Memoryless Strategies in
Stochastic Reachability Games

Stefan Kiefer[1], Richard Mayr[2], Mahsa Shirmohammadi[3], and Patrick Totzke[4]

[1] University of Oxford, UK
[2] University of Edinburgh, UK
[3] IRIF & CNRS, Université Paris cité, France
[4] University of Liverpool, UK

Abstract. We study concurrent stochastic reachability games played on finite graphs. Two players, Max and Min, seek respectively to maximize and minimize the probability of reaching a set of target states. We prove that Max has a memoryless strategy that is optimal from all states that have an optimal strategy. Our construction provides an alternative proof of this result by Bordais, Bouyer and Le Roux [4], and strengthens it, as we allow Max's action sets to be countably infinite.

Keywords: Stochastic games · Reachability · Strategy Complexity.

1   Introduction

Background. We study 2-player zero-sum stochastic games. First introduced by Shapley in his seminal 1953 work [22], they model dynamic interactions in which the environment responds randomly to players' actions. Shapley's games were generalized in [11] and [16] to allow infinite state and action sets and non-termination.

In *concurrent games*, the two players (Max and Min) jointly create an infinite path through a directed graph. In each round of the play, both independently choose an action. The next state is then determined according to a pre-defined distribution that depends on the current state and the chosen pair of actions. *Turn-based games* (also called switching-control games) are a subclass where each state is owned by some player and only this player gets to choose an action. These have received much attention by computer scientists, e.g., [12,6,5,14,2]. An even more special case of stochastic games are *Markov Decision Processes (MDPs)* where all states are owned by Max (i.e., Min is passive). MDPs are also called *games against nature*.

We consider *reachability objectives* which are defined w.r.t. a given subset of *target* states. A play is defined as winning for Max iff it visits a target state at least once. Thus Max aims to maximize the probability that the target set is reached. Dually, Min aims to minimize the probability of reaching the target. So Min pursues the dual *safety objective* of avoiding the target.

Reachability is arguably the simplest objective in games on graphs. It can trivially be encoded into reward-based objectives; i.e., every play that reaches the target gets reward 1 and all other plays get reward 0.

In turn-based reachability games over finite state spaces, there always exist optimal Max strategies that are memoryless (choices only depend on the current state and not the history of the play; this is also called positional) and deterministic (always chose on action as opposed to randomising among several) [7], [17, Proposition 5.6.c, Proposition 5.7.c].

This does not carry over to finite concurrent reachability games. E.g., in the *snowball game* (aka *Hide-or-Run* game) [9, Example 1][16,1] (also see Example 4 later in this paper), Max does not have an optimal strategy. However, by [21, Corollary 3.9], Max always has ε-optimal randomized memoryless strategies, and this holds even in countably infinite reachability games with finite action sets. (Deterministic strategies are generally useless in concurrent games, even in very simple games such as *Rock-Paper-Scissors*.)

Even though optimal Max strategies do not always exist, it is still interesting how much memory they need in those instances where they do exist. For finite concurrent reachability games with finite action sets, it was recently shown by Bordais, Bouyer and Le Roux [4] that optimal Max strategies, if they exist, can be chosen as randomized memoryless. The proof is constructive and iteratively builds the strategy on the finite state space. To show the correctness, one needs to argue about the performance of the constructed strategy. The proof in [4] heavily relies on the properties of induced finite MDPs, obtained by fixing one strategy in the game. In particular, it uses the existence of end components in these finite MDPs and their particular properties.

Our contribution. We give an alternative proof of this result by Bordais, Bouyer and Le Roux [4]. While our proof is also constructive, it is simpler and works directly from first principles on games, without using properties of induced MDPs. Moreover, it uses a construction that we call "leaky games", by which we reduce the reachability objective to its dual safety objective. Finally, our result is slightly stronger, because it holds even if Max is allowed countably infinite action sets (while Min still has finite action sets). [5]

This result requires the state space to be finite. However, for other results in this paper we allow the state space and action sets to be countably infinite, unless explicitly stated otherwise.

2   Preliminaries

A *probability distribution* over a countable set S is a function $\alpha : S \to [0,1]$ with $\sum_{s \in S} \alpha(s) = 1$. The *support* of α is the set $\{s \in S \mid \alpha(s) > 0\}$. We write $\mathcal{D}(S)$ for the set of all probability distributions over S. For $\alpha \in \mathcal{D}(S)$ and a

---

[5] It may be possible to generalize the proof in [4] to countably infinite Max actions sets, but this would require (at least) a generalization of the fixpoint theorem [4, Theorem 12] to this setting.

function $v : S \to \mathbb{R}$ we write $\langle \alpha, v \rangle \stackrel{\text{def}}{=} \sum_{s \in S} \alpha(s) v(s)$ for the expectation of v with respect to α.

Stochastic Games

We study perfect information stochastic games between two players, Max(imizer) and Min(imizer). A *(concurrent) game* $\mathcal{G}$ is played on a countable set of states S. For each state $s \in S$ there are nonempty countable *action* sets $A(s)$ and $B(s)$ for Max and Min, respectively. A *mixed action* for Max (resp. Min) in state s is a distribution over $A(s)$ (resp. $B(s)$).

Let $Z \stackrel{\text{def}}{=} \{(s,a,b) \mid s \in S, \ a \in A(s), \ b \in B(s)\}$. For every triple $(s,a,b) \in Z$ there is a distribution $p(s,a,b) \in \mathcal{D}(S)$ over successor states. We call a state $s \in S$ a *sink* state if $p(s,a,b) = s$ for all $a \in A(s)$ and $b \in B(s)$. We extend the *transition function* p to mixed actions $\alpha \in \mathcal{D}(A(s))$ and $\beta \in \mathcal{D}(B(s))$ by letting

$$p(s, \alpha, \beta) \stackrel{\text{def}}{=} \sum_{a \in A(s)} \sum_{b \in B(s)} \alpha(a)\beta(b)p(s,a,b),$$

which is a distribution over S. A *play* from an initial state s_0 is an infinite sequence in Z^ω where the first triple contains s_0. Starting from s_0, the game is played in stages $\mathbb{N} = \{0, 1, 2, \dots\}$. At every stage $n \in \mathbb{N}$, the play is in some state s_n. Max chooses a mixed action $\alpha_n \in \mathcal{D}(A(s_n))$ and Min chooses a mixed action $\beta_n \in \mathcal{D}(B(s_n))$. The next state s_{n+1} is then chosen according to the distribution $p(s_n, a_n, b_n)$.

Strategies and Probability Measures

The set of *histories* at stage n, with $n \in \mathbb{N}$, is denoted by H_n. That is, $H_0 \stackrel{\text{def}}{=} S$ and $H_n \stackrel{\text{def}}{=} Z^n \times S$ for all $n > 0$. Let $H \stackrel{\text{def}}{=} \bigcup_{n \in \mathbb{N}} H_n$ be the set of all histories; note that H is countable. For each history $h = (s_0, a_0, b_0) \cdots (s_{n-1}, a_{n-1}, b_{n-1})s_n \in H_n$, let $s_h \stackrel{\text{def}}{=} s_n$ denote the final state in h.

A *strategy* for Max is a function σ that to each history $h \in H$ assigns a mixed action $\sigma(h) \in \mathcal{D}(A(s_h))$. Denote by Σ the set of strategies for Max. Analogously, a *strategy* for Min is a function π that to each history h assigns a mixed action $\pi(h) \in \mathcal{D}(B(s_h))$, and Π denotes the set of strategies for Min. A Max strategy is called *memoryless* if $\sigma(h)$ depends only on s_h; i.e., for all $h, h' \in H$ with $s_h = s_{h'}$ we have $\sigma(h) = \sigma(h')$. A memoryless strategy σ is fully determined by $(\sigma(s))_{s \in S}$. A *memory-based strategy* bases its decisions not only on the current state, but also on the current mode of its memory, and it can update its memory at every step, depending on the observed events in this step. Memory is called *public* if the content is also observable by the opposing player and *private* otherwise. Finite-memory strategies use a memory with only finitely many different possible modes. A step counter is a special case of infinite memory in the form of a discrete clock that gets incremented at every step, independently of the actions of the players. Strategies that use just a step counter are also called *Markov strategies*.

An initial state s_0 and a pair of strategies σ, π for Max and Min induce a probability measure on sets of plays. We write $\mathcal{P}_{\mathcal{G}, s_0, \sigma, \pi}(E)$ for the probability of a measurable set of plays E starting from s_0. It is initially defined for the cylinder sets generated by the histories and then extended to the sigma-algebra by Carathéodory's unique extension theorem [3]. Given a random variable $V :$ $Z^\omega \to \mathbb{R}$, we will write $\mathcal{E}_{\mathcal{G}, s_0, \sigma, \pi}(V)$ for the expectation of V w.r.t. $\mathcal{P}_{\mathcal{G}, s_0, \sigma, \pi}$. We may drop $\mathcal{G}$ from the subscript when it is understood.

Objectives

We consider reachability and safety objectives. Given a set $T \subseteq S$ of states, the *reachability* objective $\texttt{Reach}(T)$ is the set of plays that visit T at least once, i.e., $s_h \in T$ holds for some history h that is a prefix of the play. The dual *safety* objective $\texttt{Avoid}(T) \stackrel{\text{def}}{=} Z^\omega \setminus \texttt{Reach}(T)$ consists of the plays that never visit T. We can and will assume that $T = \{\top\}$ holds for a sink state $\top \in S$ and write $\texttt{Reach}(\top)$ for $\texttt{Reach}(\{\top\})$. Similarly, we assume that there is another sink state $\bot \in S$, with $\bot \neq \top$, and write $\texttt{Avoid}(\bot)$ for $\texttt{Avoid}(\{\bot\})$. Max attempts to maximize the probability of achieving the given objective (usually $\texttt{Reach}(\top)$ or $\texttt{Avoid}(\top)$), whereas Min attempts to minimize it.

Value and Optimality

For a game $\mathcal{G}$, objective E and initial state s_0, the *lower value* and *upper value* of s_0 are respectively defined as

$$\texttt{val}_{\mathcal{G}, E}^{\downarrow}(s_0) \stackrel{\text{def}}{=} \sup_{\sigma \in \Sigma} \inf_{\pi \in \Pi} \mathcal{P}_{\mathcal{G}, s_0, \sigma, \pi}(E) \quad \text{and} \quad \texttt{val}_{\mathcal{G}, E}^{\uparrow}(s_0) \stackrel{\text{def}}{=} \inf_{\pi \in \Pi} \sup_{\sigma \in \Sigma} \mathcal{P}_{\mathcal{G}, s_0, \sigma, \pi}(E).$$

The inequality $\texttt{val}_{\mathcal{G}, E}^{\downarrow}(s_0) \leq \texttt{val}_{\mathcal{G}, E}^{\uparrow}(s_0)$ holds by definition. If $\texttt{val}_{\mathcal{G}, E}^{\downarrow}(s_0) = \texttt{val}_{\mathcal{G}, E}^{\uparrow}(s_0)$, then this quantity is called the *value*, denoted by $\texttt{val}_{\mathcal{G}, E}(s_0)$. For reachability objectives, like all Borel objectives, the value exists if all action sets are finite [18], and even if for all states s we have that $A(s)$ is finite or $B(s)$ is finite [10, Theorem 11]. We always assume the latter, so that $\texttt{val}_{\mathcal{G}, E}(s_0)$ exists. For $\varepsilon \geq 0$, a Max strategy σ is called ε-*optimal* from s_0 if for all Min strategies π we have $\mathcal{P}_{\mathcal{G}, s_0, \sigma, \pi}(E) \geq \texttt{val}_{\mathcal{G}, E}(s_0) - \varepsilon$. A 0-optimal strategy is also called *optimal*. For $E = \texttt{Reach}(\top)$ or $E = \texttt{Avoid}(\bot)$ we have

$$\texttt{val}_{\mathcal{G}, E}(s_0) = \sup_{\alpha \in \mathcal{D}(A(s_0))} \inf_{\beta \in \mathcal{D}(B(s_0))} \langle p(s_0, \alpha, \beta), \texttt{val}_{\mathcal{G}, E} \rangle. \tag{1}$$

We note in passing that the equality in (1) also holds in the states $\top, \bot$, as these are sink states.

3 Martingales for Safety

For the remainder of the paper we fix a game $\mathcal{G}$ over a countable state space S. Whenever we make finiteness assumptions on S and the action sets, we state

them explicitly. Several times we use the following lemma, a consequence of the optional-stopping theorem for submartingales.

Lemma 1. *Let σ and π be Max and Min strategies, respectively. Suppose v : $S \to [0,1]$ is a function with $v(\bot) = 0$ such that $\langle p(s_h, \sigma(h), \pi(h)), v \rangle \geq v(s_h)$ holds for all histories h. Then $\mathcal{P}_{s_0,\sigma,\pi}(\texttt{Avoid}(\bot)) \geq v(s_0)$ holds for all $s_0 \in S$.*

Proof. Let $s_0 \in S$. Define a sequence of random variables $V(0), V(1), \ldots$ with $V(i) : Z^\omega \to [0,1]$ by

$$V(i)((s_0, a_0, b_0)(s_1, a_1, b_1) \cdots) \ \stackrel{\text{def}}{=} \ v(s_i) \,.$$

Similarly, denote by $H(i) : Z^\omega \to H_i$ the function that maps each random play to its unique prefix in H_i. Recall that we can create random variables via composition of functions. Consider the function $s : H \to S$ that maps each finite history h to its final state s_h; this function composed with $H(i)$ is the random variable $s_{H(i)} : Z^\omega \to S$. Moreover, the function v composed with $s_{H(i)}$ is the random variable $v(s_{H(i)})$. Note that $V(i) = v(s_{H(i)})$. For $i \geq 0$, let $\mathcal{F}_i$ be the sigma-algebra generated by the cylinder sets corresponding to the histories $h \in H_i$. Then $V(i)$ and $H(i)$ are $\mathcal{F}_i$-measurable. By the assumption on v we have

$$
\begin{aligned}
\mathcal{E}_{s_0,\sigma,\pi}(V(i+1) \mid \mathcal{F}_i) \ &= \ \mathcal{E}_{s_0,\sigma,\pi}(v(s_{H(i+1)}) \mid \mathcal{F}_i) \\
&= \ \sum_{s \in S} p(s_{H(i)}, \sigma(H(i)), \pi(H(i))) v(s) \\
&= \ \langle p(s_{H(i)}, \sigma(H(i)), \pi(H(i))), v \rangle \\
&\geq \ v(s_{H(i)}) \ = \ V(i) \,.
\end{aligned}
$$

It follows that $V(0), V(1), \ldots$ is a submartingale with respect to the filtrations $\mathcal{F}_0, \mathcal{F}_1, \ldots$. Moreover $|V(i)| \leq 1$ holds for all i. This boundedness condition together with the optional-stopping theorem imply that, almost surely, $V(0), V(1), \ldots$ converges to a random variable $V(\infty)$ with

$$\mathcal{E}_{s_0,\sigma,\pi}(V(\infty)) \ \geq \ \mathcal{E}_{s_0,\sigma,\pi}(V(0)) \ = \ v(s_0) \,.$$

On the other hand, since $\bot$ is a sink,

$$
\begin{aligned}
\mathcal{E}_{s_0,\sigma,\pi}(V(\infty)) \ &= \ \mathcal{P}_{s_0,\sigma,\pi}(\texttt{Avoid}(\bot)) \cdot \mathcal{E}_{s_0,\sigma,\pi}(V(\infty) \mid \texttt{Avoid}(\bot)) \\
&\quad + \mathcal{P}_{s_0,\sigma,\pi}(\texttt{Reach}(\bot)) \cdot \mathcal{E}_{s_0,\sigma,\pi}(V(\infty) \mid \texttt{Reach}(\bot)) \\
&\leq \ \mathcal{P}_{s_0,\sigma,\pi}(\texttt{Avoid}(\bot)) \cdot 1 + \mathcal{P}_{s_0,\sigma,\pi}(\texttt{Reach}(\bot)) \cdot 0 \\
&= \ \mathcal{P}_{s_0,\sigma,\pi}(\texttt{Avoid}(\bot)) \,. \qquad \square
\end{aligned}
$$

4 Optimal Safety

We will construct memoryless strategies for reachability from strategies for safety. The following proposition will be useful for this purpose and is of independent interest.
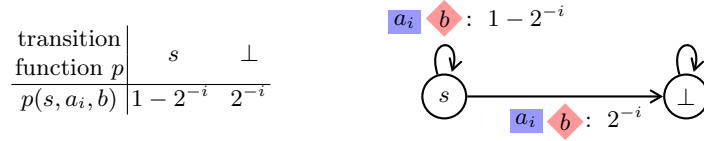
transition function p	s	$\perp$
$p(s, a_i, b)$	$1 - 2^{-i}$	2^{-i}

Fig. 1. A finite-state MDP where Max has no ε-optimal memoryless strategy for safety objective $\mathtt{Avoid}(\perp)$. The state $\perp$ is a sink state. The set $\{a_i \mid i \in \mathbb{N}\}$ of Max's actions at s is countable, while Min only has a single action b.

Proposition 2. *Suppose Max's action sets are finite. Then Max has a memoryless strategy σ that is optimal for $\mathtt{Avoid}(\perp)$ from every state. That is, for all $s \in S$ and all Min strategies π*

$$\mathcal{P}_{s,\sigma,\pi}(\mathtt{Avoid}(\perp)) \geq \mathtt{val}_{\mathtt{Avoid}(\perp)}(s).$$

Proof. Since Max has only finite action sets, the supremum in (1) is taken over a compact set of mixed actions. Therefore, it is a maximum; i.e, for every $s \in S$ there is a mixed action, say $\sigma(s) \in \mathcal{D}(A(s))$, such that for all mixed Min actions $\beta \in \mathcal{D}(B(s))$

$$\mathtt{val}_{\mathtt{Avoid}(\perp)}(s) \leq \langle p(s, \sigma(s), \beta), \mathtt{val}_{\mathtt{Avoid}(\perp)} \rangle. \tag{2}$$

Extend σ to a memoryless Max strategy in the natural way, and let π be an arbitrary Min strategy. Then the function $\mathtt{val}_{\mathtt{Avoid}(\perp)}$ satisfies the conditions of Lemma 1. Thus, $\mathcal{P}_{s,\sigma,\pi}(\mathtt{Avoid}(\perp)) \geq \mathtt{val}_{\mathtt{Avoid}(\perp)}(s)$ holds for all s. □

Example 3. The assumption that Max has finite action sets cannot be dropped from Proposition 2. This assumption is required even for finite-state MDPs, i.e., when S is finite and Min has only one action per state. In fact, memoryless ε-optimal strategies do not always exist for safety objectives. As an example, consider the game depicted in Figure 1, which was first discussed in [15]. The strategy that plays a_{i+k} at stage i is $\frac{1}{2^k}$-optimal for Max; indeed, the probability of reaching $\perp$ at stage i will be at most $2^{-(i+k+1)}$. Hence, the probability of $\mathtt{Avoid}(\perp)$ is at least $1 - \sum_{i \in \mathbb{N}} 2^{-(i+k+1)} = 1 - \frac{1}{2^k}$, as required. This implies that $\mathtt{val}_{\mathtt{Avoid}(\perp)}(s) = 1$.

Let σ be some arbitrary memoryless strategy for Max. The probability that state $\perp$ is not reached at stage i is $(1 - p(s, \sigma(s), b)(\perp))^i$. Clearly, the probability of the event $\mathtt{Avoid}(\perp)$ is 0. □

5 ε-Optimal Reachability

Max does not always have optimal strategies for reachability, even when S and all action sets are finite.

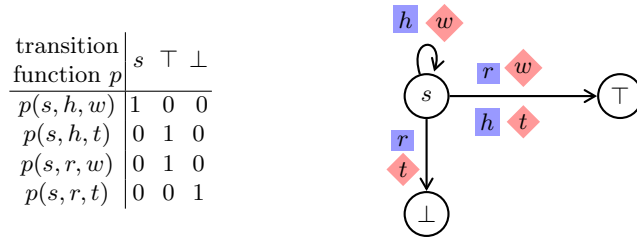| transition function p	s	$\top$	$\bot$
$p(s,h,w)$	1	0	0
$p(s,h,t)$	0	1	0
$p(s,r,w)$	0	1	0
$p(s,r,t)$	0	0	1

Fig. 2. The snowball game where Max has no optimal strategy for $\texttt{Reach}(\top)$. The states $\top$ and $\bot$ are sink states. Max's action set at s is $\{hide, run\}$, shown as h and r in the figure, while Min's action set is $\{wait, throw\}$, shown as w and t.

Example 4. Consider the game depicted in Figure 2, introduced in [16], also known as *snowball* or *hide-or-run* [8]. Following the intuition given in [8], Max initially hides behind a bush (state s) and his goal is to reach home (state $\top$) without being hit by a snowball; Min is armed with a single snowball.

Below, in order to show that $\texttt{val}_{\texttt{Reach}(\top)}(s) = 1$, for all ε, such that $0 < \varepsilon < 1$, we exhibit an ε-optimal strategy for Max. Given ε, consider the memoryless Max's strategy σ defined by $\sigma(s)(hide) = 1 - \varepsilon$ and $\sigma(s)(run) = \varepsilon$. After fixing σ in the game, by Proposition 2, Min has optimal memoryless strategies π for $\texttt{Avoid}(\top)$ in the new game. Let π be a memoryless strategy for Min, defined as

$$\pi(s)(throw) = x \qquad \text{and} \qquad \pi(s)(wait) = 1 - x.$$

for some $0 \leq x \leq 1$. Then

$$\mathcal{P}_{s,\sigma,\pi}(\texttt{Reach}(\top)) = \varepsilon(1 - x) + (1 - \varepsilon)x + (1 - \varepsilon)(1 - x)\mathcal{P}_{s,\sigma,\pi}(\texttt{Reach}(\top)).$$

Solving this for $\mathcal{P}_{s,\sigma,\pi}(\texttt{Reach}(\top))$ yields

$$\mathcal{P}_{s,\sigma,\pi}(\texttt{Reach}(\top)) \;=\; 1 - \frac{x\varepsilon}{\varepsilon + x(1 - \varepsilon)} \;\geq\; 1 - \frac{x\varepsilon}{x\varepsilon + x(1 - \varepsilon)} \;=\; 1 - \varepsilon,$$

implying that σ is an ε-optimal strategy for Max.

A straightforward argument shows that there is no optimal strategy for Max in this game. If Max always plays hide, Min can wait forever. If not, assume that at some stage Max plays run with probability $\varepsilon > 0$; then Min would throw to reach $\bot$ with a positive probability. □

In this section we prove the following proposition.

Proposition 5. *Suppose that Min's action sets are finite. Then for every $\varepsilon > 0$ and every finite subset of states $S_0 \subseteq S$ Max has a memoryless strategy σ that is ε-optimal for $\texttt{Reach}(\top)$ from all $s_0 \in S_0$ and for each $s \in S$ the support of $\sigma(s)$ is finite.*

Remark 6. The assumption that Min's action sets are finite can be replaced by the assumption that S is finite. The proof requires an extension of Lemma 1.

Proposition 5 does not carry over to countably infinite reachability games with infinite action sets for Min; see Section 7.

For the rest of the section we assume that Min's action sets are finite. Towards a proof of Proposition 5 we consider, for $n \geq 0$, the horizon-restricted reachability objective $\mathtt{Reach}_n(\top) \subseteq \mathtt{Reach}(\top)$ consisting of the plays that reach $\top$ within at most n steps. For $s \in S$ write $\mathtt{val}_n(s)$ for $\mathtt{val}_{\mathtt{Reach}_n(\top)}(s)$. We have $\mathtt{val}_0(\top) = 1$ and $\mathtt{val}_0(s) = 0$ for all $s \neq \top$. For all $n \geq 0$, since $\mathtt{Reach}_n(\top) \subseteq \mathtt{Reach}_{n+1}(\top)$, we have for all $s \in S$

$$
\begin{aligned}
\mathtt{val}_n(s) \;\leq\; \mathtt{val}_{n+1}(s) \;&=\; \sup_{\alpha \in \mathcal{D}(A(s))} \inf_{\beta \in \mathcal{D}(B(s))} \langle p(s, \alpha, \beta), \mathtt{val}_n \rangle \\
&=\; \sup_{\alpha \in \mathcal{D}(A(s))} \min_{b \in B(s)} \langle p(s, \alpha, b), \mathtt{val}_n \rangle,
\end{aligned}
$$

as Min's action sets are finite. For all $s \in S$, since $\mathtt{val}_0(s) \leq \mathtt{val}_1(s) \leq \ldots \leq 1$, there is a limit $\mathtt{val}_\infty(s) \overset{\mathrm{def}}{=} \lim_{n \to \infty} \mathtt{val}_n(s)$.

Lemma 7. *Suppose that Min's action sets are finite. For all $s \in S$ we have*

$$
\mathtt{val}_\infty(s) \;=\; \sup_{\alpha \in \mathcal{D}(A(s))} \min_{b \in B(s)} \langle p(s, \alpha, b), \mathtt{val}_\infty \rangle.
$$

Proof. Let $s \in S$. Towards the "$\leq$" inequality, for any n we have

$$
\begin{aligned}
\mathtt{val}_{n+1}(s) \;&=\; \sup_{\alpha \in \mathcal{D}(A(s))} \min_{b \in B(s)} \langle p(s, \alpha, b), \mathtt{val}_n \rangle \\
&\leq\; \sup_{\alpha \in \mathcal{D}(A(s))} \min_{b \in B(s)} \langle p(s, \alpha, b), \mathtt{val}_\infty \rangle.
\end{aligned}
$$

Thus, $\mathtt{val}_\infty(s) \leq \sup_{\alpha \in \mathcal{D}(A(s))} \min_{b \in B(s)} \langle p(s, \alpha, b), \mathtt{val}_\infty \rangle$.

Towards the "$\geq$" inequality, for any fixed α and b the function $\langle p(s, \alpha, b), \cdot \rangle : \mathbb{R}^S \to \mathbb{R}$ is a linear map and, thus, continuous. Hence, for any α

$$
\begin{aligned}
\lim_{n \to \infty} \min_{b \in B(s)} \langle p(s, \alpha, b), \mathtt{val}_n \rangle \;&=\; \min_{b \in B(s)} \lim_{n \to \infty} \langle p(s, \alpha, b), \mathtt{val}_n \rangle \\
&=\; \min_{b \in B(s)} \langle p(s, \alpha, b), \mathtt{val}_\infty \rangle.
\end{aligned}
$$

It follows that

$$
\begin{aligned}
\mathtt{val}_\infty(s) \;=\; \lim_{n \to \infty} \mathtt{val}_{n+1}(s) \;&=\; \lim_{n \to \infty} \sup_{\alpha \in \mathcal{D}(A(s))} \min_{b \in B(s)} \langle p(s, \alpha, b), \mathtt{val}_n \rangle \\
&\geq\; \sup_{\alpha \in \mathcal{D}(A(s))} \lim_{n \to \infty} \min_{b \in B(s)} \langle p(s, \alpha, b), \mathtt{val}_n \rangle \\
&=\; \sup_{\alpha \in \mathcal{D}(A(s))} \min_{b \in B(s)} \langle p(s, \alpha, b), \mathtt{val}_\infty \rangle. \qquad \square
\end{aligned}
$$

Now the following lemma follows from Lemma 1.

Lemma 8. *Suppose that Min's action sets are finite. Then $\mathtt{val}_{\mathtt{Reach}(\top)} = \mathtt{val}_\infty$.*

Proof. Since $\mathtt{Reach}(\top) \supseteq \mathtt{Reach}_n(\top)$, we have $\mathtt{val}_{\mathtt{Reach}(\top)} \geq \mathtt{val}_n$. Towards the other inequality, define a function $v : S \to [0,1]$ by $v(s) \stackrel{\mathrm{def}}{=} 1 - \mathtt{val}_\infty(s)$. We have $v(\top) = 1 - \mathtt{val}_\infty(\top) = 0$. Let σ be an arbitrary Max strategy. For every history h, let $b_h \in \arg\min_{b \in B(s_h)} \langle p(s_h, \sigma(h), b), \mathtt{val}_\infty \rangle$. Define a Min strategy π with $\pi(h)(b_h) = 1$ for all histories h. For all histories h we have

$$
\begin{aligned}
\langle p(s_h, \sigma(h), \pi(h)), v \rangle \; &= \; 1 - \langle p(s_h, \sigma(h), \pi(h)), \mathtt{val}_\infty \rangle && \text{from the definition of } v \\
&\geq \; 1 - \mathtt{val}_\infty(s_h) && \text{Lemma 7} \\
&= \; v(s_h) && \text{definition of } v.
\end{aligned}
$$

By Lemma 1, for all $s \in S$

$$
\mathcal{P}_{s,\sigma,\pi}(\mathtt{Reach}(\top)) \; = \; 1 - \mathcal{P}_{s,\sigma,\pi}(\mathtt{Avoid}(\top)) \; \leq \; 1 - v(s) = \mathtt{val}_\infty(s) \,.
$$

Since σ was arbitrary, we conclude that $\mathtt{val}_{\mathtt{Reach}(\top)} \leq \mathtt{val}_\infty$. $\qquad\square$

Let us introduce an operation on games which makes transitions "leak" to $\bot$. The intention is to "reduce" reachability to safety, in the sense that in a leaky game, $\mathtt{Avoid}(\bot)$ is included in (and, hence, equal to) $\mathtt{Reach}(\top)$ up to a measure-zero set of plays. We set up the leaky game so that Max has an optimal strategy for $\mathtt{Avoid}(\bot)$, which, according to Proposition 2, can be chosen to be memoryless.

For technical reasons, which will manifest themselves later, we associate the leaks with Min actions. Concretely, for a state s, a Min action $b \in B(s)$, and some $\varepsilon > 0$, by *making b leak ε* we refer to obtaining from p another transition function $\check{p}$ by setting, for all $a \in A(s)$ and $t \in S$,

$$
\check{p}(s,a,b)(t) \; \stackrel{\mathrm{def}}{=} \;
\begin{cases}
(1-\varepsilon)p(s,a,b)(t) & \text{if } t \neq \bot \\
(1-\varepsilon)p(s,a,b)(\bot) + \varepsilon & \text{if } t = \bot \,.
\end{cases}
$$

Intuitively, a fraction of ε of the probability mass leaks to $\bot$ whenever b is taken. We use leaks to prove Proposition 5.

Proof (of Proposition 5). Let $\varepsilon > 0$, and let $S_0 \subseteq S$ be finite. Choose $\varepsilon_1, \varepsilon_2, \varepsilon_3 > 0$ such that $\varepsilon_1 + \varepsilon_2 + \varepsilon_3 = \varepsilon$. For each $s \in S$ choose $a(s) \in A(s)$. By Lemma 8, since S_0 is finite, there is $n \geq 0$ such that $\mathtt{val}_n(s) \geq \mathtt{val}_{\mathtt{Reach}(\top)}(s) - \varepsilon_1$ holds for all $s \in S_0$. Inductively define (in general non-memoryless) Max strategies $\sigma_0, \ldots, \sigma_n$ as follows. For each history h, define $\sigma_0(h)(a(s_h)) \stackrel{\mathrm{def}}{=} 1$. For $i \in \{0, \ldots, n-1\}$ and each state $s \in S$ we have $\sup_{\alpha \in \mathcal{D}(A(s))} \min_{b \in B(s)} \langle p(s, \alpha, b), \mathtt{val}_i \rangle = \mathtt{val}_{i+1}(s)$; thus we can define $\sigma_{i+1}(s)$ so that $\min_{b \in B(s)} \langle p(s, \sigma_{i+1}(s), b), \mathtt{val}_i \rangle \geq \mathtt{val}_{i+1}(s) - \frac{\varepsilon_2}{n}$. Since $\frac{\varepsilon_2}{n} > 0$, we can assume without loss of generality that the support of $\sigma_{i+1}(s)$ is finite. Finally, define $\sigma_{i+1}(zh) \stackrel{\mathrm{def}}{=} \sigma_i(h)$ for all $z \in Z$ and $h \in H$.

We show inductively that for all $i \in \{1, \ldots, n\}$ and all $s \in S$

$$
\inf_{\pi \in \Pi} \mathcal{P}_{\mathcal{G}, s, \sigma_i, \pi}(\mathtt{Reach}_i(\top)) \; \geq \; \mathtt{val}_i(s) - i \cdot \frac{\varepsilon_2}{n} \,. \tag{3}
$$

This is immediate for $i = 0$. For $i \in \{0, \ldots, n-1\}$ we have

$$\inf_{\pi \in \Pi} \mathcal{P}_{\mathcal{G},s,\sigma_{i+1},\pi}(\texttt{Reach}_{i+1}(\top))$$

$$\geq \min_{b \in B(s)} \langle p(s, \sigma_{i+1}(s), b), \inf_{\pi \in \Pi} \mathcal{P}_{\mathcal{G},\cdot,\sigma_i,\pi}(\texttt{Reach}_i(\top)) \rangle$$

$$\geq \min_{b \in B(s)} \langle p(s, \sigma_{i+1}(s), b), \texttt{val}_i \rangle - i \cdot \frac{\varepsilon_2}{n} \qquad \text{induction hypothesis}$$

$$\geq \texttt{val}_{i+1}(s) - \frac{\varepsilon_2}{n} - i \cdot \frac{\varepsilon_2}{n} \qquad \text{definition of } \sigma_{i+1}(s) \,,$$

proving (3).

Now in every state except $\top$, make each Min action leak $\frac{\varepsilon_3}{n}$. Further, remove any Max actions that σ_n never takes. Call the resulting game $\mathcal{G}^-$. Then Max's action sets in $\mathcal{G}^-$ are finite and we have for all $s_0 \in S$ and all Min strategies π

$$\mathcal{P}_{\mathcal{G}^-,s,\sigma_n,\pi}(\texttt{Avoid}(\bot))$$

$$\geq \mathcal{P}_{\mathcal{G}^-,s,\sigma_n,\pi}(\texttt{Reach}_n(\top)) \qquad \texttt{Avoid}(\bot) \supseteq \texttt{Reach}_n(\top)$$

$$\geq \mathcal{P}_{\mathcal{G},s,\sigma_n,\pi}(\texttt{Reach}_n(\top)) - \varepsilon_3 \qquad \text{at most } n \cdot \frac{\varepsilon_3}{n} \text{ leaks in } n \text{ steps}$$

$$\geq \texttt{val}_n(s) - \varepsilon_2 - \varepsilon_3 \qquad \text{by (3)}$$

$$\geq \texttt{val}_{\mathcal{G},\texttt{Reach}(\top)}(s) - \varepsilon_1 - \varepsilon_2 - \varepsilon_3 \qquad \text{choice of } n$$

$$= \texttt{val}_{\mathcal{G},\texttt{Reach}(\top)}(s) - \varepsilon \qquad \text{choice of } \varepsilon_1, \varepsilon_2, \varepsilon_3.$$

Thus, for all $s \in S_0$

$$\texttt{val}_{\mathcal{G}^-,\texttt{Avoid}(\bot)}(s) \ \geq \ \texttt{val}_{\mathcal{G},\texttt{Reach}(\top)}(s) - \varepsilon \,. \tag{4}$$

Due to the leaks, in $\mathcal{G}^-$ the events $\texttt{Reach}(\top)$ and $\texttt{Avoid}(\bot)$ coincide up to measure zero for all Max and all Min strategies. Since Max's action sets in $\mathcal{G}^-$ are finite, by Proposition 2 Max has a memoryless strategy σ that is optimal for $\texttt{Avoid}(\bot)$ from every state. Thus, for all $s \in S_0$ and all Min strategies π

$$\mathcal{P}_{\mathcal{G},s,\sigma,\pi}(\texttt{Reach}(\top)) \ \geq \ \mathcal{P}_{\mathcal{G}^-,s,\sigma,\pi}(\texttt{Reach}(\top))$$

$$= \mathcal{P}_{\mathcal{G}^-,s,\sigma,\pi}(\texttt{Avoid}(\bot)) \qquad \text{as argued above}$$

$$\geq \texttt{val}_{\mathcal{G}^-,\texttt{Avoid}(\bot)}(s) \qquad \sigma \text{ is optimal}$$

$$\geq \texttt{val}_{\mathcal{G},\texttt{Reach}(\top)}(s) - \varepsilon \qquad \text{by (4).}$$

That is, the memoryless strategy σ is ε-optimal from all $s \in S_0$, and for each $s \in S$ the support of $\sigma(s)$ is finite. □

6   Optimal Reachability

In this section we show the following result, a generalization of Bordais et al. [4, Theorem 28], via a different proof.

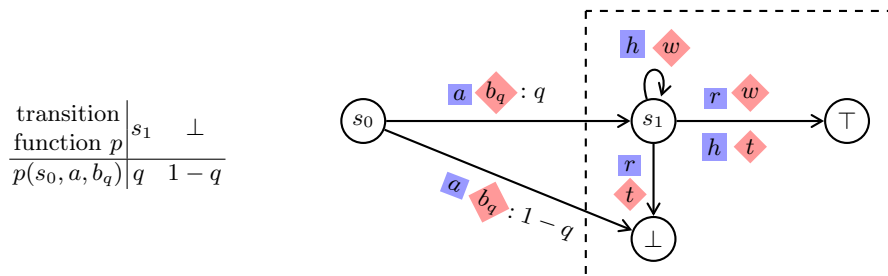The transition function table (left of the figure):

transition function p	s_1	$\bot$
$p(s_0, a, b_q)$	q	$1-q$

Fig. 3. A game where Max has no *memoryless* optimal strategy for $\texttt{Reach}(\top)$. The states $\top$ and $\bot$ are sink states. Min's action set at s_0 is $\{b_q \mid q \in (\frac{1}{2}, 1) \cap \mathbb{Q}\}$, while Max has a single action a at s_0. The game in dashed box is the snowball game from Figure 2.

Theorem 9. *Suppose that S and Min's action sets are finite. Then for every $\varepsilon > 0$ Max has a memoryless strategy σ that is ε-optimal for $\texttt{Reach}(\top)$ from all states and optimal from all states from which Max has any optimal strategy.*

The assumption that Min's action sets are finite cannot be dropped, as the following example shows.

Example 10. Consider the game depicted in Figure 3, wherein state s_1 is the start state of the snowball game from Example 4. Recall that $\texttt{val}_{\texttt{Reach}(\top)}(s_1) = 1$, but Max does not have an optimal strategy from s_1. By this, and the fact that Min can choose b_q for q arbitrary close to $\frac{1}{2}$, we deduce that $\texttt{val}_{\texttt{Reach}(\top)}(s_0) = \frac{1}{2}$. Observe that when taking action b_q at s_0, Min is increasing the value from $\frac{1}{2}$ to $q > \frac{1}{2}$. Indeed, we claim that Max has an optimal strategy from s_0. If Max observes which action b_q Min takes in s_0, and plays a $(q - \frac{1}{2})$-optimal strategy from s_1, the probability of reaching $\top$ will be at least $q \cdot (1 - (q - \frac{1}{2})) \geq q - (q - \frac{1}{2}) = \frac{1}{2}$. But Max has to "remember" Min's value increase in order to play sufficiently well in s_1. So Max does not have a memoryless optimal strategy from s_0. In fact, a step counter (discrete clock) would not help Max either, since the step counter value does not contain any information about q. □

The assumption that S is finite cannot be dropped either, not even for the subclass of turn-based games, as the following example shows.

Example 11. Consider the finitely-branching turn-based reachability game depicted in Figure 4 (from [13, Section 9]). The initial state is u_1, states u_i are Min-controlled, states s_i are Max-controlled and t is the target state.

At every state u_i, Min can either go right (red transition) or go up. At every state s_i, Max can either go right (red transition) or go down. It is easy to show that $\texttt{val}_\mathcal{G}(s_i) > \texttt{val}_\mathcal{G}(u_i)$ for all i. Thus the only locally optimal Min move is to go right. Similarly, $\texttt{val}_\mathcal{G}(s_i) > y_i$, and thus the only locally optimal Max move is to go right.

Max has an optimal strategy from u_1 (and also from every other state) as follows. First, always go right. If Min ever goes up at some u_i then go down at s_i.
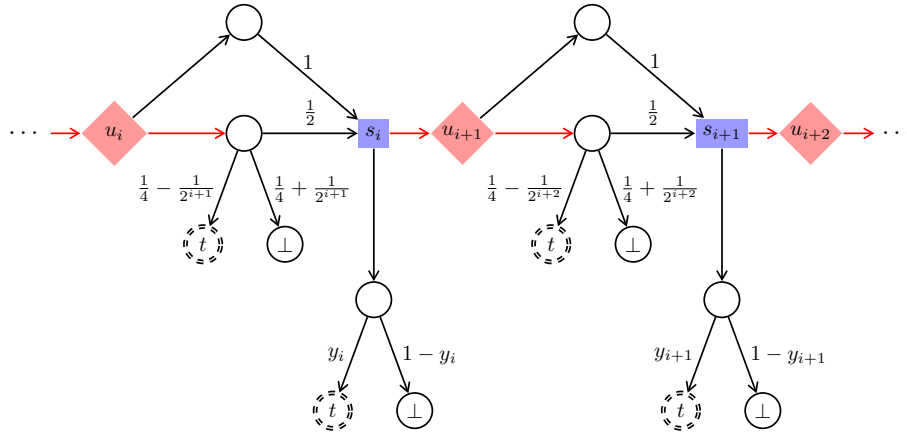
Fig. 4. A finitely-branching turn-based reachability game $\mathcal{G}$ with initial state u_1, where optimal Max strategies cannot be memoryless (and cannot even be Markov). For clarity, we have drawn several copies of the target state t. The number y_i is defined as $\frac{1}{2} - \frac{1}{2^{i+1}}$.

Intuitively, by going up, Min increases the value (i.e., gives a "gift" to Max). By going down at s_i, Max plays sub-optimally locally, but realizes most of the value of s_i, i.e., loses less than Min's previous gift.

However, no memoryless strategy (and not even any Markov strategy, i.e., any strategy that uses only a step counter) can be optimal for Max from u_1. First, a step counter gives no advantage to Max, since in this example the step counter is implicit in the current state anyway. For any memoryless Max strategy, there are to cases. Either this strategy never goes down with any positive probability at any state s_i. Then Min can avoid the target t completely by always playing up, and thus this Max strategy is not optimal. Otherwise, let s_j be first state where Max goes down with some positive probability. This Max strategy is not optimal either, since Min can always go right (locally optimal), and Max's choice at s_j is locally sub-optimal.

See [13, Section 9] for a formal proof of this example. It is also shown there that, in every countably infinite finitely-branching turn-based reachability game, optimal Max strategies (if they exist at all) can be chosen as deterministic and using a step counter plus 1 bit of public memory. So the example above is tight.

□

Now we prove Theorem 9. Let S and Min's action sets be finite. We partition the finite state space S into S_0, S_1 so that S_0 contains exactly the states from which Max has an optimal strategy for $\texttt{Reach}(\top)$, and S_1 contains exactly the states from which Max does not have an optimal strategy. Clearly, $\top, \bot \in S_0$.

By the finiteness of Min's action sets and (1), for every value-0 state s and every mixed Max action $\alpha \in \mathcal{D}(A(s))$ Min has an action b that keeps the game (surely) in a value-0 state. Therefore, an optimal Max strategy cannot rely on

Min "gifts" in value-0 states. Hence, we can assume without loss of generality that $\perp$ is the only value-0 state. We write $S_0^+ \overset{\text{def}}{=} S_0 \setminus \{\perp\}$, so that all states in S_0^+ have a positive value.

The challenge is that the play can "cross over" between S_0 and S_1. Our approach is as follows. First we fix a memoryless strategy σ_0 on S_0 that is optimal for Max as long as Min responds optimally. Then, building on the idea of the previous section, we add "leaks" to the game so that

- in S_0 the leaks do not decrease the value and Max still has an optimal strategy;
- in S_1 the leaks decrease the value only by little.

After having fixed σ_0 on S_0 and introduced suitable leaks, an optimal safety strategy for $\mathtt{Avoid}(\perp)$ will serve as the memoryless strategy σ claimed in Theorem 9. This strategy extends σ_0 to the whole state space and is optimal from S_0 (no matter what Min does).

Below, to avoid clutter whenever we write $\mathtt{val}(s)$ we mean $\mathtt{val}_{\mathcal{G},\mathtt{Reach}(\top)}(s)$ where $\mathcal{G}$ is the original game, even when other auxiliary games are discussed.

We start by defining a memoryless strategy, σ_0, only on S_0 so that σ_0 is optimal from S_0 as long as Min "does not increase the value".

For a state $s \in S_0$, call a mixed Max action $\alpha \in \mathcal{D}(A(s))$ *optimality-preserving* if for all Min actions $b \in B(s)$ we have $\langle p(s,\alpha,b), \mathtt{val} \rangle \geq \mathtt{val}(s)$ and if $\langle p(s,\alpha,b), \mathtt{val} \rangle = \mathtt{val}(s)$ then the support of $p(s,\alpha,b)$ is a subset of S_0. Note that every optimal Max strategy from S_0 is optimality-preserving at least in the first step. Therefore, every state in S_0 has an optimality-preserving mixed Max action.

For a state $s \in S_0$ and an optimality-preserving mixed action α, call a Min action $b \in B(s)$ *value-preserving* if $\langle p(s,\alpha,b), \mathtt{val} \rangle = \mathtt{val}(s)$ and *value-increasing* otherwise (i.e., $\langle p(s,\alpha,b), \mathtt{val} \rangle > \mathtt{val}(s)$).

We define inductively a non-decreasing sequence $R(0), R(1), R(2), \ldots$ of subsets of S_0. Define $R(0) \overset{\text{def}}{=} \{\top\}$. For every $n \geq 0$, define $R(n+1)$ as $R(n)$ union the set of those states $s \in S_0$ from which Max has an optimality-preserving mixed action, say α_s, and a number $\delta_s > 0$ such that for all value-preserving Min actions $b \in B(s)$ we have $p(s,\alpha_s,b)(R(n)) \geq \delta_s$. We note in passing that the existence of $\delta_s > 0$ is guaranteed due to finiteness of Min's action set. Informally speaking, $R(n+1)$ consists of those states that are already in $R(n)$ or have an optimality-preserving mixed action α_s so that in the next step $R(n)$ is entered with a positive probability unless Min takes a value-increasing action. Defining $R \overset{\text{def}}{=} R(|S|-1)$, since S is finite we have $R = R(|S|-1) = R(|S|) = R(|S|+1) = \ldots$. Moreover, we can define $\delta \overset{\text{def}}{=} \min_{s \in R} \delta_s > 0$.

Lemma 12. *We have $R = S_0^+$.*

Proof. By an easy induction, $\perp \notin R$. Hence $R \subseteq S_0^+$. Towards the reverse inclusion, suppose there is $s \in S_0 \setminus R$. It suffices to show that $s = \perp$. Let σ be an arbitrary optimal Max strategy starting from s. Note that $\sigma(s)$ is optimality-preserving. Since $s \notin R = R(|S|)$, there is a value-preserving Min action $b \in$

$B(s)$ such that the support of $p(s, \sigma(s), b)$ does not overlap with $R(|S| - 1) = R$. That is, the support of $p(s, \sigma(s), b)$ is a subset of $S_0 \setminus R$ and, thus, does not contain $\top$. Since σ is optimal, σ has to play optimally—and, thus, has to use optimality-preserving mixed actions—as long as Min has taken only value-preserving actions. Then it follows inductively that there is a deterministic Min strategy (playing b in the very first step) that keeps the play in $S_0 \setminus R$ forever. In particular, $\top$ is not reached. Since the optimal strategy σ was arbitrary, we have $\mathtt{val}(s) = 0$; i.e., $s = \bot$. $\qquad\square$

Define σ_0 to be the memoryless Max strategy on S_0 with $\sigma_0(s) = \alpha_s$ for all $s \in S_0$, where α_s is from the definition of $R(n)$. For $s \in S_0$, call $b \in B(s)$ *value-preserving* (or *value-increasing*) if b is value-preserving (or value-increasing, respectively) for s and $\sigma_0(s)$.

Lemma 13. *For any Max strategy that plays σ_0 on S_0 and any Min strategy, almost all plays that eventually remain in S_0^+ and eventually contain only value-preserving Min actions reach $\top$.*

Proof. Let σ be a Max strategy that plays σ_0 on S_0, and let π be a Min strategy. Since the set of (finite) histories is countable, it suffices to consider the set of plays which always remain in S_0^+ and contain only value-preserving Min actions. Let $s \in S_0^+$. By Lemma 12, we have $s \in R$. Let $n \leq |S| - 1$ be the smallest n with $s \in R(n)$. We have $p(s, \sigma_0(s), b)(R(n-1)) \geq \delta$ for all value-preserving Min actions b. It follows that at any time the probability that in at most $|S| - 1$ steps $R(0) = \{\top\}$ is reached is at least $\delta^{|S|-1} > 0$. Thus, almost surely $\top$ is eventually reached. $\qquad\square$

Since S_0 and Min's action sets are finite, we can choose an $\varepsilon > 0$ small enough so that for all $s \in S_0$ and all value-increasing Min actions b we have

$$\langle p(s, \sigma_0(s), b), \mathtt{val} \rangle \geq \mathtt{val}(s) + 2\varepsilon \tag{5}$$

So any value increase by a Min action is at least 2ε.

Define a game $\mathcal{G}_1$ with state space $S_1 \cup \{\top, \bot\}$. Its transition function p_1 is obtained from p by redirecting probability mass away from $S_0 \setminus \{\top, \bot\}$ as follows. Each transition to a state $s \in S_0 \setminus \{\top, \bot\}$ is redirected to $\top$ with probability $\mathtt{val}(s)$ and to $\bot$ with probability $1 - \mathtt{val}(s)$. Then the value of each state in $\mathcal{G}_1$ is equal to its value in $\mathcal{G}$.

As in the proof of Proposition 5 we obtain from $\mathcal{G}_1$ a "leaky" version, $\mathcal{G}_1^-$, with transition function p_1^-, such that Max's action sets are finite and, defining $v(s) \stackrel{\text{def}}{=} \mathtt{val}_{\mathcal{G}_1^-, \mathtt{Avoid}(\bot)}(s)$ for all $s \in S_1 \cup \{\top, \bot\}$, we have

$$v(s) \geq \mathtt{val}(s) - \varepsilon \qquad \text{for all } s \in S_1. \tag{6}$$

By Proposition 2 Max has a memoryless optimal (for safety) strategy σ_1 in $\mathcal{G}_1^-$. Thus, for all $s \in S_1$ and all $b \in B(s)$, as in Equation (2), we have

$$\langle p_1^-(s, \sigma_1(s), b), v \rangle \geq v(s).$$

Next we define a "leaky" version, $\mathcal{G}^-$, of the whole game $\mathcal{G}$ with the whole state space S and a transition function p^-. On S_1, function p^- is defined similarly to p_1^-, except that the probability mass that was directed away from $S_0 \setminus \{\top, \bot\}$ in the definition of $\mathcal{G}_1$ is now not directed away and instead enters S_0 as originally. However, the leaks in every transition from S_1 are also present in p^-. Extend the function v to the whole state space S by defining $v(s) \stackrel{\text{def}}{=} \mathtt{val}(s)$ for all $s \in S_0$. Then, for all $s \in S_1$ and all $b \in B(s)$

$$\langle p^-(s, \sigma_1(s), b), v \rangle \;=\; \langle p_1^-(s, \sigma_1(s), b), v \rangle \;\geq\; v(s) \,.$$

On S_0, the transition function p^- is obtained from p by making value-increasing Min actions leak ε, in the sense defined in Section 5. For all $s \in S_0$ and all value-increasing Min actions $b \in B(s)$

$$
\begin{aligned}
&\langle p^-(s, \sigma_0(s), b), v \rangle \\
\geq\ & \langle p^-(s, \sigma_0(s), b), \mathtt{val} \rangle - \varepsilon && \text{by (6) and the definition of } v \text{ on } S_0 \\
\geq\ & \langle p(s, \sigma_0(s), b), \mathtt{val} \rangle - \varepsilon - \varepsilon && \text{from the definition of } p^- \\
\geq\ & \mathtt{val}(s) + 2\varepsilon - \varepsilon - \varepsilon && \text{by (5)} \\
=\ & v(s) && \text{definition of } v \text{ on } S_0.
\end{aligned}
$$

For all $s \in S_0$ and all value-preserving Min actions $b \in B(s)$

$$
\begin{aligned}
&\langle p^-(s, \sigma_0(s), b), v \rangle \\
=\ & \langle p(s, \sigma_0(s), b), v \rangle && \text{from the definition of } p^- \\
=\ & \langle p(s, \sigma_0(s), b), \mathtt{val} \rangle && \text{support of } p(s, \sigma_0(s), b) \text{ is a subset of } S_0 \\
=\ & \mathtt{val}(s) && b \text{ is value-preserving} \\
=\ & v(s) && \text{definition of } v \text{ on } S_0.
\end{aligned}
$$

Define the memoryless strategy σ by naturally combining σ_0 and σ_1. We have shown above that for all $s \in S$ and all $b \in B(s)$ we have $\langle p^-(s, \sigma(s), b), v \rangle \geq v(s)$. From applying Lemma 1 we conclude that for all $s \in S$ and all Min strategies π we have

$$\mathcal{P}_{\mathcal{G}^-, s, \sigma, \pi}(\mathtt{Avoid}(\bot)) \;\geq\; v(s) \,.$$

In $\mathcal{G}^-$, due to the leaks, almost all $\bot$-avoiding plays eventually remain in S_0^+ and eventually have only value-preserving Min actions. But by Lemma 13, almost all of these plays reach $\top$. Thus, we have for all $s \in S$ and all Min strategies π

$$\mathcal{P}_{\mathcal{G}^-, s, \sigma, \pi}(\mathtt{Reach}(\top)) \;=\; \mathcal{P}_{\mathcal{G}^-, s, \sigma, \pi}(\mathtt{Avoid}(\bot)) \;\geq\; v(s) \,,$$

which, by definition, equals $\mathtt{val}(s)$ if $s \in S_0$ and, by (6), is at least $\mathtt{val}(s) - \varepsilon$ if $s \in S_1$. Since $\mathcal{P}_{\mathcal{G}, s, \sigma, \pi}(\mathtt{Reach}(\top)) \geq \mathcal{P}_{\mathcal{G}^-, s, \sigma, \pi}(\mathtt{Reach}(\top))$, this completes the proof of Theorem 9.

7   Conclusion and Related Work

We have shown that, in finite reachability games with finite action sets for Min, optimal Max strategies, where they exist, can be chosen as memoryless randomized. However, this does not carry over to countably infinite reachability games.

Intuitively, the reason for this is that Min can play sub-optimally and give "gifts" to Max that increase the expected value of the current state, but delay progress towards the target. In countably infinite reachability games, Min might give infinitely many smaller and smaller gifts and delay progress indefinitely, unless Max uses memory to keep track of these gifts in order to react correctly.

Finite reachability games are simpler, because gifts from Min to Max are universally lower bounded in size, due to the finite state space and the finiteness of Min's action sets. Therefore, Min cannot give infinitely many gifts to Max, except in a nullset of the plays. Without such distracting gifts, Max can make steady progress towards the target. Moreover, if Min does give a gift once, then Max does not need to remember how large it was, since it is universally lower bounded.

The existence of optimal strategies is also affected by the size of the state space. While finite turn-based reachability games always admit optimal Max strategies [7], even in countably infinite MDPs, optimal Max strategies for reachability need not exist. However, it was shown by Ornstein [20, Thm. B] that ε-optimal Max strategies for reachability in countably infinite MDPs can be chosen as memoryless and deterministic. These strategies can even be made uniform, i.e., independent of the start state. Moreover, if an optimal strategy does exist for Max in a countable MDP, then there also exists one that is memoryless and deterministic [20, Prop. B].

These results on countable MDPs do not carry over to countable 2-player stochastic reachability games. While Max always has ε-optimal randomized memoryless strategies in countable concurrent reachability games with finite action sets [21, Corollary 3.9], these strategies depend on the start state and cannot be made uniform [19]. This non-uniformity even holds for the subclass of countable finitely branching turn-based reachability games [13]. However, uniformity can be regained with 1 bit of public memory, i.e., there exist uniformly ε-optimal Max strategies in countable concurrent reachability games with finite action sets that are deterministic and use just 1 bit of public memory [13]. Optimal Max strategies in countable turn-based finitely branching reachability games, where they exist, can be chosen to use just a step counter and 1 bit of public memory (but not just a step counter or just finite memory). On the other hand, in concurrent games with finite action sets, a step counter plus finite private memory does not suffice for optimal Max strategies in general [13].

If Min is allowed infinite action sets (resp. infinite branching) in countably infinite reachability games, then Max generally needs infinite memory for ε-optimal, optimal and almost surely winning strategies. There exists a turn-based countable reachability game with infinite Min branching (and finite Max branching), such that every state admits an almost surely winning strategy for Max, and yet every Max strategy that uses only a step counter plus finite private memory

is still useless (in the sense that Min can make its attainment arbitrarily close to zero) [13].

References

1. de Alfaro, L., Henzinger, T.A., Kupferman, O.: Concurrent reachability games. In: Annual Symposium on Foundations of Computer Science (FOCS). pp. 564–575. IEEE Computer Society (1998). https://doi.org/10.1109/SFCS.1998.743507
2. Bertrand, N., Genest, B., Gimbert, H.: Qualitative determinacy and decidability of stochastic games with signals. Journal of the ACM **64**(5), 33:1–33:48 (2017), https://doi.org/10.1145/3107926
3. Billingsley, P.: Probability and Measure. Wiley, New York, NY (1995), third edition
4. Bordais, B., Bouyer, P., Le Roux, S.: Optimal strategies in concurrent reachability games. In: Computer Science Logic (CSL). LIPIcs, vol. 216, pp. 7:1–7:17. Schloss Dagstuhl - Leibniz-Zentrum für Informatik (2022). https://doi.org/10.4230/LIPIcs.CSL.2022.7
5. Bouyer, P., Markey, N., Randour, M., Sangnier, A., Stan, D.: Reachability in Networks of Register Protocols under Stochastic Schedulers. In: International Colloquium on Automata, Languages and Programming (ICALP). vol. 55, pp. 106:1–106:14. Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik, Dagstuhl, Germany (2016). https://doi.org/10.4230/LIPIcs.ICALP.2016.106
6. Chen, T., Forejt, V., Kwiatkowska, M., Parker, D., Simaitis, A.: Prism-games: A model checker for stochastic multi-player games. In: Tools and Algorithms for the Construction and Analysis of Systems (TACAS). pp. 185–191. Springer (2013). https://doi.org/10.1007/978-3-642-36742-7_13
7. Condon, A.: The complexity of stochastic games. Information and Computation **96**(2), 203–224 (1992). https://doi.org/10.1016/0890-5401(92)90048-K
8. De Alfaro, L., Henzinger, T.A., Kupferman, O.: Concurrent reachability games. Theor. Comput. Sci. **386**(3), 188–217 (2007). https://doi.org/10.1016/j.tcs.2007.07.008
9. Everett, H.: Recursive games. In: Contributions to the Theory of Games, Volume III, Annals of Mathematics Studies, vol. 39, pp. 47–78. Princeton University Press, Princeton, N.J. (1957). https://doi.org/10.1515/9781400882151-004
10. Flesch, J., Predtetchinski, A., Sudderth, W.: Positive zero-sum stochastic games with countable state and action spaces. Applied Mathematics and Optimization **82**, 499–516 (2020). https://doi.org/10.1007/s00245-018-9536-3
11. Gillette, D.: Stochastic games with zero stop probabilities. In: Contributions to the Theory of Games (AM-39), Volume III, pp. 179–188. Princeton University Press (dec 1958). https://doi.org/10.1515/9781400882151-011
12. Gimbert, H., Horn, F.: Solving simple stochastic tail games. In: Annual ACM-SIAM Symposium on Discrete Algorithms. pp. 847–862 (2010). https://doi.org/10.1137/1.9781611973075
13. Kiefer, S., Mayr, R., Shirmohammadi, M., Totzke, P.: Strategy complexity of reachability in countable stochastic 2-player games. Dynamic Games and Applications (2023), https://arxiv.org/abs/2203.12024, to appear
14. Kiefer, S., Mayr, R., Shirmohammadi, M., Wojtczak, D.: On strong determinacy of countable stochastic games. In: ACM/IEEE Symposium on Logic in Computer Science (LICS). pp. 1–12. IEEE Computer Society (2017). https://doi.org/10.1109/LICS.2017.8005134

15. Kiefer, S., Mayr, R., Shirmohammadi, M., Wojtczak, D.: Parity Objectives in Countable MDPs. In: ACM/IEEE Symposium on Logic in Computer Science (LICS) (2017). https://doi.org/10.1109/LICS.2017.8005100
16. Kumar, P.R., Shiau, T.H.: Existence of value and randomized strategies in zero-sum discrete-time stochastic dynamic games. SIAM Journal on Control and Optimization **19**(5), 617–634 (1981). https://doi.org/10.1137/0319039
17. Kučera, A.: Turn-based stochastic games. In: Apt, K.R., Grädel, E. (eds.) Lectures in Game Theory for Computer Scientists, pp. 146–184. Cambridge University Press (2011). https://doi.org/10.1017/CBO9780511973468.006
18. Maitra, A., Sudderth, W.: Finitely additive stochastic games with Borel measurable payoffs. International Journal of Game Theory **27**(2), 257–267 (July 1998). https://doi.org/10.1007/s001820050071
19. Nowak, A., Raghavan, T.: Positive stochastic games and a theorem of Ornstein. In: Stochastic Games And Related Topics, Theory and Decision Library (Game Theory, Mathematical Programming and Operations Research), vol. 7. Springer (1991). https://doi.org/10.1007/978-94-011-3760-7_11
20. Ornstein, D.: On the existence of stationary optimal strategies. Proceedings of the American Mathematical Society **20**(2), 563–569 (1969). https://doi.org/10.2307/2035700
21. Secchi, P.: Stationary strategies for recursive games. Math. Oper. Res. **22**(2), 494–512 (1997). https://doi.org/10.1287/moor.22.2.494
22. Shapley, L.S.: Stochastic games. Proceedings of the National Academy of Sciences **39**(10), 1095–1100 (1953). https://doi.org/10.1073/pnas.39.10.1095