Journal of Machine Learning Research 17 (2016) 1-23 Submitted 7/15; Revised 6/16; Published 8/16

An Online Convex Optimization Approach to Blackwell's Approachability

Nahum Shimkin SHIMKIN@EE.TECHNION.AC.IL
Faculty of Electrical Engineering
Technion—Israel Institute of Technology
Haifa 32000, ISRAEL

Abstract

The problem of approachability in repeated games with vector payoffs was introduced by Blackwell in the 1950s, along with geometric conditions and corresponding approachability strategies that rely on computing a sequence of *direction vectors* in the payoff space. For convex target sets, these vectors are obtained as projections from the current average payoff vector to the set. A recent paper by Abernethy, Batlett and Hazan (2011) proposed a class of approachability algorithms that rely on Online Linear Programming for obtaining alternative sequences of direction vectors. This is first implemented for target sets that are convex cones, and then generalized to any convex set by embedding it in a higher-dimensional convex cone. In this paper we present a more direct formulation that relies on general Online Convex Optimization (OCO) algorithms, along with basic properties of the support function of convex sets. This leads to a general class of approachability algorithms, depending on the choice of the OCO algorithm and the used norms. Blackwell's original algorithm and its convergence are recovered when Follow The Leader (or a regularized version thereof) is used for the OCO algorithm.

Keywords: approachability, online convex optimization, repeated games with vector payoffs

1. Introduction

Blackwell's approachability theory and the regret-based framework of online learning both address a repeated decision problem in the presence of an arbitrary (namely, unpredictable) adversary. Approachability, as introduced by Blackwell (1956), considers a fundamental feasibility issue for repeated matrix games with vector-valued payoffs. Referring to one player as the *agent* and to the other as *Nature*, a set S in the payoff space is *approachable* by the agent if it can ensure that the average payoff vector converges (with probability 1) to S, irrespectively of Nature's strategy. Blackwell's seminal paper provided geometric conditions for approachability, which are both necessary and sufficient for *convex* target sets S, and a corresponding approachability strategy for the agent. Approachability has found important applications in the theory of learning in games (Aumann and Maschler, 1995; Fudenberg and Levine,

1998; Peyton Young, 2004), and in particular in relation with no-regret strategies in repeated games as further elaborated below. A recent textbook exposition of approachability and some of it applications can be found in Maschler et al. (2013), and a comprehensive survey is provided by Perchet (2014).

Concurrently to Blackwell's paper, Hannan (1957) introduced the concept of no-regret strategies in the context of repeated matrix games. The *regret* of the agent is the shortfall of the cumulative payoff that was actually obtained relative to the one that would have been obtained with the best (fixed) action in hindsight, given Nature's observed action sequence. A no-regret strategy, or algorithm, ensures that the regret grows sub-linearly in time. The no-regret criterion has been widely adopted in the machine learning literature as a standard measure for the performance of online learning algorithms, and its scope has been greatly extended accordingly. Of specific relevance here is the Online Convex Optimization (OCO) framework, where Nature's discrete action is replaced by the choice of a convex function at each stage, and the agent's decision is a point in a convex set. The influential text of Cesa-Bianchi and Lugosi (2006) offers a broad overview of regret in online learning. Extensive surveys of OCO algorithms are provided by Shalev-Shwartz (2011); Hazan (2012, April 2016).

It is well known that no-regret strategies for repeated games can be obtained as particular instances of the approachability problem. A specific scheme was already given by Blackwell (1954), and an alternative formulation that leads to more explicit strategies was proposed by Hart and Mas-Colell (2001). The present paper considers the opposite direction, namely how no-regret algorithms for OCO can be used as a basic for an approachability strategy. Specifically, the OCO algorithm is used to generate a sequence of vectors that replace the projection-based direction vectors in Blackwell's algorithm. This results in a general class of approachability algorithms, that includes Blackwell's algorithm (and some generalizations thereof by Hart and Mas-Colell (2001)) as special cases.

The idea of using an online-algorithm to provide the sequence of direction vectors originated in the work of Abernethy et al. (2011), who showed how any no-regret algorithm for the online *linear* optimization problem can be used as a basis for an approachability algorithm. The scheme suggested in Abernethy et al. (2011) first considers target sets S that are convex cones. The generalization to any convex set is carried out by embedding the original target set as a convex cone in a higher dimensional payoff space. Here, we propose a more direct scheme that avoids the above-mentioned embedding. This construction relies on the *support function* of the target set, which is related to Blackwell's approachability conditions on the one hand, and on the other provides a variational expression for the point-to-set distance. Consequently, the full range of OCO algorithms can be used to provide a suitable sequence of direction vectors.

As we shall see, Blackwell's original algorithm is recovered from our scheme when the standard Follow the Leader (FTL) algorithm is used for the OCO part. Recovering the (known) convergence of this algorithm directly from the OCO viewpoint is a bit

more intricate. First, when the target set has a smooth boundary, we show that FTL converges at a "fast" (logarithmic) rate, hence leading to a correspondingly fast convergence of the average reward to the target set. To address the general case, we further show that Blackwell's algorithm is still exactly recovered when an appropriately *regularized* version of FTL is used, from which the standard $O(T^{-1/2})$ convergence rate may be deduced.

The basic results of approachability theory have been extended in numerous directions. These include additional theoretical results, such as the characterization of non-convex approachable sets; extended models, such as stochastic (Markov) games and games with partial monitoring; and additional approachability algorithms for the basic model. For concreteness we will expand only on the latter (below, in Subsection 2.1), and refer the reader to the above-mentioned overviews for further information.

The paper proceeds as follows. In Section 2 we recall the relevant background on Blackwell's approachability and Online Convex Optimization. Section 3 presents the proposed scheme, in the form of a meta-algorithm that relies on a generic OCO algorithm, discusses the relation to the scheme of Abernethy et al. (2011), and demonstrates a specific algorithm that is obtained by using Generalized Gradient Descent for the OCO algorithm. In Section 4 we describe the relations with Blackwell's original algorithm and its convergence. Section 5 outlines the extension of the proposed framework to general (rather than Euclidean) norms, followed by some concluding remarks.

Notation: The standard (dot) inner product in $\mathbb{R}^d$ is denoted by $\langle \cdot, \cdot \rangle$, $\| \cdot \|_2$ is the Euclidean norm, $\mathrm{d}(z, S) = \inf_{s \in S} \|z - s\|_2$ denotes the corresponding point-to-set distance, $B_2 = \{w \in \mathbb{R}^d : \|w\|_2 \leq 1\}$ denotes the Euclidean unit ball, $\Delta(I)$ is the set of probability distributions over a finite set I, $\mathrm{diam}(S) = \sup_{s,s' \in S} \|s - s'\|_2$ is the diameter of the set S, and $\|\mathcal{R} - S\|_2 = \sup_{s' \in \mathcal{R}, s \in S} \|s - s'\|_2$ denotes the maximal distance between points in sets $\mathcal{R}$ and S.

2. Model and Background

We start with brief reviews of Blackwell's approachability theory and Online Convex Programming, focusing on those aspects that are most relevant to this paper.

2.1 Approachability

Consider a repeated game with *vector-valued* rewards that is played by two players, the *agent* and *Nature*. Let I and J denote the finite action sets of these players, respectively, with corresponding mixed actions $x = (x(1), \ldots, x(|I|)) \in \Delta(I)$ and $y = (y(1), \ldots, y(|J|)) \in \Delta(J)$. Let $r : I \times J \to \mathbb{R}^d$, $d \geq 1$, be the vector-valued reward function of the single-stage game, which is extended to mixed action as usual through the bilinear function

$$r(x, y) = \sum_{i,j} x(i) y(j) r(i, j) \,.$$

Similarly, we denote $r(x,j) = \sum_i x(i) r(i,j)$. The specific meaning of $r(\cdot,\cdot)$ should be clear by its argument.

The game is repeated in stages $t = 1, 2, \ldots$, where at stage t actions i_t and j_t are chosen by the players, and the reward vector $r(i_t, j_t)$ is obtained. A pure strategy for the agent is a mapping from each possible history $(i_1, j_1, \ldots, i_{t-1}, j_{t-1})$ to an action i_t, and a mixed strategy is a probability distribution over the pure strategies. Nature's strategies are similarly defined. Any pair of strategies for the agent and Nature thus induce a probability measure on the game sequence $(i_t, j_t)_{t=1}^{\infty}$.

Let

$$\bar{r}_T = \frac{1}{T} \sum_{t=1}^{T} r(i_t, j_t)$$

denote the T-stage average reward vector. We may now recall Blackwell's definition of an approachable set.

Definition 1 (Approachability) *A set $S \subset \mathbb{R}^d$ is **approachable** if there exists a strategy for the agent such that $\bar{r}_t$ converges to S with probability 1, at a uniform rate over Nature's strategies. That is, for any $\epsilon > 0$ and $\delta > 0$ there exists $T \geq 1$ such that*

$$\text{Prob}\left\{ \sup_{t \geq T} \mathrm{d}(\bar{r}_t, S) > \epsilon \right\} \leq \delta, \tag{1}$$

for any strategy of Nature. A strategy of the agent that satisfies this property is an approachability strategy *for S.*

Remarks:

1. It is evident that approachability of a set and its closure are equivalent, hence we shall henceforth consider only closed target sets S.

2. In some treatments of approachability, convergence of the expected distance $E(\mathrm{d}(\bar{r}_t, S))$ and its rates are of central interest; see Perchet (2014). We shall consider these rates as well in the following.

3. In some models of interest, the decision variable of the agent may actually be the continuous variable x (in place of i), so that the actual reward is $r(x,j)$. All definitions and results below easily extend to this case, as long as x remains in a compact and convex set, and $r(x,j)$ is linear in x over that set.

For convex sets, approachability is fully characterized by the following result, which also provides an explicit strategy for the agent.

Theorem 2 (Blackwell, 1956) *A closed convex set $S \subset \mathbb{R}^d$ is approachable* if and only if *either one of the following equivalent conditions holds:*

(i) For each unit vector $u \in R^d$, there exists a mixed action $x = x_S(u) \in \Delta(I)$ such that

$$\langle u, r(x,j)\rangle \leq \sup_{s\in S}\langle u, s\rangle\,, \quad \textit{for all } j \in J\,. \tag{2}$$

(ii) For each $y \in \Delta(J)$ there exists $x \in \Delta(I)$ such that $r(x,y) \in S$.

If S is approachable, then the following strategy is an approachability strategy for S: For $z \notin S$, let $u_S(z)$ denote the unit vector that points to z from $\mathrm{Proj}_S(z)$, the closest point to z in S. For $t \geq 2$, if $\bar{r}_{t-1} \notin S$, choose i_t according to the mixed action $x_t = x_S(u_S(\bar{r}_{t-1}))$; otherwise, choose i_t arbitrarily.

Blackwell's approachability strategy relies on the sequence of direction vectors $u_t = u_S(\bar{r}_{t-1})$, obtained through Euclidean projections unto the set S. A number of extensions and alternative algorithms for the basic game model have been proposed since. Most related to the present paper is the use of more general direction vectors. In Hart and Mas-Colell (2001), the direction vectors are obtained as the gradient of a suitable potential function; Blackwell's algorithm is recovered when the potential is taken as the Euclidean distance to the target set, while the use of other norms provides a range of useful variants. We will relate these variants to the present work in Section 5. As mentioned in the introduction, Abernethy et al. (2011) introduced the use of no-regret algorithms to generate the sequence of direction vectors.

A different class of approachability algorithms relies on Blackwell's *dual* condition in Theorem 2(*ii*), thereby avoiding the computation of direction vectors as projections (or related operations) to the target set S. Based on that condition, one can define a *response map* that assigns to each mixed action y of Nature a mixed action x of the agent such that the reward vector $r(x,y)$ belongs to S. An approachability algorithm that applies the response map to a *calibrated forecast* of the opponents actions was proposed in Perchet (2009), and further analyzed in Bernstein et al. (2014). A computationally feasible response-based scheme that avoids the hard computation of calibrated forecasts is provided by Bernstein and Shimkin (2015). This paper also demonstrates the utility of the response-based approach for a class of generalized no-regret problems, where the set S is geometrically complicated, hence computing a projection is hard, but the response function is readily available. The response-based viewpoint is pursued further in the work of Mannor et al. (2014), which aims to approach the best-in-hindsight target set in an unknown game.

2.2 Online Convex Optimization (OCO)

OCO extends the framework of no-regret learning to function minimization. Let W be a convex and compact set in $\mathbb{R}^d$, and let $\mathcal{F}$ be a set of convex functions $f : W \to \mathbb{R}$. Consider a sequential decision problem, where at each stage $t \geq 1$ the agent chooses a point $w_t \in W$, and then observes a function $f_t \in F$. An *Algorithm* for the agent is a rule for choosing w_t, $t \geq 1$, based on the history $\{f_k, w_k\}_{k\leq t-1}$. The regret of an

algorithm $\mathcal{A}$ is defined as

$$\text{Regret}_T(\mathcal{A}) = \sup_{f_1,\ldots,f_T \in \mathcal{F}} \left\{ \sum_{t=1}^{T} f_t(w_t) - \min_{w \in W} \sum_{t=1}^{T} f_t(w) \right\}, \tag{3}$$

where the supremum is taken over all possible functions $f_t \in \mathcal{F}$. An effective algorithm should guarantee a small regret, and in particular one that grows sub-linearly in T.

The OCO problem was introduced in this generality in Zinkevich (2003), along with the following Online Gradient Descent algorithm:

$$w_{t+1} = \text{Proj}_W(w_t - \eta_t g_t), \quad g_t \in \partial f_t(w_t). \tag{4}$$

Here $\partial f_t(w_t)$ is the subdifferential of f_t at w_t, (η_t) is a diminishing gain sequence, and Proj_W denotes the Euclidean projection onto the convex set W. To state a regret bound for this algorithm, let $\text{diam}(W)$ denote the diameter of W, and suppose that all subgradients of the functions f_t are uniformly bounded in norm by a constant G.

Proposition 3 (Zinkevich, 2003) *For the Online Gradient Descent algorithm in* (4) *with gain sequence* $\eta_t = \frac{\eta}{\sqrt{t}}$, $\eta > 0$, *the regret is upper bounded as follows:*

$$\text{Regret}_T(\text{OGD}) \le \left(\frac{\text{diam}(W)^2}{\eta} + 2\eta G^2 \right) \sqrt{T}. \tag{5}$$

Several classes of OCO algorithms are now known, as surveyed in Cesa-Bianchi and Lugosi (2006); Shalev-Shwartz (2011); Hazan (2012). Of particular relevance here is the Regularized Follow the Leader (RFTL) algorithm, specified by

$$w_{t+1} = \operatorname*{argmin}_{w \in W} \left\{ \sum_{k=1}^{t} f_k(w) + R_t(w) \right\}, \tag{6}$$

where $R_t(w)$, $t \ge 1$ is a sequence of regularization functions. With $R_t \equiv 0$, the algorithm reduces to the basic Follow the Leader (FTL) algorithm, which does not generally lead to sublinear regret, unless additional requirements such as strong convexity are imposed on the functions f_t (we will revisit the convergence of FTL in Section 4). For RFTL, we will require the following standard convergence result. Recall that a function $R(w)$ over a convex set W is called *ρ-strongly convex* if $R(w) - \frac{\rho}{2}\|w\|_2^2$ is convex there.

Proposition 4 *Suppose that each function f_t is Lischitz-continuous over W, with Lipschitz coefficient L_f. Let $R_t(w) = \rho_t R(w)$, where $0 < \rho_t < \rho_{t+1}$, and the function $R : W \to [0, R_{\max}]$ is is 1-strongly convex and Lipschitz continuous with coefficient L_R. Then*

$$\text{Regret}_T(\text{RFTL}) \le 2L_f \sum_{t=1}^{T} \frac{L_f + (\rho_t - \rho_{t-1})L_R}{\rho_t + \rho_{t-1}} + \rho_T R_{\max}. \tag{7}$$

The proof of this bound is outlined in the Appendix.

3. OCO-Based Approachability

In this section we present the proposed OCO-based approachability algorithm. We start by introducing the support function and its relevant properties, and express Blackwell's separation condition in terms of this function. We then present the proposed algorithm, in the form of a meta-algorithm that incorporates a generic OCO algorithm. As a concrete example, we consider the specific algorithm obtained when Online Gradient Descent is used for the OCO part.

3.1 The Support Function

Let set $S \subset \mathbb{R}^d$ be a closed and convex set. The *support function* $h_S : \mathbb{R}^d \to \mathbb{R} \cup \{\infty\}$ of S is defined as

$$h_S(w) \triangleq \sup_{s \in S} \langle w, s \rangle, \quad w \in \mathbb{R}^d.$$

It is evident that h_S is a convex function (as a pointwise supremum of linear functions), and is positive homogeneous: $h_S(aw) = a h_S(w)$ for $a \geq 0$. Furthermore, the Euclidean distance from a point $z \in \mathbb{R}^d$ to S can be expressed as

$$\mathrm{d}(z, S) = \max_{w \in B_2} \left\{ \langle w, z \rangle - h_S(w) \right\}, \tag{8}$$

where B_2 is the closed Euclidean unit ball (see, e.g., Boyd and Vandenberghe (2004, Section 8.1.3); see also Lemma 16 below). It follows that

$$\operatorname*{argmax}_{w \in B_2} \left\{ \langle w, z \rangle - h_S(w) \right\} = \begin{cases} 0 & : \ z \in S \\ u_S(z) & : \ z \notin S \end{cases} \tag{9}$$

with $u_S(z)$ as defined in Theorem 2, namely the unit vector pointing from $\mathrm{Proj}_S(z)$ to z.

Blackwell's separation condition in (2) can now be written in terms of the support function as follows:

$$\langle w, r(x, j) \rangle \leq \sup_{s \in S} \langle w, s \rangle \equiv h_S(w) \,.$$

We can now rephrase the primal condition in Theorem 2 in the following form.

Corollary 5 *A closed and convex set S is approachable* if and only if *for every vector $w \in B_2$ there exists a mixed action $x \in \Delta(I)$ so that*

$$\langle w, r(x, j) \rangle - h_S(w) \leq 0, \quad \forall j \in J. \tag{10}$$

We note that equation (10) defines a linear inequality for x, so that a mixed action $x \in \Delta(I)$ that satisfies (10) for a given direction w can be computed using linear programming. More concretely, existence of a mixed action x that satisfies (10) can be equivalently stated as

$$\mathrm{val}(w \cdot r) \triangleq \min_{x \in \Delta(I)} \max_{j \in J} \langle w, r(x, j) \rangle \leq h_S(w) \,,$$

where $\mathrm{val}(w \cdot r)$ is the minimax value of the matrix game with a scalar payoff that is obtained by projecting the reward vectors $r(i,j)$ onto w. Consequently, the mixed action x that satisfies (10) can be taken as a minimax strategy for the agent in this game.

3.2 The General Algorithm

The proposed algorithm (see Algorithm 1 below) builds on the following idea. First, we apply an OCO algorithm to generate a sequence of *direction vectors* $w_t \in B_2$, so that

$$\sum_{t=1}^{T} (\langle w_t, r_t \rangle - h_S(w_t)) \geq T \max_{w \in B_2} \{\langle w, \bar{r}_T \rangle - h_S(w)\} - a(T), \tag{11}$$

where $r_t = r(x_t, j_t)$ is considered (within the OCO algorithm) an arbitrary vector that is revealed after w_t is specified, and $a(T)$ is of order $o(T)$. The mixed action $x_t \in \Delta(I)$, in turn, is chosen (after w_t is revealed) to satisfy (10), so that $\langle w_t, r(x_t, j_t) \rangle - h_S(w_t) \leq 0$, hence

$$\langle w_t, r_t \rangle - h_S(w_t) \leq \langle w_t, r_t \rangle - \langle w_t, r(x_t, j_t) \rangle \stackrel{\triangle}{=} \delta_t.$$

Using this inequality in (11), and observing the distance formula (8), yields

$$\mathrm{d}(\bar{r}_T, S) \leq \frac{a(T)}{T} + \Delta(T) \to 0\,,$$

where $\Delta(T) = \frac{1}{T} \sum_{t=1}^{T} \delta_t$, a stochastic term that converges to 0, as discussed below.

To secure (11), observe that the function $f(w; r) = -\langle w, r \rangle + h_S(w)$ is convex in w for each vector r. Therefore, an OCO algorithm can be applied to the sequence of convex functions $f_t(w) = -\langle w, r_t \rangle + h_S(w)$, where $r_t = r(x_t, j_t)$ is considered an arbitrary vector which is revealed only after w_t is specified. Applying an OCO algorithm $\mathcal{A}$ with $\mathrm{Regret}_T(\mathcal{A}) \leq a(T)$ to this setup, we obtain a sequence (w_t) such that

$$\sum_{t=1}^{T} f_t(w_t) \leq \min_{w \in B_2} \sum_{t=1}^{T} f_t(w) + a(T)\,,$$

where

$$\sum_{t=1}^{T} f_t(w_t) = -\sum_{t=1}^{T} (\langle w_t, r_t \rangle - h_S(w_t))\,,$$
$$\sum_{t=1}^{T} f_t(w) = -\sum_{t=1}^{T} (\langle w, r_t \rangle - h_S(w)) = -T(\langle w, \bar{r}_T \rangle - h_S(w))\,.$$

This can be seen to imply (11).

The discussion above leads to the following generic approachability algorithm.

Algorithm 1 (OCO-based Approachability Meta-Algorithm)

- Given: *A closed, convex and approachable set S; a procedure (e.g., a linear program) to compute $x \in \Delta(I)$, for a given vector w, so that* (10) *is satisfied; an OCO algorithm $\mathcal{A}$ for the functions $f_t(w) = -\langle w, r_t\rangle + h_S(w)$, with $Regret_T(\mathcal{A}) \le a(T)$.*
- Repeat for $t = 1, 2, \ldots$:
  1. *Obtain w_t from the OCO algorithm applied to the convex functions $f_k(w) = -\langle w, r_k\rangle + h_k(w)$, $k \le t-1$, so that inequality* (11) *is satisfied.*
  2. *Choose x_t according to* (10)*, so that $\langle w_t, r(x_t, j)\rangle - h_S(w_t) \le 0$ holds for all $j \in J$.*
  3. *Observe Nature's action j_t, and set $r_t = r(x_t, j_t)$.*

To state our convergence result for this algorithm, we first consider the term Δ_t that arises due to the difference between $r_t = r(i_t, j_t)$ and $r(x_t, j_t)$. The analysis follows by standard convergence results for martingale difference sequences.

Lemma 6 *Let*

$$\Delta_T = \frac{1}{T}\sum_{t=1}^{T} \delta_t, \quad \delta_t = \langle w_t, r(i_t, j_t) - r(x_t, j_t)\rangle \,.$$

Then $E(\Delta_T) = 0$, and $\Delta_T \to 0$ w.p. 1, at a uniform rate independent of Nature's strategy. Specifically,

$$P\{|\Delta_T| > \epsilon\} \le \frac{6\rho_0^2}{\epsilon^2 T}\,, \tag{12}$$

where $\rho_0 = \max_{j\in J} \max_{i,i'\in I} \|r(i,j) - r(i',j)\|_2$.

Proof Let $H_t = (i_k, j_k, w_k)_{1\le k\le t}$. Observe that w_t and j_t are chosen based only on H_{t-1}, hence do not depend on i_t, and similarly i_t is randomly and independently chosen according to x_t. It follows that $E(\delta_t|H_{t-1}) = 0$, which implies that $E(\Delta_T) = 0$. Furthermore, (δ_t) is a Martingale difference sequence, uniformly bounded by

$$|\delta_t| \le \|w_t\|_2 \|r(i_t, j_t) - r(x_t, j_t)\|_2 \le \max_j \max_{i,i'} \|r(i,j) - r(i',j)\|_2 \stackrel{\triangle}{=} \rho_0\,,$$

(where $w_t \in B_2$ was used in the second inequality). Convergence of Δ_t now follows by standard results for martingale difference sequences; the specific rate bound in (12) follows from Proposition 4.1 and Equation (4.7) in Shimkin and Shwartz (1993), upon noting that $X_t \stackrel{\triangle}{=} \sum_{k=1}^{t} \delta_k$ satisfies $E(X_{t+1}^2|H_t) = X_t^2 + 0 + E(\delta_{t+1}^2|H_t) \le X_t^2 + \rho_0^2$. ■

Convergence of Algorithm 1 may now be summarised as follows.

Theorem 7 *Under Algorithm 1, for any $T \geq 1$ and any strategy of the opponent, it holds w.p. 1 that*

$$\mathrm{d}(\bar{r}_T, S) \leq \frac{a(T)}{T} + \Delta_T \,,$$

where Δ_T is defined in Lemma 6 and is a zero-mean random variable that converges to zero at a uniform rate, as specified there. In particular,

$$E(\mathrm{d}(\bar{r}_T, S)) \leq \frac{a(T)}{T} \,.$$

Proof As observed above, application of the OCO algorithm implies (11). Recalling (8), we obtain

$$\begin{aligned}
\mathrm{d}(\bar{r}_T, S) &= \max_{w \in B_2} \{\langle w, \bar{r}_T \rangle - h_S(w)\} \\
&\leq \frac{1}{T} \sum_{t=1}^{T} (\langle w_t, r_t \rangle - h_S(w_t)) + \frac{a(T)}{T} \\
&= \frac{1}{T} \sum_{t=1}^{T} \langle w_t, r(x_t, j_t) \rangle - h_S(w_t)) + \frac{a(T)}{T} + \frac{1}{T} \sum_{t=1}^{T} \langle w_t, r_t - r(x_t, j_t) \rangle \,.
\end{aligned}$$

But since $\langle w_t, r(x_t, j_t) \rangle - h_S(w_t) \leq 0$ by choice of x_t in the algorithm, and using the definition of Δ_t, we obtain that $\mathrm{d}(\bar{r}_T, S) \leq \frac{a(T)}{T} + \Delta_T$, as claimed. The rest now follows by the properties of Δ_t, stated in Lemma 6. ■

To recap, any OCO algorithm that guarantees (11) with $\frac{a(T)}{T} \to 0$, induces an approachability strategy with rate of convergence bounded by the sum of two terms: the first is $\frac{a(T)}{T}$, related to the regret bound of the OCO algorithm, and the second is a zero-mean stochastic term of order $T^{-1/2}$ (at most), which arises due to the difference between the actual rewards $r_t = r(i_t, j_t)$ and their means $r(x_t, j_t)$.

We conclude this subsection with a few remarks. The first two concern instances where the stochastic term Δ_T is nullified.

Remark 8 (Pure Actions) *Suppose that the inequality $\max_j \langle w_t, r(x, j) \rangle - h_S(w_t) \leq 0$ in step 2 of the Algorithm can always be satisfied by pure actions (so that x_t assigns probability 1 to to a single action, i_t). Then choosing such x_t's clearly implies that $r(x_t, j_t) = r(i_t, j_t)$, hence the term Δ_t in Theorem 7 becomes identically zero.*

Remark 9 (Smooth Rewards) *In some problems, the rewards of interest may actually be the smoothed rewards $r(x_t, j_t)$ or $r(x_t, y_t)$, rather than $r(i_t, j_t)$. Focusing on the first case for concreteness, let us redefine r_t as $r(x_t, j_t)$, and assume that this reward vector can be computed or observed by the agent following each stage t. Applying Algorithm 1 with these modified rewards now leads to the same bound as in Theorem 7, but with $\Delta_T = 0$.*

Remark 10 (Convex Cones) *The approachability algorithm of Abernethy et al. (2011) starts with target sets S that are restricted to be convex cones. For S a closed convex cone, the support function is given by*

$$h_S(w) = \begin{cases} 0 & : \ w \in S^o \\ \infty & : \ w \notin S^o \end{cases}$$

where S^o is the polar cone *of S. The required inequality in* (11) *thus reduces to*

$$\sum_{t=1}^{T} \langle w_t, r_t \rangle \geq T \max_{w \in B_2 \cap S^o} \langle w, \bar{r}_T \rangle - a(T) \,.$$

The sequence (w_t) can be obtained in this case by applying an online linear *optimization algorithm restricted to $w_t \in B_2 \cap S^o$. This is the algorithm proposed by Abernethy et al. (2011). The extension to general convex sets is handled there by lifting the problem to a $(d+1)$-dimensional space, with payoff vector $r'(x,y) = (\kappa, r(x,y))$ and target set $S' = \mathrm{cone}(\{\kappa\} \times S)$, where $\kappa = \max_{s \in S} \|s\|_2$, for which it holds that $\mathrm{d}(u, S) \leq 2\mathrm{d}(u', S')$.*

3.3 An OGD-based Approachability Algorithm

As a concrete example, let us apply the Online Gradient Descent algorithm specified in (4) to our problem. With $W = B_2$ and $f_t(w) = -(\langle w, r_t \rangle - h_S(w))$, we obtain in step 1 of Algorithm 1,

$$w_{t+1} = \mathrm{Proj}_{B_2}\{w_t + \eta_t(r_t - y_t)\} \,, \quad y_t \in \partial h_S(w_t) \,.$$

Observe that $\mathrm{Proj}_{B_2}(v) = v / \max\{1, \|v\|_2\}$, and (e.g., by Corollary 8.25 in Rockafellar and Wets (1997))

$$\partial h_S(w) = \operatorname*{argmax}_{s \in S} \langle s, w \rangle \,.$$

To evaluate the convergence rate in (5), observe that $\mathrm{diam}(B_2) = 2$, and, since $y_t \in S$, $\|g_t\|_2 = \|r_t - y_t\|_2 \leq \|\mathcal{R} - S\|_2$, where $\mathcal{R} = \{r(x,y)\}_{x \in \Delta(I), y \in \Delta(J)}$ is the reward set. Assuming for the moment that the goal set S is bounded, we obtain

$$E(\mathrm{d}(\bar{r}_T, S)) \leq \frac{b(\eta)}{\sqrt{T}} \,, \quad \text{with} \ \ b(\eta) = \frac{4}{\eta} + 2\eta \|\mathcal{R} - S\|_2^2 \,.$$

For $\eta = \sqrt{2}/\|\mathcal{R} - S\|_2$, we thus obtain $b(\eta) = 4\sqrt{2}\|\mathcal{R} - S\|_2$.

If S is not bounded, it can always be intersected with $\mathcal{R}$ (without affecting its approachability), yielding $\|\mathcal{R} - S\|_2 \leq \mathrm{diam}(\mathcal{R})$. This amounts to modifying the choice of y_t in the algorithm to

$$y_t \in \partial h_{S \cap \mathcal{R}}(w_t) = \operatorname*{argmax}_{y \in S \cap \mathcal{R}} (y, w) \,.$$

Alternatively, one may restrict attention (by projection) to vectors w_t in the set $\{w \in B_2 : h_S(w) < \infty\}$, similarly to the case of convex cones mentioned in Remark 10 above; we will not go here into further details.

4. Blackwell's Algorithm and (R)FTL

We next examine the relation between Blackwell's approachability algorithm and the proposed OCO-based scheme. We first show that Blackwell's algorithm coincides with OCO-based approachability when FTL is used as the OCO algorithm. We use this equivalence to establish fast (logarithmic) convergence rates for Blackwell's algorithm when the target set S has a smooth boundary. Interestingly, this equivalence does not provide a convergence result for general convex sets. To complete the picture, we show that Blackwell's algorithm can more generally be obtained via a *regularized* version of FTL, which leads to an alternative proof of convergence of the algorithm in the general case.

4.1 Blackwell's algorithm as FTL

Recall Blackwell's algorithm as specified in Theorem 2, namely x_{t+1} is chosen as a mixed action that satisfies (2) for $u = u_S(\bar{r}_t)$ (with x_{t+1} chosen arbitrarily if $\bar{r}_t \in S$, which is equivalent to setting $u = 0$ in that case).

Similarly, in Algorithm 1, x_{t+1} is chosen as a mixed action that satisfies (2) for $u = w_{t+1}$. Using FTL (i.e., Equation (6) with $R_t = 0$) for the OCO part gives

$$w_{t+1} = \operatorname*{argmin}_{w \in B_2} \sum_{k=1}^{t} f_k(w), \quad \text{with } f_k(w) = -\langle w, r_k \rangle + h_S(w)\,.$$

Equivalence of the two algorithms now follows directly from the following observation.

Lemma 11 *With $f_k(w)$ as above,*

$$\operatorname*{argmin}_{w \in B_2} \sum_{k=1}^{t} f_k(w) = \begin{cases} u_S(\bar{r}_t) & : \ \bar{r}_t \notin S \\ 0 & : \ \bar{r}_t \in S \end{cases} .$$

Proof Observe that $\sum_{k=1}^{t} f_k(w) = -t(\langle w, \bar{r}_t \rangle - h_S(w))$, so that

$$\operatorname*{argmin}_{w \in B_2} \sum_{k=1}^{t} f_k(w) = \operatorname*{argmax}_{w \in B_2} \{\langle w, \bar{r}_t \rangle - h_S(w)\}\,.$$

The required equality now follows from (9). ■

To establish convergence of Blackwell's algorithm via this equivalence, one needs to show that FTL guarantees the regret bound in (11) for an arbitrary reward sequence $(r_t) \subset \mathcal{R}$, with a sublinear rate sequence $a(T)$. It is well know, however, that (unregularized) FTL does not guarantee sublinear regret, without some additional assumptions on the function f_t. A simple counter-example, reformulated to the present case, is devised as follows: Let $S = \{0\} \subset \mathbb{R}$, so that $h_S(w) = 0$, and suppose that

$r_1 = -1$ and $r_t = 2(-1)^t$ for $t > 1$. Since $w_t = \text{sign}(\bar{r}_{t-1})$ and $\text{sign}(r_t) = -\text{sign}(\bar{r}_{t-1})$, we obtain that $f_t(w_t) = -r_t w_t = 1$, leading to a linearly-increasing regret.

The failure of FTL in this example is clearly due to the fast changes in the predictors w_t. We now add some smoothness assumptions on the set S that can mitigate such abrupt changes.

Assumption 1 *Let S be a compact and convex set. Suppose that the boundary ∂S of S is smooth with curvature bounded by κ_0, namely:*

$$\|\vec{n}(s_1) - \vec{n}(s_2)\|_2 \le \kappa_0 \|s_1 - s_2\|_2 \quad \textit{for all} \quad s_1, s_2 \in \partial S\,, \tag{13}$$

where $\vec{n}(s)$ is the unique unit outer normal to S at $s \in \partial S$.

For example, for a closed Euclidean ball of radius ρ, (13) is satisfied with equality for $\kappa_0 = \rho^{-1}$. The assumed smoothness property may in fact be formulated in terms of an interior sphere condition: For any point in $s \in S$ there exists a ball $B(\rho) \subset S$ with radius $\rho = \kappa_0^{-1}$ such that $s \in B(\rho)$.

Proposition 12 *Let Assumption 1 hold. Consider Blackwell's algorithm as specified in Theorem 2, and denote $w_t = u_S(\bar{r}_{t-1})$ (with w_1 arbitrary). Then, for any time $T \ge 1$ such that $\bar{r}_T \notin S$, (11) holds with*

$$a(T) = C_0(1 + \ln T), \tag{14}$$

where $C_0 = \text{diam}(\mathcal{R})\, \|\mathcal{R} - S\|_2\, \kappa_0$, and $\ln(\cdot)$ is the natural logarithm. Consequently,

$$E(\text{d}(\bar{r}_T, S)) \le C_0 \frac{1 + \ln T}{T}\,, \quad T \ge 1\,. \tag{15}$$

Proof Observe first that the regret bound in (14) implies (15). Indeed, for $\bar{r}_T \notin S$, $\text{d}(\bar{r}_T, S) \le a(T)/T$ follows as in Theorem 7, while if $\bar{r}_T \in S$ then $\text{d}(\bar{r}_T, S) = 0$ and (15) holds trivially.

We proceed to establish the logarithmic regret bound in (14). Let $f_t(w) = -\langle w, r_t \rangle + h_S(w)$, $W = B_2$, and denote

$$\text{Regret}_T(f_{1:T}) = \sum_{t=1}^{T} f_t(w_t) - \min_{w \in W} \sum_{t=1}^{T} f_t(w) = \sum_{t=1}^{T} (f_t(w_t) - f_t(w_{T+1}))\,. \tag{16}$$

A standard induction argument (e.g., Lemma 2.1 in Shalev-Shwartz (2011)) verifies that

$$\sum_{t=1}^{T} (f_t(w_t) - f_t(u)) \le \sum_{t=1}^{T} (f_t(w_t) - f_t(w_{t+1})) \tag{17}$$

holds for any $u \in W$, and in particular for $u = w_{T+1}$. It remains to upper-bound the differences in the last sum.

Consider first the case where $\bar{r}_t \notin S$ for all $1 \le t \le T$. We first show that $\|w_t - w_{t+1}\|_2$ is small, which implies the same for $|f_t(w_t) - f_t(w_{t+1})|$. By its definition, $w_{t+1} = u_S(\bar{r}_t)$, the unit vector pointing to $\bar{r}_t$ from $c_t \stackrel{\triangle}{=} \mathrm{Proj}_S(\bar{r}_t)$, which clearly coincides with the outer unit normal $\vec{n}(c_t)$ to S at c_t. It follows that

$$\|w_t - w_{t+1}\|_2 = \|\vec{n}(c_{t-1}) - \vec{n}(c_t)\|_2 \le \kappa_0 \|c_{t-1} - c_t\|_2 \le \kappa_0 \|\bar{r}_{t-1} - \bar{r}_t\|_2 \,,$$

where the first inequality follows by Assumption 1, and the second due to the shrinking property of the projection. Substituting $\bar{r}_t = \bar{r}_{t-1} + \frac{1}{t}(r_t - \bar{r}_{t-1})$ obtains

$$\|w_t - w_{t+1}\|_2 \le \frac{\kappa_0}{t} \|r_t - \bar{r}_{t-1}\|_2 \le \frac{\kappa_0}{t} \mathrm{diam}(\mathcal{R}) \,. \tag{18}$$

Next, observe that for any pair of unit vectors w_1 and w_2,

$$\begin{aligned} f_t(w_1) - f_t(w_2) &= -\langle w_1 - w_2, r_t \rangle + h_S(w_1) - h_S(w_2) \\ &= -\langle w_1 - w_2, r_t \rangle + \max_{s \in S} \langle w_1, s \rangle - \max_{s \in S} \langle w_2, s \rangle \\ &\le -\langle w_1 - w_2, r_t \rangle + \langle w_1, s_1 \rangle - \langle w_2, s_1 \rangle \\ &= \langle w_1 - w_2, s_1 - r_t \rangle \;\le\; \|w_1 - w_2\|_2 \|\mathcal{R} - S\|_2 \,, \end{aligned}$$

where $s_1 \in S$ attains the first maximum. Since the same bound holds for $f_t(w_2) - f_t(w_1)$, it holds also for the absolute value. In particular,

$$|f_t(w_t) - f_t(w_{t+1})| \le \|w_t - w_{t+1}\|_2 \|\mathcal{R} - S\|_2 \,, \tag{19}$$

and together with (18) we obtain

$$|f_t(w_t) - f_t(w_{t+1})| \le \frac{\kappa_0}{t} \,\mathrm{diam}(\mathcal{R}) \,\|\mathcal{R} - S\|_2 = \frac{C_0}{t} \,.$$

Substituting in (17) and summing over t^{-1} yields the regret bound

$$\mathrm{Regret}_T(f_{1:T}) \le C_0(1 + \ln T). \tag{20}$$

We next extend this bound to case where $\bar{r}_t \in S$ for some t. In that case $w_{t+1} = 0$, and $w_t - w_{t+1}$ may not be small. However, since $f_t(0) = 0$, such terms will not affect the sum in (17). Recall that we need to establish (14) for T such that $\bar{r}_T \notin S$. In that case, any time t for which $\bar{r}_t \in S$ is follows by some time $m \le T$ with $\bar{r}_m \notin S$. Let $1 \le k < m \le T$ be indices such that $\bar{r}_k, \ldots \bar{r}_{m-1} \in S$, but $\bar{r}_{k-1} \notin S$ (or $k = 1$) and $\bar{r}_m \notin S$. Then $w_{k+1}, \ldots, w_m = 0$, and

$$\sum_{t=k}^{m} (f_t(w_t) - f_t(w_{t+1})) = f_k(w_k) - f_m(w_{m+1}) \,.$$

Proceeding as above, we obtain similarly to (18),

$$\|w_k - w_{m+1}\|_2 \leq \kappa_0 \|\bar{r}_{k-1} - \bar{r}_m\|_2 \leq \mathrm{diam}(\mathcal{R}) \sum_{t=k}^{m-1} \frac{\kappa_0}{t} \,,$$

and the regret bound in (20) may be obtained as above. ■

The last result establishes a fast convergence rate (of order $\log T/T$) for Blackwell's approachability algorithm, under the assumed smoothness of the target set. We note that conditions for fast approachability (of order T^{-1}) were derived in Perchet and Mannor (2013), but are of different nature than the above.

Logarithmic convergence rates were derived for OCO algorithms in Hazan et al. (2007), under strong convexity conditions on the function f_t. This is apparently related to the present result, especially given the equivalence between strong convexity of a function and strong smoothness of its Legendre-Fenchel transform (cf. Shalev-Shwartz (2011), Lemma 2.19). However, we observe that the support function is h_S is *not* strongly convex, so that the logarithmic regret bound in (20) does not seem to follow from existing results. Rather, a basic property which underlies both cases is insensitivity of the maximum point to small perturbations in f_t, which here leads to the inequality (18).

4.2 Blackwell's algorithm as RFTL

The smoothness requirement in Assumption 1 does not hold for important classes of target sets, such as polyhedra and cones. As observed above, in absence of such additional smoothness properties the interpretation of Blackwell's algorithm through an FTL scheme does not entail its convergence, as the regret of FTL (and the corresponding bound $a(T)$ in (11)) might increase linearly in general.

To accommodate general (non-smooth) sets, we show next that Blackwell's algorithm can be identified more generally with a *regularized* version of FTL. This algorithm does guarantee an $O(\sqrt{T})$ regret in (11), and consequently leads to the standard $O(T^{-1/2})$ rate of convergence of Blackwell's approachability algorithm.

Let us apply the RFTL algorithm in equation (6) as the OCO part in Algorithm 1, with a quadratic regularization function $R_t(w) = \frac{\rho_t}{2}\|w\|_2^2$. This gives

$$w_{t+1} = \operatorname*{argmin}_{w \in B_2} \left\{ \sum_{k=1}^{t} f_k(w) + \frac{\rho_t}{2}\|w\|_2^2 \right\} , \quad f_k(w) = -\langle w, r_k \rangle + h_S(w) \,.$$

The following equality is the key to the required equivalence. It relies essentially on the positive-homogeneity property of the support function h_S, and consequently of the funcctions f_k above.

Lemma 13 *For $\rho_t > 0$, and w_{t+1} as defined above,*

$$w_{t+1} = \begin{cases} \beta_t u_S(\bar{r}_t) & : \quad \bar{r}_t \notin S \\ 0 & : \quad \bar{r}_t \in S \end{cases} , \tag{21}$$

where $\beta_t = \min\{1, \frac{t}{\rho_t} \mathrm{d}(\bar{r}_t, S)\} > 0$.

Proof Recall that $\sum_{k=1}^{t} f_k(w) = -t(\langle w, \bar{r}_t \rangle - h_S(w))$, so that

$$\operatorname*{argmin}_{w \in B_2} \left\{ \sum_{k=1}^{t} f_k(w) + \frac{\rho_t}{2} \|w\|_2^2 \right\} = \operatorname*{argmax}_{w \in B_2} \left\{ \langle w, \bar{r}_t \rangle - h_S(w) - \frac{\rho_t}{2t} \|w\|_2^2 \right\} .$$

To compute the right-hand side, we first maximize over $\{w : \|w\|_2 = \beta\}$, and then optimize over $\beta \in [0, 1]$. Denote $z = \bar{r}_t$, and $\eta = \rho_t / t$. Similarly to Lemma 11,

$$\operatorname*{argmax}_{\|w\|_2 = \beta} \left\{ \langle w, z \rangle - h_S(w) - \frac{\eta}{2} \|w\|_2^2 \right\} = \operatorname*{argmax}_{\|w\|_2 = \beta} \{ \langle w, z \rangle - h_S(w) \} = \begin{cases} \beta u_S(z) & : \quad z \notin S \\ 0 & : \quad z \in S \end{cases} .$$

Now, for $z \notin S$,

$$\max_{\|w\|_2 = \beta} \left\{ \langle w, z \rangle - h_S(w) - \frac{\eta}{2} \|w\|_2^2 \right\} = \beta \mathrm{d}(z, S) - \frac{\eta}{2} \beta^2 .$$

Maximizing the latter over $0 \le \beta \le 1$ gives $\beta^* = \min\{1, \frac{\mathrm{d}(z,S)}{\eta}\}$. Substituting back z and η gives (21). ■

This immediately leads to the required conclusion.

Proposition 14 *Algorithm 1 with quadratically regularized FTL is equivalent to Blackwell's algorithm.*

Proof Observe that the vector w_{t+1} in Equation (21) is equal to $u_S(\bar{r}_t)$ from Blackwell's algorithm in Theorem 2, up to a positive scaling by β_t. This scaling does not affect the choice of x_{t+1} according to (10), as the support function $h_S(w)$ is positive homogeneous. ■

Compared to non-regularzied FTL (or Blackwell's algorithm), we see the direction vectors in Equation (21) are scaled by a positive constant. Essentially, the effect of this scaling is to reduce the magnitude of w_{t+1} when $\bar{r}_t$ is close to S. While such scaling does not affect the choice of action x_t, it does lead to sublinear-regret for the OLO algorithm, and consequently convergence of the approachability algorithm. This is summarized as follows.

Proposition 15 *Let S be a convex and compact set. Consider the RFTL algorithm specified in equation (21), with $\rho_t = \rho\sqrt{t}$, $\rho > 0$. The regret of this algorithm is bounded by*

$$\mathrm{Regret}_T(\mathrm{RFTL}) \le \left(\frac{2L_f^2}{\rho} + \frac{\rho}{2}\right)\sqrt{T} + \frac{L_f}{2}\ln(T) + 4L_f \;\triangleq\; a_0(T)\,,$$

where $L_f = \|\mathcal{R} - S\|_2$. Consequently, if this RFTL algorithm is used in step 1 of Algorithm 1 to compute w_t, we obtain

$$E(\mathrm{d}(\bar{r}_T, S)) \le \frac{a_0(T)}{T} = O(T^{-\frac{1}{2}})\,, \qquad T \ge 1\,. \tag{22}$$

Proof The regret bound follows from the one in Proposition 4, evaluated for $f_t(w) = -\langle r_t, w\rangle + h_S(w)$, $W = B_2$, $R(w) = \frac{1}{2}\|w\|_2^2$, and $\rho_t = \rho\sqrt{t}$. Recalling that $\partial f_t(w) = -r_t + \mathrm{argmax}_{s\in S}\langle w, s\rangle$, the Lipschitz constant of f_t is upper bounded by $\|\mathcal{R} - S\|_2 \triangleq L_f$. Furthermore, $R_{\max} = \frac{1}{2}$ and $L_R = 1$. Therefore,

$$\mathrm{Regret}_T(\mathrm{RFTL}) \le 2L_f \sum_{t=1}^{T} \frac{L_f + \rho(\sqrt{t} - \sqrt{t-1})}{\rho(\sqrt{t} + \sqrt{t-1})} + \frac{\rho}{2}\sqrt{T}\,. \tag{23}$$

To upper bound the sum, we note that

$$\sum_{t=1}^{T} \frac{1}{(\sqrt{t} + \sqrt{t-1})} = \sum_{t=1}^{T}(\sqrt{t} - \sqrt{t-1}) = \sqrt{T},$$

and

$$\begin{aligned}\sum_{t=1}^{T}\left(\frac{\sqrt{t} - \sqrt{t-1}}{\sqrt{t} + \sqrt{t-1}}\right) &= \sum_{t=1}^{T}\frac{1}{(\sqrt{t} + \sqrt{t-1})^2} \le 2 + \sum_{t=3}^{T}\frac{1}{(2\sqrt{t-1})^2} \\ &\le 2 + \frac{1}{4}\int_{t=1}^{T}\frac{1}{t}dt = 2 + \frac{1}{4}\ln(T)\,.\end{aligned}$$

Substituting in (23) gives the stated regret bound. The second part now follows directly from Theorem 7. ■

With $\rho = 2L_f$, we obtain in (22) the convergence rate

$$E(\mathrm{d}(\bar{r}_T, S)) \le \frac{2\|\mathcal{R} - S\|_2}{\sqrt{T}} + o(\frac{1}{\sqrt{T}})\,.$$

We emphasize that the algorithm discussed in this section is equivalent to Blackwell's algorithm, hence its convergence is known. The proof of convergence here

is certainly not the simplest, nor does it lead to the best constants in the convergence rate. Indeed, Blackwell's proof (which recursively bounds the square distance $\mathrm{d}(\bar{r}_T, S)^2$) leads to the bound $\sqrt{E(\mathrm{d}(\bar{r}_T, S)^2)} \leq \frac{\|\mathcal{R}-S\|_2}{\sqrt{T}}$. Rather, our main purpose was to provide an alternative view and analysis of Blackwell's algorithm, which rely on a standard OCO algorithm. That said, the logarithmic convergence rate of the expected distance that was obtained under the smoothness Assumption 1 appears to be new.

5. Extensions with General Norms

As was mentioned in Subsection 2.1, a class of approachability algorithms that generalizes Blackwell's strategy was introduced by Hart and Mas-Colell (2001). The direction vectors (u_t) in Blackwell's algorithm, that are defined through Euclidean projection, are replaced in that paper by the gradient of a smooth *potential function*; Blackwell's algorithm is recovered when the potential is taken as the Euclidean distance to the target set. Other instances of interest are obtained by defining the potential through the p-norm distance; this, in turn, was used as a basis for a general class of no-regret algorithms in repeated games.

In this section we provide an extension of the OCO-based approachability algorithm from Section 3, which relies on a general norm rather than the Euclidean one to obtain the direction vectors (w_t). The proposed algorithms coincide with those of Hart and Mas-Colell (2001) when the RFTL algorithm is used for the OCO part.

Let $\|\cdot\|$ denote some norm on $\mathbb{R}^d$. The dual norm, denoted $\|\cdot\|_*$, is defined as

$$\|x\|_* = \max_{w\in\mathbb{R}^d:\|w\|\leq 1} \langle w, x\rangle \,.$$

For example, if the primal norm is the p-norm $\|x\|_p = (\sum_{i=1}^d x_i^p)^{\frac{1}{p}}$ with $p \in (0,1)$, the dual norm is the q-norm, with $q \in (0,1)$ that satisfies $\frac{1}{p} + \frac{1}{q} = 1$.

The following relations between the support function h_S and the point-to-set distance d_* will be required. The first is needed to show convergence of the algorithm, and the second for the interpretation of the FTL-based variant.

Lemma 16 *Let S be a closed convex set with support function h_S, and let $d_*(z,S) = \min_{s\in S} \|z-s\|_*$ denote the point-to-set distance with respect to the dual norm. Then, for any $z \in \mathbb{R}^d$,*

$$\mathrm{d}_*(z,S) = \max_{\|w\|\leq 1} \{\langle w, z\rangle - h_S(w)\} \,, \tag{24}$$

and

$$\partial d_*(z,S) = \operatorname*{argmax}_{\|w\|\leq 1} \{\langle w, z\rangle - h_S(w)\} \,, \tag{25}$$

where $\partial d_*(z,S)$ is the subgradient of $d_*(\cdot,S)$ at z.

Proof We first note that the maximum in (24) is attained since h_S is a lower semi-continuous function, and $\{\|w\| \leq 1\}$ is a compact set. To establish (24) we invoke the minimax theorem. By definition of h_S,

$$\max_{\|w\|\leq 1}(\langle w, z\rangle - h_S(w)) = \max_{\|w\|\leq 1}\inf_{s\in S}\langle w, z-s\rangle\,.$$

Observe that $\{\|w\| \leq 1\}$ is a convex and compact set, S is convex by definition, and $\langle w, z-s\rangle$ is linear both in u and in s. We may thus apply Sion's minimax theorem to obtain that the last expression equals

$$\inf_{s\in S}\sup_{\|w\|\leq 1}\langle w, z-s\rangle = \inf_{s\in S}\|z-s\|_* = \mathrm{d}_*(z,S)\,,$$

where the definition of the dual norm was used in the last step, and (24) is obtained.

Proceeding to (25), we observe (24) implies that $\mathrm{d}_*(\cdot, S)$ is the Legendre-Fenchel transform of an appropriately modified function $\bar{h}_S$, namely $\mathrm{d}_*(z,S) = \max_{w\in\mathbb{R}^d}\{\langle w, z\rangle - \bar{h}_S(w)\}$ where $\bar{h}_S(w) = h_S(w)$ if $\|w\| \leq 1$ and $\bar{h}_S(w) = \infty$ for $\|w\| > 1$. Evidently $\bar{h}_S$ is a convex and lower semi-continuous function, which follows since both S and $\{\|w\| \leq 1\}$ are closed and convex sets. The equality in (25) now follows directly by Proposition 11.3 in Rockafellar and Wets (1997). ■

The algorithm: We can now repeat the blueprints of Subsection 3.2 to obtain an approachability algorithm for a convex target set S, which here relies on any norm $\|\cdot\|$. First, we apply an OCO algorithm to the functions $f_t(w) = -\langle w, r_t\rangle + h_S(w)$ over the convex compact set $\{w \in \mathbb{R}^d : \|w\| \leq 1\}$ to obtain, analogously to equation (11), a sequence of vectors (w_t) such that

$$\sum_{t=1}^{T}(\langle w_t, r_t\rangle - h_S(w_t)) \geq T\max_{\|w\|\leq 1}\{\langle w, \bar{r}_T\rangle - h_S(w)\} - a(T)\,. \tag{26}$$

Next, each w_t is used as the direction vector for stage t, and the mixed action x_t is chosen so that $\langle w_t, r_t\rangle - h_S(w_t) \leq 0$ holds for any action of Nature. Observing (24), we obtain that

$$\mathrm{d}_*(\bar{r}_T, S) \leq \frac{a(T)}{T} \to 0\,.$$

Follow the Leader: Consider the specific case that where FTL is used for the OCO algorithm. That is,

$$w_{t+1} \in \operatorname*{argmin}_{\|w\|\leq 1}\sum_{k=1}^{t} f_k(w) \equiv \operatorname*{argmax}_{\|w\|\leq 1}\{\langle w, \bar{r}_t\rangle - h_S(w)\}\,. \tag{27}$$

By (25), this is equivalent to $w_{t+1} \in \partial\mathrm{d}_*(\bar{r}_t, S)$. In particular, if $\mathrm{d}_*(z,S)$ is differentiable at $z = \bar{r}_t$ then $w_{t+1} = \nabla\mathrm{d}_*(\bar{r}_t, S)$. We therefore recover the approachability algorithm of Hart and Mas-Colell (2001) for the potential function $P(z) = \mathrm{d}_*(z,S)$.

Convergence of the approachability algorithm of Hart and Mas-Colell (2001) requires the potential function $P(z)$ to be continuously differentiable. As observed there, for $P(z) = \mathrm{d}_*(z, S)$ this holds if either the norm $\|\cdot\|_*$ is smooth (e.g., the q-norm for $1 < q < \infty$), or the boundary of S is smooth.

In our framework, convergence analysis of the FTL-based OCO algorithm can be carried out similarly to that of Section 4. In particular, similarly to the procedure of Subsection 4.2, if the norm $\|\cdot\|$ is smooth we can guarantee convergence of the OCO algorithm without affecting the induced approachability algorithm by adding an appropriate regularization term in (27), namely setting

$$w_{t+1} \in \operatorname*{argmin}_{\|w\|\le 1} \left\{ \sum_{k=1}^{t} f_k(w) - \frac{\rho_t}{2}\|w\|^2 \right\} .$$

By analogy to Lemma 13, the added regularization does not modify the direction of w_t but only its magnitude, hence the choice of actions x_t is the induced approachability algorithm remains the same. Convergence rates can be obtained along the lines of Section 4, and will not be considered in detail here.

Acknowledgments

The author wishes to thank Elad Hazan for helpful comments on a preliminary version of this work, and to the anonymous referees for many useful comments that helped improve the presentation. This research was supported by the Israel Science Foundation grant No. 1319/11.

Appendix A.

Proof of Proposition 4: We follow the outline of the proof of Lemma 2.10 in Shalev-Shwartz (2011), modified to accommodate a non-constant regularization sequence ρ_t. The starting point is the inequality, proved by induction,

$$\sum_{t=1}^{T}(f_t(w_t) - f_t(u)) \le \sum_{t=1}^{T}(f_t(w_t) - f_t(w_{t+1})) + \rho_t R(u)\,, \tag{28}$$

which holds for any $u \in W$. Therefore,

$$\sum_{t=1}^{T}(f_t(w_t) - f_t(u)) \le L_f \sum_{t=1}^{T} \|w_t - w_{t+1}\|_2 + \rho_t R(u)\,. \tag{29}$$

Denote $F_t(w) = \sum_{k=1}^{t-1} f_k(w) + \rho_{t-1}R(w)$. Then F_t is ρ_{t-1}-strongly convex, and w_t is its minimizer by definition. Hence, it holds generally that

$$F_t(u) \ge F_t(w_t) + \frac{\rho_{t-1}}{2}\|u - w_t\|_2^2\,,$$

and in particular,

$$F_t(w_{t+1}) \geq F_t(w_t) + \frac{\rho_{t-1}}{2}\|w_{t+1} - w_t\|_2^2, \tag{30}$$

$$F_{t+1}(w_t) \geq F_{t+1}(w_{t+1}) + \frac{\rho_t}{2}\|w_t - w_{t+1}\|_2^2 . \tag{31}$$

Summing and cancelling terms, we obtain

$$f_t(w_t) - f_t(w_{t+1}) + (\rho_t - \rho_{t-1})(R(w_t) - R(w_{t+1}) \geq \frac{\rho_t + \rho_{t-1}}{2}\|w_{t+1} - w_t\|_2^2 .$$

But the left-hand side is upper-bounded by $(L_f + (\rho_t - \rho_{t-1})L_R)\|w_{t+1} - w_t\|_2$, which implies that

$$\|w_{t+1} - w_t\|_2 \leq 2\frac{L_f + (\rho_t - \rho_{t-1})L_R}{\rho_t + \rho_{t-1}} .$$

Substituting in (29) gives the bound stated in the Proposition. ■

References

J. Abernethy, P. L. Bartlett, and E. Hazan. Blackwell approachability and low-regret learning are equivalent. In *Conference on Learning Theory (COLT)*, pages 27–46, June 2011.

R.J. Aumann and M. Maschler. *Repeated Games with Incomplete Information*. MIT Press, Boston, MA, 1995.

A. Bernstein and N. Shimkin. Response-based approachability with applications to generalized no-regret problems. *Journal of Machine Learning Research*, 16:747–773, 2015.

A. Bernstein, S. Mannor, and N. Shimkin. Opportunistic approachability and generalized no-regret problems. *Mathematics of Operations Research*, 39(4):1057–1093, 2014.

D. Blackwell. Controlled random walks. In *Proceedings of the International Congress of Mathematicians*, volume III, pages 335–338, 1954.

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

S. Boyd and L. Vandenberghe. *Convex Optimization*. Cambridge University Press, Cambridge, UK, 2004.

N. Cesa-Bianchi and G. Lugosi. *Prediction, Learning, and Games*. Cambridge University Press, New York, NY, 2006.

D. Fudenberg and D. K. Levine. *The Theory of Learning in Games*. MIT Press, Boston, MA, 1998.

J. Hannan. Approximation to Bayes risk in repeated play. *Contributions to the Theory of Games*, 3:97–139, 1957.

S. Hart and A. Mas-Colell. A general class of adaptive strategies. *Journal of Economic Theory*, 98:26–54, 2001.

E. Hazan. The convex optimization approach to regret minimization. In S. Sra et al., editor, *Optimization for Machine Learning*, chapter 10. MIT Press, Cambridge, MA, 2012.

E. Hazan. *Introduction to Online Convex Optimization*. Online book draft, http://ocobook.cs.princeton.edu/, April 2016.

E. Hazan, A. Agarwal, and S. Kale. Logarithmic regret algorithms for online convex optimization. *Machine Learning*, 69(2-3):169–192, 2007.

S. Mannor, V. Perchet, and G. Stoltz. Approachability in unknown games: Online learning meets multi-objective optimization. In *Conference on Learning Theory (COLT)*, pages 339–355, Barcelona, Spain, May 2014.

M. Maschler, E. Solan, and S. Zamir. *Game Theory*. Cambridge University Press, Cambridge, UK, 2013.

V. Perchet. Calibration and internal no-regret with partial monitoring. In *International Conference on Algorithmic Learning Theory (ALT)*, Porto, Portugal, October 2009.

V. Perchet. Approachability, regret and calibration: Implications and equivalences. *Journal of Dynamics and Games*, 1:181–254, 2014.

V. Perchet and S. Mannor. Approachability, fast and slow. In *Proc. COLT 2013: JMLR Workshop and Conference Proceedings*, volume 30, pages 474–488, 2013.

H. Peyton Young. *Strategic Learning and Its Limits*. Oxford University Press, 2004.

R.T. Rockafellar and R. Wets. *Variational Analysis*. Springer-Verlag, 1997.

S. Shalev-Shwartz. Online learning and online convex optimization. *Foundations and Trends in Machine Learning*, 4:107–194, 2011.

N. Shimkin and A. Shwartz. Guaranteed performance regions in Markovian systems with competing decision makers. *IEEE Transactions on Automatic Control*, 38(1): 84–95, 1993.

M. Zinkevich. Online convex programming and generalized infinitesimal gradient ascent. In *International Conference on Machine Learning (ICML)*, pages 928–936, 2003.