

SHOULD I REMEMBER MORE THAN YOU?

– ON THE BEST RESPONSE TO BOUNDED RECALL STRATEGIES –

RENÉ LEVÍNSKÝ, ABRAHAM NEYMAN, AND MIROSLAV ZELENÝ*

ABSTRACT. In this note we consider a general infinitely repeated game with N players based on a deterministic stage game with a finite set of actions. We show that in the situation where players $i \in \{1, 2, \dots, n-1\}$ employ stationary t_i -bounded recall strategies (t_i -SBR strategy) there exists a t -SBR strategy for player n being the best response to the original strategies. Thus, there is no need for perfect infinite recall when playing against “bounded recall” players. Actually, it is sufficient if the best response strategy reveals the same cognitive load as the “most complicated” original strategy, i.e., $t = \max_i t_i$. We also reprove some older results for automata. Finally, we prove that the above described condition for sufficient complexity of the best response strategy is also satisfied if we allow for behavioral strategies or time-dependent bounded recall.

1. INTRODUCTION

There are two most frequent approaches modeling strategies of bounded complexity in (infinitely) repeated games. Aumann and Sorin (1989); Neyman (1985, 1997); Aumann (1981) and Lehrer (1988) consider players with bounded recall who have imperfect consciousness of the actual stage of the game and their action in the current stage game relies only on t previous signals they observed and which they are capable to “remember”. Rubinstein (1986) and Abreu and Rubinstein (1988) deal with (infinitely) repeated games in which players are represented by finite automata (Moore machines). These two models enable to measure the complexity of the strategy. In the bounded recall approach the strategy complexity is described by “depth of recall” t ; the complexity of strategy played by automaton is measured by the minimal number of states the automaton should have to play the given strategy.

The following general question naturally arises: What is the complexity of the strategy that is the best response to a strategy with a given complexity. Abreu and Rubinstein (1988) show that for every finite automaton A_1 there exists a finite automaton A_2 such that A_2 maximizes own profit in the game against A_1 and the number of states of A_2 is less or equal the number of states of A_1 .

We address a similar question for bounded recall model. If a rational player with perfect infinite recall looks for the best response against a general t -SBR strategy, it is enough, if the best response strategy reveals the identical “memory capacity” t as the original strategy. It is true that for every strategy with a bounded recall there exists an automaton

Key words and phrases. Bounded recall strategies, Repeated games, Best response, Finite automata.

* The third author was supported by the research project MSM 0021620839 financed by MSMT.

that realizes the given strategy. Still, the complexity measure differs in the two approaches: two strategies realized by the automata with identical number of states do not generally reveal the same depth of bounded recall.

As a tool we use the theory of Markov decision processes (MDP), namely theorems on existence of the best stationary strategy for a given MDP. In fact, once we get the idea to re-phrase our problem as MDP our results are simple corollaries of results of Blackwell (1962) and Derman (1965).

The new perspective of Blackwell's optimality extends the previous results of Abreu and Rubinstein (1988) in three different directions. First, the statements can be proven for behavioral automata as well as for the behavioral SBR-strategies by the same manner. Second, we get all the results also for behavioral strategies. Third, the Blackwell's theorem gives all statements in a robust form – namely for the whole interval of discount factors $\beta \in [\beta_0, 1]$.

In the last part of our paper we show that for fixed β are our results valid also for time-dependent automata. Since in this case the state space of the corresponding MDP is not more finite, we employ the result of Derman (1965) for MDP with countable state spaces.

All relevant notions will be defined and commented in the next section. Section 3 contains the main result and its proof.

2. PRELIMINARIES

2.1. Game and supergame. The set of (positive) natural numbers is denoted by \mathbf{N} . If X is a finite set, then $\Delta(X)$ denotes the set of all probabilities on X .

Throughout the paper $G = \langle \Sigma_1, \Sigma_2, u_1, u_2 \rangle$ will be a fixed two person game in normal form, where Σ_i is a nonempty finite set of strategies for player i ($i = 1, 2$) and $u_i : \Sigma_1 \times \Sigma_2 \rightarrow \mathbf{R}$ is the payoff function of player i .

By a *supergame of G* (in notation G^∞) we mean an infinite sequence of repetitions of G . At each period $t = 1, 2, 3, \dots$ players 1 and 2 make simultaneous and independent moves $s_t^i \in \Sigma_i$, $i = 1, 2$.

Denote $\Sigma = \Sigma_1 \times \Sigma_2$. If $\alpha \in \Sigma$, then the first and the second coordinate of α are denoted by α^1 and α^2 respectively. The symbol $\Sigma^{<\mathbf{N}}$ denotes all finite sequences of elements of Σ including the empty one. If we have $a, b \in \Sigma^{<\mathbf{N}}$, then the *concatenation* of a and b is denoted by $a^{\wedge}b$.

Although we deal with two-person games only, one can immediately extend our results to the game with n players ($n > 2$) since players $1, 2, \dots, n - 1$ can be considered as one player playing actions from the space $\Sigma_1 \times \dots \times \Sigma_{n-1}$.

2.2. SBR-strategies. A *behavioral strategy for player i in G^∞* is a mapping $\omega : \Sigma^{<\mathbf{N}} \rightarrow \Delta(\Sigma_i)$. The player i following a behavioral strategy ω plays at the t -th round an action $a \in \Sigma_i$ with the probability $\omega(z_1, \dots, z_{t-1})(a)$ where $(z_1, \dots, z_{t-1}) \in \Sigma^{t-1}$ is the sequence of actions which have been already played.

Let $k \in \mathbf{N}$. By a *behavioral k -SBR strategy for player i in G^∞* we mean a pair (e, ω) , where $e = (e_1, e_2, \dots, e_k) \in \Sigma^k$ and $\omega : \Sigma^k \rightarrow \Delta(\Sigma_i)$ is a mapping. Player i following the

strategy (e, ω) plays as follows. If moves $z_1, \dots, z_l \in \Sigma$ have been played, then player i takes the sequence s , which is formed by the last k elements of the sequence $(e_1, \dots, e_k, z_1, \dots, z_l)$, and his k -th move is $a \in \Sigma_i$ with the probability $\omega(s)(a)$. It is easy to see that any behavioral k -SBR strategy can be considered as a behavioral strategy in the sense of the previous definition.

Notice that the behavioral strategies are not just convex combinations of deterministic strategies: in the run of the game where one player employs strategy mixing s_1 and s_2 with non-degenerated probabilities p_1 and p_2 we will never observe a cycle in comparison to any deterministic t -SBR strategy.

We say that a behavioral strategy σ is a *behavioral SBR strategy* if σ is a behavioral k -SBR strategy for some $k \in \mathbf{N}$. The deterministic version of the above notions, i.e., *strategy*, *k -SBR strategy*, *SBR strategy*, are obtained just by replacing $\Delta(\Sigma_i)$ by Σ_i .

2.3. Automata and behavioral automata. A *k -state behavioral automaton* (for player 1 in G) is a quadruple $\langle S, s_0, \alpha, \tau \rangle$, where S is a finite nonempty set (the state space), $s_0 \in S$ is the initial state, $\alpha : S \rightarrow \Delta(\Sigma_1)$ is a probabilistic action function, and $\tau : S \times \Sigma \rightarrow S$ is a transition function. A behavioral automaton $\langle S, s_0, \alpha, \tau \rangle$ defines a behavioral strategy σ^1 (for player 1, say) inductively: $\sigma^1(\emptyset) = \alpha(s_0)$, $\sigma^1(z_1, \dots, z_t) = \alpha(s_t)$, where $s_t = \tau(s_{t-1}, z_t)$.

Again a *k -state (deterministic) automaton* is defined by the replacement of $\Delta(\Sigma_1)$ with Σ_1 .

2.4. Profit of the supergame. There is probably no canonical way how to define a payoff function for G^∞ . We employ the following approach. Let σ^i , $i = 1, 2$, be behavioral strategies. We denote $\mu(\sigma^1, \sigma^2)$ the unique measure on $\Sigma^\mathbf{N}$ (defined on Borel sets) satisfying $\mu(w \times \Sigma^\mathbf{N}) = \pi(w)$ for every $w \in \Sigma^{<\mathbf{N}}$, where $\pi(w)$ denotes the probability that w is the initial sequence in G^∞ , where player i follow the strategy σ^i , $i = 1, 2$. Then we define the following evaluation functions of the profit of the player 2 in G^∞ , where the player i follows the strategy σ^i , $i = 1, 2$.

$$\begin{aligned} V_\beta(\sigma^1, \sigma^2) &= \int_{\Sigma^\mathbf{N}} \left(\sum_{t=1}^{\infty} \beta^{t-1} u_2(\mathbf{w}_t) \right) d\mu(\sigma^1, \sigma^2)(\mathbf{w}), \\ \overline{H}(\sigma^1, \sigma^2) &= \int_{\Sigma^\mathbf{N}} \limsup_{k \rightarrow \infty} \left(\frac{1}{k} \sum_{t=1}^k u_2(\mathbf{w}_t) \right) d\mu(\sigma^1, \sigma^2)(\mathbf{w}), \\ \underline{H}(\sigma^1, \sigma^2) &= \int_{\Sigma^\mathbf{N}} \liminf_{k \rightarrow \infty} \left(\frac{1}{k} \sum_{t=1}^k u_2(\mathbf{w}_t) \right) d\mu(\sigma^1, \sigma^2)(\mathbf{w}). \end{aligned}$$

If σ^1 and σ^2 are SBR strategies then the run in G^∞ is uniquely determined and is eventually periodic. It is not difficult to see that $\overline{H}(\sigma^1, \sigma^2) = \underline{H}(\sigma^1, \sigma^2)$ and the joint value can be computed as average over the period, which appears in $\mathbf{w} \in \Sigma^\mathbf{N}$.

2.5. Markov decision processes. Let us recall some basic facts on Markov decision processes (cf. Neyman (2003)). By a *Markov decision process (MDP, for short)* we mean a 5-tuple $\langle S, A, r, p, \mu \rangle$ such that

- S is a nonempty countable set (set of states),
- A is a nonempty finite set (set of actions),
- $r(z, a)$ is a real number for every $z \in S$ and $a \in A$ (reward function),
- $p(z, a)$ is a probability on S for every $z \in S$ and $a \in A$,
- μ is an initial probability on S .

One can interpret this structure as follows. The set A is the set of feasible actions which can be played at any state z by the decision maker. The sequence $(z_1, a_1, z_2, a_2, \dots)$ of states and actions of the process is realized in the following way. The initial state z_1 is chosen with the probability $\mu(z_1)$. If the sequence $(z_1, a_1, z_2, a_2, \dots, z_t)$ has been constructed, then the decision maker plays an action a_t , then he receives a payoff $r(z_t, a_t)$. The probability of the next state $z_{t+1} \in S$ of the process is given by the probability distribution $p(z_t, a_t)$.

A *strategy* for an MDP is a function σ that assigns to every finite sequence of states and actions $h = (z_1, a_1, z_2, a_2, \dots, z_t)$ a probability $\sigma(h)$ on A . If $\sigma(h)$ is always a Dirac measure, then σ is *deterministic*. By a *stationary deterministic strategy* for an MDP we mean a deterministic strategy depending only on the last state. Thus one can view a stationary strategy as a mapping assigning $a \in A$ to each $z \in S$.

Fix an MDP and denote $\mathfrak{T} = (S \times A)^{\mathbb{N}}$. The decision maker wants to maximize a specific evaluation of the sequence $(r(\mathfrak{s}_t))_{t=1}^{\infty}$ of payoffs, $\mathfrak{s} \in \mathfrak{T}$. We will deal with the following evaluation methods. Let σ be a strategy for the decision maker. Let $\nu(\sigma)$ be a measure on \mathfrak{T} such that $\nu(\sigma)(w \times \mathfrak{T})$, $w \in (S \times A)^{<\mathbb{N}}$, equals probability that w is the initial sequence, when the decision maker follows strategy σ . We set

$$\begin{aligned} v(\beta, \sigma) &= \int_{\mathfrak{T}} \left(\sum_{t=1}^{\infty} \beta^{t-1} r(\mathfrak{s}_t) \right) d\nu(\sigma)(\mathfrak{s}), \\ \bar{h}(\sigma) &= \int_{\mathfrak{T}} \limsup_{k \rightarrow \infty} \left(\frac{1}{k} \sum_{t=1}^k r(\mathfrak{s}_t) \right) d\nu(\sigma)(\mathfrak{s}), \\ \underline{h}(\sigma) &= \int_{\mathfrak{T}} \liminf_{k \rightarrow \infty} \left(\frac{1}{k} \sum_{t=1}^k r(\mathfrak{s}_t) \right) d\nu(\sigma)(\mathfrak{s}). \end{aligned}$$

2.6. The existence of optimal stationary strategy. The key tool in our paper is Blackwell theorem and Derman theorem. The term “strategy” is considered here in the context of MDP.

Theorem 2.1 (Derman, 1965). *Let $\langle S, A, r, p, \mu \rangle$ be an MDP and $\beta \in (0, 1)$. Then there is a stationary deterministic strategy σ such that, for every strategy τ , we have $v(\beta, \sigma) \geq v(\beta, \tau)$.*

Theorem 2.2 (Blackwell, 1962). *Let $\langle S, A, r, p, \mu \rangle$ be an MDP with finitely many states. Then there is a stationary deterministic strategy σ and a discount factor $\beta_0 \in (0, 1)$ such that*

- *for every strategy τ , we have $v(\beta, \sigma) \geq v(\beta, \tau)$ for every $\beta \in [\beta_0, 1)$;*
- *for every strategy τ , we have $\underline{h}(\sigma) \geq \bar{h}(\tau)$;*

- for every $\varepsilon > 0$ there exists $N \in \mathbf{N}$ such that, for every strategy τ and every $n \geq N$, we have

$$\int_{\mathfrak{S}} \left(\frac{1}{n} \sum_{t=1}^n r(\mathfrak{s}_t) \right) d\nu(\sigma)(\mathfrak{s}) \geq \int_{\mathfrak{S}} \left(\frac{1}{n} \sum_{t=1}^n r(\mathfrak{s}_t) \right) d\nu(\tau)(\mathfrak{s}) - \varepsilon.$$

3. MAIN RESULT

3.1. Finite state automata and k -SBR strategies.

Theorem 3.1. *Let σ^1 be a behavioral k -SBR strategy of player 1 in G^∞ (a strategy of player 1 in G^∞ that is defined by a behavioral k -state automaton respectively). Then the following holds.*

- (i) *For every $\beta \in (0, 1)$ there exists a k -SBR strategy σ^2 (a strategy σ^2 defined by a k -state automaton respectively) such that for every behavioral strategy τ in G^∞ we have*

$$V_\beta(\sigma^1, \sigma^2) \geq V_\beta(\sigma^1, \tau).$$

- (ii) *There is a k -SBR strategy σ^2 (a strategy σ^2 defined by a k -state automaton respectively) and a discount factor $\beta_0 \in (0, 1)$ such that*

– *for every behavioral strategy τ and every $\beta \in [\beta_0, 1)$, we have*

$$V_\beta(\sigma^1, \sigma^2) \geq V_\beta(\sigma^1, \tau);$$

– *for every behavioral strategy τ , we have*

$$\underline{H}(\sigma^1, \sigma^2) \geq \overline{H}(\sigma^1, \tau);$$

– *for every $\varepsilon > 0$ there exists $N \in \mathbf{N}$ such that, for every behavioral strategy τ and every $n \geq N$, we have*

$$\int_{\Sigma^{\mathbf{N}}} \left(\frac{1}{n} \sum_{t=1}^n u_2(\mathfrak{w}_t) \right) d\mu(\sigma_1, \sigma_2)(\mathfrak{w}) \geq \int_{\Sigma^{\mathbf{N}}} \left(\frac{1}{n} \sum_{t=1}^n u_2(\mathfrak{w}_t) \right) d\mu(\sigma_1, \tau)(\mathfrak{w}) - \varepsilon.$$

Proof. The only thing to do is to establish a correspondence between MDP and behavioral k -SBR strategies (behavioral k -state automata respectively). First suppose that σ^1 is a behavioral k -SBR strategy for the player 1. Write $\sigma^1 = (e, \varphi)$. We define an MDP $\langle S, A, r, p, \mu \rangle$ as follows. We set $S := \Sigma^k$, $A := \Sigma_2$, μ will be Dirac measure sitting at e and for every $z = (z_1, \dots, z_k) \in S$, $a \in \Sigma_2$ we define

$$r(z, a) := \sum_{c \in \Sigma_1} \varphi(z)(c) \cdot u_2(c, a),$$

$$p(z, a)(w) := \begin{cases} \varphi(z)(c) & w = (z_2, \dots, z_k)^\wedge([c, a]), \\ 0 & \text{otherwise.} \end{cases}$$

To prove assertion (i) we use Theorem 2.1 giving a stationary deterministic strategy ω such that for every strategy τ of the decision maker in MDP, we have $v(\beta, \omega) \geq v(\beta, \tau)$. The stationary deterministic strategy ω is a mapping $\omega : S \rightarrow \Sigma_2$. The desired strategy σ^2

is defined as the pair (e, ω) . Now let ψ be a behavioral strategy in G^∞ for player 2. One can consider ψ also as a strategy for the decision maker in our MDP. Then we have

$$V_\beta(\sigma^1, \sigma^2) = v(\beta, \omega) \geq v(\beta, \psi) = V_\beta(\sigma^1, \psi)$$

and we are done.

As for assertion (ii) we proceed in the same way but instead of Theorem 2.1 we use Theorem 2.2.

Now assume that the strategy σ^1 is defined via a k -state behavioral automaton $\langle S, s_0, \alpha, \tau \rangle$. Let $S := S$, $A := \Sigma_2$, let μ be Dirac measure sitting at s_0 , $p(z, a) := \tau(z, a)$, and

$$r(z, a) := \sum_{c \in \Sigma_1} \alpha(z)(c) \cdot u_2(c, a)$$

for every $z \in S, a \in \Sigma_2$. Applying Theorems 2.1 and 2.2 to the MDP $\langle S, A, r, p, \mu \rangle$ we get the desired conclusions. \square

Remark 3.2. A part of (ii) can be found in Abreu and Rubinstein (1988), where deterministic automata are studied.

3.2. Time-dependent automata. The literature on Markov decision processes is vast and offers different modifications of the approach presented above. Let us conclude this note by an application of Derman's theorem to time-dependent automata.

The game G and the corresponding supergame G^∞ are as above. A *k -state time-dependent behavioral automaton* (for player i in G) is a quadruple $\langle S, s_0, \alpha, \tau \rangle$, where S is a set with $k \in \mathbf{N}$ elements (the state space), $s_0 \in S$ is an initial state, $\alpha : \mathbf{N} \times S \rightarrow \Delta(\Sigma_i)$ is a probabilistic action function, and $\tau : \mathbf{N} \times S \times \Sigma \rightarrow S$ is a transition function. An automaton $\langle S, s_0, \alpha, \tau \rangle$ defines a behavioral strategy $\sigma^i : \Sigma^{<\mathbf{N}} \rightarrow \Delta(\Sigma_i)$ for player i inductively: $\sigma^i(\emptyset) = \alpha(1, s_0)$, $\sigma^i(z_1, \dots, z_t) = \alpha(t+1, s_t)$, where $s_t = \tau(t, s_{t-1}, z_t)$.

Theorem 3.3. *Let σ^1 be a strategy of player 1 in G^∞ that is defined by a k -state time-dependent behavioral automaton. Then for every $\beta \in (0, 1)$ there exists a strategy σ^2 defined by a k -state time-dependent deterministic automaton such that for every strategy τ in G^∞ we have*

$$V_\beta(\sigma^1, \sigma^2) \geq V_\beta(\sigma^1, \tau).$$

Proof. Assume that the strategy σ^1 is defined via a k -state time-dependent behavioral automaton $\langle S', s_0, \alpha, \tau \rangle$. We define an MDP $\langle S, A, r, p, \mu \rangle$ as follows: $S := \mathbf{N} \times S'$, $A := \Sigma_2$, μ is Dirac measure sitting at s_0 , and for every $(n, s) \in \mathbf{N} \times S' = S$, $a \in \Sigma$ we have

$$r(n, s, a) := \sum_{c \in \Sigma_1} \alpha(n, s)(c) \cdot u_2(c, a),$$

$$p(n, s, a) := \text{Dirac measure sitting at } (n+1, \tau(n, s, a)).$$

Applying Theorem 2.1 we get a deterministic stationary strategy $\omega : \mathbf{N} \times S' \rightarrow \Sigma_2$ such that for every strategy τ of the decision maker in our MDP, we have $v(\beta, \omega) \geq v(\beta, \tau)$.

The desired automaton $\langle S^*, s_0^*, \alpha^*, \tau^* \rangle$ is defined by $S^* := S'$, $s_0^* := (1, s_0)$, and for every $n \in \mathbf{N}$, $s \in S'$, $a \in \Sigma$, we set

$$\begin{aligned}\alpha^*(n, s) &:= \omega(n, s), \\ \tau^*(n, s, a) &:= \tau(n, s, a).\end{aligned}$$

□

REFERENCES

- ABREU, D., AND A. RUBINSTEIN (1988): “The structure of Nash equilibrium in repeated games with finite automata,” *Econometrica*, 56(6), 1259–1281.
- AUMANN, R. J. (1981): “Survey of repeated games,” in *Essays in Game Theory and Mathematical Economics in Honour of Oskar Morgenstern*, pp. 11–42. Wissenschaftsverlag, Bibliographisches Institut, Mannheim, Wien, Zurich.
- AUMANN, R. J., AND S. SORIN (1989): “Cooperation and bounded recall,” *Games and Economic Behavior*, 1(1), 5–39.
- BLACKWELL, D. (1962): “Discrete dynamic programming,” *Annals of Mathematical Statistics*, 33, 719–726.
- DERMAN, C. (1965): “Markovian Sequential Control Processes – Denumerable State Spaces,” *Journal of Mathematical Analysis and Applications*, 10, 295–302.
- LEHRER, E. (1988): “Repeated games with stationary bounded recall strategies,” *Journal of Economic Theory*, 46(1), 130–144.
- NEYMAN, A. (1985): “Bounded Complexity Justifies Cooperation in the Finitely Repeated Prisoners’s Dilemma,” *Economics Letters*, 19, 227–229.
- (1997): “Cooperation, Repetition, and Automata,” in *Cooperation: Game Theoretic Approaches, NATO ASI Series F*, ed. by S. Hart, and A. Mas-Colell, pp. 233–255. Springer-Verlag.
- (2003): “From Markov chains to stochastic games,” in *Stochastic games and applications*, ed. by A. Neyman, and S. Sorin, pp. 9–25. NATO Science Series, Kluwer, Dordrecht, Boston, London.
- RUBINSTEIN, A. (1986): “Finite automata play the repeated prisoner’s dilemma,” *Journal of Economic Theory*, 39(1), 83–96.

MAX PLANCK INSTITUTE FOR ECONOMICS, KAHLAISCHE STRASSE 10, 07745 JENA, GERMANY
E-mail address: levinsky@econ.mpg.de

INSTITUTE OF MATHEMATICS AND CENTER FOR THE STUDY OF RATIONALITY, THE HEBREW UNIVERSITY OF JERUSALEM, GIVAT RAM, JERUSALEM 91904, ISRAEL
E-mail address: aneyman@math.huji.ac.il

CHARLES UNIVERSITY, FACULTY OF MATHEMATICS AND PHYSICS, SOKOLOVSKÁ 83, 186 75, PRAHA 8, CZECH REPUBLIC
E-mail address: zeleny@karlin.mff.cuni.cz