Survey of Online Learning and Approachability Theory

Alexandre Kaiser

Abstract

Blackwell's approachability theory serves as an important foundation for current work in online learning due to the problem of regret minimization. This survey presents a comprehensive analysis of three seminal papers that have significantly advanced the field of approachability theory and online learning. The papers examined include Bernstein and Shimkin (2015), Mannor et al. (2014) and Abernethy (2011). The first paper builds upon Blackwell's approachability conditions to introduce a response-based algorithm for approachability that avoids the need for Blackwell's oracle. We then examine the extension to the response-based algorithm by Mannor et al. in the context of partial monitoring, which removes the need for any oracles. This survey then covers the link that Abernethy makes between approachability and online linear optimization problems that have proven incredible influential for applying online machine learning algorithms.

Keywords: online learning, approachability, regret

1 Introduction

Machine learning often deals with multi-objective optimization problems such as multi-class classification, calibration and constrained optimization, just to name a few. It is not clearly apparent that it would be reasonable to optimize for multiple objectives, nevertheless Blackwell's approachability theory, from his work in game theory following Von Neumann's minimax theorem [6], presents a way to consider target sets in any dimension $\mathbb{R}^d$ as the focus of our learning, and provides algorithms to approach those target sets efficiently.

In this survey, we will explore the various learning algorithms that are approachable, then we will discuss some important conversions from approachability problems to problems of interest such as no-regret learning and online linear optimization. So far, most extensions of approachability to other problems utilize Blackwell's original approachability algorithm. Perhaps the other algorithms for approachability that we discuss will make certain extensions of approachability computationally feasible for some applications.

As opposed to discussing the work that's been done to reduce the learning bounds of particular approachable online learning algorithms, this survey focuses on exploring the diversity of approachability problems that have algorithmic solutions. At the end of our survey we discuss the problem of online linear optimization, which has been used extensively in practice, notably its Follow-The-Regularized-Leader (FTRL) formulation which applies online gradient descent to a variety of learning problems. It is worth noting that there has been a considerable amount of work done on FTRL algorithms, most recently by Kwon (2021) [7] and Dann et al. (2023) [4], to tune the regularizing function to particular tasks, as well as to reduce the resulting regret bounds.

2 Blackwell

To review Blackwell's approachability problem [3], we consider a sequential two-player game with vector rewards between an agent and an opponent.* Each player has a set of actions, let $\mathcal{A}$ be the set of actions by the agent and $\mathcal{B}$ be the set of action by the opponent. At each round n, starting at 1, both players will select an action $a_n \in \mathcal{A}$ and $b_n \in \mathcal{B}$ and observe a vector of rewards $r(a_n, b_n) \in \mathbb{R}^d$.

Each player can sample their actions from what we call mixed actions $p_n \in \Delta(\mathcal{A})$ and $q_n \in \Delta(\mathcal{B})$, where $\Delta(\mathcal{A})$ and $\Delta(\mathcal{B})$ denotes the set of distributions across the sets $\mathcal{A}$ and $\mathcal{B}$ respectively. Let us denote the strategy† $\pi_n : (\mathcal{A} \times \mathcal{B})^{n-1} \to \Delta(\mathcal{A})$ to be the agent's process of determining a mixed action p_n as a function of the previous rounds of the game. Similarly, note the opponent's strategy $\sigma_n : (\mathcal{A} \times \mathcal{B})^{n-1} \to \Delta(\mathcal{B})$.

To distinguish between actual and expected rewards, let $R = r(a, b)$ be the actual rewards from actions $a \in \mathcal{A}$ and $b \in \mathcal{B}$. As a shorthand, we will write the expected rewards $r(p, q)$, where

$$r(p, q) = \sum_{a \in \mathcal{A}} \sum_{b \in \mathcal{B}} p(a)q(b)r(a, b)$$

Similarly we will denote $r(p, b) = \sum_{a \in \mathcal{A}} p(a)r(a, b)$ the expected reward from the mixed action p and the pure action b.

In approachability, we are interested in the average actualized reward, namely

$$\bar{R}_n = \frac{1}{n} \sum_{k=1}^{n} r(a_n, b_k),$$

and whether an algorithm can lead the long-run $\bar{R}_n$ into a closed target set S of rewards.

Definition 2.1. *Approachable sets*
A closed set $S \subseteq \mathbb{R}^d$ is approachable by the agent if there exists a strategy π of the agent such that average actual rewards $\bar{R}_n$ converges to S. That is, the shortest Euclidean distance between $\bar{R}_n$ and S, denoted by $d(\bar{R}_n, S)$, goes to zero almost surely for every strategy σ of the opponent at a uniform rate over the opponent's strategies. We can say, for every $\epsilon > 0$ there exists an integer N such that

$$\mathbb{P}^{\pi,\sigma}\{\sup_{n \geq N} d(\bar{R}_n, S) \geq \epsilon\} \leq \epsilon$$

for any strategy σ of the opponent.

Definition 2.2. *B-sets*
A closed set $S \subseteq \mathbb{R}^d$ will be called a B-set if for every $x \notin S$ there exists a mixed action p^* such that $r(p^*, q) \in H$ where H is the half-space tangent to the projection of x onto S and $H \supset S$.

Definition 2.3. *D-sets*
A closed set $S \subseteq \mathbb{R}^d$ will be called a D-set if for every $q \in \Delta(\mathcal{B})$ there exists a mixed strategy $p \in \Delta(\mathcal{A})$ such that $r(p, q) \in S$. We shall refer to such p as a response of the agent to q.

*In applications of approachability theory, notably in online learning, the opponent is often referred to as Nature since we are considering our learning agent's ability to learn from the arbitrary actions of Nature. See Section 6 and Section 7 for more details.

†in the context of online learning, the strategy π_n of the agent is equivalent to the learning algorithm

Consider the case where a B-set oracle $\mathcal{O} : \mathbb{R}^d \to \Delta(\mathcal{A})$ gives us the mixed action p^* as a function of a reward (in practice, it is the average actual reward) and a D-set oracle $\mathcal{D} : \Delta(\mathcal{B}) \to \Delta(\mathcal{A})$ gives us the response p^* to q.

Theorem 2.4. *Blackwell's Theorem*

1. ***Primal condition.*** *A B-set is approachable.*
2. ***Dual condition.*** *A closed set S is approachable only if it is a D-set.*
3. ***Convex sets.*** *Convex D-sets are B-sets, and thus approachable.*

From Blackwell's Primal condition, we directly obtain the following approachability algorithm.

Algorithm 1: Blackwell approachability

Input a B-set Oracle $\mathcal{O}$

Initialization: At time step n=1, use arbitrary mixed action p_1 and set an arbitrary target point $r_1^* \in S$.

At time step $n = 2, 3, \ldots$

1. $$p_n = \begin{cases} \mathcal{O}(\bar{R}_n), & \text{if } r_{n-1} \notin S \\ \text{arbitrary}, & \text{otherwise} \end{cases}$$
2. Observe $r_n = r(p_n, b)$

From this point on, we will replace the round n with the time t to have notation more consistent with online learning. For notation such as average reward up until round n (previously written as $\bar{R}_n$) we will now refer to the average reward by time T, written as $\bar{R}_T$.

3 Response-based

Let us assume that S is a closed, convex and approachable set. It follows from Theorem 2.4 that S is a D-set, so that there exists a response oracle $\mathcal{D}$ for any mixed action q by the opponent. The target set S is defined by the convex hull and its response oracle $\mathcal{D}$, where $S \triangleq \text{conv}\{r(\mathcal{D}(q), q), q \in \Delta(\mathcal{B})\}$. Bernstein and Shimkin [2] introduced an approachability algorithm that utilizes the response oracle, which, for some applications, alleviates the need to compute the projection of the reward onto the target set that is required to compute the B-set oracle used in Blackwell's algorithm.

Their algorithm (algorithm 2) makes use of a steering vector that finds a direction from the current average reward $\bar{R}_T$ to a point in S that is determined by the response oracle. The steering vector is used to formulate a scalar game, which by Von Neumann's minimax theorem [6] is solvable.

Algorithm 2: Response-based approachability

Input a D-set oracle $\mathcal{D}$

Initialization: At time step $t = 1$, use arbitrary mixed action p_1 and set an arbitrary target point $r_1^* \in S$.

At time step $t = 2, 3, \ldots$

1. Set an approachability direction

$$\lambda_{t-1} = \bar{r}_{t-1}^* - \bar{r}_{t-1},$$

where

$$\bar{r}_{t-1} = \frac{1}{t-1}\sum_{k=1}^{t-1} r(p_k, q_k), \quad \bar{r}_{t-1}^* = \frac{1}{t-1}\sum_{k=1}^{t-1} r_k^*$$

are, respectively, the average (smoother) reward vector and the average target point.

2. Solve the zero-sum matrix game with payoff matrix defined by $r(a, b)$ projected in the direction λ_{t-1}. Namely, find the equilibrium strategies p_t and q_t^* that satisfy

$$p_t \in \operatorname*{argmax}_{p\in\Delta(\mathcal{A})} \min_{q\in\Delta(\mathcal{B})} \lambda_{t-1} \cdot r(p, q),$$

$$q_t^* \in \operatorname*{argmin}_{q\in\Delta(\mathcal{B})} \max_{p\in\Delta(\mathcal{A})} \lambda_{t-1} \cdot r(p, q),$$

3. Pick $p_t^* = \mathcal{D}(q_t^*)$, and set the target point $r_t^* = r(p_t^*, q_t^*)$.

To present the convergence result from algorithm 2, let us define the reward span,

$$\rho = \max_{a,b,a',b'} ||r(a, b) - r(a', b')||$$

where $||\cdot||$ is the Euclidean norm.

Theorem 3.1. *For a closed, convex and approachable set S, if an agent follows the strategy specified in Algorithm 2 then*

$$d(\bar{r}_T, S) \le ||\lambda_T|| \le \frac{\rho}{\sqrt{T}}, \quad T \ge 1, \tag{1}$$

for any strategy of the opponent.

Proof. Suppose that at each step $t \ge 1$, the agent chooses the triple (p_t, p_t^*, q_t^*) so that,

$$\lambda_{t-1} \cdot (r(p_t, b) - r(p_t^*, q_t^*)) \ge 0, \quad \forall b \in \mathcal{B}, \tag{2}$$

Then $||\lambda_t|| \le \frac{\rho}{\sqrt{t}}$ for $t \ge 1$.

If we know p_t^* is chosen from the response oracle $\mathcal{D}$ then

$$d(\bar{r}_T, S) \le ||\lambda_T||, \quad \text{for } T \ge 1,$$

In order to prove the inequality in equation 2, we show that it is implied from the minimax theorem and our choice of (p_t, q_t^*)

$$\begin{aligned}\lambda_{t-1} \cdot r(p_t, b_t) &\geq \max_{p \in \Delta(\mathcal{A})} \min_{q \in \Delta(\mathcal{B})} \lambda_{t-1} \cdot r(p, q) \\ &= \min_{q \in \Delta(\mathcal{B})} \max_{p \in \Delta(\mathcal{A})} \lambda_{t-1} \cdot r(p, q) \\ &\triangleq \max_{p \in \Delta(\mathcal{A})} \lambda_{t-1} \cdot r(p, q_t^*),\end{aligned}$$

□

4 Unknown Games

In both previous formulations of approachability, our algorithms depended on oracles to give us our next mixed action p_t. For Blackwell's algorithm, our B-set oracle gave us p_t as function of current reward, and in our response-based algorithm our D-set oracle gave us p_t as the response to q_t^* to steer our learning. As we will discuss in Section 6, there are problems which deal with two sources of unknowness that make it infeasible to construct or obtain the required oracles for the previous algorithms. The first source of unknowness is that the target set S might be unknown. It could be that some natural target set is proven to be unachievable, or perhaps not ambitious enough. The second source of unknowness is that the structure of the game could be unknown (or even non-existent). In that case, we can not observe the opponent's set of decisions, nor their strategy in hindsight.

Let us reconsider the setting of approachability for unknown games. The game is still operating in discrete sequential rounds which we denote as t. We still have our set of actions $\mathcal{A}$ and we can still consider mixed actions $p_t \in \Delta(\mathcal{A})$ at each time t. However, since we can't observe our opponents actions, then there is no concept of the expected reward $r(p, q)$. Instead, we consider an alternative expected reward

$$r_t(p) = \sum_{a \in \mathcal{A}} p(a) m_{t,a} \triangleq p \odot m_t,$$

where $m_{t,a}$ represents the reward observed at time t for the agent's action $a \in \mathcal{A}$. We impose the restriction that $m_{t,a} \in \mathcal{K}$, where $\mathcal{K}$ is a convex and bounded set of $(\mathbb{R}^d)^{\mathcal{A}}$. Thus $r_t(p)$ is interpreted as the expected reward at time t of using a mixed action p. Note that even though we do not observe the opponent's actions, we will still operate under the assumption that we can observe the reward at time t for any action $a_t \in \mathcal{A}$.

To solve for the unknown target set, Mannor et al. start by considering the smallest target set in hindsight and then relax it. To do so, let's start by defining a base target set S and its ℓ_p-norm expansion $S_\alpha = \{r \in \mathbb{R}^d : \exists s \in S \text{ s.t. } ||r - s||_p \leq \alpha\}$ for all $\alpha \geq 0$. We can also think of the expansion set as the set of rewards with a distance α of the base target set. Next we fix a target function $\phi : \mathcal{K} \to \mathbb{R}^+$ which takes the average actual rewards $\bar{m}_T = \frac{1}{T} \sum_{t=1}^{T} m_t$ and returns the α used for the expansion of the target set, written as $S_{\phi(\bar{m}_t)}$.

Mannor et al. [5] show that the optimal ϕ^*, which minimizes the distance to the base target set with respect to the optimal $p^* \in \Delta(\mathcal{A})$ in hindsight, is unachievable (there does not exist an approachable algorithm for the target set S_{ϕ^*}). Nevertheless, if we expand the target set S_α by taking the concavification of ϕ^*, defined as the least concave function above ϕ^*, the problem becomes achievable but not ambitious enough.

Let the concavification of ϕ^* be written as

$$\phi^*(m) = \sup\left\{ d\left(\sum_{i=1}^{N} \lambda_i p^*(m_i) \odot m_i, S\right) : N \geq 1 \text{ and } \sum_{i=1}^{N} \lambda_i m_i = m \right\}$$

Proof. To prove that the concavification of ϕ^* is achievable, let us introduce the graph $\mathcal{G}_\phi$ of set-valued mapping $m \in \mathcal{K} \to S_{\phi(m)}$,

$$\mathcal{G}_\phi = \{(m, r) \in \mathcal{K} \times \mathbb{R}^d \text{ s.t. } r \in S_{\phi(m)}\}$$

Definition 4.1. *A continuous target function ϕ is called achievable if the target set $S_{\phi(\bar{m}_t)}$ is approachable by the agent. Equivalently, the target function ϕ is called achievable if the associated graph $\mathcal{G}_\phi$ is approachable for the payoff $(p, m) \in \Delta(\mathcal{A}) \times \mathcal{K} \to (m, r)$. We can write $(\bar{m}_T, \bar{r}_T) \to \mathcal{G}_\phi$ as $T \to \infty$.*

It suffices to prove that $\mathcal{G}_{\phi^*}$ is a B-set, that is for all $m \in \mathcal{K}$ there must exist $p \in \Delta(\mathcal{A})$ such that $(m, p \odot m) \in \mathcal{G}_\phi$. By construction, we have p^* that satisfies $(m, p^* \odot m) \in \mathcal{G}_\phi$ thus $\mathcal{G}_\phi$ is a B-set and equivalently it is approachable. □

Now we can present a general class of ambitious enough target function by introducing a parameter Ψ, the response function to replace the optimal p^* that we get from ϕ^*. For an arbitrary response function Ψ, we can define the target function as,

$$\phi^{\Psi}(m) = \sup\left\{ d\left(\sum_{i=1}^{N} \lambda_i \Psi(m_i) \odot m_i, S\right) : N \geq 1 \text{ and } \sum_{i=1}^{N} \lambda_i m_i = m \right\}$$

By construction, the target function ϕ^{Ψ} gives us the approachable set $S_{\phi^{\Psi}}$. We can now formulate an approachability algorithm for unknown games with respect to our choice of response function Ψ.

As in Algorithm 4 in Section 7, the following algorithm for unknown games relies on an auxiliary regret-minimizing strategy $\mathcal{R}^{\ddagger}$ for scalar payoffs m'_t. Let $\mathcal{R} : (\mathcal{K})^{t-1} \to \Delta(\mathcal{A})$ such that for bounded payoffs $m'_t \in [-B, B]$ for $B > 0$ we have,

$$\max_{p \in \Delta(\mathcal{A})} \sum_{t=1}^{T} p \odot m'_t \leq 4B\sqrt{T \ln A} + \sum_{t=1}^{T} p_t \odot m'_t$$

The algorithm $\mathcal{R}$ is said to have a sublinear worst-case guarantee, that the regret of $\mathcal{R}$ is $o(T)$.

[‡]The regret-minimizing strategy $\mathcal{R}$ is equivalent to the approachable strategy π from previous sections

Algorithm 3: Block response-based approachability

Input a regret minimizing strategy $\mathcal{R}$ (with initial action p_1) and a response function Ψ

Initialization: At time step n=1, use initial action p_1 from $\mathcal{R}$ and observe m_1

For block $n = 2, 3, \ldots$

1. Compute the total discrepancy δ_n at the beginning of black n (that is, till the end of block $n-1$),
$$\delta_n = \bar{r}_{n-1} - \bar{r}^{(\Psi)}_{n-1}$$
where
$$\bar{r}_{n-1} = \sum_{k=1}^{n-1}\sum_{t=1}^{k} p_t \odot m_t, \quad \bar{r}^{(\Psi)}_{n-1} = \sum_{k=1}^{n-1} k\psi(\bar{m}^{(k)}) \odot \bar{m}^{(k)}$$
where
$$\bar{m}^{(k)} = \frac{1}{k}\sum_{t=1}^{k} m_{(k,t)}$$

2. Run a fresh instance $\mathcal{R}_n$ of $\mathcal{R}$ for n rounds as follows:
set $p_{(n,1)} = p_1$; then, **for time** $t = 1, \ldots, n$,
   (a) play $p_{(n,t)}$ and observe $m_{(n,t),a} \in (\mathbb{R}^d)^{\mathcal{A}}$
   (b) feed $\mathcal{R}_n$ with the vector payoff $m'_{(n,t),a}$ with components given, for $a \in \mathcal{A}$, by
$$m'_{(n,t),a} = -\langle \delta_n, m_{(n,t),a} \rangle$$
   (c) obtain from $\mathcal{R}_n$ a mixed action $p_{(n,t+1)}$

Note that the subscript (n,t) refers to the time at block n for time in the block of t

Theorem 4.2. *For all response functions Ψ, and an agent following algorithm 3, we have*
$$d(\bar{r}_T, S_{\phi^\Psi(\bar{m}_T)}) \leq 10T^{-1/4}\ln A + 3\rho T^{-1/2},$$
where ρ is the span of elements in $\mathcal{K}$.

To prove Theorem 4.2, we use induction with respect to each block.

Proof. We are able to upper bound the direction δ_n by the span by a function $\beta(n)$ so that
$$||\delta_{n+1}||_2^2 \leq \beta(n).$$
For $n = 1$ we define $\beta(1) = 4\rho^2$.

After one induction step, we obtain
$$\begin{aligned} ||\delta_{n+2}||_2^2 &= ||\delta_{n+1}||_2^2 \\ &+ 2\left(-\sum_{t=1}^{n+1} p_{(n+1,t)} \odot m'_{(n+1,t)} + \sum_{t=1}^{n+1} p^{(n+1)} \odot m'_{(n+1,t)}\right) \\ &+ ||\sum_{k=1}^{n+1}\sum t = 1^k p_t \odot m_t - (n+1)\Psi(\bar{m}^{(n+1)}) \odot \bar{m}^{(n+1)}||_2^2 \end{aligned}$$

Now, by applying the Cauchy-Schwartz inequality for all a and t,

$$|m'_{(n+1,t),a} \leq ||\delta_{n+1}||_2 ||m_{(n+1,t),a}||_2 \leq \rho\sqrt{\beta(n)}$$

By putting everything together, we prove that,

$$\beta(n+1) = \beta(n) + 8\rho\sqrt{\beta(n)}\sqrt{\ln A} + 4\rho^2(n+1)^2$$

which can be bounded by

$$||\delta_{n+1}||_2^2 \leq \beta(n) \leq 32\rho^2(\ln A)n^3$$

Finally, we have approachability bounds

$$\begin{aligned} d(\bar{r}_T, S_{\phi^\Psi(\bar{m}_T)}) &\leq \frac{1}{T}||\delta_{n_T+1}|| + 2\rho\frac{n_T}{T} \\ &\leq \frac{1}{T}\rho\sqrt{32n_T^3 \ln A} + 2\rho\frac{n_T}{T} \\ &\leq 10T^{-1/4}\ln A + 3\rho T^{-1/2} \end{aligned}$$

□

5 Link to regret

It has been shown by many people (Blackwell, Hannan, and recently Abernethy [1]) that many regret minimization problems can be formulated as approachability problems, which implies the existence of a no-regret strategy.

Let's first consider the simple formulation of regret for scalar rewards (referred to as utility) $u = r(p, q)$. The decision sets $\mathcal{A}$ and $\mathcal{B}$ are the same as in previous sections, they depend on the particular no-regret problem. We will denote the average utility as $\bar{U}_T$ and the empirical distribution of q in hindsight as

$$\bar{q}_T(b) \triangleq \frac{1}{T}\sum_{t=1}^{T} \mathbb{I}\{b = b_t\}$$

where $\mathbb{I}$ is the indicator function.

Let the best average utility in hindsight be defined as the maximum utility achieved by a pure strategy, namely

$$u^*(\bar{q}_T) \triangleq \max_{a\in\mathcal{A}} \frac{1}{T}\sum_{t=1}^{T} u(a, \bar{q}_T)$$

Definition 5.1. *A strategy of the agent is termed a Hannan consistent no-regret algorithm if*

$$\lim_{T\to\infty} \mathbb{P}[\sup_{\bar{q}_T}(u^*(\bar{q}_T) - \bar{U}_T) \leq 0] = 1$$

for any opponent strategy.

Let S be the set of pairs $\{(u, \bar{q}) : u \leq u^*(\bar{q}) \text{ and } \bar{q} \in \Delta(\mathcal{B})\}$, the set of no-regret rewards for a given empirical distribution $\bar{q}$. If we define the vector of rewards as the pair of $(\bar{U}_T, \bar{q}_T)$ then we have effectively converted this simple regret minimization problem into an approachability problem.

We can expand our definition for regret in d-dimensions similarly. Let us reformulate our regret as the distance between the average rewards from strategy $\mathcal{R}$ and the negative orphant set $(-\infty, 0]^d$, namely

$$\text{Regret}_T(\mathcal{R}) \triangleq \sup_{\bar{q}_T} d(u^*(\bar{q}_T) - \bar{U}_T, (-\infty, 0]^d)$$

Since the above formulations for regret minimization can be clearly converted to an approachability problem, we can utilize the algorithms surveyed in the previous sections to implement a no-regret strategy. Note that we could use other distance metrics and other target sets in place of the ones used above.

6 Link to constrained optimization

Consider the same Blackwell game as before for a scalar reward $u : \mathcal{A} \times \mathcal{B} \to \mathbb{R}$, which we will refer to as the utility function, and a vector-valued cost function $c : \mathcal{A} \times \mathcal{B} \to \mathbb{R}^s$. Assume we are given a closed convex set $\Gamma \subseteq \mathbb{R}^s$ of allowed long-term average cost. We say that the constraint set is feasible if there exists a D-set oracle for any opponent action q such that the expected cost $c(p, q) \in \Gamma$.

The constrained optimization problem can be considered an approachability problem where the reward is a concatenation of the utility and the cost $r(p, q) = u(p, q) \oplus c(p, q) \in \mathbb{R}^{s+1}$. Then the target set S would be a concatenation of our target utility and the constraint set.

Considering the goal of maximizing utility, it might be unclear to have a feasible and ambitious target utility. Mannor et al. [5] proposed using unknown games to solve for the target set in constrained optimization problems. Algorithm 3 offers the construction of an ambitious and feasible target set as a function of our choice of response function ψ.

7 Link to Online Linear Optimization

Online Linear Optimization (OLO) is a particular online convex optimization problem where an agent is trying to minimize a cumulative linear loss function.

Let the set of decisions by the agent be a compact convex decision set $\mathcal{A} \subset \mathbb{R}^d$. Each action $x \in \mathbb{R}^d$ is a weight vector that is trying to optimize the linear loss for a no-regret algorithm. The opponent will play an action $f_t \in \mathbb{R}^d$ at time t and we will observe the reward $r(x, f) = \langle f, x \rangle$.

The goal of online linear optimization is to minimize the average reward $\bar{R}_n$ by implementing a no-regret strategy $\mathcal{R}$ defined below.

Definition 7.1. *Given an OLO algorithm $\mathcal{R}$ and a sequence of loss vectors $f_1, f_2, \cdots \in \mathbb{R}^d$ let us define the regret as*

$$\textit{Regret}(\mathcal{R}) = \sum_{k=1}^{n} \langle f_k, x_k \rangle - \min_{x \in \mathcal{A}} \sum_{k=1}^{n} \langle f_k, x \rangle$$

To prove the equivalence of approachability and regret minimization for OLO, we need to introduce some notation. Let $B_2(r)$ be the ℓ_2-norm ball of radius r. We will then convert the target set into a cone by concatenating the $x \in \mathcal{A}$ with some κ to work with the resulting cone $\tilde{\mathcal{A}} \triangleq \kappa \times \mathcal{A}$. The conversion of the target set into a cone is vital to prove that there exists a B-set oracle for the approachability problem formulated in Algorithm 4.

Definition 7.2. *Given any set $K \subset \mathbb{R}^d$, define the conic hull $C = \text{cone}(K) \triangleq \{\alpha x : \alpha \geq 0,\ x \in K\}$. We also define the polar cone $C^0 \triangleq \{\theta \in \mathbb{R}^d : \langle \theta, x \rangle \leq 0$ for all $x \in C\}$*

Lemma 7.3. *If C is a convex cone then*

1. $(C^0)^0 = C$
2. *Any given hyperplane H of C^0 can be written as $\{\theta \in \mathbb{R}^d : \langle \theta, x \rangle = 0\}$ for some unique vector $x \in C$ (up to a scaling factor).*

Lemma 7.4. *For every convex cone C in $\mathbb{R}^d$*

$$d(x, C) = \max_{\theta \in C^0 \cap B_2(1)} \langle \theta, x \rangle$$

In order to not lose information about the underlying set when considering the process of converting it into a cone, we shall embed the set into a higher dimension and instead consider the $\text{cone}(\kappa \times \mathcal{K}) \subset \mathbb{R}^{d+1}$ where κ is the diameter of $\mathcal{K}$.

Lemma 7.5. *Consider a compact convex set $\mathcal{K} \subset \mathbb{R}^d$ and $x \notin \mathcal{K}$. Let $\tilde{x} \triangleq \kappa \oplus x$ and $\tilde{\mathcal{K}} \triangleq \{\kappa\} \times \mathcal{K}$. Then we have,*

$$d(\tilde{x}, \text{cone}(\tilde{\mathcal{K}})) \leq d(x, \mathcal{K}) \leq 2d(\tilde{x}, \text{cone}(\tilde{\mathcal{K}}))$$

Algorithm 4: Conversion of approachability to OLO

Input a compact convex decision set $\mathcal{K} \subset \mathbb{R}^d$, and an approachability algorithm $\mathcal{R}$

Initialization: Set a Blackwell instance where $\mathcal{A} = \mathcal{K}$, $\mathcal{B} = B_2(1)$, $r(x, f) = \langle f, x \rangle$, and $S = \text{cone}(\{\kappa\} \times \mathcal{K})^0$.

Construct an approachability oracle $\mathcal{O}$.

At time step $t = 1, 2, \ldots$

1. Let: $p_t = \mathcal{R}^{\mathcal{O}}(f_1, f_2, \ldots, f_{t-1})$
2. Recieve: cost function f_t

Algorithm 4 requires a B-set oracle $\mathcal{O}$ to exist to implement the blackwell approachability algorithm $\mathcal{R}^{\mathcal{O}}$. We can prove that there exists a B-set oracle for the conic target set $S = \text{cone}(\{\kappa\} \times \mathcal{K})^0$.

Proof. Assume we have a halfspace H which is tangent to S and contains S. Since S is a cone, that implies $H = \{\theta : \langle \theta, z_H \rangle \leq 0\}$ for some $z_H \in \mathbb{R}^d$. Furthermore, $S \subset H$ implies that $z_H \in S^0 = (\text{cone}(\{\kappa\} \times \mathcal{K})^0)^0 = \text{cone}(\{\kappa\} \times \mathcal{K})$. That is equivalent to $z_H = \alpha(\kappa \oplus x_H)$ for some $x_H \in \mathcal{A}$ and $\alpha > 0$. Thus, by construction, there exists an oracle $\mathcal{O}(H) \to x_H$. □

Theorem 7.6. *Algorithm 4 produces the following regret bound,*

$$Regret(\mathcal{R}) \leq 2\kappa D_T(\mathcal{R})T^{-1}$$

where $D_T(\mathcal{R})$ is the distance by the approachability algorithm $\mathcal{R}$, namely

$$D_T(\mathcal{R}) \triangleq d(\bar{U}_T, S)$$

Proof. Applying Lemma 7.3 and Lemma 7.4 to D_T we can prove that

$$D_T = \max_{w \in \text{cone}(\kappa \oplus \mathcal{A}) \cap B_2(1)} \langle \frac{1}{T} \sum_{t=1}^{T} r(x_t, f_t), w \rangle$$

From there, we can assume $||w|| = 1$ so we can write $w = \frac{\kappa \oplus x}{||\kappa \oplus x||}$ for some $x \in \mathcal{K}$. Now we can show,

$$\begin{aligned} D_T(\mathcal{R}) &\geq \max_{x \in \mathcal{K}} \langle \frac{1}{T} \sum_{t=1}^{T} r(x_t, f_t), \frac{\kappa \oplus x}{||\kappa \oplus x||} \rangle \\ &= \frac{1}{T} \max_{x \in \mathcal{K}} \frac{\sum_{t=1}^{T} \langle f_t, x_t \rangle - \sum_{t=1}^{T} \langle f_t, x \rangle}{||\kappa \oplus x||} \\ &\geq \frac{\frac{1}{T} \left(\sum_{t=1}^{T} \langle f_t, x_t \rangle - \sum_{t=1}^{T} \langle f_t, x^* \rangle \right)}{||\kappa \oplus x^*||} \\ &\geq \frac{\frac{1}{T} \text{Regret}_T(\mathcal{R})}{2\kappa} \end{aligned}$$

□

Although Abernethy uses the B-set oracle from Blackwell's original approachability algorithm, it is clear that we could implement a response-based algorithm, or any other approachability algorithm to determine p_t at each time t.

8 Conclusion

The wide variety of available approachability algorithms that we surveyed enable extensions and future work on online learning algorithms with important and efficient learning bounds. As discussed in the paper, particular applications may or may not have access to certain approachability oracles, so by covering the different setups for the games that have approachability algorithms, we increase the set of available problems that can implement a computationally feasible learning algorithm.

References

[1] J. Abernethy, "Blackwell approachability and no-regret learning are equivalent," in Proceedings of COLT, volume 19, 2011.

[2] A. Bernstein and N. Shimkin, "Response-Based Approachability with Applications to Generalized No-Regret Problems," in Journal of Machine Learning Research, 2015.

[3] D. Blackwell, "An analog of the minimax theorem for vector payoffs," in Pacific Journal of Mathematics, 1956.

[4] C. Dann, Y. Mansour, M. Mohri, J. Schneider, B. Sivan, "Pseudonorm Approachability and Applications to Regret Minimization," arXiv:2302.01517v1, 2023.

[5] S. Mannor, V. Perchet, G. Stoltz, "Approachability in Unknown Games: Online Learning Meets Multi-Objective Optimization," in Proceedings of COLT, volume 35, 2014.

[6] J. Von Neumann, O. Morgenstern, H. Kuhn, A. Rubenstein, "Theory of games and economic behavior," in Princeton university press, 1947.

[7] J. Kwon, "Refined approachability algorithms and application to regret minimization with global costs," Université Pierre et Marie Curie, Paris 6, 2016.