Journal of Computer and System Sciences 79 (2013) 640–657

Contents lists available at SciVerse ScienceDirect

Journal of Computer and System Sciences

www.elsevier.com/locate/jcss

Strategy improvement for concurrent reachability and turn-based stochastic safety games [☆,☆☆]

Krishnendu Chatterjee [a,*], Luca de Alfaro [b], Thomas A. Henzinger [a]

[a] IST Austria (Institute of Science and Technology Austria), Austria
[b] University of California, Santa Cruz, United States

ARTICLE INFO

Article history:
Received 20 April 2010
Received in revised form 3 January 2012
Accepted 12 December 2012
Available online 19 December 2012

Keywords:
Game theory
Stochastic games
Concurrent games
Reachability and safety objectives
Strategy-improvement algorithms

ABSTRACT

We consider concurrent games played on graphs. At every round of a game, each player simultaneously and independently selects a move; the moves jointly determine the transition to a successor state. Two basic objectives are the safety objective to stay forever in a given set of states, and its dual, the reachability objective to reach a given set of states. First, we present a simple proof of the fact that in concurrent reachability games, 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, our proof is more elementary and more combinatorial. Second, we present a strategy-improvement (a.k.a. policy-iteration) algorithm for concurrent games with reachability objectives. Finally, we present a strategy-improvement algorithm for turn-based stochastic games (where each player selects moves in turns) with safety objectives. Our algorithms yield sequences of player-1 strategies which ensure probabilities of winning that converge monotonically (from below) to the value of the game.

1. Introduction

We consider games played between two players on graphs. At every round of the game, each of the two players selects a move; the moves of the players then determine the transition to the successor state. A play of the game gives rise to a path in the graph. We consider the two basic objectives for the players: *reachability* and *safety*. The reachability goal asks player 1 to reach a given set of target states or, if randomization is needed to play the game, to maximize the probability of reaching the target set. The safety goal asks player 2 to ensure that a given set of safe states is never left or, if randomization is required, to minimize the probability of leaving the target set. The two objectives are dual, and the games are determined: the supremum probability with which player 1 can reach the target set is equal to one minus the supremum probability with which player 2 can confine the game to the complement of the target set [14].

These games on graphs can be divided into two classes: *turn-based* and *concurrent*. In turn-based games, only one player has a choice of moves at each state; in concurrent games, at each state both players choose a move, simultaneously and

[☆] This work was partially supported in part by the NSF grants CCR-0132780, CNS-0720884, CCR-0225610, by the Swiss National Science Foundation, ERC Start Grant Graph Games (Project No. 279307), FWF NFN Grant S11407-N23 (RiSE), and a Microsoft faculty fellowship.

[☆☆] This paper is an improved version of Chatterjee et al. (2006, 2009) [4,3]. There is a serious and irreparable error in Theorem 4.3 of Chatterjee et al. (2009) [3] regarding the convergence property of the improvement algorithm for concurrent safety games. This is illustrated in an example of Chatterjee et al. (2012) [2]. In the present version the result is presented for turn-based stochastic safety games. We thank anonymous reviewers for many insightful comments that helped us immensely, and warmly acknowledge their help.

[*] Corresponding author.
E-mail addresses: krish.chat@gmail.com, krish.chat@ist.ac.at (K. Chatterjee).

0022-0000 © 2012 Elsevier Inc. Open access under CC BY-NC-ND license.
http://dx.doi.org/10.1016/j.jcss.2012.12.001

independently, from a set of available moves. For turn-based games, the solution of games with reachability and safety objectives has long been known. If each move determines a unique successor state, then the games are P-complete and can be solved in linear time in the size of the game graph. If, more generally, each move determines a probability distribution on possible successor states (called turn-based stochastic games or simple stochastic games), then the problem of deciding whether a turn-based game can be won with probability greater than a given threshold $p \in [0, 1]$ is in NP ∩ co-NP [5], and the exact value of the game can be computed by a strategy-improvement algorithm for reachability objectives [6], which works well in practice. These results all depend on the fact that in turn-based reachability and safety games, both players have optimal deterministic (i.e., no randomization is required), memoryless strategies. These strategies are functions from states to moves, so they are finite in number, and this guarantees the termination of the strategy-improvement algorithm for reachability objectives.

The situation is very different for concurrent games. The player-1 *value* of the game is defined, as usual, as the sup–inf value: the supremum, over all strategies of player 1, of the infimum, over all strategies of player 2, of the probability of achieving the reachability or safety goal. In concurrent reachability games, player 1 is guaranteed only the existence of ε-optimal strategies, which ensure that the value of the game is achieved within a specified tolerance $\varepsilon > 0$ [14]. Moreover, while these strategies (which depend on ε) are memoryless, in general they require randomization [14] (even in the special case in which the transition function is deterministic). For player 2 (the safety player), *optimal* memoryless strategies exist [22], which again require randomization (even when the transition function is deterministic). All of these strategies are functions from states to probability distributions on moves. The question of deciding whether a concurrent game can be won with probability greater than p is in PSPACE; this is shown by reduction to the theory of the real-closed fields [13].

To summarize: while strategy-improvement algorithms are available for turn-based stochastic reachability games [6], so far no strategy-improvement algorithms were known for concurrent reachability games. For turn-based stochastic safety games, one could apply the strategy-improvement algorithm for turn-based stochastic reachability games, however there were no strategy-improvement algorithm for turn-based stochastic safety games that converges from below to the value of the game and yields a sequence of improving strategies that converges to an optimal strategy.

Our results for concurrent reachability games. Concurrent reachability games belong to the family of stochastic games [24,14], and they have been studied more specifically in [10,9,11]. Our contributions for concurrent reachability games are two-fold. First, we present a simple and combinatorial proof of the existence of memoryless ε-optimal strategies for concurrent games with reachability objectives, for all $\varepsilon > 0$. Second, using the proof techniques we developed for proving existence of memoryless ε-optimal strategies, for $\varepsilon > 0$, we obtain a strategy-improvement (a.k.a. policy-iteration) algorithm for concurrent reachability games. Unlike in the special case of turn-based games the algorithm need not terminate in finitely many iterations.

It has long been known that optimal strategies need not exist for concurrent reachability games, and for all $\varepsilon > 0$, there exist ε-optimal strategies that are memoryless [14]. A proof of this fact can be obtained by considering limit of discounted games. The proof considers *discounted* versions of reachability games, where a play that reaches the target in k steps is assigned a value of α^k, for some discount factor $0 < \alpha \leqslant 1$. It is possible to show that, for $0 < \alpha < 1$, memoryless optimal strategies exist. The result for the undiscounted ($\alpha = 1$) case followed 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 from the field of real Puiseiux series [21]. This proof idea works not only for reachability games, but also for total-reward games with nonnegative rewards (see [15, Chapter 5] for details). A more recent result [13] establishes the existence of memoryless ε-optimal strategies for certain infinite-state (recursive) concurrent games, but again the proof relies on results from analysis and properties of solutions of certain polynomial functions. Another proof of existence of memoryless ε-optimal strategies for reachability objectives follows from the result of [14] and the proof uses induction on the number of states of the game. We show the existence of memoryless ε-optimal strategies for concurrent reachability games by more combinatorial and elementary means. Our proof relies only on combinatorial techniques and on simple properties of Markov decision processes [1,8]. As our proof is more combinatorial, we believe that the proof techniques will find future applications in game theory.

Our proof of the existence of memoryless ε-optimal strategies, for all $\varepsilon > 0$, is built upon a value-iteration scheme that converges to the value of the game [11]. The value-iteration scheme computes a sequence $u_0, u_1, u_2, \ldots$ of valuations, where for $i = 0, 1, 2, \ldots$ each valuation u_i associates with each state s of the game a lower bound $u_i(s)$ on the value of the game, such that $\lim_{i\to\infty} u_i(s)$ converges to the value of the game at s. The convergence is monotonic from below, but no rate of convergence was known. From each valuation u_i, we can extract a memoryless, randomized player-1 strategy, by considering the (randomized) choice of moves for player 1 that achieves the maximal one-step expectation of u_i. In general, a strategy π_i obtained in this fashion is not guaranteed to achieve the value u_i. We show that π_i is guaranteed to achieve the value u_i if it is *proper*, that is, if regardless of the strategy adopted by player 2, the play reaches with probability 1 states that are either in the target, or that have no path leading to the target. Next, we show how to extract from the sequence of valuations $u_0, u_1, u_2, \ldots$ a sequence of memoryless randomized player-1 strategies $\pi_0, \pi_1, \pi_2, \ldots$ that are guaranteed to be proper, and thus achieve the values $u_0, u_1, u_2, \ldots$. This proves the existence of memoryless ε-optimal strategies for all $\varepsilon > 0$. Our proof is completely different as compared to the proof of [14]: the proof of [14] uses induction on the number of states, whereas our proof is based on the notion of ranking function obtained from the value-iteration algorithm.

We then apply the techniques developed for the above proof to design a *strategy-improvement* algorithm for concurrent reachability games. Strategy-improvement algorithms, also known as *policy-iteration* algorithms in the context of Markov decision processes [18], compute a sequence of memoryless strategies $\pi'_0, \pi'_1, \pi'_2, \ldots$ such that, for all $k \geqslant 0$, (i) the strategy π'_{k+1} is at all states no worse than π'_k; (ii) if $\pi'_{k+1} = \pi'_k$, then π_k is optimal; and (iii) for every $\varepsilon > 0$, we can find a k sufficiently large so that π'_k is ε-optimal. Computing a sequence of strategies $\pi_0, \pi_1, \pi_2, \ldots$ on the basis the value-iteration scheme from above does not yield a strategy-improvement algorithm, as condition (ii) may be violated: there is no guarantee that a step in the value iteration leads to an improvement in the strategy. We will show that the key to obtain a strategy-improvement algorithm consists in recomputing, at each iteration, the values of the player-1 strategy to be improved, and in adopting a particular strategy-update rule, which ensures that all generated strategies are proper. Unlike previous proofs of strategy-improvement algorithms for concurrent games [6,15], which rely on the analysis of discounted versions of the games, our analysis is again more combinatorial. Hoffman and Karp [17] presented a strategy-improvement algorithm for the special case of concurrent games with ergodic property (i.e., from every state s any other state t can be guaranteed to reach with probability 1) (also see algorithm for discounted games in [23]). Observe that for concurrent reachability games, with the ergodic assumption the value at all states is trivially 1, and thus the ergodic assumption gives us the trivial case. Our results give a combinatorial strategy-improvement algorithm for the whole class of concurrent reachability games. The results of [13] present a strategy-improvement algorithm for recursive concurrent games with termination criteria: the algorithm of [13] is more involved (depends on properties of certain polynomial functions) and works for the more general class of recursive concurrent games. Differently from turn-based games [6], for concurrent games we cannot guarantee the termination of the strategy-improvement algorithm. However, for turn-based stochastic games we present a detailed analysis of termination criteria. Our analysis is based on bounds on the precision of values for turn-based stochastic games. As a consequence of our analysis, we obtain an improved upper bound for termination for turn-based stochastic games.

Our results for turn-based stochastic safety games. We present a strategy-improvement scheme that computes the value of a turn-based stochastic safety game, and the valuations computed monotonically converge *from below* to the value of the game. The strategy-improvement algorithm for reachability objectives is based on locally improving a strategy on the basis of the valuation it yields, and this approach does not suffice for safety objectives: we would obtain an increasing sequence of values, but they would not necessarily converge to the value of the game (see Example 2). Rather, we introduce a novel, non-local improvement step, which augments the standard valuation-based improvement step. Each non-local step involves the solution of the set of almost-sure winning states of an appropriately constructed turn-based game. The turn-based game constructed is polynomial in the state space of the original game. We show that the strategy-improvement algorithm with local and non-local improvement steps yields a monotonically increasing sequence of valuations that converge to the value of the game.

This paper is an improved version of Chatterjee et al. [4,3].

2. Definitions

Notation. For a countable set A, a *probability distribution* on A is a function $\delta : A \to [0,1]$ 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 $\mathit{Supp}(\delta) = \{x \in A \mid \delta(x) > 0\}$ the support set of δ.

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

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

We denote by $|\delta|$ the size of transition function, i.e., $|\delta| = \sum_{s\in S,\, a\in\Gamma_1(s),\, b\in\Gamma_2(s),\, t\in S} |\delta(s,a,b)(t)|$, where $|\delta(s,a,b)(t)|$ is the number of bits required to specify the transition probability $\delta(s,a,b)(t)$. We denote by $|G|$ the size of the game graph, and $|G| = |\delta| + |S|$. 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 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 an absorbing state s for all choices of moves of the two players, the successor state is always s.

Definition 2 *(Turn-based stochastic games).* A *turn-based stochastic game graph* ($2\frac{1}{2}$*-player game graph*) $G = \langle (S, E), (S_1, S_2, S_R), \delta\rangle$ consists of a finite directed graph (S, E), a partition (S_1, S_2, S_R) of the finite set S of states, and a probabilistic transition function $\delta : S_R \to \mathcal{D}(S)$, where $\mathcal{D}(S)$ denotes the set of probability distributions over the state space S. The states in S_1 are the *player*-1 states, where player 1 decides the successor state; the states in S_2 are the *player*-2 states, where player 2 decides the successor state; and the states in S_R are the *random or probabilistic* states, where the successor state is chosen according to the probabilistic transition function δ. We assume that for $s \in S_R$ and $t \in S$, we have $(s,t) \in E$ iff

$\delta(s)(t) > 0$, and we often write $\delta(s,t)$ for $\delta(s)(t)$. For technical convenience we assume that every state in the graph (S,E) has at least one outgoing edge. For a state $s \in S$, we write $E(s)$ to denote the set $\{t \in S \mid (s,t) \in E\}$ of possible successors. We denote by $|\delta|$ the size of the transition function, i.e., $|\delta| = \sum_{s \in S_R, t \in S} |\delta(s)(t)|$, where $|\delta(s)(t)|$ is the number of bits required to specify the transition probability $\delta(s)(t)$. We denote by $|G|$ the size of the game graph, and $|G| = |\delta| + |S| + |E|$.

Plays. 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 $k \geqslant 0$, 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 plays, and by Ω_s the set of all plays $\omega = \langle s_0, s_1, s_2, \ldots \rangle$ such that $s_0 = s$, that is, the set of plays starting from state s.

Selectors and strategies. A *selector* ξ for player $i \in \{1,2\}$ is a function $\xi : S \to \mathcal{D}(M)$ such that for all states $s \in S$ and moves $a \in M$, if $\xi(s)(a) > 0$, then $a \in \Gamma_i(s)$. A selector ξ for player i at a state s is a distribution over moves such that if $\xi(s)(a) > 0$, then $a \in \Gamma_i(s)$. We denote by Λ_i the set of all selectors for player $i \in \{1,2\}$, and similarly, we denote by $\Lambda_i(s)$ the set of all selectors for player i at a state s. The selector ξ is *pure* if for every state $s \in S$, there is a move $a \in M$ such that $\xi(s)(a) = 1$. A *strategy* for player $i \in \{1,2\}$ is a function $\pi : S^+ \to \mathcal{D}(M)$ that associates with every finite, nonempty sequence of states, representing the history of the play so far, a selector for player i; that is, for all $w \in S^*$ and $s \in S$, we have $Supp(\pi(w \cdot s)) \subseteq \Gamma_i(s)$. The strategy π is *pure* if it always chooses a pure selector; that is, for all $w \in S^+$, there is a move $a \in M$ such that $\pi(w)(a) = 1$. 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 $\bar{\xi}$ for the memoryless strategy consisting in playing forever the selector ξ. A strategy is *pure memoryless* if it is both pure and memoryless. In a turn-based stochastic game, a strategy for player 1 is a function $\pi_1 : S^* \cdot S_1 \to \mathcal{D}(S)$, such that for all $w \in S^*$ and for all $s \in S_1$ we have $Supp(\pi_1(w \cdot s)) \subseteq E(s)$. Memoryless strategies and pure memoryless strategies are obtained as the restriction of strategies as in the case of concurrent game graphs. The family of strategies for player 2 are defined analogously. We denote by Π_1 and Π_2 the sets of all strategies for player 1 and player 2, respectively. We denote by Π_i^M and Π_i^{PM} the sets of memoryless strategies and pure memoryless strategies for player i, respectively.

Destinations of moves and selectors. 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 the moves a_1 and a_2 are chosen. Given a state s, and selectors ξ_1 and ξ_2 for the two players, we denote by

$$Dest(s, \xi_1, \xi_2) = \bigcup_{\substack{a_1 \in Supp(\xi_1(s)),\\ a_2 \in Supp(\xi_2(s))}} Dest(s, a_1, a_2)$$

the set of possible successors of s with respect to the selectors ξ_1 and ξ_2.

Once a starting state s and strategies π_1 and π_2 for the two players are fixed, 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 plays. For an event $\mathcal{A} \subseteq \Omega_s$, we denote by $\Pr_s^{\pi_1,\pi_2}(\mathcal{A})$ the probability that a play belongs to $\mathcal{A}$ when the game starts from s and the players follow 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 follow the strategies π_1 and π_2. For $i \geqslant 0$, we denote by $\Theta_i : \Omega \to S$ the random variable denoting the i-th state along a play.

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 $v \leqslant w$ when $v(s) \leqslant w(s)$ for all states $s \in S$. For an event $\mathcal{A}$, we denote by $\Pr^{\pi_1,\pi_2}(\mathcal{A})$ the valuation $S \to [0,1]$ defined for all states $s \in S$ by $(\Pr^{\pi_1,\pi_2}(\mathcal{A}))(s) = \Pr_s^{\pi_1,\pi_2}(\mathcal{A})$. 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 $(\mathrm{E}^{\pi_1,\pi_2}(f))(s) = \mathrm{E}_s^{\pi_1,\pi_2}(f)$.

The *Pre* operator. 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 states $s \in S$:

$$Pre_{\xi_1,\xi_2}(v)(s) = \sum_{a,b \in M} \sum_{t \in S} v(t) \cdot \delta(s,a,b)(t) \cdot \xi_1(s)(a) \cdot \xi_2(s)(b),$$

$$Pre_{1:\xi_1}(v)(s) = \inf_{\xi_2 \in \Lambda_2} Pre_{\xi_1,\xi_2}(v)(s),$$

$$Pre_1(v)(s) = \sup_{\xi_1 \in \Lambda_1} \inf_{\xi_2 \in \Lambda_2} Pre_{\xi_1,\xi_2}(v)(s).$$

Intuitively, $Pre_1(v)(s)$ is the greatest expectation of v that player 1 can guarantee at a successor state of s. Also note that given a valuation v, the computation of $Pre_1(v)$ reduces to the solution of a zero-sum one-shot matrix game, and can be solved by linear programming. Similarly, $Pre_{1:\xi_1}(v)(s)$ is the greatest expectation of v that player 1 can guarantee at a successor state of s by playing the selector ξ_1. Note that all of these operators on valuations are monotonic: for two valuations v, w, if $v \leqslant w$, then for all selectors $\xi_1 \in \Lambda_1$ and $\xi_2 \in \Lambda_2$, we have $Pre_{\xi_1,\xi_2}(v) \leqslant Pre_{\xi_1,\xi_2}(w)$, $Pre_{1:\xi_1}(v) \leqslant Pre_{1:\xi_1}(w)$, and $Pre_1(v) \leqslant Pre_1(w)$.

Reachability and safety objectives. Given a set $F \subseteq S$ of *safe* states, the objective of a safety game consists in never leaving F. Therefore, we define the set of winning plays as the set $\mathrm{Safe}(F) = \{\langle s_0, s_1, s_2, \ldots\rangle \in \Omega \mid s_k \in F \text{ for all } k \geqslant 0\}$. Given a subset $T \subseteq S$ of *target* states, the objective of a reachability game consists in reaching T. Correspondingly, the set winning plays is $\mathrm{Reach}(T) = \{\langle s_0, s_1, s_2, \ldots\rangle \in \Omega \mid s_k \in T \text{ for some } k \geqslant 0\}$ of plays that visit T. For all $F \subseteq S$ and $T \subseteq S$, the sets $\mathrm{Safe}(F)$ and $\mathrm{Reach}(T)$ are measurable. An objective in general is a measurable set, and in this paper we consider only reachability and safety objectives. For an objective Φ, the probability of satisfying Φ from a state $s \in S$ under strategies π_1 and π_2 for players 1 and 2, respectively, is $\Pr_s^{\pi_1,\pi_2}(\Phi)$. We define the *value* for player 1 of the game with objective Φ from the state $s \in S$ as

$$\langle\langle 1\rangle\rangle_{\mathsf{val}}(\Phi)(s) = \sup_{\pi_1 \in \Pi_1} \inf_{\pi_2 \in \Pi_2} \Pr_s^{\pi_1,\pi_2}(\Phi);$$

i.e., the value is the maximal probability with which player 1 can guarantee the satisfaction of Φ against all player-2 strategies. Given a player-1 strategy π_1, we use the notation

$$\langle\langle 1\rangle\rangle_{\mathsf{val}}^{\pi_1}(\Phi)(s) = \inf_{\pi_2 \in \Pi_2} \Pr_s^{\pi_1,\pi_2}(\Phi).$$

A strategy π_1 for player 1 is *optimal* for an objective Φ if for all states $s \in S$, we have

$$\langle\langle 1\rangle\rangle_{\mathsf{val}}^{\pi_1}(\Phi)(s) = \langle\langle 1\rangle\rangle_{\mathsf{val}}(\Phi)(s).$$

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

$$\langle\langle 1\rangle\rangle_{\mathsf{val}}^{\pi_1}(\Phi)(s) \geqslant \langle\langle 1\rangle\rangle_{\mathsf{val}}(\Phi)(s) - \varepsilon.$$

The notion of values and optimal strategies for player 2 are defined analogously. Reachability and safety objectives are dual, i.e., we have $\mathrm{Reach}(T) = \Omega \setminus \mathrm{Safe}(S \setminus T)$. The quantitative determinacy result of [14] ensures that for all states $s \in S$, we have

$$\langle\langle 1\rangle\rangle_{\mathsf{val}}\big(\mathrm{Safe}(F)\big)(s) + \langle\langle 2\rangle\rangle_{\mathsf{val}}\big(\mathrm{Reach}(S \setminus F)\big)(s) = 1.$$

3. Markov decision processes

To develop our arguments, we need some facts about one-player versions of concurrent stochastic games, known as *Markov decision processes* (MDPs) [12,1]. For $i \in \{1, 2\}$, a *player-i MDP* (for short, i-MDP) is a concurrent game where, for all states $s \in S$, we have $|\Gamma_{3-i}(s)| = 1$. Given a concurrent game G, if we fix a memoryless strategy corresponding to selector ξ_1 for player 1, the game is equivalent to a 2-MDP G_{ξ_1} with the transition function

$$\delta_{\xi_1}(s, a_2)(t) = \sum_{a_1 \in \Gamma_1(s)} \delta(s, a_1, a_2)(t) \cdot \xi_1(s)(a_1),$$

for all $s \in S$ and $a_2 \in \Gamma_2(s)$. Similarly, if we fix selectors ξ_1 and ξ_2 for both players in a concurrent game G, we obtain a Markov chain, which we denote by G_{ξ_1,ξ_2}.

End components. In an MDP, the sets of states that play an equivalent role to the closed recurrent classes of Markov chains [20, Chapter 4] are called “end components” [7,8].

Definition 3 *(End components).* An *end component* of an i-MDP G, for $i \in \{1, 2\}$, is a subset $C \subseteq S$ of the states such that there is a selector ξ for player i so that C is a closed recurrent class of the Markov chain G_ξ.

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

(i) (*closed*) if a move in $M_i(s)$ is chosen by player i at state s, then all successor states that are obtained with nonzero probability lie in C; and
(ii) (*recurrent*) the graph (C, E), where E consists of the transitions that occur with nonzero probability when moves in $M_i(\cdot)$ are chosen by player i, is strongly connected.

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

Theorem 1. *(See [8].) For a player-*1 *selector ξ_1, let $\mathcal{C}$ be the set of end components of a* 2*-MDP G_{ξ_1}. For all player-*2 *strategies π_2 and all states $s \in S$, we have* $\Pr_s^{\bar{\xi}_1,\pi_2}(\mathrm{Inf}(\mathcal{C})) = 1$.

Corollary 1. *For a player-1 selector ξ_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 player-2 strategies π_2 and all states $s \in S$, we have $\Pr_s^{\bar{\xi}_1, \pi_2}(\text{Reach}(Z)) = 1$.*

3.0.0.1. MDPs with reachability objectives Given a 2-MDP with a reachability objective Reach(T) for player 2, where $T \subseteq S$, the values can be obtained as the solution of a linear program [15] (see Section 2.9 of [15] where linear program solution is given for MDPs with limit-average objectives and reachability objective is a special case of limit-average objectives). The linear program has a variable $x(s)$ for all states $s \in S$, and the objective function and the constraints are as follows:

$$\begin{aligned}
&\min \sum_{s \in S} x(s) \quad \text{subject to} \\
&x(s) \geqslant \sum_{t \in S} x(t) \cdot \delta(s, a_2)(t) \quad \text{for all } s \in S \text{ and } a_2 \in \Gamma_2(s), \\
&x(s) = 1 \quad \text{for all } s \in T, \\
&0 \leqslant x(s) \leqslant 1 \quad \text{for all } s \in S.
\end{aligned}$$

The correctness of the above linear program to compute the values follows from [15] (see Section 2.9 of [15], and also see [7] for the correctness of the linear program).

4. Existence of memoryless ε-optimal strategies for concurrent reachability games

In this section we present an elementary and combinatorial proof of the existence of memoryless ε-optimal strategies for concurrent reachability games, for all $\varepsilon > 0$ (optimal strategies need not exist for concurrent games with reachability objectives [14]).

4.1. From value iteration to selectors

Consider a reachability game with target $T \subseteq S$, i.e., objective for player 1 is Reach(T). Let $W_2 = \{s \in S \mid \langle\langle 1 \rangle\rangle_{\text{val}}(\text{Reach}(T))(s) = 0\}$ be the set of states from which player 1 cannot reach the target with positive probability. From [9], we know that this set can be computed as $W_2 = \lim_{k \to \infty} W_2^k$, where $W_2^0 = S \setminus T$, and for all $k \geqslant 0$,

$$W_2^{k+1} = \bigl\{ s \in S \setminus T \bigm| \exists a_2 \in \Gamma_2(s) \,.\, \forall a_1 \in \Gamma_1(s) \,.\, \mathit{Dest}(s, a_1, a_2) \subseteq W_2^k \bigr\}.$$

The limit is reached in at most $|S|$ iterations. Note that player 2 has a strategy that confines the game to W_2, and that consequently all strategies are optimal for player 1, as they realize the value 0 of the game in W_2. Therefore, without loss of generality, in the remainder we assume that all states in W_2 and T are absorbing.

Our first step towards proving the existence of memoryless ε-optimal strategies for reachability games consists in considering a value-iteration scheme for the computation of $\langle\langle 1 \rangle\rangle_{\text{val}}(\text{Reach}(T))$. 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 $s \notin T$. Let $u_0 = [T]$, and for all $k \geqslant 0$, let

$$u_{k+1} = \mathit{Pre}_1(u_k). \tag{1}$$

Note that the classical equation assigns $u_{k+1} = [T] \vee \mathit{Pre}_1(u_k)$, where $\vee$ is interpreted as the maximum in pointwise fashion. Since we assume that all 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 $u_k \leqslant u_{k+1}$, that is, $\mathit{Pre}_1(u_k) \geqslant u_k$, for all $k \geqslant 0$. The result of [11] establishes by a combinatorial argument that $\langle\langle 1 \rangle\rangle_{\text{val}}(\text{Reach}(T)) = \lim_{k \to \infty} u_k$, where the limit is interpreted in pointwise fashion. For all $k \geqslant 0$, let the player-1 selector ζ_k be a *value-optimal* selector for u_k, that is, a selector such that $\mathit{Pre}_1(u_k) = \mathit{Pre}_{1:\zeta_k}(u_k)$. An ε-optimal strategy π_1^k for player 1 can be constructed by applying the sequence $\zeta_k, \zeta_{k-1}, \ldots, \zeta_1, \zeta_0, \zeta_0, \zeta_0, \ldots$ of selectors, where the last selector, ζ_0, is repeated forever. It is possible to prove by induction on k that

$$\inf_{\pi_2 \in \Pi_2} \Pr^{\pi_1^k, \pi_2}\bigl(\exists j \in [0..k] \,.\, \Theta_j \in T\bigr) \geqslant u_k.$$

As the strategies π_1^k, for $k \geqslant 0$, 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 the memoryless strategy $\bar{\zeta}_k$ does not necessarily guarantee the value u_k.

Example 1. Consider the 1-MDP shown in Fig. 1. At all states except s_3, the set of available moves for player 1 is a singleton set. At s_3, the available moves for player 1 are a and b. The transitions at the various states are shown in the figure. The objective of player 1 is to reach the state s_0.

We consider the value-iteration procedure and denote by u_k the valuation after k iterations. Writing a valuation u as the list of values $(u(s_0), u(s_1), \ldots, u(s_4))$, we have

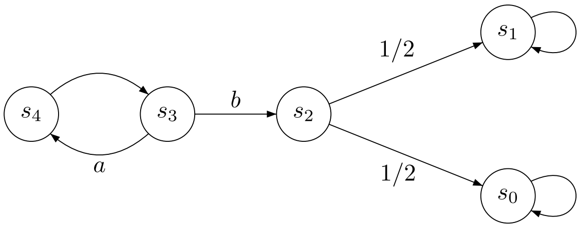

Fig. 1. An MDP with reachability objective.

$$u_0 = (1, 0, 0, 0, 0),$$
$$u_1 = Pre_1(u_0) = \left(1, 0, \frac{1}{2}, 0, 0\right),$$
$$u_2 = Pre_1(u_1) = \left(1, 0, \frac{1}{2}, \frac{1}{2}, 0\right),$$
$$u_3 = Pre_1(u_2) = \left(1, 0, \frac{1}{2}, \frac{1}{2}, \frac{1}{2}\right),$$
$$u_4 = Pre_1(u_3) = u_3 = \left(1, 0, \frac{1}{2}, \frac{1}{2}, \frac{1}{2}\right).$$

The valuation u_3 is thus a fixpoint.

Now consider the selector ξ_1 for player 1 that chooses at state s_3 the move a with probability 1. The selector ξ_1 is optimal with respect to the valuation u_3. However, if player 1 follows the memoryless strategy $\bar{\xi}_1$, then the play visits s_3 and s_4 alternately and reaches s_0 with probability 0. Thus, ξ_1 is an example of a selector that is value-optimal, but not optimal.

On the other hand, consider any selector ξ'_1 for player 1 that chooses move b at state s_3 with positive probability. Under the memoryless strategy $\bar{\xi}'_1$, the set $\{s_0, s_1\}$ of states is reached with probability 1, and s_0 is reached with probability $\frac{1}{2}$. Such a ξ'_1 is thus an example of a selector that is both value-optimal and optimal.

In the example, the problem is that the strategy $\bar{\xi}_1$ may cause player 1 to stay forever in $S \setminus (T \cup W_2)$ with positive probability. We call "proper" the strategies of player 1 that guarantee reaching $T \cup W_2$ with probability 1.

Definition 4 *(Proper strategies and selectors).* A player-1 strategy π_1 is *proper* if for all player-2 strategies π_2, and for all states $s \in S \setminus (T \cup W_2)$, we have $\Pr_s^{\pi_1,\pi_2}(\mathrm{Reach}(T \cup W_2)) = 1$. A player-1 selector ξ_1 is *proper* if the memoryless player-1 strategy $\bar{\xi}_1$ is proper.

We note that proper strategies are closely related to Condon's notion of a *halting game* [5]: precisely, a game is halting iff all player-1 strategies are proper. We can check whether a selector for player 1 is proper by considering only the pure selectors for player 2.

Lemma 1. *Given a selector ξ_1 for player* 1*, the memoryless player-*1 *strategy $\bar{\xi}_1$ is proper iff for every pure selector ξ_2 for player* 2*, and for all states $s \in S$, we have* $\Pr_s^{\bar{\xi}_1,\bar{\xi}_2}(\mathrm{Reach}(T \cup W_2)) = 1$.

Proof. We prove the contrapositive. Given a player-1 selector ξ_1, consider the 2-MDP G_{ξ_1}. If $\bar{\xi}_1$ is not proper, then by Theorem 1, there must exist an end component $C \subseteq S \setminus (T \cup W_2)$ in G_{ξ_1}. Then, from C, player 2 can avoid reaching $T \cup W_2$ by repeatedly applying a pure selector ξ_2 that at every state $s \in C$ deterministically chooses a move $a_2 \in \Gamma_2(s)$ such that $Dest(s, \xi_1, a_2) \subseteq C$. The existence of a suitable $\xi_2(s)$ for all states $s \in C$ follows from the definition of end component. □

The following lemma shows that the selector that chooses all available moves uniformly at random is proper. This fact will be used later to initialize our strategy-improvement algorithm.

Lemma 2. *Let ξ_1^{unif} be the player-*1 *selector that at all states $s \in S \setminus (T \cup W_2)$ chooses all moves in $\Gamma_1(s)$ uniformly at random. Then ξ_1^{unif} is proper.*

Proof. Assume towards contradiction that ξ_1^{unif} is not proper. From Theorem 1, in the 2-MDP $G_{\xi_1^{unif}}$ there must be an end component $C \subseteq S \setminus (T \cup W_2)$. Then, when player 1 follows the strategy $\bar{\xi}_1^{unif}$, player 2 can confine the game to C. By the definition of ξ_1^{unif}, player 2 can ensure that the game does not leave C regardless of the moves chosen by player 1, and thus, for *all* strategies of player 1. This contradicts the fact that W_2 contains all states from which player 2 can ensure that T is not reached. □

The following lemma shows that if the player-1 selector ζ_k computed by the value-iteration scheme (1) is proper, then the player-1 strategy $\overline{\zeta}_k$ guarantees the value u_k, for all $k \geqslant 0$.

Lemma 3. *Let v be a valuation such that $Pre_1(v) \geqslant v$ and $v(s) = 0$ for all states $s \in W_2$. Let ξ_1 be a selector for player* 1 *such that $Pre_{1:\xi_1}(v) = Pre_1(v)$. If ξ_1 is proper, then for all player-2 strategies π_2, we have $\Pr^{\overline{\xi}_1,\pi_2}(\mathrm{Reach}(T)) \geqslant v$.*

Proof. Consider an arbitrary player-2 strategy π_2, and for $k \geqslant 0$, 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 $v_k \geqslant v$ for all $k \geqslant 0$. In fact, $v_0 = v$, and for $k \geqslant 0$, we have

$$v_{k+1} \geqslant Pre_{1:\xi_1}(v_k) \geqslant Pre_{1:\xi_1}(v) = Pre_1(v) \geqslant v.$$

For all $k \geqslant 0$ and $s \in S$, we can write v_k as

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

Since $v(s) \leqslant 1$ when $s \in T$, the first term on the right-hand side is at most $\Pr_s^{\overline{\xi}_1,\pi_2}(\Theta_k \in T)$. For the second term, we have $\lim_{k\to\infty} \Pr^{\overline{\xi}_1,\pi_2}(\Theta_k \in S \setminus (T \cup W_2)) = 0$ by hypothesis, because $\Pr^{\overline{\xi}_1,\pi_2}(\mathrm{Reach}(T \cup W_2)) = 1$ 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 states $s \in W_2$. Hence, taking the limit with $k \to \infty$, we obtain

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

where the last inequality follows from $v_k \geqslant v$ for all $k \geqslant 0$. Note that $v_k = \Pr^{\overline{\xi}_1,\pi_2}(\Theta_k \in T)$, and since T is absorbing it follows that v_k is non-decreasing (monotonic) and is bounded by 1 (since it is a probability measure). Hence the limit of v_k is defined. The desired result follows. □

4.2. From value iteration to optimal selectors

In this section we show how to obtain memoryless ε-optimal strategies from the value-iteration scheme, for $\varepsilon > 0$. In the following section the existence such strategies would be established using a strategy-iteration scheme. The strategy-iteration scheme has been used previously to establish existence of memoryless ε-optimal strategies, for $\varepsilon > 0$ (for example see [13] and also results of Condon [5] for turn-based games). However our proof which constructs the memoryless strategies based on value-iteration scheme is new. Considering again the value-iteration scheme (1), since $\langle\langle 1\rangle\rangle_{\mathsf{val}}(\mathrm{Reach}(T)) = \lim_{k\to\infty} u_k$, for every $\varepsilon > 0$ there is a k such that $u_k(s) \geqslant u_{k-1}(s) \geqslant \langle\langle 1\rangle\rangle_{\mathsf{val}}(\mathrm{Reach}(T))(s) - \varepsilon$ at all states $s \in S$. Lemma 3 indicates that, in order to construct a memoryless ε-optimal strategy, we need to construct from u_{k-1} a player-1 selector ξ_1 such that:

(i) ξ_1 is value-optimal for u_{k-1}, that is, $Pre_{1:\xi_1}(u_{k-1}) = Pre_1(u_{k-1}) = u_k$; and
(ii) ξ_1 is proper.

To ensure the construction of a value-optimal, proper selector, we need some definitions. For $r > 0$, the *value class*

$$U_r^k = \big\{s \in S \bigm| u_k(s) = r\big\}$$

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 $\bowtie \in \{<, \leqslant, \geqslant, >\}$. For a state $s \in S$, let $\ell_k(s) = \min\{j \leqslant k \mid u_j(s) = u_k(s)\}$ be the *entry time* of s in $U_{u_k(s)}^k$, that is, the least iteration j in which the state s has the same value as in iteration k. For $k \geqslant 0$, we define the player-1 selector η_k as follows: if $\ell_k(s) > 0$, then

$$\eta_k(s) = \eta_{\ell_k(s)}(s) = \arg\max_{\xi_1 \in \Lambda_1} \inf_{\xi_2 \in \Lambda_2} Pre_{\xi_1,\xi_2}(u_{\ell_k(s)-1});$$

otherwise, if $\ell_k(s) = 0$, then $\eta_k(s) = \eta_{\ell_k(s)}(s) = \xi_1^{unif}(s)$ (this definition is arbitrary, and it does not affect the remainder of the proof). In words, the selector $\eta_k(s)$ is an optimal selector for s at the iteration $\ell_k(s)$. It follows easily that $u_k = Pre_{1:\eta_k}(u_{k-1})$, that is, η_k is also value-optimal for u_{k-1}, satisfying the first of the above conditions.

To conclude the construction, we need to prove that for k sufficiently large (namely, for k such that $u_k(s) > 0$ at all states $s \in S \setminus (T \cup W_2)$), the selector η_k is proper. To this end we use Theorem 1, and show that for sufficiently large k no end component of G_{η_k} is entirely contained in $S \setminus (T \cup W_2)$.[1] To reason about the end components of G_{η_k}, for a state $s \in S$ and a player-2 move $a_2 \in \Gamma_2(s)$, we write

$$Dest_k(s, a_2) = \bigcup_{a_1 \in Supp(\eta_k(s))} Dest(s, a_1, a_2)$$

for the set of possible successors of state s when player 1 follows the strategy $\overline{\eta}_k$, and player 2 chooses the move a_2.

Lemma 4. *Let $0 < r \leqslant 1$ and $k \geqslant 0$, and consider a state $s \in S \setminus (T \cup W_2)$ such that $s \in U_r^k$. For all moves $a_2 \in \Gamma_2(s)$, we have:*

(i) *either $Dest_k(s, a_2) \cap U_{>r}^k \neq \emptyset$,*
(ii) *or $Dest_k(s, a_2) \subseteq U_r^k$, and there is a state $t \in Dest_k(s, a_2)$ with $\ell_k(t) < \ell_k(s)$.*

Proof. For convenience, let $m = \ell_k(s)$, and consider any move $a_2 \in \Gamma_2(s)$.

- Consider first the case that $Dest_k(s, a_2) \nsubseteq U_r^k$. Then, it cannot be that $Dest_k(s, a_2) \subseteq U_{\leqslant r}^k$; otherwise, for all states $t \in Dest_k(s, a_2)$, we would have $u_k(t) \leqslant r$, and there would be at least one state $t \in Dest_k(s, a_2)$ 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 that $Dest_k(s, a_2) \cap U_{>r}^k \neq \emptyset$.
- Consider now the case that $Dest_k(s, a_2) \subseteq U_r^k$. Since $u_m \leqslant u_k$, due to the monotonicity of the Pre_1 operator and (1), we have that $u_{m-1}(t) \leqslant r$ for all states $t \in Dest_k(s, a_2)$. From $r = u_k(s) = u_m(s) = Pre_{1:\eta_k}(u_{m-1})$, it follows that $u_{m-1}(t) = r$ for all states $t \in Dest_k(s, a_2)$, implying that $\ell_k(t) < m$ for all states $t \in Dest_k(s, a_2)$. □

The above lemma states that under η_k, from each state $i \in U_r^k$ with $r > 0$ we are guaranteed a probability bounded away from 0 of either moving to a higher-value class $U_{>r}^k$, or of moving to states within the value class that have a strictly lower entry time. Note that the states in the target set T are all in U_1^0: they have entry time 0 in the value class for value 1. 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 4.

Lemma 5. *For all $k \geqslant 0$, if for all states $s \in S \setminus W_2$ we have $u_{k-1}(s) > 0$, then for all player-2 strategies π_2, we have $\Pr^{\overline{\eta}_k, \pi_2}(\text{Reach}(T \cup W_2)) = 1$.*

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 no end component of G_{η_k} is 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 $C \subseteq U_{[r_1, r_2]}^k$ with $C \cap U_{r_2} \neq \emptyset$, for some $0 < r_1 \leqslant r_2 \leqslant 1$, where $U_{[r_1, r_2]}^k = U_{\geqslant r_1}^k \cap U_{\leqslant r_2}^k$ is the union of the value classes for all values in the interval $[r_1, r_2]$. Consider a state $s \in U_{r_2}^k$ with minimal ℓ_k, that is, such that $\ell_k(s) \leqslant \ell_k(t)$ for all other states $t \in U_{r_2}^k$. From Lemma 4, it follows that for every move $a_2 \in \Gamma_2(s)$, there is a state $t \in Dest_k(s, a_2)$ such that (i) either $t \in U_{r_2}^k$ and $\ell_k(t) < \ell_k(s)$, (ii) or $t \in U_{>r_2}^k$. In both cases, we obtain a contradiction. □

The above lemma shows that η_k satisfies both requirements for optimal selectors spelt out at the beginning of Section 4.2. Hence, η_k guarantees the value u_k. This proves the existence of memoryless ε-optimal strategies for concurrent reachability games.

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

Proof. Consider a concurrent reachability game with target $T \subseteq S$. Since $\lim_{k\to\infty} u_k = \langle\langle 1 \rangle\rangle_{\mathsf{val}}(\text{Reach}(T))$, for every $\varepsilon > 0$ we can find $k \in \mathbb{N}$ such that the following two assertions hold:

$$\max_{s \in S} \bigl(\langle\langle 1 \rangle\rangle_{\mathsf{val}} \bigl(\text{Reach}(T) \bigr)(s) - u_{k-1}(s) \bigr) < \varepsilon,$$
$$\min_{s \in S \setminus W_2} u_{k-1}(s) > 0.$$

[1] In fact, the result holds for all k, even though our proof, for the sake of a simpler argument, does not show it.

Algorithm 1 Reachability strategy-improvement algorithm

```
Input: a concurrent game structure $G$ with target set $T$.
Output: a strategy $\overline{\gamma}$ for player 1.

0. Compute $W_2 = \{s \in S \mid \langle\langle 1\rangle\rangle_{\mathsf{val}}(\mathrm{Reach}(T))(s) = 0\}$.
1. Let $\gamma_0 = \xi_1^{unif}$ and $i = 0$.
2. Compute $v_0 = \langle\langle 1\rangle\rangle_{\mathsf{val}}^{\overline{\gamma}_0}(\mathrm{Reach}(T))$.
3. do {
    3.1. Let $I = \{s \in S \setminus (T \cup W_2) \mid \mathit{Pre}_1(v_i)(s) > v_i(s)\}$.
    3.2. Let $\xi_1$ be a player-1 selector such that for all states $s \in I$, we have $\mathit{Pre}_{1:\xi_1}(v_i)(s) = \mathit{Pre}_1(v_i)(s) > v_i(s)$.
    3.3. The player-1 selector $\gamma_{i+1}$ is defined as follows: for each state $s \in S$, let
            $\gamma_{i+1}(s) = \begin{cases} \gamma_i(s) & \text{if } s \notin I; \\ \xi_1(s) & \text{if } s \in I. \end{cases}$
    3.4. Compute $v_{i+1} = \langle\langle 1\rangle\rangle_{\mathsf{val}}^{\overline{\gamma}_{i+1}}(\mathrm{Reach}(T))$.
    3.5. Let $i = i + 1$.
} until $I = \emptyset$.
4. return $\overline{\gamma}_i$.
```

By construction, $\mathit{Pre}_{1:\eta_k}(u_{k-1}) = \mathit{Pre}_1(u_{k-1}) = u_k$. Hence, from Lemma 3 and Lemma 5, for all player-2 strategies π_2, we have $\Pr^{\overline{\eta}_k,\pi_2}(\mathrm{Reach}(T)) \geqslant u_{k-1}$, leading to the result. □

5. Strategy-improvement algorithm for concurrent reachability games

In the previous section, we provided a proof of the existence of memoryless ε-optimal strategies for all $\varepsilon > 0$, on the basis of a value-iteration scheme. In this section we present a strategy-improvement algorithm for concurrent games with reachability objectives. The algorithm will produce a sequence of selectors $\gamma_0, \gamma_1, \gamma_2, \ldots$ for player 1, such that:

(i) for all $i \geqslant 0$, we have $\langle\langle 1\rangle\rangle_{\mathsf{val}}^{\overline{\gamma}_i}(\mathrm{Reach}(T)) \leqslant \langle\langle 1\rangle\rangle_{\mathsf{val}}^{\overline{\gamma}_{i+1}}(\mathrm{Reach}(T))$;
(ii) if there is $i \geqslant 0$ such that $\gamma_i = \gamma_{i+1}$, then $\langle\langle 1\rangle\rangle_{\mathsf{val}}^{\overline{\gamma}_i}(\mathrm{Reach}(T)) = \langle\langle 1\rangle\rangle_{\mathsf{val}}(\mathrm{Reach}(T))$; and
(iii) $\lim_{i\to\infty} \langle\langle 1\rangle\rangle_{\mathsf{val}}^{\overline{\gamma}_i}(\mathrm{Reach}(T)) = \langle\langle 1\rangle\rangle_{\mathsf{val}}(\mathrm{Reach}(T))$.

Condition (i) guarantees that the algorithm computes a sequence of monotonically improving selectors. Condition (ii) guarantees that if a selector cannot be improved, then it is optimal. Condition (iii) guarantees that the value guaranteed by the selectors converges to the value of the game, or equivalently, that for all $\varepsilon > 0$, there is a number i of iterations such that the memoryless player-1 strategy $\overline{\gamma}_i$ is ε-optimal. Note that for concurrent reachability games, there may be no $i \geqslant 0$ such that $\gamma_i = \gamma_{i+1}$, that is, the algorithm may fail to generate an optimal selector. This is because there are concurrent reachability games that do not admit optimal strategies, but only ε-optimal strategies for all $\varepsilon > 0$ [14,10]. For *turn-based* reachability games, our algorithm terminates with an optimal selector and we will present bounds for termination.

We note that the value-iteration scheme of the previous section does not directly yield a strategy-improvement algorithm. In fact, the sequence of player-1 selectors $\eta_0, \eta_1, \eta_2, \ldots$ computed in Section 4.1 may violate condition (ii): it is possible that for some $i \geqslant 0$ we have $\eta_i = \eta_{i+1}$, but $\eta_i \neq \eta_j$ for some $j > i$. This is because the scheme of Section 4.1 is fundamentally a value-iteration scheme, even though a selector is extracted from each valuation. The scheme guarantees that the valuations $u_0, u_1, u_2, \ldots$ defined as in (1) converge, but it does not guarantee that the selectors $\eta_0, \eta_1, \eta_2, \ldots$ improve at each iteration.

The strategy-improvement algorithm presented here shares an important connection with the proof of the existence of memoryless ε-optimal strategies presented in the previous section. Here, also, the key is to ensure that all generated selectors are proper. Again, this is ensured by modifying the selectors, at each iteration, only where they can be improved.

5.1. The strategy-improvement algorithm

Ordering of strategies. We let W_2 be as in Section 4.1, and again we assume without loss of generality that all states in $W_2 \cup T$ are absorbing. We define a preorder $\prec$ on the strategies for player 1 as follows: given two player-1 strategies π_1 and π_1', let $\pi_1 \prec \pi_1'$ if the following two conditions hold: (i) $\langle\langle 1\rangle\rangle_{\mathsf{val}}^{\pi_1}(\mathrm{Reach}(T)) \leqslant \langle\langle 1\rangle\rangle_{\mathsf{val}}^{\pi_1'}(\mathrm{Reach}(T))$; and (ii) $\langle\langle 1\rangle\rangle_{\mathsf{val}}^{\pi_1}(\mathrm{Reach}(T))(s) < \langle\langle 1\rangle\rangle_{\mathsf{val}}^{\pi_1'}(\mathrm{Reach}(T))(s)$ for some state $s \in S$. Furthermore, we write $\pi_1 \preccurlyeq \pi_1'$ if either $\pi_1 \prec \pi_1'$ or $\pi_1 = \pi_1'$.

Informal description of Algorithm 1. We now present the strategy-improvement algorithm (Algorithm 1) for computing the values for all states in $S \setminus (T \cup W_2)$. The algorithm iteratively improves player-1 strategies according to the preorder $\prec$. The algorithm starts with the random selector $\gamma_0 = \xi_1^{unif}$. At iteration $i+1$, the algorithm considers the memoryless player-1 strategy $\overline{\gamma}_i$ and computes the value $\langle\langle 1\rangle\rangle_{\mathsf{val}}^{\overline{\gamma}_i}(\text{Reach}(T))$. Observe that since $\overline{\gamma}_i$ is a memoryless strategy, the computation of $\langle\langle 1\rangle\rangle_{\mathsf{val}}^{\overline{\gamma}_i}(\text{Reach}(T))$ involves solving the 2-MDP G_{γ_i}. The valuation $\langle\langle 1\rangle\rangle_{\mathsf{val}}^{\overline{\gamma}_i}(\text{Reach}(T))$ is named v_i. For all states s such that $Pre_1(v_i)(s) > v_i(s)$, the memoryless strategy at s is modified to a selector that is value-optimal for v_i. The algorithm then proceeds to the next iteration. If $Pre_1(v_i) = v_i$, the algorithm stops and returns the optimal memoryless strategy $\overline{\gamma}_i$ for player 1. Unlike strategy-improvement algorithms for turn-based games (see [6] for a survey), Algorithm 1 is not guaranteed to terminate, because the value of a reachability game may not be rational.

5.2. Convergence

Lemma 6. *Let γ_i and γ_{i+1} be the player-1 selectors obtained at iterations i and $i+1$ of Algorithm 1. If γ_i is proper, then γ_{i+1} is also proper.*

Proof. Assume towards a contradiction that γ_i is proper and γ_{i+1} is not. Let ξ_2 be a pure selector for player 2 to witness that γ_{i+1} is not proper. Then there exists a subset $C \subseteq S \setminus (T \cup W_2)$ such that C is a closed recurrent set of states in the Markov chain G_{γ_{i+1},ξ_2}. Let I be the nonempty set of states where the selector is modified to obtain γ_{i+1} from γ_i; at all other states γ_i and γ_{i+1} agree.

Since γ_i and γ_{i+1} agree at all states other than the states in I, and γ_i is a proper strategy, it follows that $C \cap I \neq \emptyset$. Let $U_r^i = \{s \in S \setminus (T \cup W_2) \mid \langle\langle 1\rangle\rangle_{\mathsf{val}}^{\overline{\gamma}_i}(\text{Reach}(T))(s) = v_i(s) = r\}$ be the value class with value r at iteration i. For a state $s \in U_r^i$ the following assertion holds: if $Dest(s, \gamma_i, \xi_2) \subsetneq U_r^i$, then $Dest(s, \gamma_i, \xi_2) \cap U_{>r}^i \neq \emptyset$. Let $z = \max\{r \mid U_r^i \cap C \neq \emptyset\}$, that is, U_z^i is the greatest value class at iteration i with a nonempty intersection with the closed recurrent set C. It easily follows that $0 < z < 1$. Consider any state $s \in I$, and let $s \in U_q^i$. Since $Pre_1(v_i)(s) > v_i(s)$, it follows that $Dest(s, \gamma_{i+1}, \xi_2) \cap U_{>q}^i \neq \emptyset$. Hence we must have $z > q$, and therefore $I \cap C \cap U_z^i = \emptyset$. Thus, for all states $s \in U_z^i \cap C$, we have $\gamma_i(s) = \gamma_{i+1}(s)$. Recall that z is the greatest value class at iteration i with a nonempty intersection with C; hence $U_{>z}^i \cap C = \emptyset$. Thus for all states $s \in C \cap U_z^i$, we have $Dest(s, \gamma_{i+1}, \xi_2) \subseteq U_z^i \cap C$. It follows that $C \subseteq U_z^i$. However, this gives us three statements that together form a contradiction: $C \cap I \neq \emptyset$ (or else γ_i would not have been proper), $I \cap C \cap U_z^i = \emptyset$, and $C \subseteq U_z^i$. □

Lemma 7. *For all $i \geqslant 0$, the player-1 selector γ_i obtained at iteration i of Algorithm 1 is proper.*

Proof. By Lemma 2 we have that γ_0 is proper. The result then follows from Lemma 6 and induction. □

Lemma 8. *Let γ_i and γ_{i+1} be the player-1 selectors obtained at iterations i and $i+1$ of Algorithm 1. Let $I = \{s \in S \mid Pre_1(v_i)(s) > v_i(s)\}$. Let $v_i = \langle\langle 1\rangle\rangle_{\mathsf{val}}^{\overline{\gamma}_i}(\text{Reach}(T))$ and $v_{i+1} = \langle\langle 1\rangle\rangle_{\mathsf{val}}^{\overline{\gamma}_{i+1}}(\text{Reach}(T))$. Then $v_{i+1}(s) \geqslant Pre_1(v_i)(s)$ for all states $s \in S$; and therefore $v_{i+1}(s) \geqslant v_i(s)$ for all states $s \in S$, and $v_{i+1}(s) > v_i(s)$ for all states $s \in I$.*

Proof. Consider the valuations v_i and v_{i+1} obtained at iterations i and $i+1$, respectively, and let w_i be the valuation defined by $w_i(s) = 1 - v_i(s)$ for all states $s \in S$. Since γ_{i+1} is proper (by Lemma 7), it follows that the counter-optimal strategy for player 2 to minimize v_{i+1} is obtained by maximizing the probability to reach W_2. In fact, there are no end components in $S \setminus (W_2 \cup T)$ in the 2-MDP $G_{\gamma_{i+1}}$. Let

$$\widehat{w}_i(s) = \begin{cases} w_i(s) & \text{if } s \in S \setminus I; \\ 1 - Pre_1(v_i)(s) < w_i(s) & \text{if } s \in I. \end{cases}$$

In other words, $\widehat{w}_i = 1 - Pre_1(v_i)$, and we also have $\widehat{w}_i \leqslant w_i$. We now show that $\widehat{w}_i$ is a feasible solution to the linear program for MDPs with the objective Reach(W_2), as described in Section 3. Since $v_i = \langle\langle 1\rangle\rangle_{\mathsf{val}}^{\overline{\gamma}_i}(\text{Reach}(T))$, it follows that for all states $s \in S$ and all moves $a_2 \in \Gamma_2(s)$, we have

$$w_i(s) \geqslant \sum_{t \in S} w_i(t) \cdot \delta_{\gamma_i}(s, a_2).$$

For all states $s \in S \setminus I$, we have $\gamma_i(s) = \gamma_{i+1}(s)$ and $\widehat{w}_i(s) = w_i(s)$, and since $\widehat{w}_i \leqslant w_i$, it follows that for all states $s \in S \setminus I$ and all moves $a_2 \in \Gamma_2(s)$, we have

$$\widehat{w}_i(s) \geqslant \sum_{t \in S} \widehat{w}_i(t) \cdot \delta_{\gamma_{i+1}}(s, a_2) \quad \bigl(\text{for } s \in (S \setminus I)\bigr).$$

Since for $s \in I$ the selector $\gamma_{i+1}(s)$ is obtained as an optimal selector for $Pre_1(v_i)(s)$, it follows that for all states $s \in I$ and all moves $a_2 \in \Gamma_2(s)$, we have

$$Pre_{\gamma_{i+1},a_2}(v_i)(s) \geqslant Pre_1(v_i)(s);$$

in other words, $1 - \mathit{Pre}_1(v_i)(s) \geqslant 1 - \mathit{Pre}_{\gamma_{i+1},a_2}(v_i)(s)$. Hence for all states $s \in I$ and all moves $a_2 \in \Gamma_2(s)$, we have

$$\widehat{w}_i(s) \geqslant \sum_{t \in S} w_i(t) \cdot \delta_{\gamma_{i+1}}(s, a_2).$$

Since $\widehat{w}_i \leqslant w_i$, for all states $s \in I$ and all moves $a_2 \in \Gamma_2(s)$, we have

$$\widehat{w}_i(s) \geqslant \sum_{t \in S} \widehat{w}_i(t) \cdot \delta_{\gamma_{i+1}}(s, a_2) \quad (\text{for } s \in I).$$

Hence it follows that $\widehat{w}_i$ is a feasible solution to the linear program for MDPs with reachability objectives. Since the reachability valuation for player 2 for $\text{Reach}(W_2)$ is the least solution (observe that the objective function of the linear program is a minimizing function), it follows that $v_{i+1} \geqslant 1 - \widehat{w}_i = \mathit{Pre}_1(v_i)$. Thus we obtain $v_{i+1}(s) \geqslant v_i(s)$ for all states $s \in S$, and $v_{i+1}(s) > v_i(s)$ for all states $s \in I$. □

Theorem 3 *(Strategy improvement). The following two assertions hold about Algorithm 1:*

(i) *For all $i \geqslant 0$, we have $\overline{\gamma}_i \preccurlyeq \overline{\gamma}_{i+1}$; moreover, if $\overline{\gamma}_i = \overline{\gamma}_{i+1}$, then $\overline{\gamma}_i$ is an optimal strategy.*
(ii) $\lim_{i\to\infty} v_i = \lim_{i\to\infty} \langle\langle 1\rangle\rangle_{\mathsf{val}}^{\overline{\gamma}_i}(\text{Reach}(T)) = \langle\langle 1\rangle\rangle_{\mathsf{val}}(\text{Reach}(T))$.

Proof. We prove the two parts as follows.

(i) The assertion that $\overline{\gamma}_i \preccurlyeq \overline{\gamma}_{i+1}$ follows from Lemma 8. If $\overline{\gamma}_i = \overline{\gamma}_{i+1}$, then $\mathit{Pre}_1(v_i) = v_i$. Let $v = \langle\langle 1\rangle\rangle_{\mathsf{val}}(\text{Reach}(T))$, and since v is the least solution to satisfy $\mathit{Pre}_1(x) = x$ (i.e., the least fixpoint) [11], it follows that $v_i \geqslant v$. From Lemma 7 it follows that $\overline{\gamma}_i$ is proper. Since $\overline{\gamma}_i$ is proper by Lemma 3, we have $\langle\langle 1\rangle\rangle_{\mathsf{val}}^{\overline{\gamma}_i}(\text{Reach}(T)) \geqslant v_i \geqslant v$. It follows that $\overline{\gamma}_i$ is optimal for player 1.

(ii) Let $v_0 = [T]$ and $u_0 = [T]$. We have $u_0 \leqslant v_0$. For all $k \geqslant 0$, by Lemma 8, we have $v_{k+1} \geqslant [T] \vee \mathit{Pre}_1(v_k)$. For all $k \geqslant 0$, let $u_{k+1} = [T] \vee \mathit{Pre}_1(u_k)$. By induction we conclude that for all $k \geqslant 0$, we have $u_k \leqslant v_k$. Moreover, $v_k \leqslant \langle\langle 1\rangle\rangle_{\mathsf{val}}(\text{Reach}(T))$, that is, for all $k \geqslant 0$, we have

$$u_k \leqslant v_k \leqslant \langle\langle 1\rangle\rangle_{\mathsf{val}}\big(\text{Reach}(T)\big).$$

Since $\lim_{k\to\infty} u_k = \langle\langle 1\rangle\rangle_{\mathsf{val}}(\text{Reach}(T))$, it follows that

$$\lim_{k\to\infty} \langle\langle 1\rangle\rangle_{\mathsf{val}}^{\overline{\gamma}_k}\big(\text{Reach}(T)\big) = \lim_{k\to\infty} v_k = \langle\langle 1\rangle\rangle_{\mathsf{val}}\big(\text{Reach}(T)\big).$$

The theorem follows. □

5.3. Termination for turn-based stochastic games

If the input game structure to Algorithm 1 is a turn-based stochastic game structure, then if we start with a proper selector γ_0 that is pure, then for all $i \geqslant 0$ we can choose the selector γ_i such that γ_i is both proper and pure: the above claim follows since given a valuation v, if a state s is a player-1 state, then there is an action a at s (or choice of an edge at s) that achieves $\mathit{Pre}_1(v)(s)$ at s. Since the number of pure selectors is bounded, if we start with a pure, proper selector then termination is ensured. Hence we present a procedure to compute a pure, proper selector, and then present termination bounds (i.e., bounds on i such that $u_{i+1} = u_i$). The construction of a pure, proper selector is based on the notion of *attractors* defined below.

Attractor strategy. Let $A_0 = W_2 \cup T$, and for $i \geqslant 0$ we have

$$A_{i+1} = A_i \cup \big\{s \in S_1 \cup S_R \;\big|\; E(s) \cap A_i \neq \emptyset\big\} \cup \big\{s \in S_2 \;\big|\; E(s) \subseteq A_i\big\}.$$

Since for all $s \in S \setminus W_2$ we have $\langle\langle 1\rangle\rangle_{\mathsf{val}}(\text{Reach}(T)) > 0$, it follows that from all states in $S \setminus W_2$ player 1 can ensure that T is reached with positive probability. It follows that for some $i \geqslant 0$ we have $A_i = S$. The pure *attractor* selector ξ^* is as follows: for a state $s \in (A_{i+1} \setminus A_i) \cap S_1$ we have $\xi^*(s)(t) = 1$, where $t \in A_i$ (such a t exists by construction). The pure memoryless strategy $\overline{\xi}^*$ ensures that for all $i \geqslant 0$, from A_{i+1} the game reaches A_i with positive probability. Hence there is no end component C contained in $S \setminus (W_2 \cup T)$ in the MDP $G_{\overline{\xi}^*}$. It follows that ξ^* is a pure selector that is proper, and the selector ξ^* can be computed in $O(|E|)$ time. We now present the termination bounds.

Termination bounds. We present termination bounds for binary turn-based stochastic games. A turn-based stochastic game is binary if for all $s \in S_R$ we have $|E(s)| \leqslant 2$, and for all $s \in S_R$ if $|E(s)| = 2$, then for all $t \in E(s)$ we have $\delta(s)(t) = \frac{1}{2}$, i.e., for all probabilistic states there are at most two successors and the transition function δ is uniform.

Lemma 9. *Let G be a binary Markov chain with $|S|$ states with a reachability objective* $\text{Reach}(T)$*. Then for all $s \in S$ we have $\langle\langle 1\rangle\rangle_{\mathsf{val}}(\text{Reach}(T)) = \frac{p}{q}$, with $p, q \in \mathbb{N}$ and $p, q \leqslant 4^{|S|-1}$.*

Proof. The results follow as a special case of Lemma 2 of [6]. Lemma 2 of [6] holds for halting turn-based stochastic games, and since Markov chain reaches the set of closed connected recurrent states with probability 1 from all states the result follows. □

Lemma 10. *Let G be a binary turn-based stochastic game with a reachability objective* $\text{Reach}(T)$*. Then for all $s \in S$ we have $\langle\langle 1\rangle\rangle_{\mathsf{val}}(\text{Reach}(T)) = \frac{p}{q}$, with $p, q \in \mathbb{N}$ and $p, q \leqslant 4^{|S_R|-1}$.*

Proof. Since pure memoryless optimal strategies exist for both players (existence of pure memoryless optimal strategies for both players in turn-based stochastic reachability games follows from [5]), we fix pure memoryless optimal strategies π_1 and π_2 for both players. The Markov chain G_{π_1,π_2} can be then reduced to an equivalent Markov chains with $|S_R|$ states (since we fix deterministic successors for states in $S_1 \cup S_2$, they can be collapsed to their successors). The result then follows from Lemma 9. □

From Lemma 10 it follows that at iteration i of the reachability strategy-improvement algorithm either the sum of the values either increases by $\frac{1}{4^{|S_R|-1}}$ or else there is a valuation u_i such that $u_{i+1} = u_i$. Since the sum of values of all states can be at most $|S|$, it follows that algorithm terminates in at most $|S| \cdot 4^{|S_R|-1}$ iterations. Moreover, since the number of pure memoryless strategies is at most $\prod_{s\in S_1} |E(s)|$, the algorithm terminates in at most $\prod_{s\in S_1} |E(s)|$ iterations. It follows from the results of [25] that a turn-based stochastic game structure G can be reduced to an equivalent binary turn-based stochastic game structure G' such that the set of player-1 and player-2 states in G and G' are the same and the number of probabilistic states in G' is $O(|\delta|)$, where $|\delta|$ is the size of the transition function in G. Thus we obtain the following result.

Theorem 4. *Let G be a turn-based stochastic game with a reachability objective* $\text{Reach}(T)$*, then the reachability strategy-improvement algorithm computes the values in time*

$$O\left(\min\left\{\prod_{s\in S_1}\bigl|E(s)\bigr|, 2^{O(|\delta|)}\right\} \cdot poly\bigl(|G|\bigr)\right);$$

where poly is polynomial function.

The results of [16] presented an algorithm for turn-based stochastic games that works in time $O(|S_R|! \cdot poly(|G|))$. The algorithm of [16] works only for turn-based stochastic games, for general turn-based stochastic games the complexity of the algorithm of [16] is better. However, for turn-based stochastic games where the transition function at all states can be expressed with constantly many bits we have $|\delta| = O(|S_R|)$. In these cases the reachability strategy-improvement algorithm (that works for both concurrent and turn-based stochastic games) works in time $2^{O(|S_R|)} \cdot poly(|G|)$ as compared to the time $2^{O(|S_R|\cdot\log(|S_R|))} \cdot poly(|G|)$ of the algorithm of [16]. A recent result of [19] presents a more refined analysis and an improved result for turn-based stochastic reachability games.

6. Existence of memoryless optimal strategies for concurrent safety games

A proof of the existence of memoryless optimal strategies for safety games can be found in [11]: the proof uses results on martingales to obtain the result. For sake of completeness we present (an alternative) proof of the result: the proof we present is similar in spirit with the other proofs in this paper and uses the results on MDPs to obtain the result. The proof is very similar to the proof presented in [13].

Theorem 5 *(Memoryless optimal strategies). Memoryless optimal strategies exist for all concurrent games with safety objectives.*

Proof. Consider a concurrent game structure G with a safety objective $\text{Safe}(F)$ for player 1. Then it follows from the results of [11] that

$$\langle\langle 1\rangle\rangle_{\mathsf{val}}\bigl(\text{Safe}(F)\bigr) = \nu X . \bigl(\min\bigl\{[F], Pre_1(X)\bigr\}\bigr),$$

where $[F]$ is the indicator function of the set F and ν denotes the greatest fixpoint. Let $T = S \setminus F$, and for all states $s \in T$ we have $\langle\langle 1\rangle\rangle_{\mathsf{val}}(\text{Safe}(F))(s) = 0$, and hence any memoryless strategy from T is an optimal strategy. Thus without loss of generality we assume all states in T are absorbing. Let $v = \langle\langle 1\rangle\rangle_{\mathsf{val}}(\text{Safe}(F))$, and since we assume all states in T are absorbing it follows that $Pre_1(v) = v$ (since v is a fixpoint). Let γ be a player-1 selector such that for all states s we have $Pre_{1:\gamma}(v)(s) = Pre_1(v)(s) = v(s)$. We show that $\overline{\gamma}$ is a memoryless optimal strategy. Consider the player-2 MDP G_γ and we

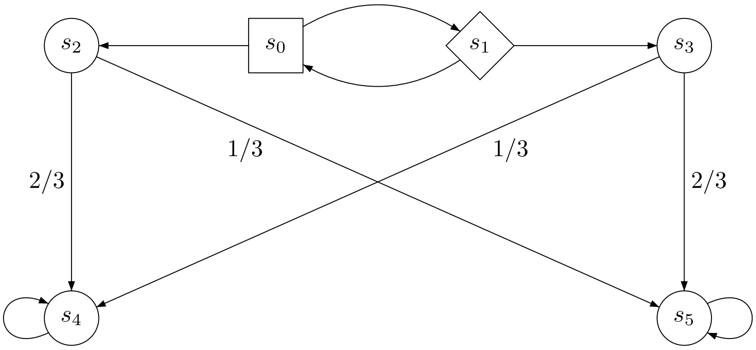

Fig. 2. A turn-based stochastic safety game.

consider the maximal probability for player 2 to reach the target set T. Consider the valuation w defined as $w = 1 - v$. For all states $s \in T$ we have $w(s) = 1$. Since $Pre_{1:\gamma}(v) = Pre_1(v)$ it follows that for all states $s \in F$ and all $a_2 \in \Gamma_2(s)$ we have

$$Pre_{\gamma,a_2}(v)(s) \geqslant Pre_1(v)(s) = v(s);$$

in other words, for all $s \in F$ we have $1 - Pre_1(v)(s) = 1 - v(s) \geqslant 1 - Pre_{\gamma,a_2}(v)(s)$. Hence for all states $s \in F$ and all moves $a_2 \in \Gamma_2(s)$, we have

$$w(s) \geqslant \sum_{t \in S} w(t) \cdot \delta_{\gamma}(s, a_2).$$

Hence it follows that w is a feasible solution to the linear program for MDPs with reachability objectives, i.e., given the memoryless strategy $\overline{\gamma}$ for player 1 the maximal probability valuation for player 2 to reach T is at most w. Hence the memoryless strategy $\overline{\gamma}$ ensures that the probability valuation for player 1 to stay safe in F against all player-2 strategies is at least $v = \langle\langle 1\rangle\rangle_{\mathsf{val}}(\mathsf{Safe}(F))$. Optimality of $\overline{\gamma}$ follows. □

7. Strategy-improvement algorithm for turn-based stochastic safety games

In this section we present a strategy-improvement algorithm for turn-based stochastic games with safety objectives. We consider a turn-based stochastic game graph with a safe set F, i.e., the objective for player 1 is $\mathsf{Safe}(F)$. The algorithm will produce a sequence of *pure* selectors $\gamma_0, \gamma_1, \gamma_2, \ldots$ for player 1, such that condition (i), condition (ii) and condition (iii) of Section 5 are satisfied. Since we consider turn-based stochastic games, we will also guarantee termination. We start with a few notations:

Value class of a valuation. Given a valuation v and a real $0 \leqslant r \leqslant 1$, the *value class* $U_r(v)$ of value r is the set of states with valuation r, i.e., $U_r(v) = \{s \in S \mid v(s) = r\}$.

Turn-based reduction. Given a turn-based stochastic game $G = \langle (S, E), (S_1, S_2, S_R), \delta\rangle$, and a valuation v such that $v = Pre_1(v)$, we construct another turn-based stochastic game $\overline{G}_v = \langle (S, \overline{E}), (S_1, S_2, S_R), \delta\rangle$ as follows: $\overline{E} = E \cap \{(s,t) \mid$ either (i) $s \in S_R$ or (ii) $s \in S_1 \cup S_2,\ t \in U_{v(s)}(v)\}$. In other words, for all player-1 and player-2 states we only retain edges that belong to the same value class. Given a turn-based stochastic game with safe set F, we refer to the above reduction as TB, i.e., $(\overline{G}_v, F) = \mathsf{TB}(G, v, F)$.

Ordering of strategies. Let G be a turn-based stochastic game and F be the set of safe states. Let $T = S \setminus F$. The set of *almost-sure winning* states is the set of states s such that the value at s is 1, i.e., $W_1 = \{s \in S \mid \langle\langle 1\rangle\rangle_{\mathsf{val}}(\mathsf{Safe}(F)) = 1\}$ is the set of almost-sure winning states. An optimal strategy from W_1 is referred as an almost-sure winning strategy. The set W_1 and an almost-sure winning strategy can be computed in linear time by the algorithm given in [9]. We assume without loss of generality that all states in $W_1 \cup T$ are absorbing. We recall the preorder $\prec$ on the strategies for player 1 (as defined in Section 5.1) as follows: given two player-1 strategies π_1 and π_1', let $\pi_1 \prec \pi_1'$ if the following two conditions hold: (i) $\langle\langle 1\rangle\rangle_{\mathsf{val}}^{\pi_1}(\mathsf{Safe}(F)) \leqslant \langle\langle 1\rangle\rangle_{\mathsf{val}}^{\pi_1'}(\mathsf{Safe}(F))$; and (ii) $\langle\langle 1\rangle\rangle_{\mathsf{val}}^{\pi_1}(\mathsf{Safe}(F))(s) < \langle\langle 1\rangle\rangle_{\mathsf{val}}^{\pi_1'}(\mathsf{Safe}(F))(s)$ for some state $s \in S$. Furthermore, we write $\pi_1 \preccurlyeq \pi_1'$ if either $\pi_1 \prec \pi_1'$ or $\pi_1 = \pi_1'$. We first present an example that shows the improvements based only on Pre_1 operators are not sufficient for safety games and then present our algorithm.

Example 2. Consider the turn-based stochastic game shown in Fig. 2, where the □ states are player-1 states, the ◇ states are player-2 states, and ◯ states are random states with probabilities labeled on edges. The safety goal is to avoid the

Algorithm 2 Safety strategy-improvement algorithm

```
Input: a turn-based stochastic game graph $G$ with safe set $F$.
Output: a pure memoryless strategy $\overline{\gamma}$ for player 1.
0. Compute $W_1 = \{s \in S \mid \langle\langle 1\rangle\rangle_{\mathsf{val}}(\mathsf{Safe}(F))(s) = 1\}$.
1. Let $\gamma_0$ be an arbitrary pure memoryless strategy and $i = 0$.
2. Compute $v_0 = \langle\langle 1\rangle\rangle_{\mathsf{val}}^{\overline{\gamma}_0}(\mathsf{Safe}(F))$.
3. do {
    3.1. Let $I = \{s \in S \setminus (W_1 \cup T) \mid Pre_1(v_i)(s) > v_i(s)\}$.
    3.2. if $I \neq \emptyset$, then
        3.2.1. Let $\xi_1$ be a player-1 pure selector such that for all states $s \in I$, we have $Pre_{1:\xi_1}(v_i)(s) = Pre_1(v_i)(s) > v_i(s)$.
        3.2.2. The player-1 selector $\gamma_{i+1}$ is defined as follows: for each state $s \in S$, let
            $\gamma_{i+1}(s) = \begin{cases} \gamma_i(s) & \text{if } s \notin I; \\ \xi_1(s) & \text{if } s \in I. \end{cases}$
    3.3. else
        3.3.1. Let $(\overline{G}_{v_i}, F) = \mathsf{TB}(G, v_i, F)$.
        3.3.2. Let $\overline{A}_i$ be the set of almost-sure winning states in $\overline{G}_{v_i}$ for $\mathsf{Safe}(F)$ and
            $\overline{\pi}_1$ be a pure memoryless almost-sure winning strategy from the set $\overline{A}_i$.
        3.3.3. if $(\overline{A}_i \setminus W_1 \neq \emptyset)$
            3.3.3.1. Let $U = \overline{A}_i \setminus W_1$.
            3.3.3.2. The player-1 selector $\gamma_{i+1}$ is defined as follows: for $s \in S$, let
                $\gamma_{i+1}(s) = \begin{cases} \gamma_i(s) & \text{if } s \notin U; \\ \overline{\pi}_1(s) & \text{if } s \in U. \end{cases}$
    3.4. Compute $v_{i+1} = \langle\langle 1\rangle\rangle_{\mathsf{val}}^{\overline{\gamma}_{i+1}}(\mathsf{Safe}(F))$.
    3.5. Let $i = i + 1$.
} until $I = \emptyset$ and $\overline{A}_{i-1} \setminus W_1 = \emptyset$.
4. return $\overline{\gamma}_i$.
```

state s_4. Consider a memoryless strategy π_1 for player 1 that chooses the successor $s_0 \to s_2$, and the counter-strategy π_2 for player 2 chooses $s_1 \to s_0$. Given the strategies π_1 and π_2, the value at s_0, s_1 and s_2 is $1/3$, and since all successors of s_0 have value $1/3$, the value cannot be improved by Pre_1. However, note that if player 2 is restricted to choose only value-optimal selectors for the value $1/3$, then player 1 can switch to the strategy $s_0 \to s_1$ and ensure that the game stays in the value class $1/3$ with probability 1. Hence switching to $s_0 \to s_1$ would force player 2 to select a counter-strategy that switches to the strategy $s_1 \to s_3$, and thus player 1 can get a value $2/3$.

Informal description of Algorithm 2. The algorithm (Algorithm 2) iteratively improves player-1 strategies according to the preorder $\prec$. Since we consider turn-based stochastic games, we will only consider pure memoryless strategies. The algorithm starts with an arbitrary pure selector γ_0. At iteration $i+1$, the algorithm considers the pure memoryless player-1 strategy $\overline{\gamma}_i$ and computes the value $\langle\langle 1\rangle\rangle_{\mathsf{val}}^{\overline{\gamma}_i}(\mathsf{Safe}(F))$. Observe that since $\overline{\gamma}_i$ is a pure memoryless strategy, the computation of $\langle\langle 1\rangle\rangle_{\mathsf{val}}^{\overline{\gamma}_i}(\mathsf{Safe}(F))$ involves solving the 2-MDP $G_{\overline{\gamma}_i}$. The valuation $\langle\langle 1\rangle\rangle_{\mathsf{val}}^{\overline{\gamma}_i}(\mathsf{Safe}(F))$ is named v_i. For all states s such that $Pre_1(v_i)(s) > v_i(s)$, the memoryless strategy at s is modified to a selector that is value-optimal for v_i. The algorithm then proceeds to the next iteration. If $Pre_1(v_i) = v_i$, then the algorithm constructs the game $(\overline{G}_{v_i}, F) = \mathsf{TB}(G, v_i, F)$, and computes $\overline{A}_i$ as the set of almost-sure winning states in $\overline{G}_{v_i}$ for the objective $\mathsf{Safe}(F)$. Let $U = \overline{A}_i \setminus W_1$. If U is nonempty, then a selector γ_{i+1} is obtained at U from a pure memoryless optimal strategy (i.e., an almost-sure winning strategy) in $\overline{G}_{v_i}$, and the algorithm proceeds to iteration $i+1$. If $Pre_1(v_i) = v_i$ and U is empty, then the algorithm stops and returns the memoryless strategy $\overline{\gamma}_i$ for player 1. We will now prove the monotonicity and optimality on termination, and for uniformity we keep the presentation and proof structure similar to the proofs for reachability games. We will use the following simple fact in the algorithm and the proofs: since we consider turn-based stochastic game, there is always a pure selector that is optimal for $Pre_1(v)$ for any valuation v. Also for a pure selector ξ_1, we will often write $t = \xi_1(s)$ to denote that $\xi_1(s)(t) = 1$.

Lemma 11. *Let γ_i and γ_{i+1} be the player-1 selectors obtained at iterations i and $i+1$ of Algorithm 2. Let $I = \{s \in S \setminus (W_1 \cup T) \mid Pre_1(v_i)(s) > v_i(s)\}$. Let $v_i = \langle\langle 1\rangle\rangle_{\mathsf{val}}^{\overline{\gamma}_i}(\mathsf{Safe}(F))$ and $v_{i+1} = \langle\langle 1\rangle\rangle_{\mathsf{val}}^{\overline{\gamma}_{i+1}}(\mathsf{Safe}(F))$. Then $v_{i+1}(s) \geqslant Pre_1(v_i)(s)$ for all states $s \in S$; and therefore $v_{i+1}(s) \geqslant v_i(s)$ for all states $s \in S$, and $v_{i+1}(s) > v_i(s)$ for all states $s \in I$.*

Proof. The proof is essentially similar to the proof of Lemma 8, and we present the details for completeness. Consider the valuations v_i and v_{i+1} obtained at iterations i and $i+1$, respectively, and let w_i be the valuation defined by $w_i(s) = 1 - v_i(s)$ for all states $s \in S$. The counter-optimal strategy for player 2 to minimize v_{i+1} is obtained by maximizing the probability to reach T. Let

$$\widehat{w}_i(s) = \begin{cases} w_i(s) & \text{if } s \in S \setminus I; \\ 1 - Pre_1(v_i)(s) < w_i(s) & \text{if } s \in I. \end{cases}$$

In other words, $\widehat{w}_i = 1 - Pre_1(v_i)$, and we also have $\widehat{w}_i \leqslant w_i$. We now show that $\widehat{w}_i$ is a feasible solution to the linear program for MDPs with the objective Reach(T), as described in Section 3. Since $v_i = \langle\langle 1\rangle\rangle_{\mathsf{val}}^{\overline{\gamma}_i}(\mathrm{Safe}(F))$, it follows that for all states $s \in S$ we have the following:

$$w_i(s) = w_i\big(\gamma_i(s)\big), \quad s \in S_1;$$
$$w_i(s) \geqslant w_i(t), \quad s \in S_2,\ t \in E(s);$$
$$w_i(s) = \sum_{t \in E(s)} w_i(t) \cdot \delta(s)(t), \quad s \in S_R.$$

Since for all states $s \in S \setminus I$, we have $\gamma_i(s) = \gamma_{i+1}(s)$ and $\widehat{w}_i(s) = w_i(s)$, and $\widehat{w}_i \leqslant w_i$, we have the following inequalities for $\widehat{w}_i$ for all states s in $S \setminus I$;

$$\widehat{w}_i(s) = \widehat{w}_i\big(\gamma_{i+1}(s)\big), \quad s \in S_1;$$
$$\widehat{w}_i(s) = w_i(s) \geqslant w_i(t) \geqslant \widehat{w}_i(t), \quad s \in S_2,\ t \in E(s);$$
$$w_i(s) = w_i(s) = \sum_{t \in E(s)} w_i(t) \cdot \delta(s)(t) \geqslant \sum_{t \in E(s)} \widehat{w}_i(t) \cdot \delta(s)(t), \quad s \in S_R.$$

Since for $s \in I$ the selector $\gamma_{i+1}(s)$ is obtained as an optimal selector for $Pre_1(v_i)(s)$, it follows that for all states $s \in I$ we have

$$\widehat{w}_i(s) \geqslant w_i\big(\gamma_{i+1}(s)\big) \geqslant \widehat{w}_i\big(\gamma_{i+1}(s)\big).$$

Observe that every state in I is a player-1 state (i.e., $I \subseteq S_1$), since in player-2 and random states the Pre_1 operator does not increase value as the valuation is obtained as the optimal valuation of an MDP.

Hence it follows that $\widehat{w}_i$ is a feasible solution to the linear program for MDPs with reachability objectives. Since the reachability valuation for player 2 for Reach(T) is the least solution (observe that the objective function of the linear program is a minimizing function), it follows that $v_{i+1} \geqslant 1 - \widehat{w}_i = Pre_1(v_i)$. Thus we obtain $v_{i+1}(s) \geqslant v_i(s)$ for all states $s \in S$, and $v_{i+1}(s) > v_i(s)$ for all states $s \in I$. □

Recall that by Example 2 it follows that improvement by only step 3.2 is not sufficient to guarantee convergence to optimal values. We now present a lemma about the turn-based reduction, and then show that step 3.3 also leads to an improvement. Finally, in Theorem 7 we show that if improvements by step 3.2 and step 3.3 are not possible, then the optimal value and an optimal strategy is obtained.

Lemma 12. *Let G be a turn-based stochastic game graph with a set F of safe states. Let v be a valuation and consider $(\overline{G}_v, F) =$ $\mathsf{TB}(G, v, F)$. Let $\overline{A}$ be the set of almost-sure winning states in $\overline{G}_v$ for the objective* Safe(F)*, and let $\overline{\pi}_1$ be a pure memoryless almost-sure winning strategy from $\overline{A}$ in $\overline{G}_v$. Consider a pure memoryless strategy $\overline{\pi}_2$ for player 2. If for all states $s \in \overline{A} \cap S_2$, we have $\overline{\pi}_2(s) \in U_{v(s)}(v)$ (i.e., player 2 selects edges in the same value class), then for all $s \in \overline{A}$, we have $\mathrm{Pr}_s^{\overline{\pi}_1,\overline{\pi}_2}(\mathrm{Safe}(F)) = 1$.*

Proof. We analyze the Markov chain arising after the player fixes the memoryless strategies $\overline{\pi}_1$ and $\overline{\pi}_2$. Since $\overline{\pi}_1$ is an almost-sure winning strategy for Safe(F) in $\overline{G}_v$ and $\overline{\pi}_2$ is a strategy in $\overline{G}_v$, it follows that in the Markov chain obtained by fixing $\overline{\pi}_1$ and $\overline{\pi}_2$ in $\overline{G}_v$, all closed connected recurrent set of states that intersect with $\overline{A}$ are contained in $\overline{A}$, and from all states of $\overline{A}$ the closed connected recurrent set of states within $\overline{A}$ are reached with probability 1. The desired result follows. □

Lemma 13. *Let γ_i and γ_{i+1} be the player-1 selectors obtained at iterations i and $i+1$ of Algorithm 2. Let $I = \{s \in S \setminus (W_1 \cup T) \mid$ $Pre_1(v_i)(s) > v_i(s)\} = \emptyset$, and $U = \overline{A}_i \setminus W_1 \neq \emptyset$. Let $v_i = \langle\langle 1\rangle\rangle_{\mathsf{val}}^{\overline{\gamma}_i}(\mathrm{Safe}(F))$ and $v_{i+1} = \langle\langle 1\rangle\rangle_{\mathsf{val}}^{\overline{\gamma}_{i+1}}(\mathrm{Safe}(F))$. Then $v_{i+1}(s) \geqslant v_i(s)$ for all states $s \in S$, and $v_{i+1}(s) > v_i(s)$ for some state $s \in U$.*

Proof. Note that since $I = \emptyset$, we have that $v_i = Pre_1(v_i)$. We first show that $v_{i+1} \geqslant v_i$. Let $w_i(s) = 1 - v_i(s)$ for all states $s \in S$. Since $v_i = \langle\langle 1\rangle\rangle_{\mathsf{val}}^{\overline{\gamma}_i}(\mathrm{Safe}(F))$, it follows that for all $s \in S$ we have

$$w_i(s) = w_i\big(\gamma_i(s)\big), \quad s \in S_1;$$
$$w_i(s) \geqslant w_i(t), \quad s \in S_2,\ t \in E(s);$$
$$w_i(s) = \sum_{t \in E(s)} w_i(t) \cdot \delta(s)(t), \quad s \in S_R.$$

Let us consider $U = \overline{A}_i \setminus W_1$. The selector $\xi_1(s)$ chosen for γ_{i+1} at $s \in U$ satisfies that $\xi_1(s) \in U_{v_i(s)}(v_i)$ (i.e., the selector chooses an edge from the same value class). It follows that for all states $s \in U$ we have

$$w_i(s) = w_i\big(\gamma_{i+1}(s)\big).$$

It follows (similar to the argument for Lemma 11) that the maximal probability with which player 2 can reach T against the strategy $\overline{\gamma}_{i+1}$ is at most w_i. It follows that $v_i(s) \leqslant v_{i+1}(s)$.

We now argue that for some state $s \in U$ we have $v_{i+1}(s) > v_i(s)$. Given the strategy $\overline{\gamma}_{i+1}$, consider a pure memoryless counter-optimal strategy π_2 for player 2 to reach T. Since the selectors $\gamma_{i+1}(s)$ at states $s \in U$ are obtained from the almost-sure strategy $\overline{\pi}$ in the turn-based game $\overline{G}_{v_i}$ to satisfy $\mathsf{Safe}(F)$, it follows from Lemma 12 that if for every state $s \in U$, we have $\pi_2(s) \in U_{v_i(s)}(v_i)$, then from all states $s \in U$, the game stays safe in F with probability 1. Since $\overline{\gamma}_{i+1}$ is a given strategy for player 1, and π_2 is counter-optimal against $\overline{\gamma}_{i+1}$, this would imply that $U \subseteq \{s \in S \mid \langle\langle 1\rangle\rangle_{\mathsf{val}}(\mathsf{Safe}(F)) = 1\}$. This would contradict that $W_1 = \{s \in S \mid \langle\langle 1\rangle\rangle_{\mathsf{val}}(\mathsf{Safe}(F)) = 1\}$ and $U \cap W_1 = \emptyset$. It follows that for some state $s^* \in U$ we have $\pi_2(s^*) \notin U_{v_i(s^*)}(v_i)$. Since for all $s \in S_2$ and $t \in E(s)$ we have $v_i(s) \leqslant v_i(t)$ (i.e. $w_i(s) \geqslant w_i(t)$), we must have that $v_i(s) < v_i(\pi_2(s^*))$ (i.e., $w_i(s) < w_i(\pi_2(s^*))$). Define a valuation z as follows: $z(s) = w_i(s)$ for $s \neq s^*$, and $z(s^*) = w_i(\pi_2(s^*))$. Given the strategy $\overline{\gamma}_{i+1}$ and the counter-optimal strategy π_2, the valuation z satisfies the inequalities of the linear program for reachability to T. It follows that the probability to reach T given $\overline{\gamma}_{i+1}$ is at most z. Thus we obtain that $v_{i+1}(s) \geqslant v_i(s)$ for all $s \in S$, and $v_{i+1}(s^*) > v_i(s^*)$. This concludes the proof. □

We obtain the following theorem from Lemma 11 and Lemma 13 that shows that the sequences of values we obtain is monotonically non-decreasing.

Theorem 6 *(Monotonicity of values). For $i \geqslant 0$, let γ_i and γ_{i+1} be the player-1 selectors obtained at iterations i and $i+1$ of Algorithm 2. If $\gamma_i \neq \gamma_{i+1}$, then (a) for all $s \in S$ we have $\langle\langle 1\rangle\rangle_{\mathsf{val}}^{\overline{\gamma}_i}(\mathsf{Safe}(F))(s) \leqslant \langle\langle 1\rangle\rangle_{\mathsf{val}}^{\overline{\gamma}_{i+1}}(\mathsf{Safe}(F))(s)$; and (b) for some $s^* \in S$ we have $\langle\langle 1\rangle\rangle_{\mathsf{val}}^{\overline{\gamma}_i}(\mathsf{Safe}(F))(s^*) < \langle\langle 1\rangle\rangle_{\mathsf{val}}^{\overline{\gamma}_{i+1}}(\mathsf{Safe}(F))(s^*)$.*

Theorem 7 *(Optimality on termination). Let v_i be the valuation at iteration i of Algorithm 2 such that $v_i = \langle\langle 1\rangle\rangle_{\mathsf{val}}^{\overline{\gamma}_i}(\mathsf{Safe}(F))$. If $I = \{s \in S \setminus (W_1 \cup T) \mid Pre_1(v_i)(s) > v_i(s)\} = \emptyset$, and $U = \overline{A}_i \setminus W_1 = \emptyset$, then $\overline{\gamma}_i$ is an optimal strategy and $v_i = \langle\langle 1\rangle\rangle_{\mathsf{val}}(\mathsf{Safe}(F))$.*

Proof. We show that for all pure memoryless strategies π_1 for player 1 we have $\langle\langle 1\rangle\rangle_{\mathsf{val}}^{\pi_1}(\mathsf{Safe}(F)) \leqslant v_i$. Since pure memoryless optimal strategies exist for turn-based stochastic games with safety objectives [5], the desired result will follow.

Let π_2 be a pure memoryless optimal strategy for player 2 in $\overline{G}_{v_i}$ for the objective complementary to $\mathsf{Safe}(F)$, where $(\overline{G}_{v_i}, F) = \mathsf{TB}(G, v_i, F)$. Consider a pure memoryless strategy π_1 for player 1. We first show that in the Markov chain obtained by fixing π_1 and π_2 in G, there is no closed connected recurrent set of states C such that $C \subseteq S \setminus (W_1 \cup T)$. Assume towards contradiction that C is a closed connected recurrent set of states in $S \setminus (W_1 \cup T)$. The following case analysis achieves the contradiction.

(i) Suppose for every state $s \in C$ we have $\pi_1(s) \in U_{v_i(s)}(v_i)$ (i.e., player 1 chooses edges in the value class defined by v_i). Since C is closed connected recurrent states, it follows by construction that for all states $s \in C$ in the game $\overline{G}_{v_i}$ we have $\mathrm{Pr}_s^{\pi_1,\pi_2}(\mathsf{Safe}(C)) = 1$. It follows that for all $s \in C$ in $\overline{G}_{v_i}$ we have $\mathrm{Pr}_s^{\pi_1,\pi_2}(\mathsf{Safe}(F)) = 1$. Since π_2 is an optimal strategy $\overline{G}_{v_i}$, it follows that $C \subseteq U = \overline{A}_i \setminus W_1$. This contradicts that $U = \emptyset$.

(ii) Otherwise for some state $s^* \in C$ we have $\pi_1(s^*) \notin U_{v_i(s)}(v_i)$. Let $r = \min\{q \mid U_q(v_i) \cap C \neq \emptyset\}$, i.e., r is the least value class with nonempty intersection with C. Hence it follows that for all $q < r$, we have $U_q(v_i) \cap C = \emptyset$. For all $s \in C \cap U_r(v_i)$ we have the following case analysis:
   (a) If $s \in S_1$, then since $Pre_1(v_i) = v_i$ (i.e., for all $t \in E(s)$ we have $v_i(t) \leqslant v_i(s)$) and $C \cap U_q(v_i) = \emptyset$ for all $q < r$, we must have that $\pi_1(s) \in U_r(v_i)$.
   (b) If $s \in S_2$, then $\pi_2(s) \in U_r(v_i)$ (since for all $s \in S_2$ in $\overline{G}_{v_i}$ we only retain edges in the same value class, i.e., $\overline{E}(s) \subseteq U_{v_i(s)}(v_i)$).
   (c) For $s \in S_R$, we must have $E(s) \subseteq C$ as C is a closed set. We argue that have $E(s) \subseteq U_r(v_i)$: the reasoning is as follows we have $Pre(v_i) = v_i$ and if $E(s) \cap (S \setminus U_r(v_i)) \neq \emptyset$, then $E(s) \cap U_q(v_i) \neq \emptyset$ for some $q < r$, leading to a contradiction that for all $q < r$ we have $C \cap U_q = \emptyset$.

   It follows that $C \subseteq U_r(v_i)$. Consider the state $s^* \in C$ such that $\pi_1(s^*) \notin U_{v_i(s)}(v_i)$. Hence we must have $\pi_1(s^*) \in U_q(v_i)$, for some $q < r$ and hence we must have $C \cap U_q(v_i) \neq \emptyset$ (since C is closed we have $\pi_1(s^*) \in C$). Thus we have a contradiction.

It follows from above that there is no closed connected recurrent set of states in $S \setminus (W_1 \cup T)$, and hence with probability 1 the game reaches $W_1 \cup T$ from all states in $S \setminus (W_1 \cup T)$. Hence the probability to satisfy $\mathsf{Safe}(F)$ is equal to the probability to reach W_1. For all states $s \in S \setminus (W_1 \cup T)$ we have

$$v_i(s) = v_i\big(\pi_2(s)\big), \quad s \in S_2 \quad \text{(by construction)};$$

$$v_i(s) \leqslant v_i(t), \quad s \in S_1,\ t \in E(s) \ \big(\text{by } Pre_1(v_i) = v_i\big);$$

$$v_i(s) = \sum_{t \in E(s)} v_i(t) \cdot \delta(s)(t), \quad s \in S_R \ \big(\text{by } Pre_1(v_i) = v_i\big).$$

It follows that given the strategies π_1 and π_2, the valuation v_i satisfies all the inequalities for linear program to reach W_1. It follows that the probability to reach W_1 from s is at most $v_i(s)$. It follows that for all $s \in S \setminus (W_1 \cup T)$ we have $\langle\langle 1\rangle\rangle_{\mathsf{val}}^{\pi_1}(\mathsf{Safe}(F))(s) \leqslant v_i(s)$. The result follows. □

Convergence. The convergence of Algorithm 2 is guaranteed by monotonicity and the fact that it only considers pure memoryless strategies (and the number of pure memoryless strategies is bounded). Hence it follows that Algorithm 2 computes a monotonically increasing sequence of valuations that converges from below to the optimal value of a turn-based stochastic game with a safety objective, and outputs a pure memoryless optimal strategy.

Retraction of Theorem 4.3 of [3]. In [3], a variant of Algorithm 2 was presented for the more general case of concurrent games and it was claimed in Theorem 4.3 that the valuations converge to the value of the concurrent safety game. Unfortunately the theorem is incorrect (with irreparable error) and we retract Theorem 4.3 of [3]. There is an explicit counter-example to Theorem 4.3 of [3] of the claim of convergence to values (this is demonstrated by Example 3 in [2]).

Acknowledgments

We thank anonymous reviewers for many helpful and insightful comments that significantly improved the presentation of the paper, and help us fix gaps in the results of the conference version. We warmly acknowledge their help.

References

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

[2] K. Chatterjee, L. de Alfaro, T.A. Henzinger, Strategy improvement in concurrent reachability and safety games, CoRR, arXiv:1201.2834, 2012.

[3] K. Chatterjee, L. de Alfaro, T.A. Henzinger, Termination criteria for solving concurrent safety and reachability games, in: SODA, ACM–SIAM, 2009, pp. 197–206.

[4] K. Chatterjee, L. de Alfaro, T.A. Henzinger, Strategy improvement in concurrent reachability games, in: QEST'06, IEEE, 2006.

[5] A. Condon, The complexity of stochastic games, Inform. and Comput. 96 (2) (1992) 203–224.

[6] A. Condon, On algorithms for simple stochastic games, in: Advances in Computational Complexity Theory, in: DIMACS Ser. Discrete Math. Theoret. Comput. Sci., vol. 13, American Mathematical Society, 1993, pp. 51–73.

[7] C. Courcoubetis, M. Yannakakis, The complexity of probabilistic verification, J. ACM 42 (4) (1995) 857–907.

[8] L. de Alfaro, Formal verification of probabilistic systems, PhD thesis, Stanford University, 1997, Technical Report STAN-CS-TR-98-1601.

[9] L. de Alfaro, T.A. Henzinger, Concurrent omega-regular games, in: Proceedings of the 15th Annual Symposium on Logic in Computer Science, IEEE Computer Society Press, 2000, pp. 141–154.

[10] L. de Alfaro, T.A. Henzinger, O. Kupferman, Concurrent reachability games, Theoret. Comput. Sci. 386 (3) (2007) 188–217.

[11] L. de Alfaro, R. Majumdar, Quantitative solution of omega-regular games, J. Comput. System Sci. 68 (2004) 374–397.

[12] C. Derman, Finite State Markovian Decision Processes, Academic Press, 1970.

[13] K. Etessami, M. Yannakakis, Recursive concurrent stochastic games, in: ICALP 06: Automata, Languages, and Programming, Springer, 2006.

[14] H. Everett, Recursive games, in: Contributions to the Theory of Games III, in: Ann. of Math. Stud., vol. 39, 1957, pp. 47–78.

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

[16] H. Gimbert, F. Horn, Simple stochastic games with few random vertices are easy to solve, in: FoSSaCS'08, 2008.

[17] A.J. Hoffman, R.M. Karp, On nonterminating stochastic games, Management Sci. 12 (5) (1966) 359–370.

[18] R.A. Howard, Dynamic Programming and Markov Processes, MIT Press, 1960.

[19] R. Ibsen-Jensen, P.B. Miltersen, Solving simple stochastic games with few coin toss positions, in: ESA, Springer, 2012, pp. 636–647.

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

[21] J.F. Mertens, A. Neyman, Stochastic games, Internat. J. Game Theory 10 (1981) 53–66.

[22] T. Parthasarathy, Discounted and positive stochastic games, Bull. Amer. Math. Soc. 77 (1) (1971) 134–136.

[23] S.S. Rao, R. Chandrasekaran, K.P.K. Nair, Algorithms for discounted games, J. Optim. Theory Appl. (1973) 627–637.

[24] L.S. Shapley, Stochastic games, Proc. Natl. Acad. Sci. USA 39 (1953) 1095–1100.

[25] U. Zwick, M.S. Paterson, The complexity of mean payoff games on graphs, Theoret. Comput. Sci. 158 (1996) 343–359.