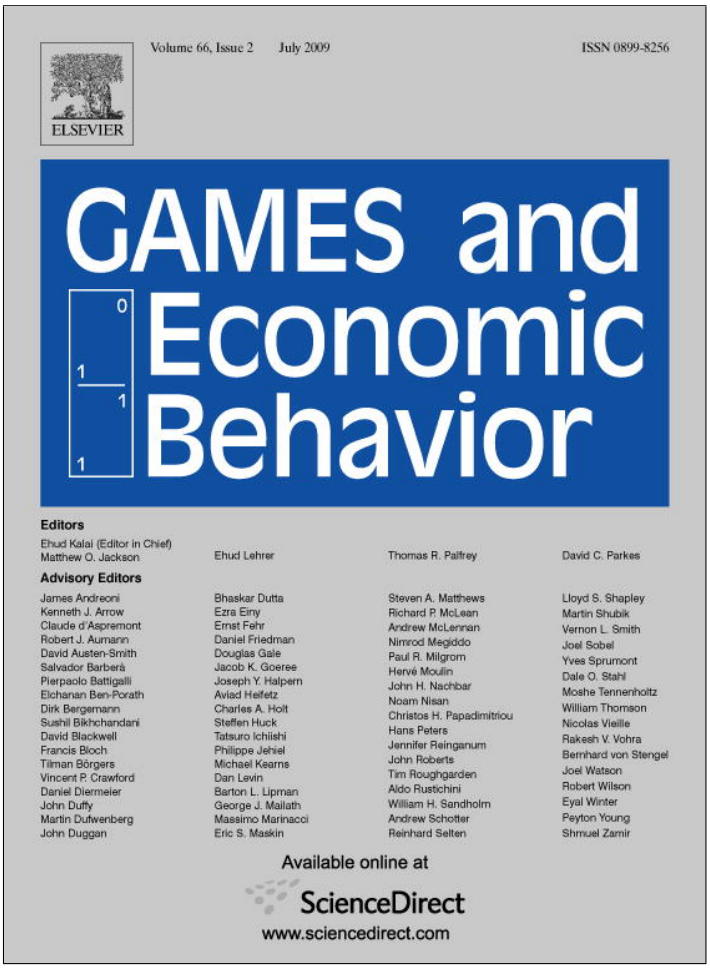


Games and Economic Behavior 66 (2009) 995–1004

Contents lists available at ScienceDirect

Games and Economic Behavior

www.elsevier.com/locate/geb

Approachability with bounded memory ☆

Ehud Lehrer, Eilon Solan *

School of Mathematical Sciences, Tel Aviv University, Tel Aviv 69978, Israel

ARTICLE INFO

Article history:
Received 8 March 2006
Available online 17 October 2008

JEL classification:
C61
C72
C73

Keywords:
Approachability
Repeated games
Vector payoffs
Bounded memory
Bounded recall
Automata
No-regret
Adaptive learning

ABSTRACT

We study Blackwell's approachability in repeated games with vector payoffs when the approaching player is restricted to use strategies with bounded memory: either strategies with bounded recall, or strategies that can be implemented by finite automata. Our main finding is that the following three statements are equivalent for closed sets. (i) The set is approachable with bounded recall strategies. (ii) The set is approachable with strategies that can be implemented with finite automata. (iii) The set contains a convex approachable set. Using our results we show that (i) there are almost-regret-free strategies with bounded memory, (ii) there is a strategy with bounded memory to choose the best among several experts, and (iii) Hart and Mas-Colell's adaptive learning procedure can be achieved using strategies with bounded memory.

1. Introduction

In a seminal paper, Blackwell (1956a) studied repeated games with vector payoffs. In such a game, a target-set is *approachable* by player 1 if he has a strategy that ensures that the long-run average payoff is in that set, regardless of the strategy that player 2 employs. A complete characterization of the family of approachable sets was given independently by Hou (1971) and Spinat (2002). Vieille (1992) studied the notions of weak-approachability and weak-excludability, which were also introduced by Blackwell (1956a), and proved that every set is either weak-approachable or weak-excludable. For partial results on weak-approachability for 2-dimensional games where both players have two actions see Hou (1969).

The strategy that Blackwell (1956a) defines to approach an approachable set depends on the current average payoff vector. To properly calculate this average the player needs to keep track of the calendar time, and to either know how many times each entry in the payoff matrix was chosen, or to be able to calculate the exact average payoff vector. When the player has bounded memory, such a strategy cannot be implemented.

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

☆ The work of the first author was supported by the Israel Science Foundation (Grant #762/045). The work of the second author was supported by the Israel Science Foundation (Grant #69/01). We thank two anonymous referees for insightful comments that improved the presentation.

* Corresponding author.
E-mail addresses: lehrer@post.tau.ac.il (E. Lehrer), eilons@post.tau.ac.il (E. Solan).

0899-8256/$ – see front matter
doi:10.1016/j.geb.2007.07.011

We fully characterize the target sets that are approachable by these two types of strategies. This characterization is expressed in geometric terms. We show that a target set is approachable by strategies with bounded memory if and only if it contains a convex set that is approachable by strategies with unbounded memory. Since in practice target set are convex, this result implies that imposing a bounded-memory restriction on strategies does not hurt efficiency. A partial characterization of target sets that are excludable against the two types of strategies with bounded memory is discussed in Lehrer and Solan (2006).

A consequence of this result is that target sets which are approachable by bounded-recall strategies are precisely those approachable by strategies that can be implemented by finite automata. Since any bounded-recall strategy can be implemented by an automaton, any target-set that is approachable by the former is also approachable by the latter. However, the family of strategies that can be implemented by an automaton is far richer than the family of bounded recall strategies. Indeed, as opposed to bounded-recall strategies, finite automata can be programmed to recall events that occurred long ago. Nevertheless, as our result shows, when considering approachable target sets, the two notions of strategies with bounded memory are equivalent.

Several applications of our findings to other models are provided. The first application concerns regret-free strategies. In the setup of sequential decision problems, a strategy is *Hannan no-regret* if it ensures a long-run average payoff that is at least as high as what the decision maker (hereafter, DM) could have achieved had he played constantly the same action. In other words, by playing a Hannan no-regret strategy, the DM feels no-regret for not having constantly played a best response vis-a-vis the empirical distribution of the state of nature.

Hannan (1957) showed in a rather complicated proof that if there are finitely many states of nature, there is always a Hannan no-regret strategy. This theorem can be derived from Blackwell's approachability theorem. Both proofs employ strategies that are not computationally bounded.

Our results imply that for every small δ there is a strategy with bounded memory guaranteeing that the DM's regret is at most δ. That is, the DM has a policy that beats, up to δ, any stationary policy which plays the same action throughout.

A second implication of our result is related to the problem of selecting an expert. Consider a sequential decision problem with several experts, each recommending one strategy. At each stage the DM may choose to adopt the advice of a different expert. If the DM knew at the outset what will be the sequence of the realized states, he could have chosen the expert who recommends the best strategy for this sequence. Our result guarantees the existence of a selection rule among experts with bounded memory, which asymptotically does as good as choosing the expert that is best suited for the (unknown) sequence of realized states. Thus, by applying this selection rule, the DM feels no-regret for not choosing the best expert at the outset, and following his or her recommendation throughout. For an elaboration on this issue the reader is referred to de Farias and Megiddo (2003).

A third implication is to learning theory. Hart and Mas-Colell (2000) (see also Zapechelnyuk, forthcoming) provide a decentralized learning procedure that converges to correlated equilibrium. This process, however, requires unbounded memory. As our results imply, players who use strategies with bounded memory can learn to play correlated equilibrium in a decentralized manner.

The paper is arranged as follows. In Section 2 we present a motivating example. We then present our model and main results in Section 3. In Section 4 we provide three implications for our results – no-regret, choosing among experts, and adaptive learning. Finally, we prove our results in Section 5.

2. An example

In this section we illustrate the model and the result by an example.

Consider the following 2×2 game with two-dimensional payoffs.

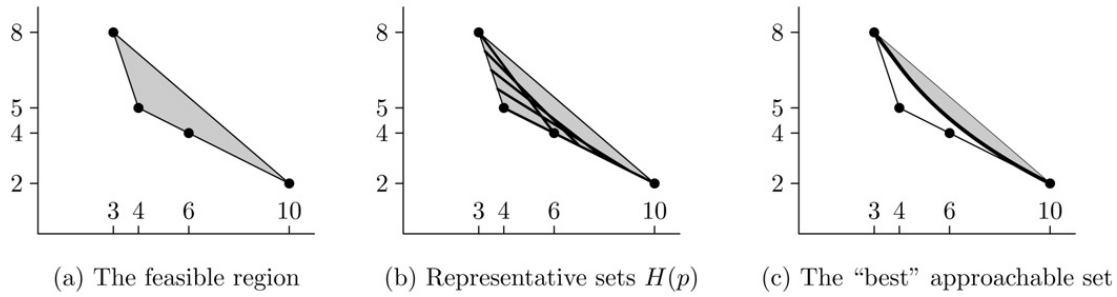
	L	R
T	$(10, 2)$	$(4, 5)$
B	$(6, 4)$	$(3, 8)$

The feasible region is depicted in Fig. 1(a). By employing the stationary strategy $[p(T), (1-p)(B)]$ player 1 guarantees that the stage payoff, and therefore also the long-run average payoff, will be on the line segment

$$H(p) = \big[(6+4p, 4-2p), (3+p, 8-3p)\big].$$

Some of the sets $H(p)$ appear in Fig. 1(b).

Can player 1 ensure that the long-run average outcome lies on the Pareto frontier of the feasible set? The approachability theorem of Blackwell (1956a) implies that the answer is negative. Actually, Blackwell's (1956a) result, together with the results of Hou (1971) or Spinat (2002), show that the "best" curve that player 1 can guarantee is the upper envelop of the intervals $H(p)$, that appears in Fig. 1(c). That is, player 1 can guarantee that the long-run average payoff lies in the gray set in Fig. 1(c). Since this set is convex, our result implies that it is also approachable by a strategy with bounded memory.



(a) The feasible region (b) Representative sets $H(p)$ (c) The "best" approachable set

Fig. 1.

3. The model and main results

In this section we present the model and the main results. We start by introducing repeated games with vector payoffs. We then define strategies with bounded memory, and finally we state the main results.

3.1. Repeated games with vector payoffs

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

At every stage $n \geqslant 1$ the two players choose, independently and simultaneously, a pair of actions $(i_n, j_n) \in I \times J$, each one in his action set. A *strategy* of player 1 (resp. player 2) is a function $\sigma : \bigcup_{n=0}^{\infty}(I \times J)^n \to \Delta(I)$ (resp. $\tau : \bigcup_{n=0}^{\infty}(I \times J)^n \to \Delta(J)$),[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. The average payoff vector up to stage n is

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

Note that for every $n \in \mathbb{N}$, $\bar{x}_n$ is a random variable with values in $\mathbb{R}^d$, 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 $\mathbb{R}^d$. For every set F in $\mathbb{R}^d$ and every $x \in \mathbb{R}^d$, let $d(x, F) = \inf_{y\in F} d(x, y)$ be the distance of x from F. For every $\delta > 0$, let $B(F, \delta) = \{x \in \mathbb{R}^d\colon\ d(x, F) \leqslant \delta\}$ 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

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

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 (1956a) provided a sufficient condition for a set to be approachable. Hou (1971) and Spinat (2002) fully characterized the family of approachable sets.

3.2. Strategies with bounded memory

In this section we define two types of strategies with bounded memory: 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 is a pair (m, σ) (resp. (m, τ)), where $m \in (I \times J)^k$, and $\sigma : (I \times J)^k \to \Delta(I)$. 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 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

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

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 according to 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 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 with an automaton and with 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. 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.

3.3. The main results

Our main result is the following.

Theorem 1. *Let $F \subseteq \mathbb{R}^d$ be a target set. The following three statements are equivalent.*

1. *F contains a convex approachable set.*
2. *F is approachable with bounded-recall strategies.*
3. *F is approachable with automata.*

Theorem 1 is a limit result. Our next two results provide the size of the memory of the bounded-recall strategy (or the size of the automaton) that is needed to approach a set up to a certain distance.

Theorem 2. *Let $F \subseteq \mathbb{R}^d$ be a set that contains a convex approachable set. For every $k \in \mathbb{N}$ there is a k-bounded-recall strategy that approaches $B(F, O(\frac{1}{\sqrt{k}}))$.*[2]

Theorem 3. *Let $F \subseteq \mathbb{R}^d$ be a set that contains a convex approachable set. For every $n \in \mathbb{N}$ there is a strategy which is implementable by an automaton with memory of size n that approaches $B(F, O(n^{-1/(2|I|\times|J|)}))$.*

Remark 3. By Remark 1 a k-bounded-recall strategy can be implemented by an automaton with $(|I| \times |J|)^k$ states. Therefore, by Theorem 2, for every k there is an automaton with $n = (|I| \times |J|)^k$ states that approaches $B(F, \frac{1}{\sqrt{k}})$. Since

$$n^{-\frac{1}{2(|I|\times|J|)}} = \big(|I| \times |J|\big)^{-\frac{k}{2(|I|\times|J|)}} < \frac{1}{\sqrt{k}} \quad \Leftrightarrow \quad k > 2\big(|I| \times |J|\big),$$

it follows that the bound given in Theorem 3 is better than the bound given in Theorem 2 when applied to automata, when k is sufficiently large.

Remark 4. Our results hold also in the setup of partial monitoring, as long as payoffs are observed by player 1. This is the case since what we actually prove is that if player 1 has a strategy that approaches a set F, then he has a strategy with bounded memory that approaches the convex hull of F.[3]

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

[3] We thank an anonymous referee for mentioning this issue.

4. Implications: No-regret, choosing an expert and adaptive learning

This section is devoted to a short description of the three applications that were mentioned in the introduction. These applications are known in the literature and are provided here in order to explicitly illustrate the implication of restricting the computational capacity of the strategies.

4.1. Regret-free strategies

The relation between various notions of no-regret and repeated games with vector payoffs is now well established (see, e.g., Cesa-Bianchi and Lugosi, 2006). Foster and Vohra (1999), using games with vector payoffs, provided a process of decentralized actions that converges to correlated equilibrium. Foster and Vohra (1999), Fudenberg and Levine (1999) and Hart and Mas-Colell (1996, 2000) introduced stronger no-regret notions than Hanann's notion, and showed that there always exists a strategy that satisfies the stronger version. Rustichini (1999), using Blackwell'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 that is immune against infinitely many replacing schemes.

In a closely related paper, Schlag (2003) examines behavioral rules that attain minimax regret in discounted repeated decision making problems. He finds minimax regret rules that can be attained by randomization using a linear function of the previous payoffs. For myopic individuals, minimax regret behavior requires only a recall of the last round. For intermediate discount factors two rounds of memory suffice to attain minimax regret.

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; the set J of possible states of nature is finite.

We denote by i_n the action chosen by DM at stage n, and by j_n the state of nature at that stage. For a stage n let $\bar{x}_n = \frac{1}{n}\sum_{t=1}^{n} u(i_t, j_t)$, and for an action $i \in I$ let $r_n^i = \frac{1}{n}\sum_{t=1}^{n} u(i, j_t)$. A *history* of length n is sequence of n states and actions, and a *strategy* is a function from the set of histories to the set of mixed action.

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

$$\mathbf{P}_{\sigma,\mathbf{j}}\Big(\liminf_{n\to\infty}(\bar{x}_n - r_n^i) \geqslant -\delta\Big) = 1.$$

A Hannan 0-no-regret strategy is also termed *Hannan no-regret* strategy. The definition of Hannan no-regret implicitly assumes that the choices of DM do not affect the evolution of nature. Indeed, otherwise the quantity $\mathbf{P}_{\sigma,\mathbf{j}}(\liminf_{n\to\infty} r_n^i)$, which represents the hypothetical payoff that the stationary strategy i yields when actually the strategy σ is followed, is meaningless.

It is well known that a strategy is Hannan no-regret if and only if it approaches the non-negative orthant $F := \{(x_1, x_2, \ldots, x_{|I|}):\ x_i \geqslant 0,\ i = 1, \ldots, |I|\}$ in the following two-player game with vector payoffs. The action sets of the two players are I and J, respectively. The payoff function $v : I \times J \to \mathbb{R}^{|I|}$ is defined by $v(i, j) = (u(i, j) - u(a, j))_{a\in I}$. Since by Hannan (1957) there is a Hannan no-regret strategy, F is approachable. Theorems 2 and 3 translate into the following conclusions.

Corollary 1. *In every sequential decision problem, for every $k \in \mathbb{N}$ there is a Hannan $O(\frac{1}{\sqrt{k}})$-no-regret k-bounded-recall strategy.*

Corollary 2. *In every sequential decision problem, for every $n \in \mathbb{N}$ there is a Hannan $O(n^{-1/2|I|\times|J|})$-no-regret strategy that can be implemented by an automaton with size n.*

A historical note. Luce and Raiffa (1957, p. 482) cite Blackwell's (1956b) proof of Hannan no-regret theorem that uses the approachability theorem. Hart and Mas-Colell (1996) were the first to note that no-regret theorems can be proven, using Blackwell's approachability theorem, by bringing the regret vector to the non-negative orthant.

In some important applications, among which is choosing among a set of experts, the stage payoff functions may not be stationary. Consider a sequential decision problem as described above. There are K experts, each recommends to DM a strategy. One expert may perform well over one sequence of states while performing badly over another. Thus, the performance quality, in terms of the long-run average payoff, varies with the sequence of realized states.

The DM needs to select at every stage an expert and to follow his advice. Formally, the action set of DM is the set of experts, $\{1, \ldots, K\}$, and the stage payoff is $u(i(k), j)$ when expert k is chosen, $i(k)$ is his recommendation (possibly mixed) and j is the realized state. A selection rule of DM is a function from the set of histories to the set of probability distributions over $\{1, \ldots, K\}$. A selection rule is *δ-no-regret* if for every sequence of states, it produces an average payoff which is asymptotically greater (up-to δ) than the average payoff that any expert can achieve.

The sufficient conditions that ensure approachability of strategies when the stage payoff functions are stationary apply also to the case of stage payoff functions that vary during the course of the game. Therefore, Corollaries 1 and 2 apply verbatim to selection rules.

4.2. Adaptive learning

A multi-player game $G = (N, (A_i)_{i\in N}, (u_i)_{i\in N})$ consists of a finite set of players N, and for every player i a finite action set A_i and a payoff function $u_i : \times_i A_i \to \mathbb{R}$. Let $\varepsilon \geqslant 0$. A probability distribution Q over $A = \times A_j$ is a *correlated ε-equilibrium* if for every player i and every $a, a' \in A_i$, $\sum_{b\in A_{-i}} Q(a,b)u_i(a,b) \geqslant \sum_{b\in A_{-i}} Q(a,b)u_i(a',b) - \varepsilon$, where $A_{-i} = \times_{j\neq i} A_j$.

Fix a player i, and define an auxiliary two-player repeated game Γ_i with vector payoffs. Player 1's set of actions is $I = A_i$, and that of player 2 is $J = A_{-i}$. For every $(a,b) \in A_i \times A_{-i}$, $v(a,b)$ is the vector payoff in $\mathbb{R}^{A_i\times A_i}$ whose (a'',a')-coordinate $((a'',a') \in A_i \times A_i)$ equals $u_i(a,b) - u_i(a',b)$ if $a = a''$, and 0 otherwise.

Hart and Mas-Colell (1996, 2000) show that the positive orthant, $\mathbb{R}_+^{A_i\times A_i}$, is approachable by player 1 in the auxiliary game. Furthermore, if *each* player i employs a strategy in Γ_i that approaches the positive orthant, the empirical frequency of the action-profiles (which is a probability distribution over A) converges to the set of correlated equilibria. This procedure is called *adaptive learning*. Theorems 2 and 3 yield the following.

Corollary 3. *For every multi-player game G and every integer k there is an adaptive learning implementable by k-bounded-recall strategies that converges to the set of $O(\frac{1}{\sqrt{k}})$-correlated equilibria.*

Corollary 4. *For every multi-player game G and every integer n there is an adaptive learning implementable by automata of size n that converges to the set of $O(\frac{1}{n^{-1/2|I|\times|J|}})$-correlated equilibria.*

5. Proofs

To prove Theorem 1 we prove the following three lemmas.

Lemma 1. *Every set that contains a convex approachable set is approachable with automata.*

Lemma 2. *Every set that contains a convex approachable set is approachable with bounded-recall strategies.*

Lemma 3. *If a closed set does not contain a convex approachable set then it is not approachable with automata.*

Since any bounded-recall strategy can be implemented by an automaton, Lemma 3 implies that if a set does not contain a convex approachable set then it is not approachable with bounded-recall strategies as well, and the proof of Theorem 1 is complete.

5.1. Proof of Lemma 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[4]

$$\forall\varepsilon,\ \exists n,\ \forall\tau,\quad \mathbf{P}_{\sigma,\tau}\left(d(\bar{x}_n, F) \geqslant \frac{\varepsilon}{2}\right) \leqslant \frac{\varepsilon}{2d}. \tag{1}$$

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 follows the strategy σ (for n stages). Denote $c = |I \times J|$. We now argue that the resulting strategy, σ_*, which can be implemented by an automaton with $1 + c + c^2 + \cdots + c^{n-1} = \frac{c^n-1}{c-1}$ 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. Since player 2 may condition his actions on the play in previous blocks, τ_k is a random variable that depends on past play. The distribution of Y_k is similar to the distribution of $\bar{x}_n$ under (σ,τ_k), so that by (1) $\mathbf{P}_{\sigma_*,\tau}(d(Y_k,F) \geqslant \frac{\varepsilon}{2}) \leqslant \frac{\varepsilon}{2d}$. Since payoffs are bounded by 1, for every $k \in \mathbb{N}$ one has

$$\mathbf{E}_{\sigma_*,\tau}\big[d(Y_k,F)\big] \leqslant \mathbf{P}_{\sigma_*,\tau}\left(d(Y_k,F) \geqslant \frac{\varepsilon}{2}\right) \times d + \mathbf{P}_{\sigma_*,\tau}\left(d(Y_k,F) < \frac{\varepsilon}{2}\right) \times \frac{\varepsilon}{2} \leqslant \varepsilon.$$

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 $(Y_k - \mathbf{E}_{\sigma_*,\tau}[Y_k \mid \mathcal{H}_k])_{k\in\mathbb{N}}$ are centered, uncorrelated, and uniformly bounded by 1. Denote

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

$\overline{Y}_k = \frac{1}{k}\sum_{l=1}^{k} Y_l$ the average payoff in the first k blocks. Since F is convex, the function $y \mapsto d(y, F)$ is convex, and by the Azuma–Hoeffding inequality (see Alon and Spencer, 1992), for every $\delta > 0$ there is $N \in \mathbb{N}$, independent of τ, such that

$$\mathbf{P}_{\sigma_*,\tau}\Big(\sup_{k\geqslant N} d(\overline{Y}_k, F) \geqslant \varepsilon + \delta\Big) < \delta.$$

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

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

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

In the strategy σ_* that we constructed, player 1 plays in blocks, and forgets past play at the beginning of each block. One may wonder whether there is a strategy that approaches $B(F, \delta)$ and depends at each stage only on the average payoff at the last k stages, for some k. Zapechelnyuk (forthcoming) shows that there is no such strategy.

5.2. Proof of Theorem 3

As in the proof of Lemma 1 it is sufficient to prove the result for convex approachable sets F. In the proof of Lemma 1 we used an arbitrary strategy σ that approaches F. Here we are going to use a specific σ, and we will bound the size of the memory that is needed to follow that σ in the first n stages.

Blackwell (1956a) devised a strategy σ_0 that (i) 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 $n \in \mathbb{N}$, $\mathbf{E}_{\sigma_0,\tau}(d(\bar{x}_n, F)) \leqslant C/\sqrt{n}$, and that (ii) depends at each stage only on the average payoff vector in the previous stages.

Fix $n \in \mathbb{N}$, and consider the construction of σ_* that appears in the proof of Lemma 1, using the strategy σ_0. Denote by Y_k the average payoff in block k. By the choice of σ_0, $\mathbf{E}_{\sigma_0,\tau}(d(Y_k, F)) \leqslant C/\sqrt{n}$, so that σ_* approaches $B(F, C/\sqrt{n})$.

Since the mixed action σ plays at each stage $k < n$ depends on the average payoff up to that stage, all histories which lead to the same empirical distributions of joint actions lead also to the same average payoff. Therefore, 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 a convex approachable set, player 1 can approach $B(F, O(n^{-1/(2c)}))$ with an automaton of size n.

5.3. Proofs of Lemma 2 and Theorem 2

By Remark 2, we can assume w.l.o.g. that F is a convex approachable set. Let σ_0 be the strategy that approaches F and was discussed in the proof of Theorem 3. Then there is $C > 0$ such that for every $n \in \mathbb{N}$, $\mathbf{E}_{\sigma_0,\tau}[d(\bar{x}_n, F)] \leqslant \frac{C}{\sqrt{n}}$, for every strategy τ of player 2.

We are going to define an n-bounded-recall strategy $\widehat{\sigma}$, which is close in spirit to the strategy σ_* that we defined in the proof of Theorem 3, and in which player 1 properly marks the beginning of each block, so that he can implement σ_0 in each block. Fix $n \geqslant C^2$. Denote by ℓ the smallest integer larger than $\sqrt{n}$. Let i_0 and i_1 be two distinct actions of player 1.

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 $\ell + \frac{n}{\ell} \leqslant 2\ell$ irregular stages in each block.

Playing at all other stages: Denote by h the partial history from the beginning of the current block to the present stage. Denote by h' the history h, after *removing* all pairs of actions that correspond to irregular stages. Under $\widehat{\sigma}$ player 1 plays after the history h the same mixed action σ_0 plays after h'.

[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.

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, and since $n \geqslant C^2$,

$$\mathbf{E}_{\sigma_*,\tau}\big[d(Y_k, H)\big] \leqslant \frac{C}{\sqrt{n-2\ell}} + 2\frac{2\ell}{n} \leqslant \frac{C+6}{\sqrt{n}}.$$

The argument provided in the proof of Lemma 1 implies that $\widehat{\sigma}$ approaches $B(F, \frac{C+6}{\sqrt{n}})$.

5.4. Proof of Lemma 3

For every $x \in \mathbb{R}^d$ and every $\delta > 0$, $B_0(x,\delta) = \{y \in \mathbb{R}^d\colon\ d(x,y) < \delta\}$ 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).

Suppose to the contrary that F is approachable with automata. We first argue that there is a subset $G \subseteq F$ that is minimal (w.r.t. set inclusion) among all closed subsets of F that are approachable with automata. Observe that the intersection of any decreasing sequence $(F_n)_{n=1}^{\infty}$ of closed sets that are approachable with automata is approachable with automata. Indeed, setting $H = \bigcap_{n=1}^{\infty} F_n$, there exists n_0 such that $B(F_{n_0}, \delta) \subset B(H, 2\delta)$, so that an automaton that approaches $B(F_{n_0}, \delta)$ also approaches $B(H, 2\delta)$. By Zorn's Lemma there is a minimal set among all closed subsets of F that are approachable with automata.

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 $d(z,G) > 3\delta$, so in particular $d(x,y) \geqslant 3\delta$.

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 $B_0(z, \frac{5}{2}\delta)$ infinitely often. Since $d(z,G) > 3\delta$ this implies that G is not approachable by A, a contradiction.[7]

For every $n \in \mathbb{N}$ define the random variable 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 $n \in \mathbb{N}$. 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 $(\tau_p^x)_{p\in P_A}$).* Our first goal is to define, for every $p \in P_A$, a strategy τ_p^x of player 2 that ensures that the 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

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

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

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

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 $N \in \mathbb{N}$ there is a strategy $\tau_{N,p}$ of player 2 satisfying

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

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

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

[7] Actually, we will show that the average payoff remains in $B_0(z, \frac{5}{2}\delta)$ from some point on.

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

$$\mathbf{P}_{(A,p),\tau_{N,p}}\left(d(\bar{x}_n,x)<\frac{5}{4}\delta \text{ for some } N\leqslant n\leqslant K_{N,p}\right)\geqslant\frac{\eta_p}{4}. \tag{6}$$

For every fixed strategy τ of player 2, every $n\in\mathbb{N}$, and every $c>0$, the function $p'\mapsto\mathbf{P}_{(A,p'),\tau}(d(\bar{x}_n,x)\geqslant c)$ is linear (and Lipschitz-1) in p'. Therefore, the strategy $\tau_{N,p}$ satisfies (6) for the automaton (A,p'), provided one replaces the quantity $\frac{5}{4}\delta$ by $\frac{5}{4}\delta+\|p-p'\|$. As the space of probability distributions over the states of the automaton is compact, one can assume that $\eta_*:=\inf_{p\in P_A}\eta_p>0$ and $K_N:=\sup_{p\in P_A}K_{N,p}<\infty$.

We will show that since Eq. (5) holds for every $p\in P_A$, for every such p there is a strategy τ_p^x that satisfies (2). The strategy 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 $n\geqslant b_l/\delta^2$ such that the average payoff in the first n stages of the block is in $B(x,\frac{3}{2}\delta)$.

By (6), in every block there is a probability greater than $\eta_*/4$ such that the average payoff at the end of the block is in $B(x,\frac{5}{4}\delta)$. Since block l lasts at least b_l/δ^2 stages, the average payoff at the last stage of block l is δ^2-close to the average payoff obtained during block l. Since $\frac{5}{4}\delta+\delta^2\leqslant\frac{3}{2}\delta$ 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

$$\mathbf{P}_{(A,p),\tau_p^y}\left(\limsup_{n\to\infty}d(\bar{x}_n,y)<\frac{3}{2}\delta\right)=1. \tag{7}$$

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

Define a strategy τ that plays in blocks of random size as follows. 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,[8] 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. a Y-block) is within $\frac{7}{4}\delta$ of x (resp. y) is at least $1-\frac{\delta}{4}$. Since the number of stages spent in X-blocks converges to λ, and since payoffs are bounded by 1, the strong law of large numbers implies that if player 2 follows τ, the long-run average payoff remains in $B_0(z,\frac{5}{2}\delta)$ from some stage on. This implies in particular that G is not approachable, and the proof is complete.

References

Alon, N., Spencer, J., 1992. The Probabilistic Method. Wiley, New York.

Aumann, R.J., 1981. Survey of repeated games. In: Essays in Game Theory and Mathematical Economics in Honor of Oskar Morgenstern. In: Bohm, V. (Ed.), Gesellschaft, Recht, Wirtschaft, Wissenschaftsverlag, vol. 4. Bibliographisches Institut, Mannheim, pp. 11–42.

Ben Porath, E., 1990. The complexity of computing a best response automaton in repeated games with mixed strategies. Games Econ. Behav. 2, 1–12.

Blackwell, D., 1956a. An analog of the minimax theorem for vector-payoffs. Pacific J. Math. 6, 1–8.

Blackwell, D., 1956b. Controlled random walks. Invited address. Institute of Mathematical Statistics (August).

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

de Farias, D.P., Megiddo, N., 2003. How to combine expert (novice) advice when actions impact the environment? Manuscript.

Feller, W., 1968. An Introduction to Probability Theory and Its Applications. John Wiley and Sons.

Foster, D., Vohra, R., 1999. Regret in the on-line decision problem. Games Econ. Behav. 29, 7–35.

Fudenberg, D., Levine, D., 1999. Conditional universal consistency. Games Econ. Behav. 29, 104–130.

Hannan, J., 1957. Approximation to Bayes risk in repeated plays. In: Contribution to the Theory of Games, vol. 3. Princeton University Press, Princeton, NJ, pp. 97–139.

Hart, S., Mas-Colell, A., 1996. A simple adaptive procedure leading to correlated equilibrium. Center for Rationality and Interactive Decision Theory, DP 126, The Hebrew University, Jerusalem, Israel.

Hart, S., Mas-Colell, A., 2000. A simple adaptive procedure leading to correlated equilibrium. Econometrica 68, 1127–1150.

Hou, T.F., 1969. Weak approachability in a two-person game. Ann. Math. Statist. 40, 789–813.

Hou, T.F., 1971. Approachability in a two-person game. Ann. Math. Statist. 42, 735–744.

Jéhiel, P., 1995. Limited horizon forecasting in repeated alternating games. J. Econ. Theory 67, 497–519.

Kalai, E., Stanford, W., 1988. Finite rationality and interpersonal complexity in repeated games. Econometrica 56, 387–410.

Lehrer, E., 1988. Repeated games with stationary bounded-recall strategies. J. Econ. Theory 46, 130–144.

[8] Recall that λ is defined so that $x=\lambda x+(1-\lambda)y$.

Lehrer, E., 2003. Wide range no-regret theorem. Games Econ. Behav. 42, 101–115.
Lehrer, E., Solan, E., 2006. Excludability with bounded computational capacity strategies. Math. Oper. Res. 31, 637–648.
Luce, R.D., Raiffa, H., 1957. Games and Decisions. John Wiley & Sons, Inc., New York.
Neyman, A., 1985. Bounded complexity justifies cooperation in finitely repeated prisoner's dilemma. Econ. Letters 19, 227–229.
Rubinstein, A., 1986. Finite automata play the repeated prisoner's dilemma. J. Econ. Theory 39, 83–96.
Rustichini, A., 1999. Minimizing regret: The general case. Games Econ. Behav. 29, 224–243.
Schlag, K., 2003. On the value of randomizing and limiting memory in repeating decision-making under minimal regret. Manuscript.
Spinat, X., 2002. A necessary and sufficient condition for approachability. Math. Oper. Res. 27, 31–44.
Vieille, N., 1992. Weak approachability. Math. Oper. Res. 17, 781–791.
Watson, J., 1994. Cooperation in the infinitely repeated prisoner's dilemma with perturbations. Games Econ. Behav. 7, 260–285.
Zapechelnyuk, A., forthcoming. Better-reply dynamics with bounded recall. Math. Oper. Res.