No Regret with Bounded Computational Capacity

Ehud Lehrer* and Eilon Solan†

May 30, 2003

Abstract

We deal with no regret and related aspects of vector-payoff games when one of the players is limited in computational capacity. We show that player 1 can almost approach with bounded-recall strategies, or with finite automata, any convex set which is approachable when no capacity bound is present. In particular we deduce that with bounded computational capacity player 1 can ensure having almost no regret.

*School of Mathematical Sciences, Tel Aviv University, Tel Aviv 69978, Israel; e-mail: lehrer@post.tau.ac.il

†School of Mathematical Sciences, Tel Aviv University, Tel Aviv 69978, Israel, and MEDS Department, Kellogg School of Management, Northwestern University, 2001 Sheridan Road, Evanston, IL 60208-2001. e-mail: eilons@post.tau.ac.il, e-solan@kellogg.northwestern.edu

1 Introduction

In a sequential decision problem, at every stage a decision maker (DM) ought to take an action. He then receives a stage-payoff that depends on both his stage-action and on the realized state of nature. The evolution of the state of nature is independent of the actions taken by DM. A strategy of DM is a prescription what he should do (possibly randomly choosing an action) after any possible history of previous actions and past realized states. A strategy is *Hannan no-regret* if it ensures a long-run average payoff that is at least as high as what DM could have achieved had he played constantly the same action. In other words, playing a Hannan no-regret strategy, DM has no regret for not constantly playing a best response against the empirical distribution of states.

Hannan (1957) showed in a rather complicated proof that if there are finitely many state of nature, there is always a Hannan no-regret strategy. This theorem can be derived from Blackwell's (1956) approachability theorem, which applies to two-player games with vector payoffs. Both proofs employ strategies that are not computationally bounded.

Two types of strategies with bounded computational capacity has been extensively studied in the literature. First, strategies with bounded recall (see, e.g., Lehrer, 1988, Watson, 1994, and Jéhiel, 1995), which can use only the recent history, and second, strategies that can be implemented by finite automata (see, e.g., Aumann, 1981, Neyman, 1985, Rubinstein, 1986, and Ben Porath, 1990).

The relation between various notions of no regret and repeated games with vector payoffs is now well established. Repeated games with vector payoffs are two-player zero-sum games in which the payoffs are not scalars, but vectors in some Euclidean space. Such games naturally arise when the players have several objective functions. Foster and Vohra (1999) used games with vector payoffs to show a process of decentralized actions that converges to correlated equilibrium. Fudenberg and Levine (1999) and Hart and Mas-

Colell (2000) introduced stronger no-regret notions than Hanann's notion, and show that there always exists a strategy that satisfies the stronger version. Rustichini (1999), using Balckwell's approachability theorem, proved a no-regret theorem when the decision maker has imperfect monitoring. Lehrer (2003) used games whose payoffs are infinite dimensional to show that there exists a strategy immunized against infinitely many replacing schemes.

To find the simplest structure of Hannan no-regret strategies we will analyze repeated games with vector payoffs. For a given sequential decision problem one can define a certain repeated game with vector payoffs such that (a) the set of strategies of DM in the decision problem stands in 1-1 relation with the set of strategies of player 1 in the repeated game with vector payoffs, and (b) a strategy of DM is Hannan no-regret if and only if the corresponding strategy in the repeated game with payoff vectors ensures that the long-run average payoff remains close to the non-negative orthant.

Thus, we are led to a more general question in the setup of repeated games with vector payoffs, namely, the characterization of sets in a Euclidean space with the property that player 1 can guarantee that the long-run average payoff remains close to the set.

In a repeated game with vector payoffs, a target set is *approachable* by player 1 if he has a strategy that ensures that the long-run average payoff is as close as required to the target set, regardless of the strategy player 2 employs (see Blackwell, 1956). The target set is *approachable with bounded-recall strategies* by player 1 if for any vicinity of the target set there is a bounded-recall strategy of player 1, which guarantees that the long-run average payoff remains in this vicinity. The set is *approachable with automata* by player 1 if the same can be done using strategies that can be implemented by finite automata. Since any bounded-recall strategy can be implemented by an automaton, any set which is approachable with bounded-recall strategies is also approachable by automaton. However, it is not clear that the converse holds. As we prove below, the two notions of approachability with bounded computational capacity strategies are equivalent. Moreover, we fully characterize the family of sets which are approachable with bounded-recall strategies. A minimal (with respect to set inclusion) approachable set is approachable with bounded-recall strategies if and only if it is convex. Furthermore, a set is approachable with bounded-recall strategies if and only if it contains a convex approachable set. A complete characterization of the family of minimal approachable set was given by Spinat (2002).

Back to the issue of Hannan no-regret strategies, since the non-negative orthant is approachable in the corresponding repeated game with vector-payoffs, and is a convex set, our results imply that in any sequential decision problem, DM has a bounded-recall strategy that is almost Hannan no-regret. That is, for every given $\delta > 0$, there is a bounded-recall strategy for which the regret is at most δ, in the long run.

2 The Model and the Main Results

2.1 Repeated games with vector payoffs

In this section we define repeated games with vector-payoffs.

A two-player *repeated game with vector-payoffs* is a triplet (I, J, M), where I and J are finite sets of actions for the two players, and $M = (m_{i,j})_{i \in I, j \in J}$ is a vector-payoff matrix, so that $m_{i,j} \in \mathbf{R}^d$ for every $i \in I$ and $j \in J$. We assume throughout that $\|M\|_\infty \leq 1$; that is, all payoffs are bounded by 1. We also assume that $|I| \geq 2$: player 1 has at least two available actions.

At every stage n the two players choose, independently and simultaneously, a pair of actions (i_n, j_n), each one in his action set. A *strategy* of player 1 (resp. player 2) is a function $\sigma : \cup_{n=0}^{\infty}(I \times J)^n \to \Delta(I)$ (resp. $\tau : \cup_{n=0}^{\infty}(I \times J)^n \to \Delta(J)$),[1] where $\Delta(A)$ is the space of probability distributions over $A = I, J$. We denote by $\mathcal{S}$ and $\mathcal{T}$ the strategy spaces of the players 1 and 2, respectively.

[1] For every finite set B we identify B^0 with a set that contains a single element.

For every stage t denote by i_t and j_t the actions of the players 1 and 2 at stage t, respectively. The average payoff vector up to stage n is

$$\bar{x}_n = \frac{\sum_{t=1}^{n} m_{i_t,j_t}}{n}.$$

Note that for every $n \in \mathbf{N}$, $\bar{x}_n$ is a random variable whose distribution is determined by the strategies of both players.

Let $d(x,y)$ denote the Euclidean distance between the points x and y in $\mathbf{R}^d$. For every set F in $\mathbf{R}^d$ and every $x \in \mathbf{R}^d$, let $d(x,F) = \inf_{y \in F} d(x,y)$ be the distance of x from F. For every $\delta > 0$, let $B(F,\delta) = \{x \in \mathbf{R}^d \colon d(x,F)) \leq \delta\}$ be the set of all points which are δ-close to F.

Definition 1 *A set F is* approachable *by player 1 if there exists a strategy $\sigma \in \mathcal{S}$ such that*

$$\forall \varepsilon > 0, \forall \eta > 0, \exists N, \forall \tau \in \mathcal{T}, \qquad \mathbf{P}_{\sigma,\tau}(\sup_{n \geq N} d(\bar{x}_n, F) \geq \varepsilon) < \eta.$$

In this case we say that σ approaches *F.*

A set F is approachable if player 1 can guarantee with arbitrarily high probability that the long-run average payoff will be as close to F as he wishes.

Blackwell (1956) provided a sufficient condition for a set to be approachable. Spinat (2002) fully characterized the family of approachable sets.

2.2 On bounded-capacity strategies

In this section we define two types of bounded-capacity strategies, namely, strategies with bounded recall and strategies that can be implemented by automata. We then combine the notion of approachability with those two types of strategies.

Let k be a natural number. A *k-bounded-recall strategy* of player 1 (resp. player 2) is a pair (m, σ) (resp. (m, τ)), where $m \in (I \times J)^k$, and

$\sigma : (I \times J)^k \to \Delta(I)$ (resp. $\tau : (I \times J)^k \to \Delta(J)$). When playing a k-bounded-recall strategy (m, σ), at any stage player 1 plays $\sigma(x)$, where x is the string of the last k joint actions. He starts the game with the (virtual) memory of m. Thus, at the first stage he plays the mixed action $\sigma(m)$, at the second stage he plays $\sigma(m', i_1, j_1)$, where m' are the first $k-1$ coordinates of m and (i_1, j_1) is the realized pair of actions of the two players at the first stage, and so on.

We denote by $\mathcal{S}_{BR}$ the set of all bounded-recall strategies of player 1.

A (non-deterministic) *automaton* A is given by (i) a finite set of states, (ii) a probability distribution over the set of states, according to which the initial state is chosen, (iii) a finite set of inputs, (iv) a finite set of outputs, (v) a function that assigns to every state a probability distribution over outputs, and (vi) a transition rule, that assigns to every state and every input a probability distribution over states. The number of states of the automaton is the *size* of the automaton.

The initial state of the automaton is chosen according to the initial distribution given in (ii). At every stage, as a function of the current state and of the input, an output is chosen by the probability distribution given in (v), and a new state is chosen according to the probability distribution given in (vi).

The literature usually assumes that one state is designated as the initial state. Since the state of the automaton is not observed, as the automaton evolves an outside observer may only infer a posterior probability over the current state of the automaton using past inputs and outputs. It is therefore more convenient to assume that the initial state is chosen at random.

When the set of inputs of the automaton is the set $I \times J$ of pairs of actions, and the set of outputs is the set I of actions of player 1, an automaton defines a strategy for player 1: at each stage player 1 simply plays the action which is the output of the automaton at that stage, and the input for the automaton is the pair of actions just played by both players.

We denote by $\mathcal{S}_A$ the set of all strategies of player 1 that can be implemented by an automaton.

Remark 1 *Every k-bounded-recall strategy can be implemented by an automaton with $|I \times J|^k$ states.*

We are interested in studying when a given set is approachable by an automaton and by a bounded-recall strategy.

Definition 2 *A set F is* approachable with bounded-recall strategies *(resp.* approachable with automata*) by player 1 if for every $\delta > 0$ there exists a strategy $\sigma \in \mathcal{S}_{BR}$ (resp. $\sigma \in \mathcal{S}_A$) that approaches $B(F, \delta)$.*

Remark 2 *Observe that if F is approachable with bounded-recall strategies (or with automata), then so is any set that contains F. Also, if the closure of F is approachable with bounded-recall strategies (or with automata), then so is F.*

Remark 3 *By Remark 1, every set approachable with bounded-recall strategies is also approachable with automata. As stated in Corollary 1 below, the converse is also true.*

2.3 No-Regret and Approachability

Consider a sequential decision problem, where the decision maker (DM) chooses at every stage n an action from a finite set I. When DM chooses i, he receives a stage-reward $u_{i,j}$, where j is the current state of nature. We assume that the set J of possible states of nature is finite, and that the evolution of the state of nature is independent of the actions chosen by DM.

We denote by i_t the action chosen by DM at stage t, and by j_t the state of nature at that stage. At each stage n, DM can compare the actual average payoff up to stage n, $\bar{x}_n = \frac{\sum_{t=1}^n u_{i_t,j_t}}{n}$, with the payoff he would have gotten had be played constantly the action $i \in I$, $r_n^i = \frac{\sum_{t=1}^n u_{i,j_t}}{n}$. DM has no regret at stage n if $\bar{x}_n \geq r_n^i$, for each $i \in I$. A strategy σ is Hannan no-regret if for every sequence $j_1, j_2, \ldots$ of states of nature and every n sufficiently large, DM has no regret at stage n. Formally,

Definition 3 *Let $\delta \geq 0$. A strategy σ is* Hannan δ-no-regret *if for every sequence $\mathbf{j} = (j_t)$ of states of nature and every action $i \in I$,*

$$\liminf_{n\to\infty} \mathbf{E}_{\sigma,\mathbf{j}}(\bar{x}_n - r_n^i) \geq -\delta.$$

A Hannan 0-no-regret strategy is also termed *Hannan no-regret* strategy.

The fact that σ is Hannan no-regret means that against any sequence of states of nature, the long-run average payoff is at least the average payoff DM could have achieved had he known in advance the empirical frequencies of the various states of nature, and played a best response against it.

Observe that the set of strategies of DM in a sequential decision problem coincides with the set of strategies of player 1 in a two-player game where the actions of the two players are I and J respectively.

It is well known that a strategy is Hannan no-regret if and only if it approaches the non-negative orthant in a proper two-player game with vector payoffs. Define the following vector-payoff game. The action sets of the two players are I and J respectively. The payoff matrix M, with entries in $\mathbf{R}^{|I|}$, is given by $m_{i,j} = (u_{i,j} - u_{a,j})_{a\in I}$.

Let $F = \{(x_1, x_2, ..., x_{|I|}) \colon x_i \geq 0, i = 1, ..., |I|\}$ be the non-negative orthant. Then if a strategy σ of player 1 approaches F, it is Hannan no-regret, and vice-versa. Since by Hannan (1957) there is a Hannan no-regret strategy, F is approachable. However, showing directly that F is approachable and thereby proving the existence of a Hannan no-regret strategy is easier than Hannan's original proof.

2.4 The Main Results

We are now ready to state our main results. The first two propositions refer to approachability.

Proposition 1 *A set that contains a convex approachable set is approachable with automata.*

Proposition 2 *Let F be a set that contains a convex approachable set. Then for every $k \in \mathbf{N}$ there is a k-bounded-recall strategy that approaches $B(F, O(\frac{1}{\sqrt{k}}))$.*[2]*.*

The following Proposition completes the characterization of sets which are approachable with bounded-recall strategies, or with automata.

Proposition 3 *A closed set that does not contain any convex approachable set is not approachable with automata.*

Propositions 1 and 3 imply the following characterization of convex minimal approachable sets in terms of approachability with bounded computational capacity strategies.

Theorem 1 *A minimal closed approachable set is approachable with bounded-recall strategies, or with automata, if and only if it is convex.*

Remark 3, together with Propositions 2 and 3, implies the following.

Corollary 1 *A set is approachable with automata if and only if it is approachable with bounded-recall strategies.*

Translated to the setup of sequential decision problems, we can derive the following results.

Corollary 2 *In every sequential decision problem, for every $\delta > 0$ there is a Hannan δ-no-regret strategy that can be implemented by an automaton.*

Corollary 3 *In every sequential decision problem, there is a constant C such that for every $k \in \mathbf{N}$ there is a Hannan $\frac{C}{\sqrt{k}}$-no-regret k-bounded-recall strategy.*

[2]Formally, there is a constant $C > 0$, independent of F, such that for every $k \in \mathbf{N}$, there is a k-bounded-recall strategy σ that approaches $B(F, \frac{C}{\sqrt{k}})$

3 Approachability with Bounded Computational Capacity

3.1 Approachability with automata

Here we prove Proposition 1, which states that any set F that contains a convex approachable set is approachable with automata. We then bound the size of the smallest automaton that approaches $B(F, \delta)$, as a function of δ.

Proof of Proposition 1: By Remark 2, it is sufficient to prove that every convex approachable set is approachable with automata.

Let F be a convex approachable set. Then, there is a strategy σ of player 1 such that[3]

$$\forall \varepsilon, \exists n, \forall \tau, \ \mathbf{P}_{\sigma,\tau}(d(\bar{x}_n, F) \geq \varepsilon/2) \leq \varepsilon/2. \tag{1}$$

Fix $\varepsilon > 0$, and let n be the minimal integer that satisfies (1) for that ε.

Suppose that player 1 plays in blocks of size n. At the beginning of each block he forgets past play, and plays the strategy σ (for n stages). We now argue that the resulting strategy, σ_*, which can be implemented by an automaton with $\frac{|I \times J|^n - 1}{|I \times J| - 1}$ states, approaches $B(F, \varepsilon)$.

Let Y_k be the average payoff in block k. Let τ_k be the strategy of player 2 used in that block. τ_k is a random variable that depends on the play in previous blocks. The distribution of Y_k is similar to the distribution of $\bar{x}_n$ under (σ, τ_k), so that by (1) $\mathbf{P}_{\sigma_*,\tau}(d(Y_k, H) \geq \varepsilon/2) \leq \varepsilon/2$. Since payoffs are bounded by 1, $\mathbf{E}_{\sigma_*,\tau}[d(Y_k, F)] \leq \varepsilon$ for every $k \in \mathbf{N}$.

Denote by $\mathcal{H}_k$ the algebra over the space of infinite plays spanned by all the cylinders that are defined by histories up to block k. The random variables $(Y_k - \mathbf{E}_{\sigma_*,\tau}[Y_k \mid \mathcal{H}_k])_{k \in \mathbf{N}}$ are centered, uncorrelated, and uniformly bounded by 1. Denote $\overline{Y}_k = \frac{1}{k} \sum_{l=1}^{k} Y_l$ the average payoff in the first l blocks. By the strong law of large numbers, and since F is convex, for every $\delta > 0$

[3] Observe that this statement is much weaker than the one given in Definition 1.

there is $N \in \mathbf{N}$, independent of τ, such that

$$\mathbf{P}_{\sigma*,\tau}\left(\sup_{k\geq N} d(\overline{Y}_k, F) \geq \varepsilon + \delta\right) < \delta.$$

Choosing $N \geq n/\delta$ we get, since payoff are bounded by 1,

$$\mathbf{P}_{\sigma*,\tau}\left(\sup_{k\geq n\times N} d(\bar{x}_k, F) \geq \varepsilon + 2\delta\right) < \delta.$$

In particular, σ_* approaches $B(F, \varepsilon)$. ■

For every pair (p, q) of mixed actions (i.e., distributions over the respective action sets), denote by $m_{p,q} = \sum_{i,j} p_i m_{i,j} q_j$ the expected vector payoff. This is the expected stage-payoff when player 1 plays the mixed action p and player 2 plays the mixed action q.

For every vector $x \in \mathbf{R}^d$ and every set A, denote by $p_A(x)$ a mixed action of player 1 that satisfies, for some $y \in A$, (i) $d(x, y) = d(x, A)$, and (ii) for every mixed action q of player 2, $\langle x - y, m_{p_A(x),q} - y\rangle \leq 0$.[4] Graphically, if y is a closest point to x in A, the condition means that the hyperplane that passes through y, and is perpendicular to the line that connects x and y, separates x from the set $H(p_A(x)) = \{m_{p_A(x),q} \colon q \in \Delta(J)\}$. Thus, if the average payoff so far, $\bar{x}_n$, is not in A, and player 1 plays the mixed action $p_A(\bar{x}_n)$, the expected stage-payoff is in $H(p_A(\bar{x}_n))$, regardless of the mixed action of player 2. Since the hyperplane that passes through y and is perpendicular to the line that connects $\bar{x}_n$ and y, separates $\bar{x}_n$ from the set $H(p_A(\bar{x}_n))$, on average, the new average payoff will get closer to A.

Remark 4 *Blackwell (1956) showed that if F is a convex approachable set, then the strategy σ that plays at every stage n the mixed action $p_F(\bar{x}_n)$ approaches F at a rate $O(1/\sqrt{n})$, that is, there is a constant $C > 0$, independent of F, such that for every strategy τ of player 2, and every $n \in \mathbf{N}$, $\mathbf{E}_{\sigma,\tau}(d(\bar{x}_n, F)) \leq C/\sqrt{n}$.*

[4]For $x, y \in \mathbf{R}^d$, $\langle x, y\rangle = \sum_{i=1}^d x_i y_i$ is the standard inner product.

Using this specific strategy σ in the proof of Proposition 1 with $n > \frac{1}{\sqrt{\varepsilon}}$, yields an explicit automaton that approaches $B(F, \varepsilon)$.

Remark 5 *A naïve estimate for the number of states of the automaton needed to implement the strategy σ_* that was suggested in the proof of Proposition 1 and approaches $B(F, O(\frac{1}{\sqrt{n}}))$, given the strategy σ that approaches H and discussed in Remark 4, is c^n, where $c = |I \times J|$. As we argue now, the number of states is much smaller.*

Since the mixed action σ plays at each stage $k < n$ depends on the average payoff up to that stage, the number of states of an automaton needed to implement the prescription of σ_* at stage k of the block is bounded by the number of different empirical distributions of joint actions. By Feller (1968, Eq. (II.5.2)), the number of different empirical distributions of joint actions after k stages is $\binom{c-1+k}{c-1}$. Therefore, by Feller (1968, Eq. (II.12.8)), one can implement this strategy using an automaton with size $\sum_{k=0}^{n-1} \binom{c-1+k}{c-1} = \binom{n-1+c}{c}$, which is of the order of n^c. Consequently, for any F that contains the convex hull of an approachable set, player 1 can approach $B(F, O(n^{-\frac{1}{2c}}))$ with an automaton of size n.

Corollary 2 can therefore be strengthened as follows. In every sequential decision problem, for every $\delta > 0$ there is a Hannan δ-no-regret strategy that can be implemented by an automaton of size $O(\log \frac{1}{\delta})$.

3.2 Approachability with bounded-recall strategies

Here we prove Proposition 2, which states that any set F that contains a convex approachable set is approachable with bounded-recall strategies.

Proof of Proposition 2: By Remark 2, we can assume w.l.o.g. that F is a convex approachable set.

Let σ be the strategy that approaches F and was discussed in Remark 4. Then there is $C > 0$ such that for every $n \in \mathbf{N}$, $\mathbf{E}_{\sigma,\tau}[d(\bar{x}_n, F)] \leq \frac{C}{\sqrt{n}}$, for every strategy τ of player 2.

Fix $n \in \mathbf{N}$. Let σ_* be the strategy we constructed in the proof of Proposition 1 given σ, with blocks of size n. Our goal is to show that σ_*, when properly modified, is an n-bounded-recall strategy, and that it approaches $B(F, O(\frac{1}{\sqrt{n}}))$.

Denote by ℓ the smallest integer larger than $\sqrt{n}$. Let i_0 and i_1 be two distinct actions of player 1.

Given the strategy σ_*, that plays in blocks of size n, we now construct an augmented n-bounded-recall strategy $\widehat{\sigma}$, that, in a sense, also plays in blocks of size n.

Marking the beginning of the block: The beginning of each block is marked by a sequence of ℓ consecutive actions i_0 of player 1. Thus, if the past $n-1$ actions of player 1 do not contain a sequence of ℓ consecutive i_0's, player 1 plays the action i_0.

Marking the end of the block: The end of the block is marked by the action i_1 of player 1. Thus, if the past $n-1$ actions of player 1 end with a sequence of ℓ consecutive i_0's, player 1 plays the action i_1.[5]

To ensure that the only sequence of ℓ consecutive i_0's appears at the beginning of the block, whenever the past $n-1$ actions of player 1 contain a sequence of ℓ consecutive i_0's, and the last $\ell-1$ actions player 1 played are all i_0, player 1 plays the action i_1.[6]

Each stage in which player 1 plays the action i_1 instead of some other action, as well as each of the ℓ stages in which he plays i_0 to mark the beginning of the block, is called an *irregular stage*. Observe that there are at most 2ℓ irregular stages in each block.

Playing at all other stages: Denote by h the partial history from the beginning of the block to the present stage. Denote by h' the history h, after *removing* all pairs of actions that correspond to irregular stages. Under $\widehat{\sigma}$

[5] The role of this part is to ensure that the block does not end with action i_0, in which case we will count this action as part of the beginning of the next block.

[6] In particular, at stage $\ell+1$ of the block the action i_1 is played by player 1.

player 1 plays after the history h the same mixed action σ_* plays after h'.

The virtual memory: The virtual memory of the strategy may be any sequence in $(I \times J)^n$ which contains only at its end a sequence of ℓ consecutive stages in which player 1 played i_0. This ensures that the first block starts at stage 1, and that apart from this fact, the virtual memory does not affect the play.

Let Y_k be the expected payoff vector during block k. Since there are at most 2ℓ irregular stages in block k,

$$\mathbf{E}_{\sigma_*,\tau}[d(Y_k, H)] \leq \frac{C}{\sqrt{n}} + \frac{2\ell}{n} \leq \frac{C+3}{\sqrt{n}}.$$

The argument provided in the proof of Proposition 1 implies that σ_* approaches $B(F, \frac{C+3}{\sqrt{n}})$. ■

Remark 6 *It would be more elegant to prove Proposition 2 by a strategy that prescribes to play at each stage the mixed action $p_H(\hat{x}_k)$, where $\hat{x}_k = \frac{1}{n}\sum_{c=1}^{k} x_{k-c}$ is the average payoff in the last n stages. However, we have not been able to prove that this strategy approaches $B(H, O(\frac{1}{\sqrt{n}}))$.*

3.3 Completing the Characterization of Approachable Sets

Here we prove Proposition 3, which states that a closed set F that does not contain any convex approachable set is not approachable with automata.

For every $x \in \mathbf{R}^d$ and every $\delta > 0$, $B_0(x, \delta) = \{y \in \mathbf{R}^d \colon d(x, y) < \delta\}$ is the open ball with radius δ around x.

When A is an automaton, and p is a probability distribution over the states of the automaton, we denote by (A, p) the automaton that is similar to A, except that its initial probability distribution is p (rather than the one indicated by A).

Proof of Proposition 3: Suppose to the contrary that F is approachable with automata. Let $G \subseteq F$ be minimal (w.r.t. set inclusion) among all closed subsets of F that are approachable with automata. As the intersection of any decreasing sequence of closed sets that are approachable with automata is approachable with automata, Zorn's Lemma guarantees the existence of such a set.

Step 1: G is not convex. G is approachable with automata and therefore it is an approachable set. Since F does not contain a convex approachable set, G is not convex. Therefore, there are $x, y \in G$ and $\lambda \in [0,1]$ such that $z := \lambda x + (1-\lambda)y \notin G$. Since G is closed, one can choose $\delta \in (0, 1/4)$ such that $d(x,y) \geq \delta$ and $d(z,G) > 3\delta$.

The set $G \setminus B_0(x,\delta)$ is non-empty (as it contains y), closed, and a strict subset of G. Since G is minimal among all closed sets which are approachable with automata, $G \setminus B_0(x,\delta)$ is not approachable with automata. Similarly, the set $G \setminus B_0(y,\delta)$ is not approachable with automata.

This implies that there is $\delta_0 < \delta/4$ such that there is no automaton that approaches the sets $B(G \setminus B_0(x,\delta), \delta_0)$ and $B(G \setminus B_0(y,\delta), \delta_0)$.

As G is approachable with automata, there is an automaton A that approaches $B(G, \frac{\delta_0}{2})$. We will define a strategy τ of player 2 that, when plays against A, guarantees that the average payoff visits $B_0(z, 2\delta)$ infinitely often. This implies that G is not approachable by A, a contradiction.[7]

For every $n \in \mathbf{N}$ define the random variables p_n as the posterior probability distribution over the states of the automaton at stage n, given past play. Denote by P_A the union of the range of p_n, over all $n \in \mathbf{N}$. P_A contains all possible beliefs player 2 may have along the game about the current state of the automaton.

Since A approaches $B(G, \frac{\delta_0}{2})$, the definition of approachability implies that so does (A, p), for every $p \in P_A$.

Step 2: Constructing a family of strategies $(\tau^x_p)_{p \in P_A}$. Our first goal is to define, for every $p \in P_A$, a strategy τ^x_p of player 2 that ensures that the

[7]Actually, we will show that the average payoff remains in $B_0(z, 2\delta)$ from some point on.

average payoff gets close to x when playing against (A,p).

Formally, we will define for every $p \in P_A$ a strategy τ_p^x of player 2 satisfying

$$\mathbf{P}_{(A,p),\tau_p^x}(\limsup_{n\to\infty} d(\bar{x}_n, x) \leq \frac{3}{2}\delta) = 1. \tag{2}$$

That is, the long-run average payoff under $((A,p),\tau_p^x)$ gets arbitrarily close to $B(x, \frac{3}{2}\delta)$.

Fix $p \in P_A$. Since the automaton (A,p) approaches $B(G, \frac{\delta_0}{2})$, for every $\eta > 0$ there is a positive integer $N_{p,\eta}$ such that

$$\forall \tau, \quad \mathbf{P}_{(A,p),\tau}(\sup_{n \geq N_{p,\eta}} d(\bar{x}_n, G) \geq \delta_0) \leq \eta. \tag{3}$$

Since the automaton (A,p) does not approach $B(G \setminus B_0(x,\delta), \delta_0)$, there is $\eta_p > 0$ such that for every $N \in \mathbf{N}$ there is a strategy $\tau_{N,p}$ of player 2 satisfying

$$\mathbf{P}_{(A,p),\tau_{N,p}}(\sup_{n\geq N} d(\bar{x}_N, B(G \setminus B_0(x,\delta))) \geq \delta_0) \geq \eta_p. \tag{4}$$

By substituting $\eta = \eta_p/2$ in (3), and $N \geq N_{p,\eta_p/2}$ in (4), we obtain

$$\mathbf{P}_{(A,p),\tau_{N,p}}(\sup_{n\geq N} d(\bar{x}_n, x) \leq \delta + \delta_0 \leq \frac{5}{4}\delta) \geq \frac{\eta_p}{2}. \tag{5}$$

In particular, there is $K_{N,p}$ such that

$$\mathbf{P}_{(A,p),\tau_{N,p}}(d(\bar{x}_n, x) \leq \frac{5}{4}\delta \text{ for some } N \leq n \leq K_{N,p}) \geq \frac{\eta_p}{4}.$$

For every fixed strategy τ of player 2, every $n \in \mathbf{N}$, and every $c > 0$, the function $p' \mapsto \mathbf{P}_{(A,p'),\tau}(d(\bar{x}_\nu, x) \geq c)$ is linear (and Lipschitz-1) in p'. As the space of probability distributions over the states of the automaton is compact, one can assume w.l.o.g. that $\eta_* := \inf_{p\in P_A} \eta_p > 0$, $N_* := \sup_{p\in P_A} N_{p,\eta_p/2} < \infty$, and $K_N := \sup_{p\in P_A} N_{N,p} < \infty$.

Since (5) holds for every $p \in P_A$, we conclude that for every $p \in P_A$ there is a strategy τ_p^x, satisfying (2). Indeed, τ_p^x plays in blocks of varying size. Let b_l be the stage in which block l starts, so that $pi_l := p_{b_l}$ is the posterior probability over the states of the automaton at the beginning of block l, given past play. At the beginning of each block, τ_p^x forgets past play, and during block l it follows $\tau_{b_l/\delta^2,\pi_l}$. The length of block l is the minimum between K_{b_l/δ^2} and the minimal $n \geq b_l/\delta^2$ such that the average payoff in the first n stages along the block is in $B(x, \frac{3}{2}\delta)$.

Since $\frac{5}{4}\delta + \delta^2 \leq \frac{3}{2}\delta$, in every block there is a probability greater than $\eta_*/4$ such that the average payoff at the end of the block is in $B(x, \frac{3}{2}\delta)$. Since block l lasts at least b_l/δ^2 stages, the average payoff at the end of block l is close to the average payoff during block l. The claim follows.

Step 3: Constructing the family $(\tau_p^y)_{p\in P_A}$**.** Replacing x by y in **Step 2** we conclude that for every $p \in P_A$ there is a strategy τ_p^y of player 2 satisfying

$$\mathbf{P}_{(A,p),\tau_p^y}(\limsup_{n\to\infty} d(\bar{x}_n, y) \leq \frac{3}{2}\delta))) = 1. \tag{6}$$

Step 4: Constructing the strategy τ**.** For every $p \in P_A$ choose $N_p \in \mathbf{N}$ such that $\mathbf{P}_{(A,p),\tau_p^x}(d(\bar{x}_n, x) \leq \frac{7}{4}\delta)$ for some $n \leq N_p) \geq 1 - \frac{\delta}{4}$ and $\mathbf{P}_{(A,p),\tau_p^y}(\ d(\bar{x}_n, y) \leq \frac{7}{4}\delta)$ for some $n \leq N_p) \geq 1 - \frac{\delta}{4}$. As before, since the space of probabilities over the states of the automaton is compact, we can assume w.l.o.g. that $N_* = \sup_{p\in P_A} N_p < +\infty$.

Define a strategy τ that plays in blocks of random size as follows. Denote by π_l the posterior probability over the states of the automaton, given past play, at the beginning of block l. At block l player 2 forgets past play, and either follows τ_p^x (in which case we call the block an X-block), or τ_p^y (in which case we call the block a Y-block). The block terminates when either (i) the average-payoff along the block is within $\frac{7}{4}\delta$ of x (in an X-block) or of y (in a Y-block), or (ii) as soon as the block lasts for N_* stages, whichever comes first. The decision whether the new block is an X-block or a Y-block is done according to the proportion of past stages that were spent in X-blocks. If the proportion is smaller than λ, the present block will be an X-block, whereas if it at least λ, the present block will be a Y-block.

The probability that the average payoff in an X-block (resp. Y-block) is within $\frac{7}{4}\delta$ of x (resp. y) is at least $1-\frac{\delta}{4}$. Since payoffs are bounded by 1, and by the strong law of large numbers, this construction ensures that if player 2 follows τ the long-run average payoff remains in $B_0(z, 2\delta)$ from some stage on.

References

[1] Aumann, R. J. (1981) Survey of Repeated Games. In Aumann et al. (eds.) *Essays in game theory and mathematical economics in honor of Oskar Morgenstern,* Bibliographishes Institut, 11-42.

[2] Ben Porath, E. (1990) The Complexity of Computing a Best Response Automaton in Repeated Games with Mixed Strategies. *Games Econ. Behavior*, **2**, 1-12.

[3] Blackwell, D. (1956) An Analog of the Minimax Theorem for Vector Payoffs. *Pacific J. Math.* **6**, 1-8.

[4] Feller, W. (1968) *An Introduction to Probability Theory and its Applications*, John Wiley and Sons.

[5] Foster, D. and R. Vohra (1999) Regret in the On-Line Decision Problem. *Games Econ. Behavior*, **29**, 7-35.

[6] Fudenberg, D. and D. Levine (1999) Conditional Universal Consistency. *Games Econ. Behavior*, **29**, 104-130.

[7] Hannan, J. (1957) Approximation to Bayes Risk in Repeated Plays. in *Contribution to the Theory of Games*, **3**, 97-139. Princeton University Press, Princeton, NJ.

[8] Hart, S. and A. Mas-Colell (2000) A Simple Adaptive Procedure Leading to Correlated Equilibrium. *Econometrica*, **68**, 1127-1150.

[9] Jéhiel, P. (1995) Limited Horizon Forecasting in Repeated Alternating Games. *J. Econ. Th.*, **67**, 497-519.

[10] Lehrer, E. (1988) Repeated Games with Stationary Bounded Recall Strategies. *J. Econ. Th.*, **46**, 130-144.

[11] Lehrer, E. (2003) Wide Range No-Regret Theorem. *Games Econ. Behavior*, **42**, 101-115.

[12] Neyman, A. (1985) Bounded Complexity Justifies Cooperation in Finitely Repeated Prisoner's Dilemma. *Econ. Let.*, **19**, 227-229.

[13] Rubinstein, A. (1986) Finite Automata Play the Repeated Prisoner's Dilemma. *J. Econ. Th.*, **39**, 83-96.

[14] Rustichini, A. (1999) Minimizing Regret: The General Case. *Games Econ. Behavior*, **29** 224-243.

[15] Spinat, X. (2002) A Necessary and Sufficient Condition for Approachability. *Math. Oper. Res.* **27**, 31-44.

[16] Watson, J. (1994) Cooperation in the Infinitely Repeated Prisoner's Dillema with Perturbations. *Game Econ. Behavior*, **7**, 260-285.