1}^{T} u(p_t, \ell_t), S\right) \to 0 \quad \text{as} \quad T \to \infty \tag{1}$$

A closed convex set S is *approachable* if the learner can guarantee this convergence property against any strategy of the adversary. Blackwell's theorem provides a fundamental characterization of approachable sets: a closed convex set $S \subseteq \mathbb{R}^d$ is approachable if and only if there exists a *response function* $p^* : \mathcal{L} \to \mathcal{P}$ such which satisfies $u(p^*(\ell), \ell) \in S$.

This characterization motivates the framework of *response-based approachability* (Bernstein and Shimkin, 2015), where we define an approachability instance via a response function $p^* : \mathcal{L} \to \mathcal{P}$. Given such a function, the goal of the player is to approach the convex hull of the set of responses to each $\ell \in \mathcal{L}$, that is to approach the set

$$S(\mathcal{L}) = \text{conv}\{u(p^*(\ell), \ell) \mid \ell \in \mathcal{L}\}$$

Opportunistic approachability (Bernstein et al., 2014) is a refinement of response-based approachability that also begins with a desired response function $p^* : \mathcal{L} \to \mathcal{P}$ for the learner, but where the learner's target set is contingent on the adversary's actual behavior over time. For any closed, convex set Q, define

$$S(Q) = \text{conv}\{u(p^*(\ell), \ell) \mid \ell \in Q\}$$

The goal of the learner in opportunistic approachability is to guarantee that, if Q is a convex set containing (asymptotically) all points of the adversary's action sequence, then the average payoff vector $\frac{1}{T}\sum_{t=1}^{T} u(p_t, \ell_t)$ converges to the set $S(Q)$. That is, the learner is "opportunistic" – they steer their average payoff to the smallest achievable set $S(Q)$ that is consistent with the adversary's actual, possibly restricted, play. Note that $S(Q) \subseteq S(\mathcal{L})$, so the target is indeed potentially smaller (better) when the adversary restricts their play ($Q \subseteq \mathcal{L}$). Formally, Bernstein et al. (2014) define the objective of opportunistic approachability as follows.

Definition 1 (Statistically/Strictly Q-Restricted Adversary (Bernstein et al., 2014)) *Let Q be a closed convex subset of $\mathcal{L}$. We say that the sequence of pure actions $\{\ell_t\}_{t=1}^{\infty}$ played by an adversary is statistically Q-restricted if the average distance of the actions to Q converges to zero:*

$$\lim_{T\to\infty} \frac{1}{T}\sum_{t=1}^{T} \texttt{dist}(\ell_t, Q) = 0.$$

We say the adversary's play is strictly Q-restricted if $\ell_t \in Q$ for all t.

Definition 2 (Statistically/Strictly Opportunistic Learner Strategy (Bernstein et al., 2014)) *Let $p^* : \mathcal{L} \to \mathcal{P}$ be a desired response function. We say that a learner's strategy is statistically opportunistic with respect to p^* if, for any $Q \subseteq \mathcal{L}$ and any sequence of pure actions $\{\ell_t\}_{t=1}^{\infty}$ played by a statistically Q-restricted adversary, the learner's strategy guarantees that the average payoff converges to the set $S(Q) = \text{conv}\{u(p^*(\ell), \ell) \mid \ell \in Q\}$. That is,*

$$\lim_{T\to\infty} \texttt{dist}\left(\frac{1}{T}\sum_{t=1}^{T} u(p_t, \ell_t), S(Q)\right) = 0,$$

Similarly, a learner's strategy is strictly opportunistic with respect to p^* if this guarantee holds against any strictly Q-restricted adversary.

2.3. Finite Time Opportunistic Approachability

Bernstein et al. (2014) define their notion of statistically Q-restricted opportunistic approachability only in the limit as $T \to \infty$. In this paper, we are interested in finite-time guarantees, and therefore need to modify these definitions for finite time horizons. It is straightforward to adapt the definition of a strictly opportunistic learner strategy to the finite time setting by simply requiring every adversary action ℓ_t to lie in the restriction set Q. Formally:

Definition 3 (Finite-time Strictly Opportunistic Learner) *We say an algorithm is strictly opportunistic at a rate of $\gamma(T)$ if, for any sequence of adversary actions $\{\ell_t\}_{t=1}^T$, its average payoff vector $\bar{u} = \frac{1}{T}u(p_t, \ell_t)$ satisfies*

$$\texttt{dist}\left(\bar{u}_T, S(Q)\right) \leq \gamma(T)$$

for $Q = \text{conv}(\{\ell_t\}_{t\in[T]})$.

On the other hand, there is some subtlety in how exactly to adapt the limit-based definition of a statistically Q-restricted adversary to finite time. For example, the following lemma shows that a straightforward translation of Definition 1 to finite time is not possible[2].

Lemma 4 *There exists an approachability problem instance such that for any learner, any $T > 0$, and any $\varepsilon > 0$, there exists an oblivious adversary $\{\ell_t\}_{t=1}^T$ and convex set Q such that $\frac{1}{T}\sum_{t=1}^T \texttt{dist}(\ell_t, Q) \leq \varepsilon$, but $\texttt{dist}(\bar{u}_T, S(Q)) \geq \frac{1}{2}$.*

Proof Define $\mathcal{P} = \mathcal{L} = [-2, 2]$, the utility to be $u(p, \ell) = p\ell \in \mathbb{R}$, and define the response function via

$$p^*(\ell) = \begin{cases} -1 & \text{if } \ell \in [-2, 1) \\ 1 & \text{if } \ell \in [1, 2] \end{cases}$$

The adversary will pick the loss sequence $\ell_t = 1$ for all $t \in [T]$.

Now consider the (singleton) sets $Q_- = \{1 - \varepsilon\}$ and $Q_+ = \{1 + \varepsilon\}$. Note that both $Q-$ and Q_+ satisfy $\frac{1}{T}\sum_{t=1}^T \texttt{dist}(\ell_t, Q) \leq \varepsilon$. However, note that $S(Q_-) = \{-1\}$ and $S(Q_+) = \{1\}$, and so at least one of $\texttt{dist}(\bar{u}_T, S(Q_-))$ and $\texttt{dist}(\bar{u}_T, S(Q_+))$ must be larger than $1/2$. ■

To work around this we consider a stronger, but still natural, notion of Q-restrictedness, which will guarantee the existence of a statistically opportunistic learner strategy for sufficiently small ε. We do this by considering the closed convex subsets of $\mathcal{L}$ which contain all but an ε-fraction of $\{\ell_t\}_{t\in[T]}$, that is $\frac{1}{T}\sum_{t=1}^T \mathbb{I}[\ell_t \notin Q] \leq \varepsilon$. Helly's theorem (Danzer et al., 1963) then implies that for sufficiently small $\varepsilon = O(1/d)$, the intersection of all such subsets will be non-empty (Lemma 29). Let this intersection be $Q^\varepsilon_{\text{int}}$. If $Q^\varepsilon_{\text{int}}$ were known to the learner, then employing standard Blackwell approachability algorithms to approach the set $S(Q^\varepsilon_{\text{int}})$ (ignoring rounds where $u(p^*(\ell_t), \ell_t) \notin S(Q^\varepsilon_{\text{int}})$) guarantees approachability to $S(Q^\varepsilon_{\text{int}})$ with a rate of $O(1/\sqrt{T})$. Next, we formally define this intersection and the target set for the approachability problem.

2. Intuitively, the issue is that the response function p^* can be discontinuous, causing two ε-restricted sets to have very different target sets.

Definition 5 (Adversary Q-sets) *Given a sequence of adversary losses $\{\ell_t\}_{t\in\mathcal{T}}$ and a parameter $\varepsilon \geq 0$, let $\mathcal{T}'$ be any subset of $\mathcal{T}$ such that its size $|\mathcal{T}|$ is at least $(1-\varepsilon)|\mathcal{T}|$. Let*

$$\mathbf{Q}^{\varepsilon}(\{\ell_t\}_{t\in\mathcal{T}}) = \{\mathrm{conv}(\{\ell_t\}_{t\in\mathcal{T}'}) \mid \mathcal{T}' \subseteq \mathcal{T} \wedge |\mathcal{T}'| \geq (1-\varepsilon)|\mathcal{T}|\} \tag{2}$$

Definition 6 (Intersection of Adversary Q-sets) *Given a sequence of adversary actions $\{\ell_t\}_{t\in\mathcal{T}}$ and a parameter $\varepsilon \geq 0$: The intersection set, $Q^{\varepsilon}_{\mathrm{int}}(\{\ell_t\}_{t\in\mathcal{T}})$, is the intersection of all possible adversary Q-sets:*

$$Q^{\varepsilon}_{\mathrm{int}}(\{\ell_t\}_{t\in\mathcal{T}}) = \bigcap_{Q\in\mathbf{Q}^{\varepsilon}(\{\ell_t\}_{t\in\mathcal{T}})} Q$$

Definition 7 (Opportunistic Target Set) *Given the intersection set $Q^{\varepsilon}_{\mathrm{int}}(\{\ell_t\}_{t\in\mathcal{T}})$ from Definition 6, the opportunistic target set is the set of ideal payoffs corresponding to the actions within the intersection set:*

$$S^{\varepsilon}_{\mathrm{int}}(\{\ell_t\}_{t\in\mathcal{T}}) = S\left(Q^{\varepsilon}_{\mathrm{int}}(\{\ell_t\}_{t\in\mathcal{T}})\right) = \mathrm{conv}\{u(p^*(\ell),\ell) \mid \ell \in Q^{\varepsilon}_{\mathrm{int}}(\{\ell_t\}_{t\in\mathcal{T}})\}.$$

When we take the Opportunistic Target Set over the full sequence $(\{\ell_t\}_{t=1}^{T})$, the explicit dependence on this will be omitted for brevity (e.g., $S^{\varepsilon}_{\mathrm{int}}$).

Note that aiming for $S(Q^{\varepsilon}_{\mathrm{int}})$ is a stronger objective than what is strictly required for an opportunistic strategy. A sufficient condition would be to guarantee that the average payoff lies in the intersection of all target sets, $\bigcap_{Q\in\mathbf{Q}^{\varepsilon}} S(Q)$. This is the approach taken in Bernstein and Shimkin (2015). However, since $Q^{\varepsilon}_{\mathrm{int}} \subseteq Q$ for all $Q \in \mathbf{Q}^{\varepsilon}$, it follows that $S(Q^{\varepsilon}_{\mathrm{int}}) \subseteq S(Q)$. Therefore, $S(Q^{\varepsilon}_{\mathrm{int}}) \subseteq \bigcap_{Q\in\mathbf{Q}^{\varepsilon}} S(Q)$. By successfully approaching the smaller set $S(Q^{\varepsilon}_{\mathrm{int}})$, a learner automatically guarantees closeness to every set $S(Q)$ for $Q \in \mathbf{Q}^{\varepsilon}$.

Definition 8 (Finite-time Statistically Opportunistic Learner) *Given a parameter $\varepsilon \geq 0$, we say an algorithm is ε-statistically opportunistic at a rate of $\gamma(T)$ and error rate $\alpha(\varepsilon)$ if, for any sequence of adversary actions $\{\ell_t\}_{t=1}^{T}$, its average payoff vector $\bar{u}_T = \frac{1}{T}\sum_t u(p_t,\ell_t)$ satisfies*

$$\mathtt{dist}\left(\bar{u}_T, S^{\varepsilon}_{\mathrm{int}}\right) \leq \gamma(T) + \alpha(\varepsilon).$$

2.4. Online Optimization Oracles

Our algorithm will rely on online optimization learners as subroutines.

Definition 9 (Online Optimization (OO) Oracle) *An OO oracle is an algorithm that operates over a convex action set $\mathcal{A}$. It interacts with an environment over T rounds as follows: in each round $t = 1,\ldots,T$, the oracle first chooses an action $x_t \in \mathcal{A}$, and then receives a loss function $f_t : \mathcal{A} \to \mathbb{R}$. The oracle provides guarantees on the regret*

$$\mathrm{Reg}_T = \sum_{t=1}^{T} f_t(x_t) - \min_{x\in\mathcal{A}} \sum_{t=1}^{T} f_t(x)$$

Typically, this guarantee is on the order of $\mathrm{Reg}_T = O(\sqrt{T})$. If the loss functions f_t are convex, we call this an Online Convex Optimization (OCO) oracle.

One important example of an OO Oracle is that of *Online Gradient Descent*.

Lemma 10 (OCO Guarantee Theorem 3.1 of Hazan (2023)) *Online Gradient Descent implements an online convex optimization oracle with a regret bound that scales as $O\left(GD\sqrt{T}\right)$, where G is a uniform bound on the norm of the gradients of the loss functions (i.e., $\|\nabla f_t(x)\| \leq G$) and D is the diameter of the convex action set $\mathcal{A}$.*

3. An Algorithmic Framework for Opportunistic Approachability

The algorithms in this paper share a common structure. This section introduces a general, epoch-based framework that serves as a blueprint for these algorithms. The framework is epoch-based and operates over L epochs, each lasting $T_e = T/L$ timesteps. It involves two collaborating online learners: an outer learner and an inner learner. The outer learner operates on the slow timescale of epochs (i.e each timestep corresponds to a single epoch). At the start of each epoch e, it chooses a direction vector, $\lambda_e \in \mathcal{B}_d(1)$. This vector defines the objective for the inner learner. A new inner learner is instantiated at the start of each epoch with a fixed direction λ_e from the outer learner and its goal is to choose actions p_t to minimize a sequence of loss functions derived from λ_e. At the end of an epoch, the combination of the adversary and learner's choices is used to generate a loss for the outer learner, which in turn updates its strategy to select a new direction λ_{e+1}. To make this framework adaptable, we introduce two abstract components: a *Loss Rule* for the inner learner and a *Target Function* for the outer learner.

A *Loss Rule* F is a mapping that takes a history of adversary actions $(\ell_{t'})_{t' \leq t}$ and a direction vector λ and outputs a loss function $f_t : \mathcal{P} \to \mathbb{R}$. For the framework to be valid, we require any such rule to produce losses that are upper bounds on the true linear loss.

Assumption 1 (Valid Loss Rule) *A loss rule F is valid if for any history and any $\lambda \in \mathcal{B}_d(1)$, the resulting loss function $f_t = F((\ell_{t'})_{t' \leq t}; \lambda)$ satisfies: $\forall p \in \mathcal{P}, \quad f_t(p) \geq \langle \lambda, u(p, \ell_t) \rangle$.*

The simplest example of a valid loss rule is $F((\ell_{t'})_{t' \leq t}; \lambda) = \langle \lambda, u(\cdot, \ell_t) \rangle$, which satisfies the assumption with equality.

Similarly, we abstract the process of generating a target for the outer learner. A *Target Function* U is a mapping that takes a history of adversary actions observed during an epoch, $\mathcal{L}_e = (\ell_t)_{t \in \mathcal{T}_e}$, the direction λ, and outputs a target payoff vector $U(\mathcal{L}_e; \lambda_e) = u_e^* \in \mathbb{R}^d$. For this target to be meaningful, it must be *consistent* with the overall goal of the algorithm, which is to approach some true opportunistic target set $S_{\text{int}}^{\varepsilon}$.

Definition 11 (Consistent Target Function) *A target function U is ε-**consistent** if, for any epoch e, its output $U(\mathcal{L}_e; \lambda_e)$ is guaranteed to be an element of the final target set, i.e., $U(\mathcal{L}_e; \lambda_e) \in S_{\text{int}}^{\varepsilon}(\{\ell_t\}_{t=1}^{T})$.*

The inner learner needs to be able to minimize its losses relative to the target and the direction vector provided by the outer learner.

Remark. A computationally simple choice in the strict case ($\varepsilon = 0$) for the target function is $U(\mathcal{L}_e) = u(p^*(\bar{\ell}_e), \bar{\ell}_e)$, where $\bar{\ell}_e$ is the average action in the epoch. While this is consistent, this may not minimize the error term. The theoretically optimal choice for a given λ_e given the largest set Q_e such that $Q_e \subset Q_{\text{int}}^{\varepsilon}$ would be $u_e^* = \arg\max_{u \in S(Q_e)} \langle \lambda_e, u \rangle$, however, it may be difficult to compute this efficiently.

Overview of Algorithm 1 The algorithm is an epoch-based algorithm that involves an outer OCO (Online Convex Optimization) learner, an inner OO (Online Optimization) learner, a loss rule F and a target function U. There are L epochs, each lasting $T_e = T/L$. The outer OCO runs over L timesteps, one for each epoch. A new inner OO is initialized for each epoch. The outcome of the interaction in the inner OO is then used to construct the loss vector for the outer OCO.

Specifically, the outer oracle operates in the dual space of payoffs, $\mathcal{B}_d(1)$, and selects a direction vector λ_e for each epoch. This direction is passed to the inner oracle, which plays the game against the adversary in the learner's action space $\mathcal{P}$. The inner oracle's objective for the epoch is simply to choose actions p_t that minimize the loss sequence f_t generated by the loss rule, which are all upper bounds on the scalar loss projected along the given direction, $\langle \lambda_e, u(p_t, \ell_t)\rangle$.

At the end of an epoch the outer oracle's loss vector is the aggregated average difference between the realized utility and this adaptive target, $g_e = \frac{1}{T_e}\sum_{t\in\mathcal{T}_e}(u(p_t,\ell_t) - u_e^*)$. This loss informs the outer oracle's choice of a new direction for the next epoch, repeating the cycle.

Algorithm 1 Epoch-based Opportunistic Learner

Input: OO oracles $\texttt{OO}_{\text{Inner}}$, $\texttt{OCO}_{\text{Outer}}$, target function U, loss rule F, number of epochs L

Initialize epoch length $T_e = T/L$
Query $\texttt{OCO}_{\text{Outer}}$ on $\mathcal{B}_d(1)$ for initial direction λ_1
for *epoch* $e \leftarrow 1$ **to** L **do**
  Initialize fresh $\texttt{OO}_{\text{Inner}}$ on action space $\mathcal{P}$ **for** $t \leftarrow (e-1)T_e + 1$ **to** eT_e **do**
    `// Inner loop plays against the adversary`
    Play p_t from $\texttt{OO}_{\text{Inner}}$ and observe ℓ_t
    Receive $f_t = F((\ell_{t'})_{t'\le t}; \lambda_e)$
    Update $\texttt{OO}_{\text{Inner}}$ with loss f_t
  end
  `// Outer loop updates the direction vector`
  Compute epoch target $u_e^* = U((\ell_t)_{t\in\mathcal{T}_e}; \lambda_e)$
  Define outer loss $g_e = \frac{1}{T_e}\sum_{t\in\mathcal{T}_e}(u(p_t,\ell_t) - u_e^*)$
  Update $\texttt{OCO}_{\text{Outer}}$ with linear loss $\langle\cdot, g_e\rangle$ to get λ_{e+1}
end

Theorem 12 *Given a loss rule F satisfying Assumption 1 and an ε-consistent target function U, Algorithm 1 satisfies*

$$\texttt{dist}(\bar{u}_T, S^{\varepsilon}_{int}) \le \frac{\mathrm{Reg}^{in}_{T_e}}{T_e} + \frac{\mathrm{Reg}^{out}_{L}}{L} + \mathrm{err}\,,$$

where $\mathrm{err} = \max_{e\in[L]}\left(\frac{1}{T_e}\min_{p\in\mathcal{P}}\sum_{t\in\mathcal{T}_e} f_t(p) - \langle\lambda_e, u_e^*\rangle\right)$.

Proof of [Theorem 12]: let $\lambda^* \in \mathcal{B}_d(1)$ be the direction vector defining the distance, such that $\texttt{dist}(\bar{u}_T, S_T) = \langle\lambda^*, \bar{u}_T\rangle - \sigma_{S_T}(\lambda^*)$, where $\sigma_S(\lambda)$ denotes the support function of S at λ. The proof bounds the total deviation, $T\cdot\texttt{dist}(\bar{u}_T, S_T)$, by chaining the regret guarantees of the two oracles.

From the consistency of the target function (Definition 11) we have $u_e^* \in S_T$, which implies $\langle \lambda^*, u_e^* \rangle \le \sigma_{S_T}(\lambda^*)$. This allows us to bound the total deviation as follows:

$$
\begin{aligned}
T \cdot \texttt{dist}(\bar{u}_T, S_T) &= \langle \lambda^*, T\bar{u}_T \rangle - T\sigma_{S_T}(\lambda^*) \\
&\le \sum_{e=1}^{L} \sum_{t \in \mathcal{T}_e} \langle \lambda^*, u(p_t, \ell_t) - u_e^* \rangle && \text{(Since } u_e^* \in S_T\text{)} \\
&\le \sum_{e=1}^{L} \sum_{t \in \mathcal{T}_e} \langle \lambda_e, u(p_t, \ell_t) - u_e^* \rangle + T_e \mathrm{Reg}_L^{\mathrm{out}} && \text{(Outer Oracle Regret)} \\
&\le \sum_{e=1}^{L} \sum_{t \in \mathcal{T}_e} f_t(p_t) - \langle \lambda_e, u_e^* \rangle + T_e \mathrm{Reg}_L^{\mathrm{out}} && \text{(Assumption 1)} \\
&\le \sum_{e=1}^{L} \Big(\min_{p^* \in \mathcal{P}} \sum_{t \in \mathcal{T}_e} (f_t(p^*) - \langle \lambda_e, u_e^* \rangle) + \mathrm{Reg}_{T_e}^{\mathrm{in}} \Big) + T_e \mathrm{Reg}_L^{\mathrm{out}} && \text{(Inner Oracle Regret)} \\
&\le \mathrm{err} \cdot \mathrm{T} + \mathrm{L}\mathrm{Reg}_{\mathrm{T_e}}^{\mathrm{in}} + \mathrm{T_e}\mathrm{Reg}_{\mathrm{L}}^{\mathrm{out}}
\end{aligned}
$$

Dividing both sides by T completes the proof. ■

Using Online Gradient Descent as the outer learner yields the following corollary.

Corollary 13 *In the same setup as Theorem 12, if the outer OCO is an instance of Online Gradient Descent with optimal tuning, the guarantee is*

$$\texttt{dist}(\bar{u}_T, S_T) \le \frac{1}{T_e} \mathrm{Reg}_{T_e}^{in} + O(L^{-\frac{1}{2}}) + \mathrm{err}$$

3.1. Strictly Opportunistic Approachability

As a warm-up, we present a simple and efficient instantiation of Algorithm 1 that results in a strictly opportunistic learning algorithm. Let $f_t = F(\ldots; \lambda_e)$ and $u_e^* = U(\mathcal{L}_e)$ be defined by

$$f_t(p) = \langle \lambda, u(p, \ell_t) \rangle, \qquad u_e^* = u(p^*(\bar{\ell}_e), \bar{\ell}_e)\,. \tag{3}$$

The choice of f_t is the typical loss used by an OCO oracle in the classical reduction of Blackwell approachability to regret minimization while the choice of u^* is just the best response to the average play of the adversary in an epoch. Algorithm 1 instantiated in this way is computationally efficient as long as projecting onto the learner's set, $\mathcal{P}$, is efficient and querying the response function, p^*, is efficient.

Theorem 14 *[Efficient Strictly Opportunistic Learner] Instantiating Algorithm 1 with the loss rule and target function defined in Equation (3) and OGD for both OO oracles yields a strictly opportunistic learning algorithm with rate $\gamma(T) = \tilde{O}(d_{\mathcal{P}}^{\frac{1}{2}} T^{-\frac{1}{4}})$.*

The proof is deferred to the Appendix A.1. It shows that the err term is non-positive in the strict case, applies the $\sqrt{T_e}$ OGD bound to the inner regret and tunes the parameters.

Our rates have a dependency on the dimension of the players action set, which is absent in regular approachability. The dimension dependency can be improved at the cost of reduced computational efficiency (e.g. using continuous exponential weights (Bubeck, 2011) as the inner OO), but some dependency is unavoidable for the opportunistic approachability problem, as the following lower bound shows.

Lemma 15 *For any learner, there exists an approachability problem for action set $\mathcal{P}, \mathcal{L} \subset \mathbb{R}^d$, $u(\mathcal{P}, \mathcal{L}) \subset [-1, 1]$ and a sequence of losses $(\ell_t)_{t=1,\dots d}$ such that* $\texttt{dist}(\bar{u}_d, S^0_{\text{int}}(\{\ell_t\}_{t\in[d]})) = \Omega(1)$.

Proof of [Lemma 15] We let the action set $\mathcal{P}, \mathcal{L}$ be the ℓ_∞ and ℓ_1 balls in d-dimensions respectively. The payoff function is the inner product of players and adversaries action. The response function is $p^*(\ell)_i = \text{sign}(\ell_i)$. Let the adversary pick the sequence, where they randomly select $\pm\mathbf{e}_t$ for $t = 1, \dots, d$. The expected average payoff vector for any learner is 0, but $S^0_{\text{int}} = \{1\}$.

■

Any rate that only depends of the maximum norm of payoffs would violate this lower bound.

3.2. Statistically Opportunistic Approachability

We now present an instantiation of Algorithm 1 that yields a statistically opportunistic learner. The target function of Equation (3) is not always consistent in this case, since $\bar{\ell}_e$ is not guaranteed to be in $Q^\varepsilon_{\text{int}}$. We solve this problem by finding a "close" point to $\bar{\ell}_e$ that is guaranteed to be in $Q^\varepsilon_{\text{int}}$, where closeness is defined in the following way using the total variation of a discrete distribution $\text{TV}(\alpha, \text{unif}) := \sum_{t\in\mathcal{T}_e} |\alpha_t - \frac{1}{T_e}|$

$$\alpha^*_e = \underset{\substack{\alpha\in\Delta(\mathcal{T}_e)\\ \sum_{t\in\mathcal{T}_e}\alpha_t\ell_t\in Q^\varepsilon_e}}{\arg\min} \ \text{TV}(\alpha, \text{unif}), \qquad \bar{\ell}^*_e = \sum_{t\in\mathcal{T}_e} \alpha_t \ell_t .$$

This is well defined with bounded divergence.

Lemma 16 *For $T_e > d_{\mathcal{P}}\varepsilon$, α^*_e exists and satisfies* $\text{TV}(\alpha^*_e, \textit{unif}) \le 2\varepsilon d_{\mathcal{P}} L$.

The proof is deferred to the Appendix A. We instantiate Algorithm 1 with the following choices. Let $f_t = F(\dots; \lambda_e)$ and $u^*_e = U(\mathcal{L}_e)$ be defined by

$$f_t(p) = \langle \lambda, u(p, \ell_t)\rangle, \qquad u^*_e = u(p^*(\bar{\ell}^*_e), \bar{\ell}^*_e) . \tag{4}$$

Theorem 17 *[(Efficient) Statistically Opportunistic Learner] Instantiating Algorithm 1 with the loss rule and target function defined in Equation (4) and OGD for both OO oracles yields an ε-statistically opportunistic learner with rate $\gamma(T) = \tilde{O}(d_{\mathcal{P}}^{\frac{1}{2}} T^{-\frac{1}{4}})$, $\alpha(\varepsilon) = (\varepsilon d_{\mathcal{P}})^{\frac{1}{3}}$.*

This algorithm is computationally efficient as long as finding the target U can be efficiently implemented.

The proof is again deferred to Appendix A.1. It is equivalent to the proof of Theorem 14, except that it uses Lemma 16 to obtain an error bound of $\text{err} \le O(\varepsilon d_{\mathcal{P}} L)$.

One simple condition under which Algorithm 1 is computationally efficient for the Statistically Opportunistic setting is that the response function is Lipschitz-continuous as demonstrated by the following lemma.

Lemma 18 *If the response function satisfies that $g_\lambda(\ell) = \langle \lambda, u(p^*(\ell), \ell)\rangle$ is c-Lipschitz for any $\bar{u}$, then setting $u_e^* = u(p^*(\bar{\ell}_e), \bar{\ell}_e)$, guarantees that $\gamma(T) = \tilde{O}(d_{\mathcal{P}}^{\frac{1}{2}} T^{-\frac{1}{4}})$, $\alpha(\varepsilon) = 3(\varepsilon d_{\mathcal{P}})^{\frac{1}{3}}$*

4. Improved Rates

In this section, we show that it is possible to obtain improved rates at the cost of an increase in computational complexity.

Starting from Corollary 13 the main term to improve is $\mathrm{Reg}^{\mathrm{in}}_{T_e}$. This requires the careful design of a new loss rule, since the worst-case regret bound for general linear losses as used in eq. (3) is $\sqrt{T_e}$. First, we convert the inner OO into an expert problem with arbitrary losses. We assume without loss of generality that $\mathcal{P}$ contains a ball of radius 1 and is contained in a ball of radius $d_{\mathcal{P}}$. Assume further that the loss functions f_t are 1-Lipschitz, which is for example guaranteed if $f_t(p)$ is piecewise identical to $\langle \lambda_e, u(p, \ell_{t'})\rangle$. Then there exists a finite cover $\mathcal{N}$ of size $|\mathcal{N}| \le (3/(d_{\mathcal{P}}^2 T))^{d_{\mathcal{P}}}$ such that $\min_{p\in\mathcal{N}} \sum_{t=1}^T f_t(p) \le 1 + \min_{p\in\mathcal{P}} \sum_{t=1}^T f_t(p)$ (Lattimore and Szepesvári, 2020, Lemma 20.1). This means that we can use any optimization oracle that solves the expert with loss vectors $y_t \in [-1,1]^{|\mathcal{N}|}, y_{t,p} = f_t(p)$, even if the loss functions f_t are not convex.

Definition 19 (Expert Oracle) *An Expert oracle is an algorithm that operates over a finite action set $\mathcal{N}$. It interacts with an environment over T rounds as follows: in each round $t = 1, \dots, T$, the oracle first chooses a mixed action $x_t \in \Delta(\mathcal{N})$, and then receives a loss function vector $y_t \in [-1,1]^{|\mathcal{N}|}$. The oracle provides guarantees on the regret*

$$\mathrm{Reg}_T = \sum_{t=1}^T \langle x_t, y_t\rangle - \min_{x^*\in\mathcal{N}} \sum_{t=1}^T y_{t,x^*}\,.$$

Specifically, we make use of the following following oracle.

Lemma 20 (Expert with path-length bound (Chen et al., 2021)) *Given a set of N experts in the expert game with losses bounded in $[-1,1]$. The Multi-scale Multiplicative-weight with Correction (MsMwC) algorithm instantiated with optimistic terms $m_t = \ell_{y-1}$ satisfies*

$$\sum_{t=1}^T \langle x_t - e_{i^*}, y_t\rangle \le O\left(\sqrt{P_T^* \log(NT)} + \log(NT)\right), \text{ where } P_T^* = \sum_{t=1}^{T-1} |y_{t+1,i^*} - y_{t,i^*}|\,.$$

All what remains is to design loss rules and target functions with small path-length and err term.

4.1. Improved Rates for Strictly Opportunistic Approachability

Define for $Q_e = \mathrm{conv}(\{\ell_t\}_{t\in\mathcal{T}_e})$ the solution to the maximin problem $\ell_e^* := \mathrm{argmax}_{\ell\in Q_e} \min_{p\in\mathcal{P}} \langle \lambda_e, u(p,\ell)\rangle$. We instantiate Algorithm 1 with the following choices. Let $f_t = F(\dots;\lambda_e)$ and $u_e^* = U(\mathcal{L}_e)$ be defined by

$$f_t(p) = \max_{t'\le t} \langle \lambda, u(p, \ell_t')\rangle, \qquad u_e^* = u(p^*(\ell_e^*), \ell_e^*)\,. \tag{5}$$

Theorem 21 *[Strictly Opportunistic Learner with fast rates] Algorithm 1 with the loss rule and target function defined in eq. (5), OGD as outer OCO and MsMwC as inner OO with optimal tuning is a strictly opportunistic learner with rate $\gamma(T) = \tilde{O}(d_{\mathcal{P}}^{\frac{1}{3}} T^{-\frac{1}{3}})$.*

The proof is deferred to the Appendix B.1. It uses that the err is non-negative and establishes logarithmic regret for the inner regret due to a constant upper bound on the path length with our choice of f_t.

4.2. Improved Rates for Statistically Opportunistic Approachability

In the statistical opportunistic case, we have to first adapt the target function to be consistent. Similarly to the previous section, let $Q_e^\varepsilon = Q_{\text{int}}^{\varepsilon L}(\{\ell_t\}_{t\in\mathcal{T}_e})$ and let

$$\ell_e^* := \text{argmax}_{\ell\in Q_e^\varepsilon} \min_{p\in\mathcal{P}} \langle \lambda_e, u(p,\ell)\rangle \,.$$

This change is not sufficient to control the error term (err) since a single outlier at the beginning can increase all successive function values. We have to operate on a generalized max function that allows to discard a certain number of outliers.

Definition 22 *We define the N-offset maximum over a set $\mathcal{N}$ of finite cardinality as follows.*

$$\text{off-max}_{\mathcal{N}}^N(f) = \min_{\substack{\mathcal{N}'\subset\mathcal{N} \\ |\mathcal{N}'|\geq|\mathcal{N}|-N}} \max_{x\in\mathcal{N}'} f(x)\,, \tag{6}$$

where by convention, $\text{off-max}_{\mathcal{N}}^N = 0$ *for* $|\mathcal{N}| \leq N$.

Further define $h_{p,\lambda}(t) = \langle \lambda, u(p,\ell_t)\rangle$. We instantiate Algorithm 1 with the following choices. Let $f_t = F(\ldots;\lambda_e)$ and $u_e^* = U(\mathcal{L}_e)$ be defined by

$$f_t(p) = \max\{\text{off-max}_{\{t'\in\mathcal{T}_e\,|\,t'\leq t\}}^{\varepsilon dT}(h_{p,\lambda}), \langle\lambda, u(p,\ell_t)\rangle\}, \qquad u_e^* = u(p^*(\ell_e^*),\ell_e^*)\,. \tag{7}$$

In words, we generally take the $\varepsilon d_{\mathcal{P}} T$-th largest loss we have seen so far for point p. Though when the loss of the current round is larger, default to this value to satisfy Assumption 1.

Theorem 23 (Statistically Opportunistic Learner with fast rates) *Algorithm 1 with the loss rule and target function defined in eq. (7), OGD as outer OCO and MsMwC as inner OO with optimal tuning is a ε-stochastically opportunistic learner with rate $\gamma(T) = \tilde{O}(d_{\mathcal{P}}^{\frac{1}{3}} T^{-\frac{1}{3}})$ and $\alpha(\varepsilon) = (\varepsilon d_{\mathcal{P}})^{\frac{1}{3}}$.*

The proof is deferred to the Appendix B.1. Bounding the Hedge regret relies on a careful bound of the Path-Length, while the error term is bounded by $O(\varepsilon d_{\mathcal{P}} L)$ again due to Helly's theorem.

5. Discussion on Optimality

The general algorithmic framework presented in Section 3 with instantiations from Section 4 achieves a convergence rate of $O(d_{\mathcal{P}}^{1/3} T^{-1/3})$ for the statistically opportunistic problem. In Lemma 15, we show that the dependence in d is not avoidable. It is natural to ask if the dependence on d and T is optimal? This section explores the question of optimality. We demonstrate that faster $O(T^{-1/2})$ rates are indeed possible in low-dimensional cases of the strict opportunistic approachability problem. However, we also show that these low-dimensional techniques do not readily extend to higher dimensions, leading to a discussion of current bottlenecks.

5.1. The One-Dimensional Case

In the simplest setting, where payoffs are one-dimensional ($d = 1$), the opportunistic target set $S(Q)$ is simply an interval (the only possible bounded convex set in one dimension). This structure permits a simple yet efficient algorithm that achieves the optimal $O(T^{-1/2})$ rate. The strategy, outlined in Algorithm 2, uses a no-regret learner that is reset whenever its average payoff appears to have crossed the target.

Algorithm 2 An algorithm for one-dimensional strictly opportunistic approachability

Input: Action set $\mathcal{P}$, utility function u, horizon T

Play an arbitrary action $p_1 \in \mathcal{P}$, observe ℓ_1, and set the target $c \leftarrow u(p^\star(\ell_1), \ell_1)$
Initialize a no-regret algorithm `OCO` over $\mathcal{P}$
for $t \leftarrow 2$ **to** T **do**
  Determine the side of the target: $s_t \leftarrow \operatorname{sign}\Big(\frac{1}{t-1}\sum_{i=1}^{t-1} u(p_i, \ell_i) - c\Big)$
  if $s_t \neq s_{t-1}$ *for some* s_{t-1} *in the history* **then**
    Reset `OCO`
  end
  Play action p_t from `OCO`
  Observe ℓ_t and feed the signed loss $f_t(\cdot) \leftarrow s_t\, u(\cdot, \ell_t)$ to `OCO`
end

The intuition behind this algorithm is that the sign s_t remains constant for long periods. If it were to flip frequently, the average payoff would oscillate around c. A no-regret learner, when fed losses that consistently push it in one direction (e.g., losses are always positive), will quickly learn to counteract them, preventing such oscillation. Consequently, any resets must be infrequent. The analysis of the long final epoch after the last reset, where the sign s_N is fixed, yields the desired $O(\sqrt{\log|\mathcal{P}|/T})$ rate on the distance to the target set $S(Q)$.

Theorem 24 *For an instance with payoff dimension $d = 1$ and a strictly Q-restricted adversary, Algorithm 2 with target $c \in S(Q)$ ensures* $\mathtt{dist}(\bar{u}_T, S(Q)) \leq O\Big(\sqrt{\log|\mathcal{P}|/T}\Big)$.

The proof of this theorem is deferred to Appendix C.

5.2. Fast Rates for Strictly Opportunistic Approachability in Low Dimensions

For dimensions $d > 1$, the primary challenge is that the learner must discover the geometry of the unknown adversary set $Q = \operatorname{conv}(\{\ell_t\}_{t=1}^T)$ over time. A natural approach for the strictly opportunistic case is to maintain an estimate of this set, $Q_t = \operatorname{conv}(\{\ell_1, \ldots, \ell_{t-1}\})$, and play as if $S(Q_t)$ were the true target. The total error then accumulates based on how much the set expands when a new point ℓ_t arrives, an error proportional to $\mathtt{dist}(\ell_t, Q_t)$. Remarkably, in low dimensions, the sum of these distances is bounded. The following lemma establishes this for $d = 2$.

Lemma 25 *Let $\{x_t\}_{t=1}^T$ be a sequence of points in $\mathbb{R}^2$ with $\|x_t\|_2 \leq 1$. Let $K_t = \operatorname{conv}(\{x_1, \ldots, x_{t-1}\})$ for $t \geq 2$. Then the cumulative distance is bounded:* $\sum_{t=2}^T \mathtt{dist}(x_t, K_t) = O(\sqrt{T})$.

The proof of this lemma is deferred to Appendix C. By running an approachability algorithm against the expanding target set $S(Q_t)$ and absorbing the error from the projections, we can prove the following theorem.

Theorem 26 *For strictly opportunistic approachability with a two-dimensional* [3] *action space $\mathcal{L}$, there exists an efficient algorithm with a convergence rate of $O(1/\sqrt{T})$.*

The proof of this lemma is deferred to Appendix C. The geometric approach breaks down for $d > 3$. A straightforward extension of Lemma 25 to d dimensions bounds the sum of the d-th powers of the distances, which only gives a cumulative distance of $O(T^{1-1/d})$. This rate is worse than $O(\sqrt{T})$ when $d > 3$.

Acknowledgments

We thank our colleagues and funding agencies; `\documentclass[anon]{alt2026}` automatically hides this text.

References

Jacob Abernethy and Shie Mannor. Does an efficient calibrated forecasting strategy exist? In *Proceedings of the 24th Annual Conference on Learning Theory*, pages 809–812. JMLR Workshop and Conference Proceedings, 2011.

Jacob D. Abernethy, Peter L. Bartlett, and Elad Hazan. Blackwell approachability and no-regret learning are equivalent. In *Proceedings of COLT*, volume 19 of *JMLR Proceedings*, pages 27–46, 2011.

Eshwar Ram Arunachaleswaran, Natalie Collina, Yishay Mansour, Mehryar Mohri, Jon Schneider, and Balasubramanian Sivan. Swap regret and correlated equilibria beyond normal-form games. In *Proceedings of the 26th ACM Conference on Economics and Computation*, pages 130–157, 2025.

Andrey Bernstein and Nahum Shimkin. Response-based approachability with applications to generalized no-regret problems. *Journal of Machine Learning Research*, 16(24):747–773, 2015. URL http://jmlr.org/papers/v16/bernstein15a.html.

Andrey Bernstein, Shie Mannor, and Nahum Shimkin. Opportunistic approachability and generalized no-regret problems. *Mathematics of Operations Research*, 39(4):1057–1083, 2014. ISSN 0364765X, 15265471. URL http://www.jstor.org/stable/24541001.

David Blackwell. An analog of the minimax theorem for vector payoffs. *Pacific Journal of Mathematics*, 6(1):1 – 8, 1956. doi: pjm/1103044235. URL https://doi.org/.

Sébastien Bubeck. Introduction to online optimization. *Lecture notes*, 2:1–86, 2011.

Liyu Chen, Haipeng Luo, and Chen-Yu Wei. Impossible tuning made possible: A new expert algorithm and its applications. In *Conference on Learning Theory*, pages 1216–1259. PMLR, 2021.

3. We believe this geometric argument, based on the trade-off between volume and surface area, can be extended to $d = 3$, also yielding an $O(T^{-1/2})$ rate.

Evgenii Chzhen, Christophe Giraud, and Gilles Stoltz. A unified approach to fair online learning via blackwell approachability. *Advances in Neural Information Processing Systems*, 34:18280–18292, 2021.

Christoph Dann, Yishay Mansour, Mehryar Mohri, Jon Schneider, and Balasubramanian Sivan. Pseudonorm approachability and applications to regret minimization. In Shipra Agrawal and Francesco Orabona, editors, *International Conference on Algorithmic Learning Theory, February 20-23, 2023, Singapore*, volume 201 of *Proceedings of Machine Learning Research*, pages 471–509. PMLR, 2023.

Ludwig Danzer, Branko Grünbaum, and Victor Klee. Helly's theorem and its relatives[1]. In *Convexity: Proceedings of the Seventh Symposium in Pure Mathematics of the American Mathematical Society*, volume 7, page 101. American Mathematical Soc., 1963.

Constantinos Daskalakis, Gabriele Farina, Maxwell Fishelson, Charilaos Pipis, and Jon Schneider. Efficient learning and computation of linear correlated equilibrium in general convex games. In *Proceedings of the 57th Annual ACM Symposium on Theory of Computing*, pages 542–553, 2025.

Maxwell Fishelson, Noah Golowich, Mehryar Mohri, and Jon Schneider. High-dimensional calibration from swap regret. *arXiv preprint arXiv:2505.21460*, 2025.

Dean P Foster and Rakesh V Vohra. Asymptotic calibration. *Biometrika*, 85(2):379–390, 1998.

Elad Hazan. Introduction to online convex optimization, 2023. URL https://arxiv.org/abs/1909.05207.

Joon Kwon. *Mirror descent strategies for regret minimization and approachability*. PhD thesis, Université Pierre et Marie Curie, Paris 6, 2016.

Tor Lattimore and Csaba Szepesvári. *Bandit algorithms*. Cambridge University Press, 2020.

Renato Paes Leme and Jon Schneider. Contextual search via intrinsic volumes. In *2018 IEEE 59th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 268–282. IEEE, 2018.

Shie Mannor and Vianney Perchet. Approachability, fast and slow. In *Journal of Machine Learning Research: Workshop and Conference Proceedings*, volume 30, pages 1–16, 2013.

Shie Mannor, Vianney Perchet, and Gilles Stoltz. Approachability in unknown games: Online learning meets multi-objective optimization. In *Proceedings of COLT*, volume 35, pages 339–355, 2014.

Sobhan Miryoosefi, Kianté Brantley, Hal Daumé III, Miroslav Dudík, and Robert E. Schapire. Reinforcement learning with convex constraints. In *Proceedings of NIPS*, pages 14070–14079, 2019.

Princewill Okoroafor, Bobby Kleinberg, and Wen Sun. Faster recalibration of an online predictor via approachability. In *International Conference on Artificial Intelligence and Statistics*, pages 4690–4698. PMLR, 2024.

Princewill Okoroafor, Robert Kleinberg, and Michael P Kim. Near-optimal algorithms for omniprediction. *arXiv preprint arXiv:2501.17205*, 2025.

Binghui Peng. High dimensional online calibration in polynomial time. *arXiv preprint arXiv:2504.09096*, 2025.

Appendix A. Deferred Proofs from Section 3

Proof of [Corollary 13] The outer OCO operates over the action set $\mathcal{B}_d(1)$, so $D = 2$. The norm of the gradient $\|g_e\|$ is bounded by twice the maximum norm of payoffs, which implies $G = 2$. Applying Lemma 10 to $\mathrm{Reg}_L^{\mathrm{in}}$ finishes the proof. ■

Proof of [Lemma 16] Define $Q_{\mathrm{close}} = \mathrm{conv}(\{\frac{1}{|\mathcal{T}'|}\sum_{t\in\mathcal{T}'}\ell_t \,|\, \mathcal{T}' \subset \mathcal{T}_e, |\mathcal{T}'| \geq T_e - \varepsilon d_{\mathcal{P}} T\})$. If we can show that $Q_{\mathrm{close}} \cap Q_e^{\varepsilon} \neq \emptyset$, then this implies the Lemma since for every $\mathcal{T}' \subset \mathcal{T}_e, |\mathcal{T}'| \geq T_e - \varepsilon d_{\mathcal{P}} T$ we have

$$\sum_{t\in\mathcal{T}'}\left(\frac{1}{|\mathcal{T}|'} - \frac{1}{T_e}\right) + \sum_{t\in\mathcal{T}_e\setminus\mathcal{T}'}\frac{1}{T_e}\Big) \leq 1 - \frac{T_e - \varepsilon d_{\mathcal{P}} T}{T_e} + \frac{\varepsilon d_{\mathcal{P}} T}{T_e} = 2\varepsilon d_{\mathcal{P}} L\,,$$

by using $T = LT_e$. By convexity of TV, the total variation to the uniform sample is bounded for every element in Q_{close}.

Finally we show that the intersection is non-empty. Assume the intersection is empty and let H be an open half-space containing Q_{close} but not Q_e^{ε}. By Lemma 29, the set H contains no more than εdT points. Take $\bar{\ell}$ as the mean of all points not in H. By definition, $\bar{\ell} \in Q_{\mathrm{close}}$. Hence such a half-space cannot exist, which implies that $Q_{\mathrm{close}} \cap Q_e^{\varepsilon} \neq \emptyset$. ■

Proof of [Lemma 18] While u_e^* is not a consistent target function it is sufficiently close to one as demonstrated by the following

$$\begin{aligned}
&\max_{\lambda\in B_d(1)} \min_{\ell'\in Q_{\mathrm{int}}^{\varepsilon}} \langle \bar{u}, u(p^*(\bar{\ell}_e^*), \bar{\ell}_e^*) - u(p^*(\ell'), \ell')\rangle \\
&\max_{\lambda\in B_d(1)} \langle \bar{u}, u(p^*(\bar{\ell}^*), \bar{\ell}^*) - u(p^*(\bar{\ell}_e), \bar{\ell}_e)\rangle \\
&\leq c\|\bar{\ell}_e^* - \bar{\ell}_e\| \\
&\leq c\mathrm{TV}(\alpha_{\mathrm{e}}^*, \mathrm{unif}) \max_{\mathrm{t}} \|\ell_{\mathrm{t}}\| \leq 2\mathrm{c}\varepsilon\mathrm{d}_{\mathcal{P}}\mathrm{L}
\end{aligned}$$

■

A.1. Proof of Theorems

Proof of [Theorem 14] Since $\bar{\ell}_e \in Q_{\mathrm{int}}^0$, this target function is consistent with S_{int}^0 and the loss rule satisfies Assumption 1. We additionally have err ≤ 0, as we see by taking the definition

$$\begin{aligned}
\min_{p\in\mathcal{P}}\left(\sum_{t\in\mathcal{T}_e} f_t(p) - \langle\lambda_e, u_e^*\rangle\right) &= \min_{p\in\mathcal{P}}\left(\sum_{t\in\mathcal{T}_e}\langle\lambda_e, u(p,\ell_t) - u(p^*(\bar{\ell}_e), \bar{\ell}_e))\rangle\right) \\
&= T_e \min_{p\in\mathcal{P}}\left(\langle\lambda_e, u(p,\bar{\ell}_e) - u(p^*(\bar{\ell}_e), \bar{\ell}_e))\rangle\right) \leq 0\,.
\end{aligned}$$

Applying Corollary 13 yields

$$\texttt{dist}(\bar{u}_T, S_{\text{int}}) \leq \frac{1}{T_e}\text{Reg}^{\text{in}}_{T_e} + O(L^{-\frac{1}{2}})\,.$$

Using OGD as the inner OO, we have $\text{Reg}^{\text{in}}_{T_e} \leq \tilde{O}(d_{\mathcal{P}}\sqrt{T_e})$. Tuning $L = \sqrt{T}/d_{\mathcal{P}}$ completes the proof. ■

Proof of [Theorem 17] We obtain the following bound on the error term.

$$\begin{aligned}\frac{1}{T_e}\min_{p\in\mathcal{P}}\sum_{t\in\mathcal{T}_e} f_t(p) - \langle\lambda_e, u^*_e\rangle &\leq \frac{1}{T_e}\sum_{t\in\mathcal{T}_e}\langle\lambda_e, u(p^*(\bar{\ell}^*_e), \ell_t) - \langle\lambda_e, u(p^*(\bar{\ell}^*_e), \bar{\ell}^*_e)\rangle \\ &= \sum_{t\in\mathcal{T}_e}(T_e^{-1} - \alpha^*_e(t))\langle\lambda_e, u(p^*(\bar{\ell}^*_e), \ell_t\rangle \\ &\leq \text{TV}(\alpha^*_{\text{e}}, \text{unif}) \leq 2\varepsilon \text{d}_{\mathcal{P}}\text{L}\,. \qquad \text{(Lemma 16)}\end{aligned}$$

Plugging in the OGD bound of Theorem 10 and tuning $L = O(\min\{\sqrt{T}/d_{\mathcal{P}}, (\varepsilon d_{\mathcal{P}})^{-\frac{2}{3}}\}$ completes the proof. ■

Appendix B. Deferred Proofs from Section 4

We require the following technical Lemmas.

Lemma 27 *The offmax satisfies for any $\mathcal{M} \subset \mathcal{N}$:* $\text{off-max}^N_{\mathcal{M}}(f) \geq \text{off-max}^{N+|\mathcal{N}|-|\mathcal{M}|}_{\mathcal{N}}(f)$ *as well as* $\sum_{s=0}^{N}\text{off-max}^s_{\mathcal{N}}(f) = \sum_{x\in\mathcal{N}} f(x)$.

Proof For the first statement, let $\mathcal{N}'$ be the argmin subset in the definition of $\text{off-max}^N_{\mathcal{M}}$. $\mathcal{N}'$ is also a valid subset in the minimization for $\text{off-max}^{N+|\mathcal{N}|-|\mathcal{M}|}_{\mathcal{N}}$, hence the value cannot be larger.

For the second, note that the offset max is iterating through all elements ordered by size, hence we are simply summing over the set. ■

We have the following geometric properties. We first recall the definition of an intersection set

Definition 28 *Let $\mathcal{N} \subset \mathbb{R}^d$ with $|\mathcal{N}| = T$, and fix $\varepsilon < 1/(d+1)$. Define*

$$\mathbf{Q}_\varepsilon := \Big\{\text{conv}(\mathcal{N}\setminus R) : R \subseteq \mathcal{N},\ |R| \leq \varepsilon T\Big\}, \qquad Q_{\text{int}} := \bigcap_{Q\in\mathbf{Q}_\varepsilon} Q.$$

Lemma 29 *Let $\mathcal{N} \subset \mathbb{R}^d$ with $|\mathcal{N}| = T$, and fix $\varepsilon < 1/(d+1)$. Let $H \subset \mathbb{R}^d$ be an open halfspace with $H \cap Q^\varepsilon_{\text{int}} = \emptyset$. Then*

$$|H \cap \mathcal{N}| \leq d\,\varepsilon T.$$

Proof Let $\mathcal{H}_\varepsilon$ denote the family of open halfspaces each containing at most εT points of $\mathcal{N}$. By construction of Q_{int}, for any $x \notin Q_{\text{int}}$ there exists $H_x \in \mathcal{H}_\varepsilon$ with $x \in H_x$. In particular, $H \subseteq \bigcup_{H_x\in\mathcal{H}_\varepsilon} H_x$.

Consider the family $\mathcal{H}_\varepsilon \cup \{H^c\}$, which covers $\mathbb{R}^d$ (since H^c covers Q_{int} and the H_x cover H). By the Helly–Danzer–Grünbaum–Klee theorem ((Danzer et al., 1963, Thm. 2.3)), there exists a subfamily of at most $d+1$ halfspaces that still covers $\mathbb{R}^d$. One of these must be H^c (the only halfspace covering Q_{int}), leaving at most d halfspaces from $\mathcal{H}_\varepsilon$ that cover H. Each of these contains at most εT points, so

$$|H \cap \mathcal{N}| \leq d \cdot \varepsilon T,$$

■

Lemma 30 *Let $B : \mathcal{A} \times \mathcal{L} \to \mathcal{R}$ be a bi-affine function. Let $\mathcal{N}$ be a set of points in $\mathcal{L}$ and Q_{int} be the intersection set of N points removed. It holds*

$$\min_{p \in \mathcal{P}} \max_{\ell \in Q_{int}} B(p, \ell) \geq \min_{p \in \mathcal{P}} \text{off-max}_{\mathcal{N}}^{dN}(B(p, \cdot)) \,.$$

Proof The inequality follows from Lemma 29. Define the half-space $H_p = \{\ell \in \mathbb{R}^d \,|\, B(p, \ell) > \max_{\ell' \in Q_{\text{int}}} B(p, \ell')\}$. This is a half-space due to B being bi-affine. H contains no more than dN points by Lemma 29. Hence taking the set $\mathcal{N}'$ that removes all points in H_p (and additional points at random to fill up the dN points) satisfies $\max_{\ell \in \mathcal{N}'} B(p, \ell) \leq \max_{\ell \in Q_{\text{int}}} B(p, \ell)$ This holds for any p, specifically for the argmin of the LHS. ■

B.1. Main proofs

Proof of [Theorem 23] We first ensure that U is consistent. In the construction of $Q_e^\varepsilon = Q_{\text{int}}^{\varepsilon T}(\mathcal{T}_e)$, we are removing an εL fraction of points from $\mathcal{T}$. Since $|\mathcal{T}| = T/L$, we remove a total of εT points, which is the same as in $Q_{\text{int}}^\varepsilon$. Since $\mathcal{T}_e \subset (\ell_t)_{t \in [T]}$, we have $Q_e^\varepsilon \subset Q_{\text{int}}^\varepsilon$, which ensures consistency of U.

Applying Corollary 13, with the regret bound of MsMwC, we need to bound the path length and the error term.

Path Length. Recall, given $h_p(t) = \langle \lambda_e, u(p, \ell_t) \rangle$, the path length is defined by

$$P^\star = \sum_{t=(e-1)T_e+1}^{eT_e-1} |f_{t+1}(p) - f_t(p)|$$
$$f_t(p) = \max\{\text{off-max}_{\mathcal{T}_e}^{\varepsilon dT}(h_p), h_p(t)\} \,.$$

We bound the Path length of this loss as follows. Let $\mathcal{T}_e^t = \{\ell_{t'} \,|\, t' \in [(e-1)T_e + 1, t]\}$:

$$\begin{aligned}
&|f_{t+1}(p) - f_t(p)| \\
&= |\max\{h_p(t+1), \text{off-max}_{\mathcal{T}_e^t}^{\varepsilon dT}(h_p)\} - \max\{h_p(t), \text{off-max}_{\mathcal{T}_e^{t-1}}^{\varepsilon dT}(h_p)\}| \\
&\leq \max\{h_p(t+1) - \text{off-max}_{\mathcal{T}_e^{t-1}}^{\varepsilon dT}(h_p), h_p(t) - \text{off-max}_{\mathcal{T}_e^t}^{\varepsilon dT}(h_p), \text{off-max}_{\mathcal{T}_e^t}^{\varepsilon dT}(h_p) - \text{off-max}_{\mathcal{T}_e^{t-1}}^{\varepsilon dT}(h_p)\} \\
&\leq (h_p(t+1) - \text{off-max}_{\mathcal{T}_e^{t-1}}^{\varepsilon dT}(h_p))_+ + (h_p(t) - \text{off-max}_{\mathcal{T}_e^t}^{\varepsilon dT}(h_p))_+ + \text{off-max}_{\mathcal{T}_e^t}^{\varepsilon dT}(h_p) - \text{off-max}_{\mathcal{T}_e^{t-1}}^{\varepsilon dT}(h_p)
\end{aligned}$$

Summing over t and using $(h_p(t) - \text{off-max}^{\varepsilon dT}_{\mathcal{T}_e^t}(h_p))_+ \le (h_p(t) - \text{off-max}^{\varepsilon dT}_{\mathcal{T}_e^{t-2}}(h_p))_+$ yields

$$\sum_{t=(e-1)T_e+1}^{eT_e-1} |f_{t+1}(p) - f_t(p)| \le 2\sum_{t=1}^{T}(h_p(t) - \text{off-max}^{\varepsilon dT}_{\mathcal{T}_e^{t-2}}(h_p))_+ + \underbrace{\text{off-max}^{\varepsilon dT}_{\mathcal{T}_e}(h)}_{\le 1}$$

To bound the remaining term, notice that we can assume wlog that $(h_p(t) - \text{off-max}^{\varepsilon dT}_{\mathcal{T}_e^{t-2}}(h_p)) \ge 0$. Otherwise take the subset of $\mathcal{T}_e$ for which this holds and repeat the argument below on this subset. Then

$$\begin{aligned}
\sum_{t=(e-1)T_e}^{eT_e}(h_p(t) - \text{off-max}^{\varepsilon dT}_{\mathcal{T}_e^{t-2}}(h_p)) &\le \sum_{t=(e-1)T_e}^{eT_e}\left(f_a(i_s) - \text{off-max}^{\varepsilon dT}_{\mathcal{T}_e^{t-2}}(h_p)\right) && \text{(by Lemma 27)}\\
&\le \sum_{s=1}^{T_e}\left(\text{off-max}^{s-1}_{\mathcal{T}_e}(h_p) - \text{off-max}^{\varepsilon dT+|\mathcal{T}_e|+2-s}_{\mathcal{T}_e}(h_p)\right)\\
&= \sum_{s'=0}^{\varepsilon dT+1}\text{off-max}^{s'}_{\mathcal{T}_e}(h_p) \le \varepsilon dT + 2
\end{aligned}$$

Hence the path length is bounded by $O(\varepsilon d_{\mathcal{P}} T)$. By AM-GM inequality, we have $\text{Reg}^{\text{in}}_{T_e} = O(\varepsilon d_{\mathcal{P}} T + \log|\mathcal{N}|)$.

Error term.

$$\begin{aligned}
\text{err} &= \max_e \min_{p\in\mathcal{P}} \frac{1}{T_e}\sum_{t\in\mathcal{T}_e}\left(f_t(p) - \langle\lambda_e, u_e^\star\rangle\right)\\
&= \max_e \min_{p\in\mathcal{P}} \frac{1}{T_e}\sum_{t\in\mathcal{T}_e}\left(\max\{\text{off-max}^{\varepsilon d_{\mathcal{P}} T}_{\mathcal{T}_e^t}(h_p), \langle\lambda_e, u(p,\ell_t)\rangle\} - \langle\lambda_e, u(p^\star(\bar{\ell}_e^\varepsilon), \bar{\ell}_e^\varepsilon)\rangle\right)\\
&\le \max_e \frac{1}{T_e}\sum_{t\in\mathcal{T}_e}\left(\max\{\text{off-max}^{\varepsilon d_{\mathcal{P}} T}_{\mathcal{T}_e}(h_{p^\star(\bar{\ell}_e^\varepsilon)}), \langle\lambda_e, u(p^\star(\bar{\ell}_e^\varepsilon), \ell_t)\rangle\} - \langle\lambda_e, u(p^\star(\bar{\ell}_e^\varepsilon), \bar{\ell}_e^\varepsilon)\rangle\right)\\
&\le \max_e \text{off-max}^{\varepsilon d_{\mathcal{P}} T}_{\mathcal{T}_e}(h_{p^\star(\bar{\ell}_e^\varepsilon)}) - \langle\lambda_e, u(p^\star(\bar{\ell}_e^\varepsilon), \bar{\ell}_e^\varepsilon)\rangle + \frac{1}{T_e}\sum_{t\in\mathcal{T}_e}\mathbb{I}\left(h_{p^\star(\bar{\ell}_e^\varepsilon)} > \text{off-max}^{\varepsilon d_{\mathcal{P}} T}_{\mathcal{T}_t}(h_{p^\star(\bar{\ell}_e^\varepsilon)})\right)\\
&\le \frac{1}{T_e}\sum_{t\in\mathcal{T}_e}\mathbb{I}\left(h_{p^\star(\bar{\ell}_e^\varepsilon)} > \text{off-max}^{\varepsilon d_{\mathcal{P}} T}_{\mathcal{T}_t}(h_{p^\star(\bar{\ell}_e^\varepsilon)})\right) && \text{(by Lemma 30)}\\
&\le \frac{\varepsilon d_{\mathcal{P}} T}{T_e} = \varepsilon d_{\mathcal{P}} L\,.
\end{aligned}$$

Applying Corollary 13 with the bound on the regret and error term, we have

$$\texttt{dist}(\bar{u}_T, S^\varepsilon_{\text{int}}) \le \tilde{O}\left(\frac{d_{\mathcal{P}} L}{T} + \frac{1}{\sqrt{L}} + \varepsilon d_{\mathcal{P}} L\right).$$

Setting $L = \min\left\{(T/d_{\mathcal{P}})^{\frac{2}{3}}, (\varepsilon d_{\mathcal{P}})^{-\frac{2}{3}}\right\}$ completes the proof. ■

Proof of [Theorem 21] The choice of loss rule satisfies Assumption 1 and the target function is consistent with S^0_{int}, hence we can apply Corollary 13 and only need to bound $\text{Reg}^{\text{In}}_{T_e}$ and err. For

the error, we have

$$\begin{aligned}\min_{p\in\mathcal{P}} \sum_{t\in\mathcal{T}_e} f_t(p) &\le T_e \min_{p\in\mathcal{P}} \max_{t\in\mathcal{T}_e} \langle \lambda_e, u(p,\ell_{t'})\rangle \\ &= T_e \max_{\ell\in Q_e} \min_{p\in\mathcal{P}} \langle \lambda_e, u(p,\ell)\rangle \\ &= T_e \min_{p\in\mathcal{P}} \langle \lambda_e, u(p,\ell_e^*)\rangle \\ &\le T_e \langle \lambda_e, u(p^*(\ell_e^*),\ell_e^*)\rangle\,,\end{aligned}$$

hence err ≤ 0. For the inner optimization, we need to bound the path length. By monotonicity of the losses, we have $f_{t+1} \ge f_t$, hence for any $p \in \mathcal{N}$: $\sum_{t=(e-1)T_e+1}^{eT_e-1} |f_{t+1}(p) - f_t(p)| = \sum_{t=(e-1)T_e+1}^{eT_e-1} (f_{t+1}(p) - f_t(p)) = f_{eT_e}(p) - f_{(e-1)T_e+1}(p) \le 2$. The inner regret is $\mathrm{Reg}^{\mathrm{in}}_{T_e} = O(\log|\mathcal{N}|)$. Plugging these in, we have

$$\texttt{dist}(\bar{u}_T, S_t) \le \frac{1}{T_e}\log(|\mathcal{N}|) + \sqrt{1/L}\,.$$

Using a discretization of size $\log|\mathcal{N}| = O(d_{\mathcal{P}}\log(T))$ and tuning $L = (T/d_{\mathcal{P}})^{\frac{2}{3}}$ completes the proof. ■

Appendix C. Deferred Proofs from Section 5

Proof of [Theorem 24] Let N be the time of the last reset, making $s_t = s_N$ for all $t \ge N$. Let $c^* \in S(Q)$ be the point closest to $\bar{u}_T$.

The no-regret guarantee of the OCO instance over the final epoch $\{N,\dots,T\}$ of length $T' = T - N + 1$ implies:

$$s_N \sum_{t=N}^{T} (u(p_t,\ell_t) - c^*) \le \mathrm{Reg}_{T'} = O(\sqrt{\log|\mathcal{P}|T'})$$

We use this to bound one side of the total deviation. Since resets are infrequent for a no-regret learner, we can treat $N = O(1)$, which means the initial sum $\sum_{t=1}^{N-1}(u_t - c^*)$ is also $O(1)$ and $\mathrm{Reg}_{T'} = O(\sqrt{\log|\mathcal{P}|T})$.

$$\begin{aligned}s_N T(\bar{u}_T - c^*) &= s_N\left(\sum_{t=1}^{N-1}(u_t - c^*) + \sum_{t=N}^{T}(u_t - c^*)\right) \\ &\le O(1) + \mathrm{Reg}_{T'} = O(\sqrt{\log|\mathcal{P}|T})\end{aligned}$$

This gives the upper bound on the signed distance: $s_N(\bar{u}_T - c^*) \le O\left(\sqrt{\log|\mathcal{P}|/T}\right)$.

For the lower bound, the no-reset condition at $t = T$ is $s_N(\bar{u}_{T-1} - c) \ge 0$.

$$\begin{aligned}s_N(\bar{u}_T - c) &= s_N\left(\frac{T-1}{T}\bar{u}_{T-1} + \frac{u_T}{T} - c\right) \\ &= \frac{T-1}{T}s_N(\bar{u}_{T-1} - c) + \frac{s_N(u_T - c)}{T} \ge \frac{s_N(u_T - c)}{T} \ge -O(1/T)\end{aligned}$$

Assuming the chosen target c is representative of the true target set $S(Q)$, the upper and lower bounds together imply that the absolute distance $|\bar{u}_T - c^*|$ is small. The two bounds combine to show:

$$\texttt{dist}(\bar{u}_T, S(Q)) = |\bar{u}_T - c^*| \leq O\left(\sqrt{\frac{\log|\mathcal{P}|}{T}}\right)$$

■

Proof of [Lemma 25] We will show that $\sum_{t=2}^{T} d(x_t, K_t)^2 = O(1)$. It then follows by Cauchy-Schwartz that $\sum_{t=2}^{T} d(x_t, K_t) = O(\sqrt{T})$.

For any $t \geq 2$, let y_t be the Euclidean projection of x_t onto K_t. Let h be the line passing through y_t perpendicular to x_t. The line h intersects the next set $K_{t+1} = \text{conv}(K_t \cup \{x_t\})$ in some line segment S of width w. We consider two cases:

- **Case 1:** $w \geq d(x_t, K_t)$**.** In this case, we will argue that the area of K_t increases by at least $\Omega(d(x_t, K_t)^2)$ upon adding the point x_t. In particular, note that $K_{t+1} \setminus K_t$ contains the triangle with base S and apex x_{t+1}, which has width w and height $d(x_t, K_t)$. The area of this triangle is therefore at least $\Omega(w \cdot d(x_t, K_t)) = \Omega(d(x_t, K_t)^2)$.

- **Case 2:** $w < d(x_t, K_t)$**.** In this case, we will argue that the perimeter of K_t increases by at least $\Omega(d(x_t, K_t))$ upon adding the point x_t. In particular, let $K' = \text{conv}(K_t \cup S)$; note that $K_t \subseteq K' \subseteq K_{t+1}$, and so $\text{Perim}(K_t) \leq \text{Perim}(K') \leq \text{Perim}(K_{t+1})$. But now, $\text{Perim}(K_{t+1}) - \text{Perim}(K') \geq 2 \cdot d(x_t, K_t) - w$ (since it gains the two legs of the previously described triangle, each of which have length at least $d(x_t, K_T)$ but loses the base S).

Since both the area and perimeter of any convex subset of the unit disk is upper bounded by a constant, this implies that $\sum_{t=2}^{T} d(x_t, K_t)^2$ is upper bounded by a constant, as desired. ■

Proof of [Theorem 26] The algorithm maintains an evolving estimate of the adversary's action set, $Q_t = \text{conv}(\{\ell_1, \ldots, \ell_{t-1}\})$. At each step t, it runs a standard approachability algorithm, feeding it the projection of the true action, $\hat{\ell}_t = \Pi_{Q_t}(\ell_t)$, instead of ℓ_t itself.

The total error of the true average payoff $\bar{u}_T$ is decomposed using the triangle inequality, where $\hat{u}_T$ is the average payoff from the projected game:

$$\texttt{dist}(\bar{u}_T, S(Q_{T+1})) \leq \underbrace{\|\bar{u}_T - \hat{u}_T\|}_{\text{Projection Error}} + \underbrace{\texttt{dist}(\hat{u}_T, S(Q_{T+1}))}_{\text{Approachability Error}}$$

The projection error is bounded by the cumulative distance of the projections. By the bilinearity property of the payoff function and Lemma 25:

$$\|\bar{u}_T - \hat{u}_T\| \leq \frac{1}{T}\sum_{t=1}^{T} \texttt{dist}(\ell_t, Q_t) = \frac{1}{T}O(\sqrt{T}) = O\left(\frac{1}{\sqrt{T}}\right).$$

The approachability error follows the standard rate of $O(1/\sqrt{T})$. ■