Blackwell's Approachability Theorem: A Generalization in a Special Case

Amy Greenwald, Amir Jafari and Casey Marks

Department of Computer Science
Brown University
Providence, Rhode Island 02912

CS-06-01
January 06

Brown University Technical Report (2006) CS-06-01

Blackwell's Approachability Theorem: A Generalization in a Special Case

Amy Greenwald AMY@BROWN.EDU
Department of Computer Science
Brown University
Providence, RI 02912

Amir Jafari AMIR@IAS.EDU
Institute of Advanced Studies
Einstein Drive
Princeton, NJ 08540

Casey Marks CASEY@CS.BROWN.EDU
Department of Computer Science
Brown University
Providence, RI 02912

1. Introduction

Blackwell's seminal approachability theorem provides a sufficient condition to ensure that, in a vector-valued repeated game, a learner's average rewards approach any closed and convex set $U \subseteq \mathbb{R}^n$ [1, 4]. In this paper, we prove a close cousin of Blackwell's theorem. On the one hand, our theorem specializes Blackwell's theorem: it provides a sufficient condition for the negative orthant $\mathbb{R}^n_- \subseteq \mathbb{R}^n$ to be approachable, rather than an arbitrary closed and convex subset of finite-dimensional Euclidean space. On the other hand, our sufficient condition (Equation 3) is weaker than Blackwell's original condition: our condition need only hold for some $c \in \mathbb{R}$, rather than precisely for $c = 0$. Moreover, in our framework, the opponents (i.e., not the learner) have at their disposal an arbitrary, not merely a finite, set of actions.

The main theorem in this paper was first established in Jafari's Master's thesis [5]. Our proof technique is related to that of Foster and Vohra [3], which, unlike the proof of Blackwell's theorem in Cesa-Bianchi and Lugosi [2], proves all lemmas from first principles.

2. Statement of Main Theorem

Consider an agent with a finite set of actions A playing a game against a set of opponents who play actions in the (arbitrary) joint action space A'. (The opponents' joint action space can be interpreted as the product of independent action sets.) Associated with each possible outcome is a vector given by the function $\rho : A \times A' \to V$, where V is a vector space over $\mathbb{R}$ with an inner product $\cdot$ and a distance metric d defined by the inner product in the standard manner: i.e., $d(x, y) = \|x - y\|_2$, for all $x, y \in V$.

Definition 1 (Game) *A* vector-valued game *is a 4-tuple* $\Gamma = (A, A', V, \rho)$*.*

We study *infinitely-repeated* vector-valued games Γ^∞ in which the agent interacts with its opponents repeatedly and indefinitely. Recall that the agent's action set A is assumed to be finite. We denote by $\Delta(A)$ the set of probability distributions over the set A, and we allow agents to play *mixed strategies*, which means that rather than selecting an action $a \in A$ to play at each round, the agent selects a probability distribution $q \in \Delta(A)$. More specifically, an arbitrary round t (for $t \geq 1$) proceeds as follows:

1. the agent selects a mixed strategy $q_t \in \Delta(A)$,
2. the agent plays an action $a_t \in A$ (which is sampled according to the distribution q_t); simultaneously, the opponents play action a'_t
3. the agent observes reward vector $\rho(a_t, a'_t)$

Definition 2 (Action History) *Given an infinitely-repeated vector-valued game Γ^∞ the* set of action histories of length t*, for $t \geq 0$, is denoted by H_t. For $t \geq 1$, H_t is given by $(A \times A')^t$: e.g., $h = \{a_\tau, a'_\tau\}_{\tau=1}^t \in H_t$. The set H_0 is defined to be a singleton.*

Definition 3 (Learning Algorithm) *Given an infinitely-repeated vector-valued game Γ^∞, a* learning algorithm *is a sequence of functions $\mathcal{L} = \{L_t\}_{t=1}^\infty$, where $L_t : H_{t-1} \to \Delta(A)$.*

We are interested in the properties of learning algorithms employed by an agent playing an infinitely-repeated vector-valued game Γ^∞. Given our agent's learning algorithm $\mathcal{L}$, and WLOG a single opposing learning algorithm $\mathcal{L}'$, we define a probability space whose universe consists of sequences of pairs of actions and whose measure can be defined inductively as:

$$\begin{aligned} & P\left[(a_t, a'_t) = (\alpha, \alpha') \mid a_\tau = \alpha_\tau, a'_\tau = \alpha'_\tau \ \forall \tau = 1, \ldots, t-1\right] \\ = \ & (L_t((\alpha_1, \alpha'_1), \ldots, (\alpha_{t-1}, \alpha'_{t-1}))(\alpha))(L'_t((\alpha_1, \alpha'_1), \ldots, (\alpha_{t-1}, \alpha'_{t-1}))(\alpha')) \end{aligned} \tag{1}$$

for all $\alpha \in A, \alpha' \in A'$. In this probability space, we define two sequences of random variables: cumulative rewards $R_t = \sum_{\tau=1}^t \rho(a_\tau, a'_\tau)$ and average rewards $\bar{\rho}_t = \frac{R_t}{t}$.

Now, following Blackwell, we define the notion of approachability as follows:

Definition 4 (Approachability) *Given an infinitely-repeated vector-valued game Γ^∞, a set $U \subseteq V$, and a learning algorithm $\mathcal{L}$, the set U is said to be* approachable *by $\mathcal{L}$, if for all $\delta > 0$, there exists t_0 such that for any opposing learning algorithm $\mathcal{L}'$,*

$$P\left[\exists t \geq t_0 \text{ s.t. } d\left(U, \bar{\rho}_t\right) \geq \delta\right] < \delta \tag{2}$$

Hence, if a learning algorithm $\mathcal{L}$ approaches a set $U \subseteq V$, then $d(U, \overline{\rho}_t) \to 0$ almost surely.

The following theorem, which is the main result of this paper, gives a sufficient condition for the negative orthant, that is, the set $\mathbb{R}^d_- = \{x \in \mathbb{R}^d \mid x_i \leq 0, \text{ for all } 1 \leq i \leq d\} \subseteq \mathbb{R}^d$, to be approachable by a learning algorithm $\mathcal{L}$ in an infinitely-repeated vector-valued game $(A, A', \mathbb{R}^d, \rho)^\infty$ where $d \in \mathbb{N}$ and $\rho(A \times A')$ is bounded.

For $x \in \mathbb{R}^d$, define x^+ by $(x^+)_i = \max\{x_i, 0\}$, for all $1 \leq i \leq d$.

Theorem 5 (Jafari) *Given an infinitely-repeated vector-valued game $(A, A', \mathbb{R}^d, \rho)^\infty$ with $d \in \mathbb{N}$ and $\rho(A \times A')$ bounded and a learning algorithm $\mathcal{L} = \{L_t\}_{t=1}^\infty$, the negative orthant $\mathbb{R}^d_- \subseteq \mathbb{R}^d$ is approachable by $\mathcal{L}$ if there exists a constant $c \in \mathbb{R}$ such that for all times $t \geq 1$, for all action histories $h \in H_{t-1}$ of length $t-1$, and for all opposing actions a',*

$$(R_{t-1}(h))^+ \cdot \mathbb{E}_{a \sim L_t(h)}\left[\rho(a, a')\right] \leq c \tag{3}$$

where $R_t(h) \equiv \sum_{\tau=1}^t \rho(a_\tau, a'_\tau)$ and $\mathbb{E}_{a \sim q}\left[\rho(a, a')\right] \equiv \sum_{a \in A} q(a)\rho(a, a')$.

3. Preliminary Linear Algebra Lemmas

Lemma 6 *For $x, y \in \mathbb{R}^n$,*

$$\|(x+y)^+\|_2^2 \leq \|x^+ + y\|_2^2 \tag{4}$$

Equivalently,

$$\|(x+y)^+\|_2^2 \leq \|x^+\|_2^2 + 2(x^+ \cdot y) + \|y\|_2^2 \tag{5}$$

Proof Observe that $\|(x+y)^+\|_2^2 = \sum_i ((x_i + y_i)^+)^2$ and $\|x^+ + y\|_2^2 = \sum_i (x_i^+ + y_i)^2$. Thus, it suffices to show that $((x_i + y_i)^+)^2 \leq (x_i^+ + y_i)^2$ for all i.

Case 1: $x_i + y_i \leq 0$. Here, $((x_i + y_i)^+)^2 = 0 \leq (x_i^+ + y_i)^2$.

Case 2: $x_i + y_i > 0$, $x_i \geq 0$. Here, $((x_i + y_i)^+)^2 = (x_i + y_i)^2 = (x_i^+ + y_i)^2$.

Case 3: $x_i + y_i > 0$, $x_i < 0$. Here, $0 < x_i + y_i < y_i$, so that $((x_i + y_i)^+)^2 = (x_i + y_i)^2 < y_i^2 = (x_i^+ + y_i)^2$. ∎

Lemma 7 *For $x, y \in \mathbb{R}^n$,*

$$\|(x+y)^+\|_2^2 \geq 2(x^+ \cdot y) + \|x^+\|_2^2 \tag{6}$$

Proof

$$\begin{aligned}
& \|(x+y)^+\|_2^2 - 2(x^+ \cdot y) - \|x^+\|_2^2 && (7)\\
= \; & \sum_{i=1}^n \left[((x_i + y_i)^+)^2 - 2x_i^+ y_i - (x_i^+)^2\right] && (8)\\
= \; & \sum_{i=1}^n \left[((x_i + y_i)^+)^2 - 2x_i^+ y_i - 2x_i^+ x_i + (x_i^+)^2\right] && (9)\\
= \; & \sum_{i=1}^n \left[((x_i + y_i)^+)^2 - 2x_i^+ (x_i + y_i) + (x_i^+)^2\right] && (10)\\
\geq \; & \sum_{i=1}^n \left[((x_i + y_i)^+)^2 - 2x_i^+ (x_i + y_i)^+ + (x_i^+)^2\right] && (11)\\
= \; & \sum_{i=1}^n ((x_i + y_i)^+ - x_i^+)^2 && (12)\\
\geq \; & 0 && (13)
\end{aligned}$$

Line (9) follows from the observation that $a^+ a = (a^+)^2$, for all $a \in \mathbb{R}$. ∎

4. Preliminary Probability Lemmas

Let $(\Omega, \mathcal{F}, P)$ be a probability space with a filtration $(\mathcal{F}_t : t \geq 0)$: that is, a sequence of σ-algebras with $\mathcal{F}_t \subseteq \mathcal{F}$ for all t and $\mathcal{F}_s \subseteq \mathcal{F}_t$ for all $s < t$. A stochastic process $(Z_t : t \geq 0)$ is said to be adapted to a filtration $(\mathcal{F}_t : t \geq 0)$ if Z_t is $\mathcal{F}_t$-measurable, for all times t: i.e., if the value of Z_t is determined by $\mathcal{F}_t$, the information available at time t.

We denote by $\mathbb{E}_t$ the conditional expectation with respect to $\mathcal{F}_t$: i.e., $\mathbb{E}_t[\cdot] = \mathbb{E}[\cdot \mid \mathcal{F}_t]$.

Lemma 8 (Product Lemma) *Assume the following:*

1. $(Z_t : t \geq 0)$ *is an adapted process such that* $\forall t$, $0 \leq Z_t < k$ *a.s., for some* $k \in \mathbb{R}$*;*
2. $\mathbb{E}_{t-1}[Z_t] \leq c_t$ *a.s., for* $t \geq 1$*, and* $\mathbb{E}[Z_0] \leq c_0$*, where* $c_t \in \mathbb{R}$*, for all* t*.*

For fixed T*,*

$$\mathbb{E}\left[\prod_{t=0}^{T} Z_t\right] \leq \prod_{t=0}^{T} c_t \tag{14}$$

Proof The proof is by induction. The claim holds for $T = 0$ by assumption. We assume it also holds for T and show it holds for $T + 1$:

$$\begin{aligned}
\mathbb{E}\left[\prod_{t=0}^{T+1} Z_t\right] &= \mathbb{E}\left[\mathbb{E}_T\left[\prod_{t=0}^{T+1} Z_t\right]\right] & (15)\\
&= \mathbb{E}\left[\mathbb{E}_T\left[\left(\prod_{t=0}^{T} Z_t\right) Z_{T+1}\right]\right] & (16)\\
&= \mathbb{E}\left[\left(\prod_{t=0}^{T} Z_t\right) \mathbb{E}_T\left[Z_{T+1}\right]\right] & (17)\\
&\leq \mathbb{E}\left[\left(\prod_{t=0}^{T} Z_t\right) c_{T+1}\right] & (18)\\
&= c_{T+1}\mathbb{E}\left[\prod_{t=0}^{T} Z_t\right] & (19)\\
&\leq \prod_{t=0}^{T+1} c_t & (20)
\end{aligned}$$

Line (15) follows from the tower property, also known as the law of iterated expectations: If a random variable X satisfies $\mathbb{E}\left[|X|\right] < \infty$ and $\mathcal{H}$ is a sub-σ-algebra of $\mathcal{G}$, which in turn is a sub-σ-algebra of $\mathcal{F}$, then $\mathbb{E}[\mathbb{E}[X \mid \mathcal{G}] \mid \mathcal{H}] = \mathbb{E}[X \mid \mathcal{H}]$ almost surely [6]. Note that $\mathbb{E}\left[\prod_{t=0}^{T+1} Z_t\right] < \infty$. Line (17) follows because $\prod_{t=0}^{T} Z_t$ is $\mathcal{F}_T$-measurable and $\mathbb{E}\left[\prod_{t=0}^{T} Z_t\right] < \infty$. Line (18) follows by assumption, since Z_t, for $t \geq 0$, is nonnegative with probability 1. Line (20) follows from the induction hypothesis. ∎

Lemma 9 (Supermartingale Lemma) *Assume the following:*

1. $(M_t : t \geq 0)$ *is a supermartingale, i.e.* $(M_t : t \geq 0)$ *is an adapted process* s.t. *for all* t*,* $\mathbb{E}[|M_t|] < \infty$ *and for* $t \geq 1$*,* $\mathbb{E}_{t-1}[M_t] \leq M_{t-1}$ *a.s.;*

2. f is a nondecreasing positive function s.t. *for $t \geq 1$, $|M_t - M_{t-1}| \leq f(t)$ a.s..*

If $M_0 = m \in \mathbb{R}$ a.s., then for fixed T, $P[M_T \geq 2\epsilon T f(T)] \leq e^{\epsilon m/f(0) - \epsilon^2 T}$, for all $\epsilon \in [0,1]$.

Proof For $t \geq 0$, let

$$Y_t = \frac{M_t}{f(T)} \tag{21}$$

and for $t \geq 1$, let $X_t = Y_t - Y_{t-1}$, so that

$$Y_t = \sum_{\tau=0}^{t} X_\tau. \tag{22}$$

Note that $X_0 = Y_0 = m/f(0)$ a.s..

Because M_t is supermartingale and $f(t)$ is positive, for $t \geq 1$,

$$\mathbb{E}_{t-1}[X_t] = \mathbb{E}_{t-1}[Y_t] - Y_{t-1} = \frac{\mathbb{E}_{t-1}[M_t] - M_{t-1}}{f(T)} \leq 0 \quad \text{a.s.} \tag{23}$$

Because f is nondecreasing, $f(t) \leq f(T)$ for all t; hence, for $1 \leq t \leq T$,

$$|X_t| = |Y_t - Y_{t-1}| = \left| \frac{M_t}{f(T)} - \frac{M_{t-1}}{f(T)} \right| \leq \frac{|M_t - M_{t-1}|}{f(t)} \leq 1 \quad \text{a.s.} \tag{24}$$

Thus, for $t \geq 1$,

$$\mathbb{E}_{t-1}\left[e^{\epsilon X_t}\right] \leq 1 + \epsilon \mathbb{E}_{t-1}[X_t] + \epsilon^2 \mathbb{E}_{t-1}[X_t^2] \leq 1 + \epsilon^2 \quad \text{a.s.} \tag{25}$$

The first inequality follows from the fact that $e^y \leq 1 + y + y^2$ for $y \leq 1$, and $\epsilon X_t \leq 1$ a.s., since $\epsilon \in [0,1]$ and $|X_t| \leq 1$ a.s. by Line (24). The second inequality follows from Line (23).

Therefore,

$$\begin{aligned}
P\left[M_T \geq 2\epsilon T f(T)\right] &= P\left[Y_T \geq 2\epsilon T\right] && (26)\\
&= P\left[e^{\epsilon Y_T} \geq e^{2\epsilon^2 T}\right] && (27)\\
&\leq \frac{\mathbb{E}\left[e^{\epsilon Y_T}\right]}{e^{2\epsilon^2 T}} && (28)\\
&= \frac{\mathbb{E}\left[e^{\epsilon \sum_{t=0}^{T} X_t}\right]}{e^{2\epsilon^2 T}} && (29)\\
&= \frac{\mathbb{E}\left[\prod_{t=0}^{T} e^{\epsilon X_t}\right]}{e^{2\epsilon^2 T}} && (30)\\
&\leq \frac{e^{\epsilon m/f(0)}(1+\epsilon^2)^T}{e^{2\epsilon^2 T}} && (31)\\
&\leq \frac{e^{\epsilon m/f(0)} e^{\epsilon^2 T}}{e^{2\epsilon^2 T}} && (32)\\
&= e^{\epsilon m/f(0) - \epsilon^2 T} && (33)
\end{aligned}$$

Line (28) follows from Markov's inequality. Line (31) follows from the Product Lemma (since $(X_t : t \geq 0)$ is an adapted process), Line (25), and the assumption that $M_0 = m$ a.s.. Line (32) follows from the fact that $(1 + x) \leq e^x$. ∎

Lemma 10 (Convergence Lemma) *Given a function $f : \mathbb{R} \to \mathbb{R}$ that maps the positive reals onto the positive reals and a stochastic process $(X_t : t \geq 0)$, if for all $\epsilon > 0$, there exists T such that for all $t \geq T$, $P[X_t \geq f(\epsilon)] \leq e^{-\epsilon t}$, then for all $\delta > 0$, there exists t_0 such that $P[\exists t \geq t_0 \text{ s.t. } X_t \geq \delta] < \delta$.*

Proof Given an arbitrary $\delta > 0$, choose $\epsilon \in f^{-1}(\delta)$. By assumption, there exists T such that for all $t > T$, $P[X_t \geq \delta] \leq e^{-\epsilon t}$. Now, for all $t' \geq T$,

$$\begin{aligned} P\left[\exists t \geq t' s.t.\ X_t \geq \delta\right] &= P\left[\bigcup_{t \geq t'} (X_t \geq \delta)\right] && (34) \\ &\leq \sum_{t \geq t'} P[X_t \geq \delta] && (35) \\ &\leq \sum_{t \geq t'} e^{-\epsilon t} && (36) \\ &= \frac{e^{-\epsilon t'}}{1 - e^{-\epsilon}} && (37) \end{aligned}$$

Hence, for sufficiently large t_0, $P\left[\exists t \geq t_0\ s.t.\ X_t \geq \delta\right] < \delta$. ∎

5. Proof of Main Theorem

To prove our main theorem, we define a stochastic process and we use Lemmas 6 and 7 (together with the assumptions of the theorem) to show that this process satisfies the assumptions of the Supermartingale Lemma. Finally, we apply the Convergence Lemma.

Theorem Given an infinitely-repeated vector-valued game $(A, A', \mathbb{R}^d, \rho)^\infty$ with $d \in \mathbb{N}$ and $\rho(A \times A')$ bounded and a learning algorithm $\mathcal{L} = \{L_t\}_{t=1}^\infty$, the negative orthant $\mathbb{R}^d_- \subseteq \mathbb{R}^d$ is approachable by $\mathcal{L}$ if there exists a constant $c \in \mathbb{R}$ such that for all times $t \geq 1$, for all action histories $h \in H_{t-1}$ of length $t-1$, and for all opposing actions a',

$$(R_{t-1}(h))^+ \cdot \mathbb{E}_{a \sim L_t(h)}\left[\rho(a, a')\right] \leq c \tag{38}$$

where $R_t(h) \equiv \sum_{\tau=1}^t \rho(a_\tau, a'_\tau)$ and $\mathbb{E}_{a \sim q}\left[\rho(a, a')\right] \equiv \sum_{a \in A} q(a)\rho(a, a')$.

Proof Recall how the learning algorithms $\mathcal{L}$ and $\mathcal{L}'$ induce a probability space over sequences of pairs of actions. We view the agent's sequence of actions $a_1, a_2, \ldots$ as a stochastic process with $(\mathcal{F}_t : t \geq 0)$ as its natural filtration. Observe that $(R_t : t \geq 0)$ is adapted to this filtration because $\rho(a_t, a'_t)$, and hence R_t, is $\mathcal{F}_t$-measurable.

The proof relies on the following observations:

1. By assumption, for all histories $h \in H_{t-1}$ of length $t-1$, for all opposing actions a'_t,

$$(R_{t-1}(h))^+ \cdot \mathbb{E}_{a_t \sim L_t(h)}\left[\rho(a_t, a'_t)\right] \leq c \tag{39}$$

If h is a random variable that takes values in H_{t-1}, then h is $\mathcal{F}_{t-1}$ measurable. For all opposing actions a'_t,

$$\mathbb{E}_{a_t \sim L_t(h)}\left[\rho(a_t, a'_t)\right] = \mathbb{E}\left[\rho(a_t, a'_t) \mid \mathcal{F}_{t-1}\right] = \mathbb{E}_{t-1}\left[\rho(a_t, a'_t)\right] \tag{40}$$

Therefore,

$$R_{t-1}^{+} \cdot \mathbb{E}_{t-1}\left[\rho(a_t, a_t')\right] \le c \quad \text{a.s.} \tag{41}$$

2. By assumption, $\rho(A \times A')$ is bounded. WLOG, assume $\rho : A \times A' \to [0,1]^d$ so that

$$\|\rho(a, a')\|_2^2 \le d \tag{42}$$

for all $a \in A$ and $a' \in A'$.

We define $M_t = \|R_t^+\|_2^2 - (2c+d)t$, for all t, and we show that $(M_t : t \ge 0)$ satisfies the assumptions of the Supermartingale Lemma.

Assumption 1: First, observe that $(M_t : t \ge 0)$ is an adapted process because $(R_t : t \ge 0)$ is an adapted process. Second, since $\rho(A \times A')$ is bounded, R_t, and hence M_t, is bounded, for all t. In particular, $\mathbb{E}[|M_t|] < \infty$, for all t. Third, for $t \ge 1$,

$$\begin{aligned}
\mathbb{E}_{t-1}[M_t] &= \mathbb{E}_{t-1}\left[\|R_t^+\|_2^2\right] - (2c+d)t && (43)\\
&\le \mathbb{E}_{t-1}\left[\|R_{t-1}^+\|_2^2 + 2R_{t-1}^+ \cdot \rho(a_t, a_t') + \|\rho(a_t, a_t')\|_2^2\right] - (2c+d)t && (44)\\
&\le \|R_{t-1}^+\|_2^2 + 2c + d - (2c+d)t \quad \text{a.s.} && (45)\\
&= M_{t-1} \quad \text{a.s.} && (46)
\end{aligned}$$

Line (44) follows from Lemma 6. Line (45) follows from Lines 41 and 42.

Assumption 2: Let $t \ge 1$. Define $f(t) = 2c + 2dt$. Note the following:

$$\begin{aligned}
\left|R_{t-1}^+ \cdot \rho(a_t, a_t')\right| &\le \|R_{t-1}^+\|_2 \|\rho(a_t, a_t')\|_2 && (47)\\
&\le \left(\sum_{\tau=1}^{t-1} \|\rho(a_\tau, a_\tau')\|_2\right) \|\rho(a_t, a_t')\|_2 && (48)\\
&\le (t-1)\sqrt{d}\sqrt{d} && (49)\\
&= (t-1)d && (50)
\end{aligned}$$

Line (47) follows from the Cauchy-Schwarz inequality. Line (48) follows from the triangle inequality. Line (49) follows from Line (42).

Two cases arise.

Case 1: $M_t - M_{t-1} \ge 0$.

$$\begin{aligned}
|M_t - M_{t-1}| &= M_t - M_{t-1} && (51)\\
&= \|R_t^+\|_2^2 - \|R_{t-1}^+\|_2^2 - (2c+d) && (52)\\
&\le 2R_{t-1}^+ \cdot \rho(a_t, a_t') + \|\rho(a_t, a_t')\|_2^2 - (2c+d) && (53)\\
&\le 2\left|R_{t-1}^+ \cdot \rho(a_t, a_t')\right| + \|\rho(a_t, a_t')\|_2^2 - (2c+d) && (54)\\
&\le 2(t-1)d + d - (2c+d) && (55)\\
&= 2dt - 2(c+d) && (56)\\
&< f(t) && (57)
\end{aligned}$$

Line (53) follows from Lemma 6. Line (55) follows from Lines (42) and (50).

Case 2: $M_t - M_{t-1} < 0$.

$$\begin{aligned}
|M_t - M_{t-1}| &= M_{t-1} - M_t && (58)\\
&= (2c+d) - \|R_t^+\|_2^2 + \|R_{t-1}^+\|_2^2 && (59)\\
&\leq (2c+d) - 2R_{t-1}^+ \cdot \rho(a_t, a_t') && (60)\\
&\leq (2c+d) + 2\left|R_{t-1}^+ \cdot \rho(a_t, a_t')\right| && (61)\\
&\leq (2c+d) + 2(t-1)d && (62)\\
&= 2c + 2dt - d && (63)\\
&< f(t) && (64)
\end{aligned}$$

Line (60) follows from Lemma 7. Line (62) follows from Line (50).

Hence, M_t satisfies the assumption of the Supermartingale Lemma, so that

$$P\left[\|R_t^+\|_2^2 - (2c+d)t \geq 4t(c+dt)\sqrt{\epsilon}\right] \leq e^{-\epsilon t} \tag{65}$$

for all $\epsilon \in [0,1]$ (note: $M_0 = 0$ a.s.). Observe that $d\left(\mathbb{R}_-^d, \bar{\rho}_t\right) = d\left(\mathbb{R}_-^d, \frac{R_t}{t}\right) = \frac{\|R_t^+\|_2}{t}$. Thus,

$$P\left[d\left(\mathbb{R}_-^d, \bar{\rho}_t\right) \geq \sqrt{\frac{2c+d}{t} + \frac{4c\sqrt{\epsilon}}{t} + 4d\sqrt{\epsilon}}\right] \leq e^{-\epsilon t} \tag{66}$$

For sufficiently large t (just how large depends on c, d, and ϵ), $\frac{2c+d}{t} + \frac{4c\sqrt{\epsilon}}{t} \leq d\sqrt{\epsilon}$, so that

$$P\left[d\left(\mathbb{R}_-^d, \bar{\rho}_t\right) \geq \sqrt{5d}\sqrt[4]{\epsilon}\right] \leq e^{-\epsilon t} \tag{67}$$

Finally, we apply the Convergence Lemma to obtain: for all $\delta > 0$, there exists t_0 such that $P\left[\exists t \geq t_0 \ s.t.\ d\left(\mathbb{R}_-^d, \bar{\rho}_t\right) \geq \delta\right] < \delta$. Because t_0 depends only on c, d, and δ, this inequality holds for the probability space regardless of the opposing learning algorithm $\mathcal{L}'$. ∎

Acknowledgments

We gratefully acknowledge Dean Foster for sparking our interest in this topic, and for ongoing discussions that helped to clarify many of the technical points in this paper. We also thank David Gondek, Zheng Li, and Martin Zinkevich for their feedback.

This research was supported by NSF Career Grant #IIS-0133689 and NSF IGERT Grant #9870676.

References

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

[2] Nicolò Cesa-Bianchi and Gábor Lugosi. *Prediction, Learning, and Games.* Cambridge University Press, Cambridge, England, Forthcoming.

[3] D. Foster and R. Vohra. Regret in the on-line decision problem. *Games and Economic Behavior*, 21:40–55, 1997.

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

[5] A. Jafari. *On the Notion of Regret in Infinitely Repeated Games*. Master's Thesis, Brown University, Providence, May 2003.

[6] D. Williams. *Probability with Martingales*. Cambridge University Press, Cambridge, 1991.