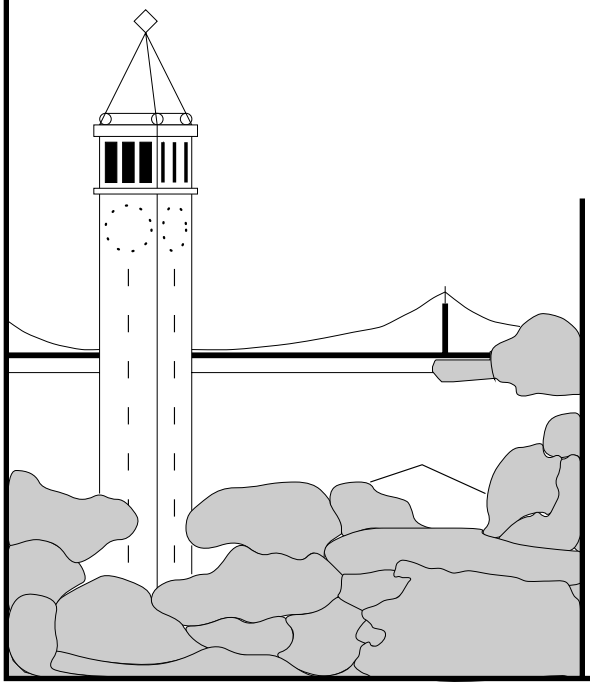
Memoryless Strategies in Concurrent Games with Reachability Objectives

Krishnendu Chatterjee, Luca de Alfaro and Thomas A. Henzinger



Report No. UCB/CSD-5-1406

August 2005

Computer Science Division (EECS)
University of California
Berkeley, California 94720

Memoryless Strategies in Concurrent Games with Reachability Objectives *

Krishnendu Chatterjee[†] Luca de Alfaro[§] Thomas A. Henzinger[†,‡]

[†] EECS, University of California, Berkeley, USA
[§] CE, University of California, Santa Cruz
[‡] EPFL, Switzerland
{c_krish,tah}@eecs.berkeley.edu, luca@soe.ucsc.edu

August 2005

Abstract

We present a simple proof of the fact that in concurrent games with reachability objectives, for all $\varepsilon > 0$, memoryless ε-optimal strategies exist. A memoryless strategy is independent of the history of plays; and an ε-optimal strategy achieves the objective with probability within ε of the value of the game. In contrast to previous proofs of this fact, which rely on the limit behavior of discounted games using advanced Puisieux series analysis, our proof is elementary and combinatorial.

1 Introduction

We consider concurrent reachability games played by two players over finite state spaces. The configuration of such a game is called a *state*. At each round, the two players choose their moves concurrently and independently; the two moves and the current state determine a successor state, or in general, a probability distribution over the successor states. A *play* of the game consists in the infinite sequence of states visited while playing the game. The goal of player 1 consists in forcing the game to a specified set of target states; the goal of player 2 consists in preventing the game from reaching a

*This research was supported in part by the ONR grant N00014-02-1-0671, the AFOSR MURI grant F49620-00-1-0327, and the NSF ITR grant CCR-0225610.

target state. Consequently, we assign value 1 to all plays that reach the target set, and value 0 to all other plays. The players can adopt strategies that are both randomized and history-dependent. Player 1 can *guarantee* a value v for the game from a state s if player 1 has a strategy that ensures that the expected value of a play from s is at least v, regardless of the strategy chosen by player 2. The *value at s of the reachability game with target T* is the supremum of the set of values that player 1 can guarantee from s. An *optimal strategy* for player 1 is a strategy that guarantees the value of the game from each state s. For $\varepsilon > 0$, an *ε-optimal strategy* for player 1 is a strategy that guarantees the objective is satisfied with a probability within ε of the value of the game for each state s.

Concurrent reachability games belong to the family of repeated games [11, 7], and they have been studied more specifically in [5, 4, 6]. It has long been known that optimal strategies need not exist for concurrent reachability games [7], so that one must settle for ε-optimality. It is also known that, for $\varepsilon > 0$, there always exist ε-optimal strategies that are memoryless, i.e., such that the probability distribution over moves depends only on the current state, and not on the past history of the game [8].

Unfortunately, the only previous proof is rather complex. The proof considered *discounted* versions of reachability games, where a play that reaches the goal in k steps is assigned a value of α^k, for some discount factor $0 < \alpha \leq 1$, rather than value 1. It is possible to show that, for $0 < \alpha < 1$, memoryless optimal strategies always exist. The result for the undiscounted ($\alpha = 1$) case follows from an analysis of the limit behavior of such optimal strategies for $\alpha \to 1$; the limit behavior is studied with the help of results on the field of real Puisieux series [8]. This proof idea works not only for reachability games, but also for total-reward games with non-negative rewards (see [8] again).

We show that the existence of memoryless ε-optimal strategies for concurrent reachability games can be established by more elementary means, which do not require the consideration of discounted versions of the games, nor results on real Puisieux series. In particular, we present a proof that relies only on combinatorial techniques, and on simple results on Markov decision processes [1, 3]. As our proof is easily accessible, we believe that the proof techniques we use may find future applications in game theory.

2 Concurrent Games with Reachability Objectives

Notation. For a countable set A, a *probability distribution* on A is a function $\delta : A \mapsto [0,1]$ such that $\sum_{a \in A} \delta(a) = 1$. We denote the set of probability distributions on A by $\mathcal{D}(A)$. Given a distribution $\delta \in \mathcal{D}(A)$, we denote by $Supp(\delta) = \{x \in A \mid \delta(x) > 0\}$ the *support* of δ.

Definition 1 (Concurrent Games) *A (two-player)* concurrent game structure $G = \langle S, Moves, \Gamma_1, \Gamma_2, \delta \rangle$ *consists of the following components:*

- *A finite state space S and a finite set $Moves$ of moves.*
- *Two move assignments $\Gamma_1, \Gamma_2 : S \mapsto 2^{Moves} \setminus \emptyset$. For $i \in \{1,2\}$, assignment Γ_i associates with each state $s \in S$, the non-empty set $\Gamma_i(s) \subseteq Moves$ of moves available to player i at state s.*
- *A probabilistic transition function $\delta : S \times Moves \times Moves \to \mathcal{D}(S)$, that gives the probability $\delta(s, a_1, a_2)(t)$ of a transition from s to t when player 1 plays move a_1 and player 2 plays move a_2, for all $s, t \in S$ and $a_1 \in \Gamma_1(s)$, $a_2 \in \Gamma_2(s)$.* ∎

At every state $s \in S$, player 1 chooses a move $a_1 \in \Gamma_1(s)$, and simultaneously and independently player 2 chooses a move $a_2 \in \Gamma_2(s)$. The game then proceeds to the successor state t with probability $\delta(s, a_1, a_2)(t)$, for all $t \in S$. A state s is called an *absorbing state* if for all $a_1 \in \Gamma_1(s)$ and $a_2 \in \Gamma_2(s)$ we have $\delta(s, a_1, a_2)(s) = 1$. In other words, at s for all choice of moves of the players, the next state is always s. For all states $s \in S$ and moves $a_1 \in \Gamma_1(s)$ and $a_2 \in \Gamma_2(s)$, we indicate by $Dest(s, a_1, a_2) = Supp(\delta(s, a_1, a_2))$ the set of possible successors of s when moves a_1, a_2 are selected.

A *path* or a *play* ω of G is an infinite sequence $\omega = \langle s_0, s_1, s_2, \ldots \rangle$ of states in S such that for all $k \geq 0$, there are moves $a_1^k \in \Gamma_1(s_k)$ and $a_2^k \in \Gamma_2(s_k)$ with $\delta(s_k, a_1^k, a_2^k)(s_{k+1}) > 0$. We denote by Ω the set of all paths and by Ω_s the set of all paths $\omega = \langle s_0, s_1, s_2, \ldots \rangle$ such that $s_0 = s$, i.e., the set of plays starting from state s.

Strategies. A *selector* ξ for player $i \in \{1, 2\}$ is a function $\xi : S \mapsto \mathcal{D}(Moves)$ such that for all $s \in S$ and $a \in Moves$, if $\xi(s)(a) > 0$, then $a \in \Gamma_i(s)$. We denote by Λ_i the set of all selectors for player $i \in \{1, 2\}$. A *strategy* for player 1 is a function $\pi : S^+ \to \Lambda_1$ that associates with every finite non-empty sequence of states, representing the history of the play so

far, a selector; we define strategies for player 2 similarly. A *memoryless* strategy is independent of the history of the play and depends only on the current state. Memoryless strategies correspond to selectors; we write $\overline{\xi}_1$ for the memoryless strategy consisting in playing forever the selector ξ_1. We denote by Π_1 and Π_2 the sets of all strategies for player 1 and player 2, respectively. We denote by Π_1^M and Π_2^M the family of memoryless strategies for player 1 and player 2, respectively.

Once the starting state s and the strategies π_1 and π_2 for the two players have been chosen, the game is reduced to an ordinary stochastic process. Hence, the probabilities of events are uniquely defined, where an *event* $\mathcal{A} \subseteq \Omega_s$ is a measurable set of paths. For an event $\mathcal{A} \subseteq \Omega_s$, we denote by $\Pr_s^{\pi_1,\pi_2}(\mathcal{A})$ the probability that a path belongs to $\mathcal{A}$ when the game starts from s and the players follows the strategies π_1 and π_2. Similarly, for a measurable function $f : \Omega_s \to \mathbb{R}$, we denote by $\mathrm{E}_s^{\pi_1,\pi_2}(f)$ the expected value of f when the game starts from s and the players follows the strategies π_1 and π_2. For $i \geq 0$, we denote by $\Theta_i : \Omega \to S$ the random variable denoting the i-th state along a path.

Valuations. A *valuation* is a mapping $v : S \to [0,1]$ associating a real number $v(s) \in [0,1]$ with each state s. Given two valuations $v, w : S \to \mathbb{R}$, we write $v \leq w$ when $v(s) \leq w(s)$ for all $s \in S$. For an event $\mathcal{A}$, we denote by $\Pr^{\pi_1,\pi_2}(\mathcal{A})$ the valuation $S \to [0,1]$ defined for all $s \in S$ by $\big(\Pr^{\pi_1,\pi_2}(\mathcal{A})\big)(s) = \Pr_s^{\pi_1,\pi_2}(\mathcal{A})$; similarly, for a measurable function $f : \Omega_s \to [0,1]$, we denote by $\mathrm{E}^{\pi_1,\pi_2}(f)$ the valuation $S \to [0,1]$ defined for all $s \in S$ by $\big(\mathrm{E}^{\pi_1,\pi_2}(f)\big)(s) = \mathrm{E}_s^{\pi_1,\pi_2}(f)$.

Given a valuation v, and two selectors $\xi_1 \in \Lambda_1$ and $\xi_2 \in \Lambda_2$, we define the valuations $Pre_{\xi_1,\xi_2}(v)$, $Pre_{1:\xi_1}(v)$, and $Pre_1(v)$ as follows, for all $s \in S$:

$$
\begin{aligned}
Pre_{\xi_1,\xi_2}(v)(s) &= \sum_{a,b \in Moves} \sum_{t \in S} v(t)\, \delta(s,a,b)(t)\, \xi_1(a)\, \xi_2(b) \\
Pre_{1:\xi_1}(v)(s) &= \inf_{\xi_2 \in \Lambda_2} Pre_{\xi_1,\xi_2}(v)(s) \\
Pre_1(v)(s) &= \sup_{\xi_1 \in \Lambda_1} \inf_{\xi_2 \in \Lambda_2} Pre_{\xi_1,\xi_2}(v)(s).
\end{aligned}
$$

Note that all of these valuations are monotonic: for two valuations v, w, if $v \leq w$, then for all selectors $\xi_1 \in \Lambda_1$ and $\xi_2 \in \Lambda_2$ we have $Pre_{\xi_1,\xi_2}(v) \leq Pre_{\xi_1,\xi_2}(w)$, $Pre_{1:\xi_1}(v) \leq Pre_{1:\xi_1}(w)$, and $Pre_1(v) \leq Pre_1(w)$.

Reachability objectives. Given a subset $T \subseteq S$ of *target states,* the goal of a reachability game consists in reaching T. Therefore, we define the

set *winning plays* as the set $\mathrm{Reach}(T) = \{\, \omega = \langle s_0, s_1, s_2, \ldots \rangle \in \Omega \mid s_k \in T \text{ for some } k \geq 0 \,\}$ of plays that reach T. For any $T \subseteq S$, the set $\mathrm{Reach}(T)$ is measurable for any choice of strategies for the two-players [12]; we denote the probability that a path is in $\mathrm{Reach}(T)$ starting from state $s \in S$, given strategies π_1 and π_2 for players 1 and 2, respectively, by $\Pr_s^{\pi_1,\pi_2}(\mathrm{Reach}(T))$. Given a state $s \in S$ and a reachability objective, $\mathrm{Reach}(T)$, we define the *value of the game* at s for player 1 as

$$\langle\!\langle 1 \rangle\!\rangle_{val}(\mathrm{Reach}(T))(s) = \sup_{\pi_1 \in \Pi_1} \inf_{\pi_2 \in \Pi_2} \Pr_s^{\pi_1,\pi_2}(\mathrm{Reach}(T)).$$

The quantitative determinacy result of [10] ensures that

$$\sup_{\pi_1 \in \Pi_1} \inf_{\pi_2 \in \Pi_2} \Pr_s^{\pi_1,\pi_2}(\mathrm{Reach}(T)) + \sup_{\pi_2 \in \Pi_2} \inf_{\pi_1 \in \Pi_1} \Pr_s^{\pi_1,\pi_2}(\Omega \setminus \mathrm{Reach}(T)) = 1.$$

A strategy π_1 for player 1 is *optimal* if for all $s \in S$ we have

$$\inf_{\pi_2 \in \Pi_2} \Pr_s^{\pi_1,\pi_2}(\mathrm{Reach}(T)) = \langle\!\langle 1 \rangle\!\rangle_{val}(\mathrm{Reach}(T))(s).$$

For $\varepsilon > 0$, a strategy π_1 for player 1 is *ε-optimal* if for all $s \in S$ we have

$$\inf_{\pi_2 \in \Pi_2} \Pr_s^{\pi_1,\pi_2}(\mathrm{Reach}(T)) \geq \langle\!\langle 1 \rangle\!\rangle_{val}(\mathrm{Reach}(T))(s) - \varepsilon.$$

3 Existence of Memoryless ε-Optimal Strategies

3.1 Markov decision processes

In our proof, we need some facts about one-player versions of concurrent stochastic games, known as *Markov decision processes* (MDPs) [1]. For $i \in \{\, 1, 2 \,\}$, an i-MDP is a concurrent game where, for all $s \in S$, we have $|\Gamma_{3-i}(s)| = 1$. Given a concurrent game G, if we fix a memoryless strategy corresponding to selector ξ for player 1, the game is equivalent to a 2-MDP G_ξ with transition function

$$\delta_\xi(s, b)(t) = \sum_{a \in \Gamma_1(s)} \delta(s, a, b)(t) \cdot \xi(s)(a),$$

for all $s \in S$ and $b \in \Gamma_2(s)$. Similarly, if we fix selectors ξ_1, ξ_2 for both players in a concurrent game G, we obtain a Markov chain, which we denote by G_{ξ_1,ξ_2}. In an MDP, the sets of states that play an equivalent role to the closed recurrent classes of Markov chains [9] are called *end-components* [2, 3].

Definition 2 (End components) *An* end-component *(EC) of a 2-MDP is a subset $C \subseteq S$ such that there is a selector ζ for player 2 under which C forms a closed recurrent class of the resulting Markov chain.*

It is not difficult to see that an equivalent characterization of an end-component C is the following. For each $s \in C$, there is a subset of moves $M(s) \subseteq \Gamma_2(s)$ such that:

1. when a move in $M(s)$ is chosen at s, all the states that can be reached with non-zero probability are in C;
2. the graph (C, E), where E consists of the transitions that occur with non-zero probability when moves in $M(\cdot)$ are taken, is strongly connected.

Given a path ω, denote by $\mathrm{Infi}(\omega)$ the set of states that occurs infinitely often along ω. Given a set $\mathcal{F} \subseteq 2^S$ of subset of states we denote by $\mathrm{Infi}(\mathcal{F})$ the event $\{\omega \mid \mathrm{Infi}(\omega) \in \mathcal{F}\}$. The following theorem states that in a 2-MDP, for any strategy of player 2, the set of states visited infinitely often is an EC with probability 1. Corollary 1 follows easily from Theorem 1.

Theorem 1 ([3]) *Let $\mathcal{C}$ be the set of end-components of a 2-MDP G_{ξ_1}. For all strategies $\pi_2 \in \Pi_2$ and all states $s \in S$, we have $\Pr_s^{\overline{\xi}_1,\pi_2}(\mathrm{Infi}(\mathcal{C})) = 1$.*

Corollary 1 *Let $\mathcal{C}$ be the set of end-components of a 2-MDP G_{ξ_1} and let $Z = \bigcup_{C \in \mathcal{C}} C$ be the set of states of all end-components. For all strategies $\pi_2 \in \Pi_2$ and all states $s \in S$, we have $\Pr_s^{\overline{\xi}_1,\pi_2}(\mathit{Reach}(Z)) = 1$.*

3.2 From value iteration to selectors

Consider a reachability game with target $T \subseteq S$. Let $W_2 = \{ s \in S \mid \langle\langle 1 \rangle\rangle_{val}(\mathrm{Reach}(T))(s) = 0 \}$ be the set of states from which player 1 cannot reach the goal with positive probability; from [4, 6] we know that player 2 has a strategy that confines the game in W_2. An arbitrary strategy for player 1 is ε-optimal for a state $s \in W_2 \cup T$; hence, without loss of generality we assume that every state $s \in W_2 \cup T$ is an absorbing state.

Our first step towards the proof of memoryless ε-optimal strategies for reachability games consists in considering a value-iteration scheme for the computation of $\langle\langle 1 \rangle\rangle_{val}(\mathrm{Reach}(T))$. Let $[T] : S \to [0, 1]$ be the indicator function of T, defined by $[T](s) = 1$ for $s \in T$, and $[T](s) = 0$ for $s \notin T$. We then define:

$$u_0 = [T] \qquad \forall k \geq 0 : \quad u_{k+1} = Pre_1(u_k) \tag{1}$$

Note that the classical equation assigns $u_{k+1} = [T] \vee Pre_1(u_k)$, where $\vee$ is interpreted as the maximum in pointwise fashion. Since we assume that states in T are absorbing, the classical equation reduces to the simpler equation given by (1). From the monotonicity of Pre_1 it follows that $u_k \leq u_{k+1}$, that is, $Pre_1(u_k) \geq u_k$, for all $k \geq 0$. The result of [6] establishes by a combinatorial argument that $\langle\langle 1 \rangle\rangle_{val}(\mathrm{Reach}(T)) = \lim_{k \to \infty} u_k$, where the limit is interpreted in pointwise fashion. A witness for an ε-optimal strategy is constructed by letting ζ_k be a selector such that $Pre_1(u_k) = Pre_{1:\zeta_k}(u_k)$, for all $k \geq 0$, and by considering the strategy σ_k for player 1 consisting in the sequence of selectors $\zeta_k, \zeta_{k-1}, \ldots, \zeta_1, \zeta_0, \zeta_0, \zeta_0, \ldots$, where the last selector, ζ_0, is repeated forever. It is then possible to prove by induction on k that

$$\inf_{\pi_2 \in \Pi_2} \Pr^{\sigma_k, \pi_2}(\exists j \in [0..k].\Theta_j \in T) \geq u_k.$$

As the strategies σ_k, for $k \geq 0$, are not necessarily memoryless, this proof does not suffice for showing the existence of memoryless ε-optimal strategies. On the other hand, the following example shows that a memoryless strategy $\overline{\zeta}_k$ does not necessarily guarantee the value u_k.

Example 1 *Consider the* 1*-MDP shown in Fig 1. At all states except state* s_3*, the set of available moves for player 1 is singleton, and at* s_3 *the available moves for player 1 is* a *and* b*. The transition at various states is shown in the Fig 1. The objective of player 1 is to reach the state* s_0*, i.e.,* $Reach(\{s_0\})$*. Given the MDP we consider the value-iteration procedure and denote by* u_k *the valuation after* k*-iterations. We have* $u_0 = (1, 0, 0, 0, 0)$ *and hence we have* $u_1 = Pre_1(1, 0, 0, 0, 0) = (1, 0, 1/2, 0, 0)$*. Similarly iterating the* Pre_1 *operator we get* $u_2 = (1, 0, 1/2, 1/2, 0)$ *and* $u_3 = (1, 0, 1/2, 1/2, 1/2)$*. This is a fix-point and we have* $u_4 = u_3$*. Now consider the selector* ζ_k *for player 1 that chooses at state* s_3 *the action* a *with probability 1. The selector* ζ_k *is optimal w.r.t. to the valuation* u_3*. However if player 1 plays the memoryless strategy* $\overline{\zeta}_k$*, then the game visits* s_3 *and* s_4 *alternately and reaches* s_0 *with probability 0. Any memoryless strategy* $\overline{\zeta}'_k$ *for player 1 that plays action* b *at state* s_3 *with positive probability ensures that the set* $\{ s_0, s_1 \}$ *of states is reached with probability 1, and* s_0 *is reached with probability* $1/2$*; and hence is an optimal strategy.* ∎

In the example, the problem is that the strategy $\overline{\zeta}_k$ may cause player 1 to stay forever in $S \setminus (T \cup W_2)$ with positive probability. The following lemma shows that, in the cases where the strategy $\overline{\zeta}_k$ guarantees reaching $T \cup W_2$ with probability 1, then $\overline{\zeta}_k$ also guarantees the value u_k.

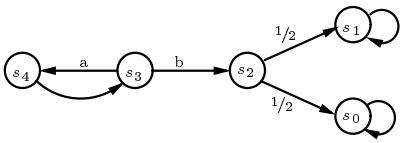


Figure 1: A MDP with reachability objective.

Lemma 1 *Let v be a valuation such that $Pre_1(v) \geq v$ and $v(s) = 0$ for all $s \in W_2$. Let ξ_1 be a selector for player 1 such that $Pre_{1:\xi_1}(v) = Pre_1(v)$. For all player 2 strategies π_2, if $\Pr^{\overline{\xi}_1,\pi_2}(Reach(T \cup W_2)) = 1$, then $\Pr^{\overline{\xi}_1,\pi_2}(Reach(T)) \geq v$.*

Proof. Consider an arbitrary $\pi_2 \in \Pi_2$, and for $k \geq 0$ let

$$v_k = \mathrm{E}^{\overline{\xi}_1,\pi_2}\big(v(\Theta_k)\big)$$

be the expected value of v after k steps under $\overline{\xi}_1$ and π_2. By induction on k, we can prove $v_k \geq v$ for all $k \geq 0$: in fact, $v_0 = v$, and for $k \geq 0$ we have $v_{k+1} \geq Pre_{1:\xi_1}(v_k) \geq Pre_{1:\xi_1}(v) = Pre_1(v) \geq v$. For all $k \geq 0$ and $s \in S$, we can write v_k as:

$$\begin{aligned} v_k(s) = {} & \mathrm{E}_s^{\overline{\xi}_1,\pi_2}\big(v(\Theta_k) \mid \Theta_k \in T\big) \cdot \Pr_s^{\overline{\xi}_1,\pi_2}\big(\Theta_k \in T\big) \\ & + \mathrm{E}_s^{\overline{\xi}_1,\pi_2}\big(v(\Theta_k) \mid \Theta_k \in S \setminus (T \cup W_2)\big) \cdot \Pr_s^{\overline{\xi}_1,\pi_2}\big(\Theta_k \in S \setminus (T \cup W_2)\big) \\ & + \mathrm{E}_s^{\overline{\xi}_1,\pi_2}\big(v(\Theta_k) \mid \Theta_k \in W_2\big) \cdot \Pr_s^{\overline{\xi}_1,\pi_2}\big(\Theta_k \in W_2\big). \end{aligned}$$

Since $v(s) \leq 1$ when $s \in T$, the first term on the right hand side is at most $\Pr_s^{\overline{\xi}_1,\pi_2}\big(\Theta_k \in T\big)$. For the second term, we have $\lim_{k\to\infty} \Pr^{\overline{\xi}_1,\pi_2}\big(\Theta_k \in S \setminus (T \cup W_2)\big) = 0$ by hypothesis, since $\Pr^{\overline{\xi}_1,\pi_2}(\mathrm{Reach}(T \cup W_2)) = 1$ and every state $s \in T \cup W_2$ is absorbing. Finally, the third term on the right hand side is 0, as $v(s) = 0$ for all $s \in W_2$. Hence, taking the limit with $k \to \infty$, we obtain

$$\Pr^{\overline{\xi}_1,\pi_2}\big(\mathrm{Reach}(T)\big) = \lim_{k\to\infty} \Pr^{\overline{\xi}_1,\pi_2}\big(\Theta_k \in T\big) \geq \lim_{k\to\infty} v_k \geq v,$$

where the last inequality follows from $v_k \geq v$ for all $k \geq 0$. ∎

3.3 From value iteration to optimal selectors

Since $\langle\langle 1 \rangle\rangle_{val}(\mathrm{Reach}(T)) = \lim_{k\to\infty} u_k$, for every $\varepsilon > 0$, there exists k, such that for all states s, we have $u_k(s) \geq u_{k-1}(s) \geq \langle\langle 1 \rangle\rangle_{val}(\mathrm{Reach}(T))(s) -$

ε. Lemma 1 indicates that, in order to construct a memoryless ε-optimal strategy, we need to construct from u_{k-1} a selector ξ_1 with the following properties:

1. $Pre_{1:\xi_1}(u_{k-1}) = Pre_1(u_{k-1}) = u_k$;
2. For all $\pi_2 \in \Pi_2$, we have $\Pr^{\overline{\xi}_1,\pi_2}(\mathrm{Reach}(T \cup W_2)) = 1$.

The first of the above conditions is easily met: it states simply that ξ_1 is an optimal selector for $Pre_1(u_{k-1})$. To meet the second condition, however, not every optimal selector suffices, as shown by Example 1.

To construct a suitable selector, we need some definitions. For $r > 0$, the *value class* $U_r^k = \{ s \in S \mid u_k(s) = r \}$, consists of the states with value r under the valuation u_k. Similarly we define $U_{\bowtie r}^k = \{ s \in S \mid u_k(s) \bowtie r \}$, for $\bowtie \in \{ <, \leq, \geq, > \}$. For a state $s \in S$, let $\ell_k(s) = \min\{ j \leq k \mid u_j(s) = u_k(s) \}$ be the *entry time* of s in $U_{u_k(s)}^k$, i.e., the least iteration j in which the state s has the same value as in iteration k. For $k \geq 0$, we define the selector η_k as follows:

$$\eta_k(s) = \eta_{\ell_k(s)} = \arg \sup_{\xi_1 \in \Lambda_1} \big[\inf_{\xi_2 \in \Lambda_2} Pre_{\xi_1,\xi_2}(u_{\ell_k(s)-1}) \big].$$

In words, the selector $\eta_k(s)$ is an optimal selector for s at the iteration $\ell_k(s)$. It follows easily that $u_k = Pre_{1:\eta_k}(u_{k-1})$. We denote by $\overline{\eta}_k^\omega$ the memoryless player-1 strategy that always follows η_k. Once we fix the selector η_k, the game is equivalent to a 2-MDP G_{η_k}, and we can analyze its behavior with the help of Corollary 1; the goal is to prove the second condition, i.e., that for all $\pi_2 \in \Pi_2$ we have $\Pr^{\overline{\eta}_k,\pi_2}(\mathrm{Reach}(T \cup W_2)) = 1$. To reason about the end-components of G_{η_k}, for a state $s \in S$, and a player-2 action $b \in \Gamma_2(s)$, we denote by

$$Dest_k(s,b) = \{ Dest(s,a,b) \mid a \in \Gamma_1(s) \wedge \eta_k(a) \geq 0 \}$$

the set of possible successors of state s when player 1 plays according to η_k, and player 2 plays according to b.

Lemma 2 *For all $k \geq 0$, consider a state $s \in S \setminus (T \cup W_2)$, and let $s \in U_r^k$, for $0 < r < 1$. For all moves $b \in \Gamma_2(s)$, we have:*

1. *either $Dest_k(s,b) \cap U_{>r}^k \neq \emptyset$,*
2. *or $Dest_k(s,b) \subseteq U_r^k$, and there is $t \in Dest_k(s,b)$ with $\ell_k(t) < \ell_k(s)$.*

Proof. For convenience, let $m = \ell_k(s)$, and consider any $b \in \Gamma_2(s)$.

- Consider first the case in which $Dest_k(s,b) \not\subseteq U^k_r$. Then, it cannot be $Dest_k(s,b) \subseteq U^k_{\leq r}$: otherwise, for all states $t \in Dest_k(s,b)$ we would have $u_k(t) \leq r$, and there would be at least one $t \in Dest_k(s,b)$ such that $u_k(t) < r$, contradicting $u_k(s) = r$ and $Pre_{1:\eta_k}(u_{k-1}) = u_k$. So, it must be $Dest_k(s,b) \cap U^k_{>r} \neq \emptyset$.
- Consider now the case in which $Dest_k(s,b) \subseteq U^k_r$. Since $u_m \leq u_k$, due to the monotonicity of the Pre_1 operator and (1), we have that $u_{m-1}(t) \leq r$ for all $t \in Dest_k(s,b)$. From $r = u_k(s) = u_m(s) = Pre_{1:\eta_k}(u_{m-1})$, we have that $u_{m-1}(t) = r$ for all $t \in Dest_k(s,b)$, implying that $\ell_k(t) < m$ for all $t \in Dest_k(s,b)$. ∎

The above lemma states that under η_k, from each state $i \in U^k_r$ we are guaranteed a probability bounded away from 0 of either moving to a higher-value class $U^k_{>r}$, or of moving to states within the value class that have a strictly lower entry time. This implies that every state in $S \setminus W_2$ has a probability bounded above zero of reaching T in at most $n = |S|$ steps, so that the probability of staying forever in $S \setminus (T \cup W_2)$ is 0. To prove this fact formally, we analyze the end components of G_{η_k} in light of Lemma 2.

Lemma 3 *For $k \geq 0$, if for all $s \in S \setminus W_2$ we have $u_{k-1}(s) > 0$, then for all $\pi_2 \in \Pi_2$, we have $\Pr^{\eta_k,\pi_2}\big(Reach(T \cup W_2)\big) = 1$.*

Proof. Since every state $s \in T \cup W_2$ is absorbing, to prove this result, in view of Corollary 1, it suffices to show that there is no end component of G_{η_k} entirely contained in $S \setminus (T \cup W_2)$. Towards the contradiction, assume there is such an end component $C \subseteq (S \setminus T \cup W_2)$; then, we have $C \subseteq U^k_{[r_1,r_2]}$ with $C \cap U_{r_2} \neq \emptyset$, for some $0 < r_1 \leq r_2 \leq 1$, where $U^k_{[r_1,r_2]} = U^k_{\geq r_1} \cap U^k_{\leq r_2}$ is the union of the value classes for values in the interval $[r_1, r_2]$. Consider a state $s \in U^k_{r_2}$ with minimal ℓ_k, i.e., such that $\ell_k(s) \leq \ell_k(t)$ for all other $t \in U^k_{r_2}$. From Lemma 2, we are guaranteed that for any $b \in \Gamma_2(s)$, there is $t \in Dest_k(s,b)$ such that (i) either $t \in U^k_{r_2}$ and $\ell_k(t) < \ell_k(s)$, (ii) or $t \in U^k_{>r_2}$. In both cases, we reach a contradiction. ∎

The above lemma shows that η_k satisfies both the requirements for optimal selectors spelt out at the beginning of Section 3.3: hence, η_k guarantees value u_k. This proves the existence of memoryless ε-optimal strategies for concurrent reachability games.

Theorem 2 (Existence of memoryless ε-optimal strategies) *For every $\varepsilon > 0$, memoryless ε-optimal strategies exist for all concurrent games with reachability objectives.*

Proof. Consider a reachability game with target $T \subseteq S$. Since $\lim_{k\to\infty} u_k = \langle\langle 1\rangle\rangle_{val}(\mathrm{Reach}(T))$, for every $\varepsilon > 0$ we can find $k \in \mathbb{N}$ such that $\max_{s\in S}\big(\langle\langle 1\rangle\rangle_{val}(\mathrm{Reach}(T))(s) - u_{k-1}(s)\big) < \varepsilon$. By construction, $Pre_{1:\eta_k}(u_{k-1}) = Pre_1(u_{k-1}) = u_k$. Hence, from Lemmas 1 and 3, for all $\pi_2 \in \Pi_2$ we have $\Pr^{\overline{\eta}_k,\pi_2}(\mathrm{Reach}(T)) \geq u_{k-1}$, leading to the result. ∎

References

[1] D.P. Bertsekas. *Dynamic Programming and Optimal Control.* Athena Scientific, 1995. Volumes I and II.

[2] C. Courcoubetis and M. Yannakakis. The complexity of probabilistic verification. *Journal of the ACM*, 42(4):857–907, 1995.

[3] L. de Alfaro. *Formal Verification of Probabilistic Systems.* PhD thesis, Stanford University, 1997. Technical Report STAN-CS-TR-98-1601.

[4] L. de Alfaro and T.A. Henzinger. Concurrent omega-regular games. In *Proc. 15th IEEE Symp. Logic in Comp. Sci.*, pages 141–154, 2000.

[5] L. de Alfaro, T.A. Henzinger, and O. Kupferman. Concurrent reachability games. In *Proc. 39th IEEE Symp. Found. of Comp. Sci.*, pages 564–575. IEEE Computer Society Press, 1998.

[6] L. de Alfaro and R. Majumdar. Quantitative solution of omega-regular games. *Journal of Computer and System Sciences*, 68:374–397, 2004. A preliminary version appeared in STOC 01: 33rd Annual ACM Symposium on Theory of Computing, 2001.

[7] H. Everett. Recursive games. In *Contributions to the Theory of Games III*, volume 39 of *Annals of Mathematical Studies*, pages 47–78, 1957.

[8] J. Filar and K. Vrieze. *Competitive Markov Decision Processes.* Springer-Verlag, 1997.

[9] J.G. Kemeny, J.L. Snell, and A.W. Knapp. *Denumerable Markov Chains.* D. Van Nostrand Company, 1966.

[10] D.A. Martin. The determinacy of Blackwell games. *The Journal of Symbolic Logic*, 63(4):1565–1581, 1998.

[11] L.S. Shapley. Stochastic games. *Proc. Nat. Acad. Sci. USA*, 39:1095–1100, 1953.

[12] M.Y. Vardi. Automatic verification of probabilistic concurrent finite-state systems. In *Proceedings of the 26th Annual Symposium on Foundations of Computer Science*, pages 327–338. IEEE Computer Society Press, 1985.