Strategy Complexity of Reachability in Countable Stochastic 2-Player 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.

Contributing authors: stekie@cs.ox.ac.uk; mayr@cs.ed.ac.uk; mahsa@irif.fr; totzke@liverpool.ac.uk;

Abstract

We study countably infinite stochastic 2-player games with reachability objectives. Our results provide a complete picture of the memory requirements of ε-optimal (resp. optimal) strategies. These results depend on the size of the players' action sets and on whether one requires strategies that are uniform (i.e., independent of the start state). Our main result is that ε-optimal (resp. optimal) Maximizer strategies require infinite memory if Minimizer is allowed infinite action sets. This lower bound holds even under very strong restrictions. Even in the special case of infinitely branching turn-based reachability games, even if all states allow an almost surely winning Maximizer strategy, strategies with a step counter plus finite private memory are still useless. Regarding *uniformity*, we show that for Maximizer there need not exist positional (i.e., memoryless) uniformly ε-optimal strategies even in the special case of finite action sets or in finitely branching turn-based games. On the other hand, in games with finite action sets, there always exists a uniformly ε-optimal Maximizer strategy that uses just one bit of public memory.

Keywords: Stochastic Games, Discrete-time games, Strategy Complexity

MSC Classification: 91A15 , 60J05 , 91A60 , 60G40 , 60G40 , 60J05

1 Introduction

Stochastic 2-player games, first introduced by Shapley in his seminal 1953 work Shapley (1953), model dynamic interactions in which the environment responds randomly to players' actions. Shapley's games were generalized by Gillette Gillette (1958) and Kumar and Shiau Kumar and Shiau (1981) to allow infinite state and action sets and non-termination. They play a central role in the solution of many problems in economics Sorin (1992); Nowak and Szajowski (2005); Jaskiewicz and Nowak (2011); Solan and Vieille (2015); Bacharach (2019), evolutionary biology Raghavan et al (2012), and computer science de Alfaro and Henzinger (2001); Neyman and Sorin (2003); Altman et al (2005, 2007); Solan and Vieille (2015); Svoreňová and Kwiatkowska (2016); Bouyer et al (2016) among others.

In general concurrent games, in each state both players independently choose an action and the next state is determined according to a pre-defined distribution that depends on 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 games were studied first in the sequence of works Filar (1980, 1981); Vrieze et al (1983); Vrieze (1987); Condon (1992), but have recently received much attention by computer scientists Gimbert and Horn (2010); Chen et al (2013); Bouyer et al (2016); Kiefer et al (2017a); Bertrand et al (2017). An even more special case of stochastic games are *Markov Decision Processes (MDPs)*: MDPs are turn-based games where all controlled states are Maximizer states. Since Minimizer is passive, they are also called games against nature.

In order to get the strongest results, we will show that our lower bound results hold even for the special subclass of turn-based games (aka switching-control), while our upper bounds hold even for general games.

A strategy for a player is a function that, given a history of a play, determines the next action of the player. Objectives are defined via functions that assign numerical rewards to plays, and the Maximizer (resp. Minimizer) aim to maximize (resp. minimize) the expected reward. A central result in zero-sum 2-player stochastic games with finite action sets is the existence of a *value* for the large class of Borel measurable objectives Martin (1998); Maitra and Sudderth (1998) (i.e., that $\sup_{Max} \inf_{Min} = \mathit{value} = \inf_{Min} \sup_{Max}$ over Maximizer/Minimizer strategies). In particular, this implies the existence of ε-optimal strategies for every $\varepsilon > 0$ and either player, i.e., strategies that enforce that the outcome of a game is ε-close to its value, regardless of the behavior of the other player. Optimal strategies (ε-optimal for $\varepsilon = 0$) need not exist in general, but their properties can be studied in those cases where they do exist, e.g., Puterman (1994); Kučera (2011); Kiefer et al (2017a, 2020a).

The nature of good strategies in stochastic games – that is ε-optimality vs. optimality, and their memory requirements – is relevant in computer science Brázdil et al (2014); Kiefer et al (2017a, 2020a), in particular, in the sense of computability Kučera (2011). It is also recognised as central notion in branches of mathematics and economics, especially operations research Maitra and Sudderth (2007), probability theory Flesch et al (2018) game theory Flesch et al (2021); Laraki et al (2013); Maitra and Sudderth (2007) and economic theory Aumann (1981); Bacharach (2019); Kalai (1990).

The simplest type of strategy bases its decisions only on the current state, and not on the history of the play. Such strategies are called *memoryless* or *positional*. [1]

Sometimes we also need to distinguish between deterministic (pure) and randomized (mixed) strategies. A *memoryless deterministic (MD)* strategy defines for each state s a single pure action that the player always chooses when the game is in s. A slightly more general strategy is *memoryless randomized (MR)*, also known as *positional*: for each state s it defines a fixed probability distribution over the player's actions from which he always samples when the game is in s.

More complex strategies might use some amount of finite memory. The strategy chooses an action depending only on the current state and the current memory mode. The memory mode can be updated in every round according to the current state, the observed chosen actions and the next state. We assume perfect information games, so the actions and states are observable at the end of every round. In general, for strategies that are not deterministic but use randomization, this memory update may also be randomized. Therefore, in the case of games, a player does not necessarily know for sure the current memory mode of the other player. It may be advantageous for a player to keep his memory mode hidden from the other player. We distinguish between *public memory*, where the strategies' memory mode is public knowledge, and *private memory*, which is hidden from the opponent. A step counter is a discrete clock that is incremented after every round. We consider this to be a type of public memory because the update is deterministic and the memory mode can be deduced by the opponent. Strategies that use only a step counter are called *Markov strategies*. Combinations of the above are possible, e.g., a strategy that uses a step counter and an additional finite public/private general purpose memory. The amount/type of memory and randomization required for a good (ε-optimal, resp. optimal) strategy for a given objective is also called its *strategy complexity*.

The Reachability Objective

With a reachability objective, a play is defined as winning for Maximizer iff it visits a defined target state (or a set of target states) at least once. Thus Maximizer aims to maximize the probability that the target is reached. Dually, Minimizer aims to minimize the probability of reaching the target. So, from

[1] A closely related concept is a *stationary* strategy, which also bases decisions only on the current state. However, some authors call a strategy "stationary ε-optimal" if it is ε-optimal from every state, and call it "semi-stationary" if it is ε-optimal only from the fixed initial state. Since this difference is important in our work, we avoid the term "stationary" here. Instead, if a strategy is ε-optimal from every state then we call it *uniformly ε-optimal*. I.e., ε-optimal stationary strategies are uniformly ε-optimal positional strategies.

Minimizer's point of view, this is the dual *safety objective* of avoiding the target.

Reachability is arguably the simplest objective in games on graphs. It can trivially be encoded into the usual reward-based objectives, i.e., every play that reaches the target gets reward 1 and all other plays get reward 0. Moreover, it can be encoded into many other objectives, by using the following intuition: "First you need to reach the target, and only then can you satisfy the other objective." E.g., this works for Büchi, Parity and average-payoff objectives.

Despite their apparent simplicity, reachability games are not trivial. While both players have optimal MD strategies in finite-state turn-based reachability games Condon (1992) (see also (Kučera, 2011, Proposition 5.6.c, Proposition 5.7.c)), this does not carry over to finite-state concurrent reachability games. A counterexample is the *Hide-or-Run* game by Kumar and Shiau Kumar and Shiau (1981); de Alfaro et al (1998), where Maximizer has no optimal strategy.

In countably infinite reachability games, Maximizer does not have an optimal strategy even if the game is turn-based, in fact not even in countably infinite finitely-branching MDPs Kiefer et al (2017b). On the other hand, Ornstein Ornstein (1969) showed that Maximizer has ε-optimal MD strategies in countably infinite MDPs. Better yet, the MD strategies can be made uniform, i.e., independent of the start state. This led to the question whether Ornstein's results can be generalized from MDPs to countably infinite stochastic games. Secchi (Secchi, 1997, Corollary 3.9) proved the following.

Proposition 1 *Maximizer has ε-optimal positional (MR) strategies in countably infinite concurrent reachability games with finite action sets.*

However, these MR strategies are not uniform, i.e., they depend on the start state. In fact, Nowak and Raghavan Nowak and Raghavan (1991) showed that there cannot exist any uniformly ε-optimal positional Maximizer strategies in countably infinite concurrent reachability games with finite action sets. Their counterexample is called the *Big Match on* $\mathbb{N}$ (which, in turn, is inspired by the *Big Match* Gillette (1958); Solan and Vieille (2015); Blackwell and Ferguson (1968); Hansen et al (2018)). Several fundamental questions remained open:

Q1. Does the negative result of Nowak and Raghavan Nowak and Raghavan (1991) still hold in the special case of countable *turn-based* (finitely branching) reachability games?
Q2. If uniformly ε-optimal Maximizer strategies cannot be positional, do they exist at all, e.g., by using more memory, and if yes, how much memory do they need?
Q3. Does the positive result of Secchi (Proposition 1) still hold if the restriction to finite action sets is dropped? The question is meaningful, since concurrent games where only one player has infinite action sets are still determined (Flesch et al, 2020, Theorem 11) (though not if both players have infinite action sets, unless one imposes other restrictions). Moreover, what about

Maximizer	countable MDPs	turn-based games finite-branching	turn-based games infinite-branching	concurrent games finite action sets
ε-optimal	MD (Ornstein, 1969, Thm. B)	MD (Kučera, 2011, Proposition 5.7.c), [Lemma 5]	∞-memory [Theorem 15]	MR (Secchi, 1997, Cor. 3.9)
uniform ε-optimal	MD (Ornstein, 1969, Thm. B)	no MR [Theorem 7]; det. public 1-bit [Theorem 6]	∞-memory [Theorem 15]	no MR (Nowak and Raghavan, 1991); rand. public 1-bit, [Theorem 6]
Optimal	MD (Ornstein, 1969, Prop. B)	no FR (Kučera, 2011, Prop. 5.7.b) No Markov [Proposition 23] step counter + det. public 1-bit, [Theorem 24]	∞-memory [Theorem 15]	∞-memory [Proposition 21]
Almost sure	MD (Ornstein, 1969, Prop. B)	MD (Kiefer et al, 2017a, Theorem 5.3)	∞-memory [Theorem 15]	MR [Theorem 22]

Table 1 The strategy complexity of Maximizer for the reachability objective. Since optimal (and a.s. winning) strategies are not guaranteed to exist, the results in the two bottom rows are conditioned upon their existence. "∞-memory" means that even randomized strategies with a step counter plus an arbitrarily large finite private memory do not suffice. Deterministic strategies are useless in concurrent games, regardless of memory.

Minimizer	turn-based games finite-branching	turn-based games infinite-branching	concurrent games finite action sets
(uniform) ε-optimal	MD (Brázdil et al, 2011, Thm. 3.1)	no FR, (Kiefer et al, 2017b, Thm. 3) det. Markov [Theorem 27]	MR, (Nowak and Raghavan, 1991, Thm. 1)
Optimal	MD, (Brázdil et al, 2011, Thm. 3.1)	∞-memory [Proposition 28]	MR, (Nowak and Raghavan, 1991, Thm. 1)

Table 2 The strategy complexity of Minimizer for the reachability objective. Since optimal Minimizer strategies do not need to exist for infinitely branching games (unlike in the other cases), the result of Proposition 28 is conditioned upon their existence. Deterministic strategies are useless in concurrent games, regardless of memory.

infinitely branching turn-based reachability games? How much memory do good Maximizer strategies need in these cases?

Our Contribution

Our results, summarized in Tables 1 and 2, provide a comprehensive view on the strategy complexity of (uniformly) ε-optimal strategies for reachability (and also about optimal strategies when they exist).

Our first result strengthens the negative result of Nowak and Raghavan Nowak and Raghavan (1991) to the turn-based case.

First Lower-Bound result (Q1): *(Theorem 7) There exists a finitely branching turn-based version of the Big Match on $\mathbb{N}$ where Maximizer still does not have any uniformly ε-optimal MR strategy.*

Our second result solves the open question about uniformly ε-optimal Maximizer strategies. While uniformly ε-optimal Maximizer strategies cannot be positional, 1 bit of memory is enough.

Main Upper-Bound result (Q2): *(Theorem 6) In concurrent games with finite action sets and reachability objective, for any $\varepsilon > 0$, Maximizer has a uniformly ε-optimal public-memory 1-bit strategy. This strategy can be chosen as deterministic if the game is turn-based and finitely branching.*

Our main contribution (Theorem 2) addresses Q3. It determines the strategy complexity of Maximizer in infinitely branching reachability games.

Our result is a strong lower bound, and we present the path towards it via a sequence of (unfulfilled) hopes for an upper bound.

Hope 1: *In turn-based reachability games, Maximizer has ε-optimal MD strategies.*

This is motivated by the fact that the property holds if the game is finitely branching ((Kučera, 2011, Proposition 5.7.c) and Lemma 5) or if it is just an MDP Ornstein (1969).

One might even have hoped for *uniformly* ε-optimal MD strategies, i.e., strategies that do not depend on the start state of the game, but this hope was crushed by the answer to Q1.

Let us mention a concern about Hope 1 as stated (i.e., disregarding uniformity). Consider any turn-based reachability game that is finitely branching, and let $x \in [0, 1]$ be the value of the game. The proof of Proposition 1 actually shows that for every $\varepsilon > 0$, Maximizer has both a strategy and a time horizon $n \in \mathbb{N}$ such that for all Minimizer strategies, the game visits the target state with probability at least $x - \varepsilon$ *within the first n steps* of the game. There is no hope that such a guarantee on the time horizon can be given in infinitely branching games. Indeed, consider the infinitely many states $f_0, f_1, f_2, \ldots$, where f_0 is the target state and for $i > 0$ state f_i leads to f_{i-1} regardless of the players' actions, and an additional Minimizer state, u, where Minimizer chooses, by her action, one of the f_i as successor state. In this game, starting from u, Maximizer wins with probability 1 (he is passive in this game). Minimizer cannot avoid losing, but her strategy determines when f_0 is visited. This shows that a proof of Hope 1 would require different methods.

In case Hope 1 turns out to be false, there are various plausible weaker versions. Let us briefly discuss their motivation.

Hope 2: *Hope 1 is true if MD is replaced by MR.*

This is motivated by Proposition 1, i.e., that in concurrent games with finite action sets for both players, Maximizer has ε-optimal MR but not MD strategies. In fact, (Flesch et al, 2020, Theorem 12.3) implies that this holds even under the weaker assumption that just Minimizer has finite action sets (while Maximizer is allowed infinite action sets).

Hope 3: *Hope 1 is true if Maximizer has an optimal strategy.*

This is motivated by the fact that in MDPs with *Büchi* objective (i.e., the player tries to visit a set of target states infinitely often), if the player has an optimal strategy, he also has an MD optimal strategy. The same is not true for ε-optimal strategies (they require memory) Kiefer et al (2017b). This example shows that although optimal strategies do not always exist, if they do exist, they may be simpler.

Hope 4: *Hope 1 is true if Maximizer has an almost-surely winning strategy, i.e., a strategy that guarantees him to visit the target state with probability* 1.

This is weaker than Hope 3, because an almost-surely winning strategy is necessarily optimal.

In a turn-based game, let us associate to each state s its *value*, which is the value of the game when started in s. We call a controlled step $s \to s'$ *value-decreasing* (resp., *value-increasing*), if the value of s' is smaller (resp., larger) than the value of s. It is easy to see that Maximizer cannot do value-increasing steps and Minimizer cannot do value-decreasing steps, but the opposite is possible in general.

Hope 5: *Hope 1 is true if Maximizer does not have value-decreasing steps.*

Hope 6: *Hope 1 is true if Minimizer does not have value-increasing steps.*

Hopes 5 and 6 are motivated by the fact that sometimes the absence of Maximizer value-decreasing steps or the absence of Minimizer value-increasing steps implies the existence of optimal Maximizer strategies and then Hope 3 might apply. For example, in finitely branching turn-based reachability games, the absence of Maximizer value-decreasing steps or the absence of Minimizer value-increasing steps implies the existence of optimal Maximizer strategies, and they can be chosen MD (Kiefer et al, 2017a, Theorem 5).

Hope 7: *Hope 1 is true for games with acyclic game graph.*

This is motivated, e.g., by the fact that in *safety* MDPs (where the only active player tries to *avoid* a particular state f) with acyclic game graph and infinite action sets the player has ε-optimal MD strategies (Kiefer et al, 2020a, Corollary 26). The same does not hold without the acyclicity assumption (Kiefer et al, 2017b, Theorem 3).

Hope 8: *Hope 1 is true if Maximizer can additionally use a step counter to choose his actions.*

This is weaker than Hope 7, because by using a step counter Maximizer effectively makes the game graph acyclic. However, the reverse does not hold. Not every acyclic game graph has an implicit step counter.

Hope 9: *In turn-based reachability games, Maximizer has ε-optimal strategies that use only finite memory.*

This is motivated, e.g., by the fact that in MDPs with acyclic game graph and Büchi objective, the player has ε-optimal deterministic strategies that require only 1 bit of memory, but no ε-optimal MR strategies Kiefer et al (2019).

It might be advantageous for Maximizer to keep his memory mode private. This motivates the following final weakening of Hope 9.

Hope 10: *In turn-based reachability games, Maximizer has ε-optimal strategies that use only private finite memory.*

The main contribution of this paper is to crush all these hopes. That is, Hope 1 is false, even if all weakenings proposed in Hopes 2-10 are imposed *at the same time*. Specifically, we show the following theorem.

Theorem 2 *There is a turn-based reachability game (necessarily, by Proposition 1, with infinite action sets for Minimizer) with the following properties:*

1. *for every Maximizer state, Maximizer has at most two actions to choose from;*
2. *for every state Maximizer has a strategy to visit the target state with probability* 1*, regardless of Minimizer's strategy;*
3. *for every Maximizer strategy that uses only a step counter and private finite memory and randomization, for every* $\varepsilon > 0$*, Minimizer has a strategy so that the target state is visited with probability at most* ε*.*

This lower bound trivially carries over to concurrent stochastic games with infinite Minimizer action sets, and for all Borel objectives that subsume reachability, e.g., Büchi, co-Büchi, Parity, average-reward and total-reward.

To put this result into perspective, we show in Section 8 that it is crucial that Minimizer can use infinite branching (resp. infinite actions sets) *infinitely often*. If the game is restricted such that Minimizer can use infinite actions sets only *finitely often* in any play then Maximizer still has uniformly ε-optimal public 1-bit strategies.

While optimal Maximizer strategies need not exist in general, it is still relevant to study the case where they do exist. If Minimizer can use infinite branching (resp. infinite actions sets) then Theorem 2 shows that Maximizer needs infinite memory even in that case. However, optimal Maximizer strategies in finitely branching turn-based games require exactly a step counter plus 1 bit of public memory; cf. Section 9.

Finally, in Section 10, we show the strategy complexity of Minimizer for all cases. In particular, Minimizer has uniformly ε-optimal positional strategies in turn-based games that are infinitely branching but acyclic.

2 Preliminaries and Notations

A *probability distribution* over a countable set S is a function $f : S \to [0,1]$ with $\sum_{s \in S} f(s) = 1$. Let $\mathtt{supp}(f) \stackrel{\text{def}}{=} \{s \mid f(s) > 0\}$ denote the support of f. We write $\mathcal{D}(S)$ for the set of all probability distributions over S.

We study perfect information 2-player stochastic games between two players $\Box$ (Maximizer) and $\Diamond$ (Minimizer).

Concurrent Stochastic Games

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 Maximizer and Minimizer, respectively. 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. The set of *plays* from an initial state s_0 is given by the infinite sequences in Z^ω where the first triple contains s_0. The game from s_0 is played in stages $\mathbb{N} = \{0,1,2,\dots\}$. At every stage $t \in \mathbb{N}$, the play is in some state s_t. Maximizer chooses an action $a_t \in A(s_t)$ and Minimizer chooses an action $b_t \in B(s_t)$. The next state s_{t+1} is then chosen according to the distribution $p(s_t,a_t,b_t)$. (Since we just consider the reachability objective here, we don't define a reward function.)

Turn-based Stochastic Games

A special subclass of concurrent stochastic games are turn-based games, where in each round either Maximizer or Minimizer has a single action to play. Turn-based games $\mathcal{G} = (S, (S_\Box, S_\Diamond, S_\bigcirc), \longrightarrow, P)$ are defined such that the countable set of states S is partitioned into the set $S_\Box$ of states of player $\Box$, the set $S_\Diamond$ of states of player $\Diamond$ and *random states* $S_\bigcirc$. The relation $\longrightarrow \subseteq S \times S$ is the transition relation. We write $s{\longrightarrow}s'$ if $(s,s') \in \longrightarrow$, and we assume that each state s has a *successor* state s' with $s{\longrightarrow}s'$. The probability function $P : S_\bigcirc \to \mathcal{D}(S)$ assigns to each random state $s \in S_\bigcirc$ a probability distribution over its successor states.

The game $\mathcal{G}$ is called *finitely branching* if each state has only finitely many successors; otherwise, it is *infinitely branching*. A game is *acyclic* if the underlying graph $(S, \longrightarrow)$ is acyclic. Let $\odot \in \{\Box, \Diamond\}$. At each stage t, if the game is in state $s_t \in S_\odot$ then player $\odot$ chooses a successor state s_{t+1} with $s_t{\longrightarrow}s_{t+1}$; otherwise the game is in a random state $s_t \in S_\bigcirc$ and proceeds randomly to s_{t+1} with probability $P(s_t)(s_{t+1})$. If $S_\odot = \emptyset$, we say that player $\odot$ is *passive*, and the game is a *Markov decision process (MDP)*. A Markov chain is an MDP where both players are passive.

Strategies and Probability Measures

The set of histories at stage t is denoted by H_t, i.e., $H_0 \stackrel{\text{def}}{=} S$ and $H_t \stackrel{\text{def}}{=} Z^t \times S$. Let $H \stackrel{\text{def}}{=} \bigcup_{t\in\mathbb{N}} H_t$ be the set of all histories. For each history h, let s_h denote the final state in h. (In the special case of turn-based games, the history can be represented by the sequence of states $s_0 s_1 \cdots s_h$, where $s_i{\longrightarrow}s_{i+1}$ for all $i \in \mathbb{N}$.) We say that a history $h \in H_t$ *visits* the set of states $T \subseteq S$ at stage t if $s_h \in T$.

A mixed action for Maximizer (resp. Minimizer) in state s is a distribution over $A(s)$ (resp. $B(s)$). A *strategy* for Maximizer (resp. Minimizer) is a function σ (resp. π) that to each history $h \in H$ assigns a mixed action $\sigma(h) \in \mathcal{D}(A(s_h))$ (resp. $\pi(h) \in \mathcal{D}(B(s_h))$ for Minimizer). For turn-based games this means instead a distribution $\sigma(h) \in \mathcal{D}(S)$ over successor states if $s_h \in S_\Box$ (and similarly for Minimizer with $s_h \in S_\Diamond$). Let Σ (resp. Π) denote the set of strategies for Maximizer (resp. Minimizer).

An initial state s_0 and a pair of strategies σ, π for Maximizer and Minimizer induce a probability measure on sets of plays. We write $\mathcal{P}_{\mathcal{G},s_0,\sigma,\pi}(\mathfrak{R})$ for the probability of a measurable set of plays $\mathfrak{R}$ starting from s_0. It is initially defined for the cylinder sets and extended to the sigma algebra by Carathéodory's unique extension theorem Billingsley (1995). We will write $\mathcal{E}_{\mathcal{G},s_0,\sigma,\pi}$ for the expectation w.r.t. $\mathcal{P}_{\mathcal{G},s_0,\sigma,\pi}$.

Objectives

We consider the reachability objective for Maximizer. Given a set $T \subseteq S$ of states, the *reachability* objective $\mathtt{Reach}(T)$ is the set of plays that visit T at least once. From Minimizer's point of view, this is the dual *safety* objective $\mathtt{Safety}(T) \stackrel{\text{def}}{=} Z^\omega \setminus \mathtt{Reach}(T)$ of plays that never visit T. Maximizer (resp. Minimizer) attempts to maximize (resp. minimize) the probability of $\mathtt{Reach}(T)$.

Value and Optimality

For a game $\mathcal{G}$, initial state s_0 and objective $\mathfrak{R}$ the *lower value* is defined as

$$\alpha(s_0) \stackrel{\text{def}}{=} \sup_{\sigma\in\Sigma} \inf_{\pi\in\Pi} \mathcal{P}_{\mathcal{G},s_0,\sigma,\pi}(\mathfrak{R})$$

Similarly, the *upper value* is defined as

$$\beta(s_0) \stackrel{\text{def}}{=} \inf_{\pi\in\Pi} \sup_{\sigma\in\Sigma} \mathcal{P}_{\mathcal{G},s_0,\sigma,\pi}(\mathfrak{R})$$

The inequality $\alpha(s_0) \le \beta(s_0)$ trivially holds. If $\alpha(s_0) = \beta(s_0)$, then this quantity is called the *value* of the game, denoted by $\mathtt{val}_{\mathcal{G},\mathfrak{R}}(s_0)$. Reachability objectives, like all Borel objectives, have value by Maitra and Sudderth (1998). For $\varepsilon > 0$, a strategy $\sigma \in \Sigma$ from s_0 for Maximizer is called ε*-optimal* if $\forall \pi \in \Pi.\, \mathcal{P}_{\mathcal{G},s_0,\sigma,\pi}(\mathfrak{R}) \ge \mathtt{val}_{\mathcal{G},\mathfrak{R}}(s_0) - \varepsilon$. Similarly, a strategy $\pi \in \Pi$ from s_0 for Minimizer is called ε-optimal if $\forall \sigma \in \Sigma.\, \mathcal{P}_{\mathcal{G},s_0,\sigma,\pi}(\mathfrak{R}) \le \mathtt{val}_{\mathcal{G},\mathfrak{R}}(s_0) + \varepsilon$. If a strategy is 0-optimal we simply call it optimal.

Memory-based Strategies

A *memory-based strategy* σ of Maximizer is a strategy that can be described by a tuple $(\mathsf{M}, \mathsf{m}_0, \sigma_\alpha, \sigma_\mathsf{m})$ where M is the set of memory modes, $\mathsf{m}_0 \in \mathsf{M}$ is the initial memory mode, and the functions σ_α and σ_m describe how actions are chosen and memory modes updated; see below. A play according to σ generates a random sequence of memory states $\mathsf{m}_0, \ldots, \mathsf{m}_t, \mathsf{m}_{t+1}, \ldots$ from a given set of memory modes M, where m_t is the memory mode at stage t. The strategy σ selects the action at stage t according to a distribution that depends only on the current state s_t and the memory m_t. Maximizer's action a_t is chosen via a distribution $\sigma_\alpha(s_t, \mathsf{m}_t) \in \mathcal{D}(A(s_t))$. (Minimizer's action is b_t). The next memory mode m_{t+1} of Maximizer is chosen according to a distribution $\sigma_m(s_t, a_t, b_t, s_{t+1}) \in \mathcal{D}(\mathsf{M})$ that depends on the chosen actions

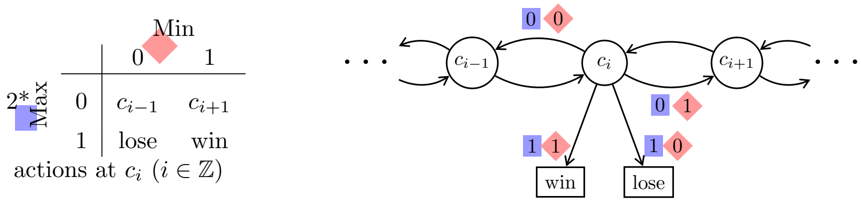

Fig. 1 Concurrent Big Match on $\mathbb{Z}$; see Definition 1

and the observed outcome. The memory is *private* if the other player cannot see the memory mode. Otherwise, it is *public*.

Let $\sigma[\mathsf{m}]$ denote the memory-based strategy σ that starts in memory mode m. In cases where the time is relevant and the strategy has access to the time (by using a step counter) $\sigma[\mathsf{m}](t)$ denotes the strategy σ in memory mode m at time t.

A *finite-memory strategy* is one where $|\mathsf{M}| < \infty$. A *k-memory strategy* is a memory-based strategy with at most k memory modes, i.e., $|\mathsf{M}| \leq k$. A 2-memory strategy is also called a *1-bit strategy*. A strategy is *memoryless* (also called positional or stationary) if $|\mathsf{M}| = 1$. A strategy is called *Markov* if it uses only a step counter but no additional memory. A strategy is *deterministic* (also called pure) if the distributions for the action and memory update are Dirac. Otherwise, it is called *randomized* (or mixed). Pure stationary strategies are also called MD (memoryless deterministic) and mixed stationary strategies are also called MR (memoryless randomized). Similarly, pure (resp. mixed) finite-memory strategies are also called FD (resp. FR).

A finite-memory strategy σ is called *uniformly ε-optimal* for an objective $\mathfrak{R}$ if there is a memory mode $\mathsf{m}_0 \in \mathsf{M}$ such that $\forall s \in S. \forall \pi. \mathcal{P}_{\mathcal{G},s,\sigma[\mathsf{m}_0],\pi}(\mathfrak{R}) \geq \mathtt{val}_{\mathcal{G},\mathfrak{R}}(s) - \varepsilon$, i.e., the strategy performs well from every state.

The definitions above carry over directly to the simpler turn-based games where we have chosen/observed transitions instead of actions.

3 Uniform Strategies in Concurrent Games

First, we consider the concurrent version of the Big Match on the integers, in the formulation of Fristedt et al (1995).

Definition 1 (Concurrent Big Match on $\mathbb{Z}$) *This game is shown in Figure 1. The state space is $\{c_i \mid i \in \mathbb{Z}\} \cup \{\text{win}, \text{lose}\}$, where the states* win *and* lose *are absorbing. Both players have the action set $\{0, 1\}$ at each state. If Maximizer chooses action 1 in c_i then the game is decided in this round: If Minimizer chooses 0 (resp. 1) then the game goes to* lose *(resp.* win*). If Maximizer chooses action 0 in c_i and Minimizer chooses action 0 (resp. 1) then the game goes to c_{i-1} (resp. c_{i+1}). Maximizer wins iff state* win *is reached or $\liminf\{i \mid c_i \text{ visited}\} = -\infty$.*

Theorem 3 ((Fristedt et al, 1995, Theorem 1.1)) *In the concurrent Big Match on $\mathbb{Z}$, shown in Figure 1, every state c_i has value $1/2$. An optimal strategy for Minimizer*

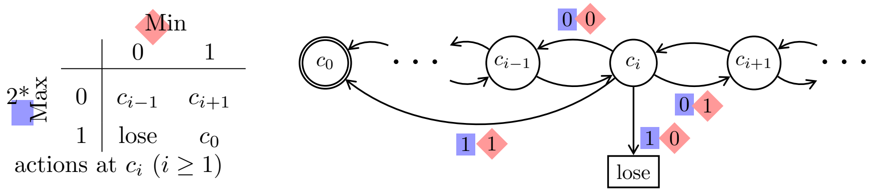

Fig. 2 Concurrent Big Match on $\mathbb{N}$; see Definition 2.

is to toss a fair coin at every stage. Maximizer has no optimal strategy, but for any start state c_x and any positive integer N, he can win with probability $\geq N/(2N+2)$ by choosing action 1 with probability $1/(n+1)^2$ whenever the current state is c_i with $i = x + N - n$ for some $n \geq 0$.

The concurrent Big Match on $\mathbb{Z}$ is not a reachability game, due to its particular winning condition. However, the following slightly modified version (played on $\mathbb{N}$) is a reachability game.

Definition 2 (Concurrent Big Match on $\mathbb{N}$) *This game is shown in Figure 2. The state space is $\{c_i \mid i \in \mathbb{N}\} \cup \{\text{lose}\}$ where* lose *and c_0 are absorbing. Both players have the action set $\{0, 1\}$ at each state. If Maximizer chooses action 1 in c_i then the game is decided in this round: If Minimizer chooses 0 (resp. 1) then the game goes to* lose *(resp. c_0). If Maximizer chooses action 0 in c_i and Minimizer chooses action 0 (resp. 1) then the game goes to c_{i-1} (resp. c_{i+1}).*

Maximizer wins iff c_0 is reached, i.e., we have the reachability objective $\texttt{Reach}(\{c_0\})$.

The following theorem summarizes results on the concurrent Big Match on $\mathbb{N}$ by combining results from Fristedt et al (1995) and Nowak and Raghavan (1991).

Theorem 4 *Denote by $\mathcal{G}$ the concurrent Big Match game on $\mathbb{N}$, as shown in Figure 2, and let $x \geq 0$. Then,*

1. $\texttt{val}_{\mathcal{G}}(c_x) = (x+2)/(2x+2) \geq 1/2$.
2. *For every start state c_x and $N \geq 0$, Maximizer can win with probability $\geq N/(2N+2)$ by choosing action 1 with probability $1/(n+1)^2$ whenever the current state is c_i with $i = x + N - n$ for some $n \geq 0$.*
3. *For any $\varepsilon < 1/2$ there is no uniformly ε-optimal positional (MR) strategy for Maximizer. Every MR Maximizer strategy σ attains arbitrarily little from c_x as $x \to \infty$. Formally,* $\limsup_{x\to\infty} \inf_\pi \mathcal{P}_{\mathcal{G},c_x,\sigma,\pi}(\texttt{Reach}(\{c_0\})) = 0$.

Proof Item 1 follows directly from (Fristedt et al, 1995, Proposition 5.1).

Item 2 follows from Theorem 3, since it is easier for Maximizer to win in the game of Definition 2 than in the game of Definition 1.

Towards item 3, we follow the proof of (Nowak and Raghavan, 1991, Lemma 4). Let σ be an MR Maximizer strategy and $f(x)$ the probability that σ picks action 1 at state c_x. There are two cases.

In the first case $\sum_{x\geq 1} f(x) < \infty$. Let π be the strategy of Minimizer that always picks action 1. For all $x \geq 1$ we have

$$\begin{aligned}\mathcal{P}_{\mathcal{G},c_x,\sigma,\pi}(\mathtt{Reach}(\{c_0\})) \quad \leq \quad & f(x) + (1-f(x))f(x+1) \\ & + (1-f(x))(1-f(x+1))f(x+2) \\ & + \ldots \;\leq \sum_{k=x}^{\infty} f(k) < \infty.\end{aligned}$$

Thus $\limsup_{x\to\infty} \mathcal{P}_{\mathcal{G},c_x,\sigma,\pi}(\mathtt{Reach}(\{c_0\})) = 0$.

In the second case $\sum_{x\geq 1} f(x) = \infty$. Let π be the strategy of Minimizer that always picks action 0. For all $x \geq 1$ we have

$$\begin{aligned}\mathcal{P}_{\mathcal{G},c_x,\sigma,\pi}(\mathtt{Reach}(\{c_0\})) &= (1-f(x))(1-f(x-1))\cdots(1-f(1)) \\ &= \prod_{k=1}^{x}(1-f(k)) \\ &\leq \frac{1}{1+\sum_{k=1}^{x} f(k)} \qquad \text{by Proposition 29}\end{aligned}$$

Thus $\limsup_{x\to\infty} \mathcal{P}_{\mathcal{G},c_x,\sigma,\pi}(\mathtt{Reach}(\{c_0\})) = 0$.

Since, by item 1, $\mathtt{val}_{\mathcal{G}}(c_x) = (x+2)/(2x+2) \geq 1/2$ for every $x \geq 0$, the MR strategy σ cannot be uniformly ε-optimal for any $\varepsilon < 1/2$. □

While uniformly ε-optimal Maximizer strategies cannot be positional, we show that they can be chosen with just 1 bit of public memory in Theorem 6.

First we need an auxiliary lemma that is essentially known; see, e.g., (Maitra and Sudderth, 1996, Section 7.7) and (Flesch et al, 2020, Theorem 12.1). We extend it slightly to fit our purposes, i.e., for the proof of Theorem 6 below.

Lemma 5 *Consider a concurrent game with countable state space S,* finite *action sets for Minimizer at every state and unrestricted (possibly infinite) action sets for Maximizer.*

For the reachability objective $\mathtt{Reach}(T)$, for every finite set $S_0 \subseteq S$ of initial states, and for every $\varepsilon > 0$, there exists a memoryless strategy σ and a finite set of states $R \subseteq S$ such that for all $s_0 \in S_0$

$$\inf_{\pi\in\Pi} \mathcal{P}_{s_0,\sigma,\pi}(\mathtt{Reach}_R(T)) \;\geq\; \mathtt{val}_{\mathtt{Reach}(T)}(s_0) - \varepsilon\,,$$

where $\mathtt{Reach}_R(T)$ denotes the objective of visiting T while remaining in R before visiting T. If the game is turn-based and finitely branching at Minimizer-controlled states, there is a deterministic (i.e., MD) such strategy σ.

Proof Since Minimizer's action sets are finite, by (Flesch et al, 2020, Theorem 11.1), the game has value. Moreover, using the finiteness of Minimizer's action sets again, it follows from (Flesch et al, 2020, Theorem 12.1) that for all $s \in S$

$$\lim_{n\to\infty} \mathtt{val}_{\mathtt{Reach}_n(T)}(s) \;=\; \mathtt{val}_{\mathtt{Reach}(T)}(s)\,, \tag{1}$$

where $\mathtt{Reach}_n(T)$ denotes the objective of visiting T within at most n rounds of the game.

To achieve the uniformity (across the set S_0 of initial states) required by the statement of the lemma, we add a fresh "random" state (i.e., a state in which each player has only a single action available) that branches uniformly at random to a state in S_0. Call this state $\hat{s}_0$. The value of $\hat{s}_0$ is the arithmetic average of the values of the states in S_0. It follows that every $(\varepsilon/|S_0|)$-optimal memoryless strategy for Maximizer in $\hat{s}_0$ must be ε-optimal in every state in S_0. So it suffices to prove the statement of the lemma under the assumption that S_0 is a singleton, say $S_0 = \{s_0\}$.

Fix $\varepsilon > 0$ and let $\varepsilon' \stackrel{\text{def}}{=} \varepsilon/4$. By Equation (1) there is a number n such that $\mathtt{val}_{\mathtt{Reach}_n(T)}(s_0) = \mathtt{val}_{\mathtt{Reach}(T)}(s_0) - \varepsilon'$. Let σ be a Maximizer strategy such that

$$\inf_{\pi\in\Pi} \mathcal{P}_{s_0,\sigma,\pi}(\mathtt{Reach}_n(T)) \;\geq\; \mathtt{val}_{\mathtt{Reach}(T)}(s_0) - 2\varepsilon' . \tag{2}$$

For each m with $0 \leq m \leq n$ we will inductively construct a finite subset $H'_m \subseteq H_m$ of the m-step histories of plays from s_0 that are compatible with σ such that, for every Minimizer strategy π, the plays in $H'_m Z^\omega$ have probability $\geq 1 - \frac{m}{n}\varepsilon'$, where the event $H'_m Z^\omega$ is defined as the set of continuations of the the m-step histories in H'_m. Formally,

$$\inf_{\pi\in\Pi} \mathcal{P}_{s_0,\sigma,\pi}(H'_m Z^\omega) \;\geq\; 1 - \frac{m}{n}\varepsilon' \tag{3}$$

The base case of $m = 0$ is trivial. Now we show the inductive step from m to $m+1$. For any of the finitely many histories $h \in H'_m$ ending in some state s, consider the chosen mixed actions $a \in \mathcal{D}(A(s))$ and $b \in \mathcal{D}(B(s))$ by Maximizer and Minimizer, respectively. Since $B(s)$ is finite, b has finite support. However, the distribution a can have infinite support. We fix a sufficiently large finite subset A' of the support of a that has probability mass $\geq 1 - \frac{\varepsilon'}{2n}$. Consider the set $\gamma(s)$ of possible successor states of s. Since the size of the support of b is upper bounded by the finite number $|B(s)|$ independently of π, we can pick a finite subset $\gamma'(s) \subseteq \gamma(s)$ sufficiently large such that both Maximizer's chosen action is inside A' and the chosen successor state is inside $\gamma'(s)$ with probability $\geq 1 - \frac{1}{n}\varepsilon'$. We then define H'_{m+1} as the finitely many one-round extensions of histories in H'_m with Maximizer action in A' and successor state in $\gamma'(s)$. Using the induction hypothesis and the properties above, we obtain that

$$\inf_{\pi\in\Pi} \mathcal{P}_{s_0,\sigma,\pi}(H'_{m+1} Z^\omega) \;\geq\; 1 - \frac{m}{n}\varepsilon' - \frac{\varepsilon'}{n} = 1 - \frac{m+1}{n}\varepsilon'$$

This completes the induction step.

For every $0 \leq m \leq n$ let R_m be the finite set of states that are visited during the first m steps of the histories in H'_m. Then $R \stackrel{\text{def}}{=} R_n$ is a finite set of states. Using (2) and (3), it follows that $\inf_{\pi\in\Pi} \mathcal{P}_{s_0,\sigma,\pi}(\mathtt{Reach}_R(T)) \geq \inf_{\pi\in\Pi} \mathcal{P}_{s_0,\sigma,\pi}(H'_n Z^\omega \cap \mathtt{Reach}_R(T)) \geq \inf_{\pi\in\Pi} \mathcal{P}_{s_0,\sigma,\pi}(\mathtt{Reach}_n(T)) - \varepsilon' \geq \mathtt{val}_{\mathtt{Reach}(T)}(s_0) - 3\varepsilon'$. (Note that the restriction on the time horizon, n, has been lifted here). The restriction of the objective to the (finitely many) states in R means that we have effectively another reachability game. It is known (see, e.g., (Secchi, 1997, Corollary 3.9)) that for concurrent games with finite action sets and reachability objective, Maximizer has a memoryless ε'-optimal strategy. In turn-based games with finitely many states, he even has an MD optimal strategy Condon (1992). So Maximizer has a memoryless (in the turn-based case: MD) strategy σ' so that

$$\inf_{\pi\in\Pi} \mathcal{P}_{s_0,\sigma',\pi}(\mathtt{Reach}_R(T)) \geq \mathtt{val}_{\mathtt{Reach}(T)}(s_0) - 4\varepsilon' = \mathtt{val}_{\mathtt{Reach}(T)}(s_0) - \varepsilon.$$

$\square$

Theorem 6 *For any concurrent game with finite action sets and reachability objective, for any $\varepsilon > 0$, Maximizer has a uniformly ε-optimal public 1-bit strategy. If the game is turn-based and finitely branching, Maximizer has a deterministic such strategy.*

Proof Denote the game by $\hat{\mathcal{G}}$, over state space S. It is convenient to describe the 1-bit strategy in terms of a game $\mathcal{G}$ with state space $S \times \{0, 1\}$, where the second component (0 or 1) reflects the current memory mode of Maximizer. Accordingly, we think of the state space of $\mathcal{G}$ as organized in two *layers*, the "inner" and the "outer" layer, with the memory mode being 0 and 1, respectively. In these terms, our goal is to construct, for the layered game $\mathcal{G}$, a *memoryless* strategy for Maximizer. Upon reaching the target, the memory mode is irrelevant, so for notational simplicity we denote the objective as $\mathtt{Reach}(T)$, also in the layered game $\mathcal{G}$ (instead of $\mathtt{Reach}(T \times \{0, 1\})$). The current state of the layered game is known to both players (to Minimizer in particular); this corresponds to the (1-bit) memory being public in the original game $\hat{\mathcal{G}}$: at each point in the game, Minimizer knows the distribution of actions that Maximizer is about to play. Notice that the values of states $(s, 0)$ and $(s, 1)$ in $\mathcal{G}$ are equal to the value of s in $\hat{\mathcal{G}}$; this is because the definition of value does not impose restrictions on the memory of strategies and so the players could, in $\hat{\mathcal{G}}$, simulate the two layers of $\mathcal{G}$ in their memory if that were advantageous.

In general Maximizer does not have uniform ε-optimal memoryless strategies in reachability games; cf. Theorem 4. So our construction will exploit the special structure in the layered game, namely, the symmetry of the two layers. The memoryless Maximizer strategy we construct will be ε-optimal from each state $(s, 0)$ in the inner layer, but not necessarily from the states in the outer layer.

As building blocks we use the non-uniform memoryless strategies we get from Lemma 5; in the turn-based finitely branching case they are even MD. We combine them by "plastering" the state space (of the layered game). This is inspired by Ornstein's construction Ornstein (1969); see (Kiefer et al, 2020b, Section 3.2) for a recent description.[2]

In the general concurrent case, a memoryless strategy prescribes for each state (s, i) a probability distribution over Maximizer's actions. We define a memoryless strategy by successively *fixing* such distributions in more and more states. Technically, one can *fix* a state s by replacing the actions $A(s)$ available to Maximizer by a single action which is a convex combination over $A(s)$. Visually, we "plaster" the whole state space by the fixings. This is in general an infinite (but countable) process; it defines a memoryless strategy for Maximizer in the limit.

In the turn-based and finitely branching case, an MD strategy prescribes one outgoing transition for each Maximizer state. Accordingly, *fixing* a Maximizer state means restricting the outgoing transitions to a single such outgoing transition. The plastering proceeds similarly as in the concurrent case; it defines an MD strategy for Maximizer in the limit.

Put the states of $\hat{\mathcal{G}}$ in some order, i.e., $s_1, s_2, \ldots$ with $S = \{s_1, s_2, \ldots\}$. The plastering proceeds in *rounds*. In round $i \geq 1$ we fix the states in $S_i[0] \times \{0\}$ and in $S_i[1] \times \{1\}$, where $S_1[0], S_2[0], \ldots \subseteq S$ are pairwise disjoint and $S_1[1], S_2[1], \ldots \subseteq S$ are pairwise disjoint; see Figure 3 for an example of sets $S_i[0]$ and $S_i[1]$ in a two-

[2] These papers consider MDPs, i.e., Minimizer is passive. In MDPs Maximizer has uniform ε-optimal MD strategies (even without layering the system) Ornstein (1969).

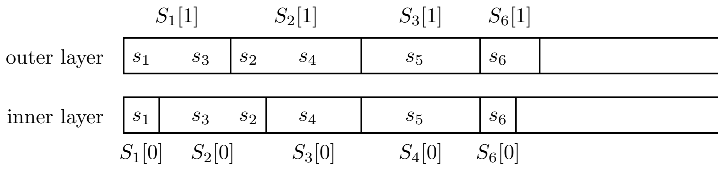

Fig. 3 Example of sets $S_i[0]$ and $S_i[1]$ in the two layers of $\mathcal{G}$, where the outer and inner layers are $S \times \{1\}$ and $S \times \{0\}$, respectively.

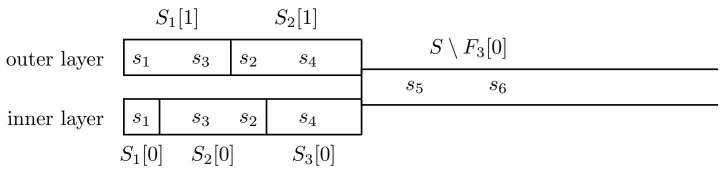

Fig. 4 Example of a game $\bar{\mathcal{G}}_3$.

layers game. Define $F_i[0] \stackrel{\text{def}}{=} \bigcup_{j \leq i} S_j[0]$ and $F_i[1] \stackrel{\text{def}}{=} \bigcup_{j \leq i} S_j[1]$. So $F_i[0] \times \{0\}$ and $F_i[1] \times \{1\}$ are the states that have been fixed by the end of round i. We will keep an invariant $F_i[0] \subseteq F_i[1] \subseteq F_{i+1}[0]$.

Let $\mathcal{G}_i$ be the game obtained from $\mathcal{G}$ after the fixings of the first $i-1$ rounds (with $\mathcal{G}_1 = \mathcal{G}$). Define

$$S_i[0] \stackrel{\text{def}}{=} (\{s_i\} \cup S_{i-1}[1]) \setminus F_{i-1}[0]$$

(and $S_1[0] \stackrel{\text{def}}{=} \{s_1\}$), the set of states to be fixed in round i. In particular, round i guarantees that states $(s,0)$ whose "outer sibling" $(s,1)$ has been fixed previously are also fixed, ensuring $F_{i-1}[1] \subseteq F_i[0]$. Note also that $\bigcup_{j=1}^{\infty} S_j[0] = \bigcup_{j=1}^{\infty} S_j[1] = S$. The set $S_i[1]$ will be defined below.

In round i we fix the states in $(S_i[0] \times \{0\}) \cup (S_i[1] \times \{1\})$ in such a way that

1. starting from any $(s,0)$ with $s \in S_i[0]$, the (infimum over all Minimizer strategies π) probability of reaching T using only fixed states is not much less than the value $\mathtt{val}_{\mathcal{G}_i,\mathtt{Reach}(T)}((s,0))$; and
2. for all states $(s,0) \in S \times \{0\}$ in the inner layer, the value $\mathtt{val}_{\mathcal{G}_{i+1},\mathtt{Reach}(T)}((s,0))$ is almost as high as $\mathtt{val}_{\mathcal{G}_i,\mathtt{Reach}(T)}((s,0))$.

The purpose of goal (A) is to guarantee good progress towards the target when starting from any state $(s,0)$ in $S_i[0] \times \{0\}$. The purpose of goal (B) is to avoid fixings that would cause damage to the values of other states in the inner layer.

We want to define the fixings in round i. First we define an auxiliary game $\bar{\mathcal{G}}_i$ with state space $\bar{S}_i \stackrel{\text{def}}{=} (F_i[0] \times \{0,1\}) \cup (S \setminus F_i[0])$. Game $\bar{\mathcal{G}}_i$ is obtained from $\mathcal{G}_i$ by collapsing, for all $s \in S \setminus F_i[0]$, the siblings $(s,0),(s,1)$ (neither of which have been fixed yet) to a single state s. See Figure 4. The game $\bar{\mathcal{G}}_i$ inherits the fixings from $\mathcal{G}_i$. The values remain equal; in particular, for $s \in S \setminus F_i[0]$, the values of $(s,0)$ and $(s,1)$ in $\mathcal{G}_i$ and the value of s in $\bar{\mathcal{G}}_i$ are all equal.

Let $\varepsilon_i > 0$. We apply Lemma 5 to $\bar{\mathcal{G}}_i$ with set of initial states $S_i[0] \times \{0\}$. So Maximizer has a memoryless strategy σ_i for $\bar{\mathcal{G}}_i$ and a finite set of states $R \subseteq \bar{S}_i$ so that for all $s \in S_i[0]$ we have $\inf_\pi \mathcal{P}_{\bar{\mathcal{G}}_i,(s,0),\sigma_i,\pi}(\texttt{Reach}_R(T)) \geq \texttt{val}_{\mathcal{G}_i,\texttt{Reach}(T)}((s,0)) - \varepsilon_i$.

Now we carry the strategy σ_i from $\bar{\mathcal{G}}_i$ to $\mathcal{G}_i$ by suitably adapting it (see below). Then we obtain $\mathcal{G}_{i+1}$ from $\mathcal{G}_i$ by fixing (the adapted version of) σ_i in $\mathcal{G}_i$.

More precisely, we adapt σ_i to $\mathcal{G}_i$ by treating states $s \in S \setminus F_i[0]$ in $\bar{\mathcal{G}}_i$ as states in the *outer* layer $(s,1)$ of $\mathcal{G}_i$. In particular, transitions that in $\bar{\mathcal{G}}_i$ go from a state $(s,j) \in F_i[0] \times \{0,1\}$ to a state $s' \in S \setminus F_i[0]$ (or vice versa, respectively), go in $\mathcal{G}_i$ from (s,j) to $(s',1)$ (or vice versa, respectively).

Accordingly, define $S_i[1] \stackrel{\text{def}}{=} (S_i[0] \setminus F_{i-1}[1]) \cup ((S \setminus F_i[0]) \cap R)$ (this ensures that $F_i[0] \subseteq F_i[1]$), and obtain $\mathcal{G}_{i+1}$ from $\mathcal{G}_i$ by fixing the adapted version of σ_i in $(S_i[0] \times \{0\}) \cup (S_i[1] \times \{1\})$. This yields, for all $s \in S_i[0]$,

$$\inf_{\sigma,\pi} \mathcal{P}_{\mathcal{G}_{i+1},(s,0),\sigma,\pi}(\texttt{Reach}_{(F_i[0]\times\{0\})\cup(F_i[1]\times\{1\})}(T)) \geq \texttt{val}_{\mathcal{G}_i,\texttt{Reach}(T)}((s,0)) - \varepsilon_i, \tag{4}$$

achieving goal (A) above. Notice that the fixings in $\mathcal{G}_{i+1}$ "lock in" a good attainment from $S_i[0] \times \{0\}$, regardless of the Maximizer strategy σ. Now we extend (4) to achieve goal (B) from above: for all $s \in S$ we have

$$\texttt{val}_{\mathcal{G}_{i+1},\texttt{Reach}(T)}((s,0)) \ \geq \ \texttt{val}_{\mathcal{G}_i,\texttt{Reach}(T)}((s,0)) - \varepsilon_i\,. \tag{5}$$

Indeed, consider any Maximizer strategy σ in $\mathcal{G}_i$ from any $(s,0)$. Without loss of generality we can assume that σ is such that the play enters the outer layer only (if at all) after having entered $F_i[0] \times \{0\}$. Now change σ to a strategy σ' in $\mathcal{G}_{i+1}$ so that as soon as $F_i[0] \times \{0\}$ is entered, σ' respects the fixings (and plays arbitrarily afterwards). By (4) this decreases the (infimum over Minimizer strategies π) probability by at most ε_i. Thus,

$$\inf_\pi \mathcal{P}_{\mathcal{G}_{i+1},(s,0),\sigma',\pi}(\texttt{Reach}(T)) \ \geq \ \inf_\pi \mathcal{P}_{\mathcal{G}_i,(s,0),\sigma,\pi}(\texttt{Reach}(T)) - \varepsilon_i\,.$$

Taking the supremum over strategies σ in $\mathcal{G}_i$ yields (5).

For any $\varepsilon > 0$ choose $\varepsilon_i \stackrel{\text{def}}{=} 2^{-i}\varepsilon$; thus, $\sum_{i\geq 1} \varepsilon_i = \varepsilon$. Let σ be the memoryless strategy that respects all fixings in all $\mathcal{G}_i$. Then, by (5), for all $s \in S$ we have

$$\inf_\pi \mathcal{P}_{\mathcal{G},(s,0),\sigma,\pi}(\texttt{Reach}(T)) \ \geq \ \texttt{val}_{\mathcal{G},\texttt{Reach}(T)}((s,0)) - \sum_{i=1}^{\infty} \varepsilon_i\,,$$

so σ is ε-optimal in $\mathcal{G}$ from all $(s,0)$. Hence, the corresponding public 1-bit memory strategy (with initial memory mode 0, corresponding to the inner layer) is uniform ε-optimal in $\hat{\mathcal{G}}$. ☐

4 Uniform Strategies in Turn-based Games

We now define a finitely branching turn-based game in Definition 3 that is very similar to the concurrent Big Match on $\mathbb{N}$, as shown in Figure 2. The difference is that at each c_i Maximizer has to announce his mixed choice of actions first, rather than concurrently with Minimizer. Note that Maximizer only announces his distribution over the actions $\{0,1\}$, not any particular action. Since a good Maximizer strategy needs to work even if Minimizer knows it in advance, this makes no difference with respect to the attainment of positional Maximizer

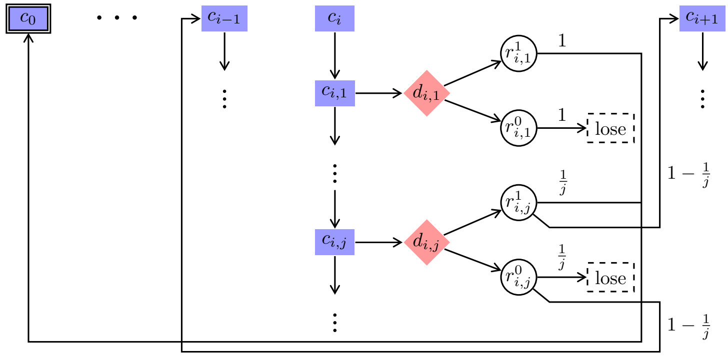

Fig. 5 Turn-based Big Match on $\mathbb{N}$.

strategies. Another slight difference is that Maximizer is restricted to choosing distributions with only *rational* probabilities where the probability of picking action 1 is of the form $1/k$ for some $k \in \mathbb{N}$. However, since we know that there exist good Maximizer strategies of this form (cf. Theorem 4), it is not a significant restriction.

Definition 3 (Turn-based Big Match on $\mathbb{N}$) *This game is shown in Figure 5. Maximizer controls the set* $\{c_i \mid i \in \mathbb{N}\} \cup \{c_{i,j} \mid i,j \in \mathbb{N}\} \cup \{\text{lose}\}$ *of states, whereas Minimzer controls only the states in* $\{d_{i,j} \mid i,j \in \mathbb{N}\}$*. The remaining set* $\{r^0_{i,j}, r^1_{i,j} \mid i,j \in \mathbb{N}\}$ *of states are random. For all* $i,j \in \mathbb{N}$*, there are following transitions*

$$c_i \longrightarrow c_{i,1} \qquad c_{i,j} \longrightarrow c_{i,j+1} \qquad c_{i,j} \longrightarrow d_{i,j}$$
$$d_{i,j} \longrightarrow r^0_{i,j} \qquad d_{i,j} \longrightarrow r^1_{i,j}$$

and lose$\longrightarrow$lose*. Intuitively, by going from* c_i *to* $d_{i,j}$*, Maximizer chooses action* 1 *with probability* $1/j$ *and action* 0 *with probability* $1-1/j$*. Minimizer chooses actions* 0 *or* 1 *by going from* $d_{i,j}$ *to* $r^0_{i,j}$ *or* $r^1_{i,j}$*, respectively. The probabilistic function is defined by*

$$P(r^0_{i,j})(\text{lose}) = 1/j \qquad P(r^0_{i,j})(c_{i-1}) = 1 - 1/j$$
$$P(r^1_{i,j})(c_0) = 1/j \qquad P(r^1_{i,j})(c_{i+1}) = 1 - 1/j$$

where $i,j \in \mathbb{N}$*. The objective is* $\texttt{Reach}(\{c_0\})$*.*

This finitely branching turn-based game mimics the behavior of the game in Definition 2.

Theorem 7 *Consider the turn-based Big Match game* $\mathcal{G}$ *on* $\mathbb{N}$ *from Definition 3 and let* $x \geq 0$*.*

1. *For every start state* c_x *and* $N \geq 0$*, Maximizer can win with probability* $\geq N/(2N+2)$ *by choosing the transitions* $c_i \longrightarrow \ldots d_{i,j}$ *where* $j = (n+1)^2$ *whenever he is in state* c_i *with* $i = x + N - n$ *for some* $n \geq 0$*.*

   In particular, $\texttt{val}_{\mathcal{G}}(c_x) \geq 1/2$*.*
2. *For any* $\varepsilon < 1/2$ *there does not exist any uniformly* ε*-optimal positional (MR) strategy for Maximizer.*

Every MR Maximizer strategy σ attains arbitrarily little from c_x as $x \to \infty$. Formally, $\limsup_{x\to\infty} \inf_{\pi} \mathcal{P}_{\mathcal{G},c_x,\sigma,\pi}(\texttt{Reach}(\{c_0\})) = 0$.

Proof Let $\mathcal{G}'$ be the concurrent game from Definition 2.

Towards item 1, consider the concurrent game $\mathcal{G}'$ and the turn-based game $\mathcal{G}$ from Definition 3. Let c_x be our start state. After fixing the Maximizer strategy from Theorem 4(2) in $\mathcal{G}'$, we obtain an MDP $\mathcal{M}'$ from Minimizer's point of view. Similarly, after fixing the strategy described above in $\mathcal{G}$, we obtain an MDP $\mathcal{M}$. Then $\mathcal{M}'$ and $\mathcal{M}$ are almost isomorphic (apart from linear chains of steps $c_{j,j} \longrightarrow c_{i,j+1} \dots$ in $\mathcal{M}'$), and thus the infimum of the chance of winning, over all Minimizer strategies are the same. Therefore the result follows from Theorem 4(2).

Towards item 2, note that every Maximizer MR strategy σ in $\mathcal{G}$ corresponds to a Maximizer MR strategy σ' in $\mathcal{G}'$. First, forever staying in states $c_{i,j}$ is losing, since the target is never reached. Thus, without restriction, we assume that σ almost surely moves from c_i to some $d_{i,j}$ eventually. Let $p_{i,j}$ be the probability that σ moves from c_i to $d_{i,j}$. Thus the corresponding strategy σ' in $\mathcal{G}'$ in c_i plays action 1 with probability $\sum_j p_{i,j}(1/j)$ and action 0 otherwise. Again the MDPs resulting from fixing the respective strategies in $\mathcal{G}$ and $\mathcal{G}'$ are (almost) isomorphic, and thus the result follows from Theorem 4(3). $\square$

In the rest of this section we describe an alternative construction of a turn-based finitely branching reachability game without uniform ε-optimal Maximizer strategies via applying Theorem 6 to the concurrent game of Definition 2.

Definition 4 *Consider the concurrent reachability game from Definition 2 and let $\varepsilon = 1/4$. By Theorem 6, Maximizer has a uniform ε-optimal 1-bit strategy $\hat{\sigma}$. Let $p_{i,0}$ (resp. $p_{i,1}$) be the probability that $\hat{\sigma}$ picks action* 1 *at state c_i when in memory mode* 0 *(resp. memory mode* 1*).*

We construct a turn-based reachability game $\mathcal{G}$ with branching degree two where Maximizer can pick randomized actions according to these probabilities $p_{i,0}, p_{i,1}$, but nothing else. Let $\mathcal{G} = (S, (S_\Box, S_\Diamond, S_\bigcirc), \longrightarrow, P)$ where $S_\Box = \{c_i \mid i \in \mathbb{N}\} \cup \{\text{lose}\}$, $S_\Diamond = \{d_{i,0}, d_{i,1} \mid i \in \mathbb{N}\}$, and $S_\bigcirc = \{r_{i,0,0}, r_{i,0,1}, r_{i,1,0}, r_{i,1,1} \mid i \in \mathbb{N}\}$. We have controlled transitions $c_i \longrightarrow d_{i,j}$, $d_{i,j} \longrightarrow r_{i,j,k}$ for all $i \in \mathbb{N}$ and $j, k \in \{0,1\}$ and lose$\longrightarrow$lose. *Intuitively, by going from c_i to $d_{i,j}$, Maximizer chooses action* 1 *with probability $p_{i,j}$ and action* 0 *otherwise. Minimizer chooses action k by going from $d_{i,j}$ to $r_{i,j,k}$. The random transitions are defined by $P(r_{i,j,0})(\text{lose}) = p_{i,j}$, $P(r_{i,j,0})(c_{i-1}) = 1 - p_{i,j}$, $P(r_{i,j,1})(c_0) = p_{i,j}$, $P(r_{i,j,1})(c_{i+1}) = 1 - p_{i,j}$.*

The objective is $\texttt{Reach}(\{c_0\})$.

Theorem 8 *Consider the turn-based reachability game $\mathcal{G}$ of branching degree two from Definition 4 and let $x \geq 0$.*

1. $\texttt{val}_{\mathcal{G}}(c_x) \geq 1/4$.
2. *There does not exist any uniformly ε-optimal positional (MR) strategy for Maximizer.*

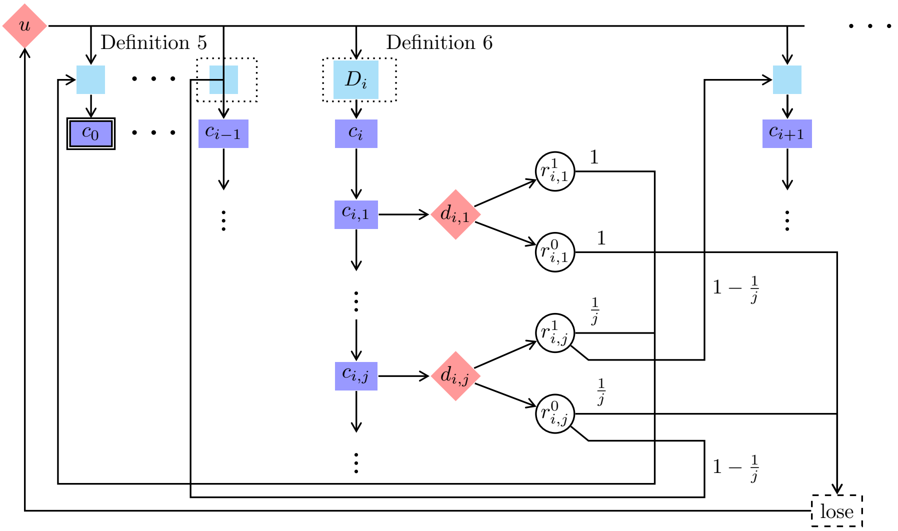

Fig. 6 The scheme of games defined in Definition 5 and in Definition 6, where in the former the teal-coloured boxes are replaced with a line connecting u directly to the c_i, whereas in the latter such a teal-coloured box is replaced with a Delay gadget D_i illustrated in Figure 7.

Every MR Maximizer strategy σ attains arbitrarily little from c_x as $x \to \infty$. Formally, $\limsup_{x\to\infty} \inf_\pi \mathcal{P}_{\mathcal{G},c_x,\sigma,\pi}(\mathtt{Reach}(\{c_0\})) = 0$.

Proof Let $\mathcal{G}'$ be the concurrent game from Definition 2. We have $\mathtt{val}_{\mathcal{G}'}(c_x) \geq 1/2$ by Theorem 4. Consider the (1/4)-optimal 1-bit Maximizer strategy $\hat{\sigma}$ used in $\mathcal{G}'$ in Definition 4. We can define a corresponding 1-bit Maximizer strategy σ in $\mathcal{G}$. In every state c_i, it picks the move $c_i{\longrightarrow}d_{i,j}$ whenever its memory mode is j, and it updates its memory in the same way as $\hat{\sigma}$. Then

$$\begin{aligned}\mathtt{val}_{\mathcal{G}}(c_x) &\geq \inf_\pi \mathcal{P}_{\mathcal{G},c_x,\sigma,\pi}(\mathtt{Reach}(\{c_0\}))\\ &= \inf_\pi \mathcal{P}_{\mathcal{G}',c_x,\hat{\sigma},\pi}(\mathtt{Reach}(\{c_0\}))\\ &\geq \mathtt{val}_{\mathcal{G}'}(c_x) - 1/4 \geq 1/4.\end{aligned}$$

For item 2 the argument is exactly the same as in Theorem 7(2). □

5 No Positional Strategies for Reachability in Infinitely Branching Games

In finitely branching turn-based stochastic 2-player games with reachability objectives, Maximizer has ε-optimal MD strategies (Lemma 5). We go on to show that this does not carry over to infinitely branching turn-based reachability games. In this case, there are not even ε-optimal MR strategies, i.e., good Maximizer strategies need memory. The reason for this is the infinite branching of Minimizer, since infinite branching of Maximizer states and random states can easily be encoded into finite branching in the case of reachability objectives.

Definition 5 *Consider the finitely branching turn-based reachability game from Definition 3. We construct an infinitely branching game $\mathcal{G}$ by adding a new Minimizer-controlled initial state u, Minimizer-transitions $u \longrightarrow c_i$ for all $i \in \mathbb{N}$ and* lose$\longrightarrow u$. *See Figure 6 for a scheme of this game. The objective is still* $\mathtt{Reach}(\{c_0\})$.

Theorem 9 *Let $\mathcal{G}$ be the infinitely branching turn-based reachability game from Definition 5.*

1. *All states in $\mathcal{G}$ are almost sure winning. I.e., for every state s there exists a Maximizer strategy σ such that* $\inf_\pi \mathcal{P}_{\mathcal{G},s,\sigma,\pi}(\mathtt{Reach}(\{c_0\})) = 1$.
2. *For each MR Maximizer strategy σ we have*

$$\inf_\pi \mathcal{P}_{\mathcal{G},u,\sigma,\pi}(\mathtt{Reach}(\{c_0\})) = 0.$$

I.e., for any $\varepsilon < 1$ there does not exist any ε-optimal MR Maximizer strategy σ from state u.

Proof Towards item 1, first note that by playing

$$c_i \longrightarrow c_{i,1} \longrightarrow d_{i,1}$$

Maximizer can enforce that he either wins (if Minimizer goes to $r^1_{i,1}$) or the game returns to state u (via state lose if Minimizer goes to $r^0_{i,1}$). Thus it suffices to show that Maximizer can win almost surely from state u. We construct a suitable strategy σ (which is not MR). By Theorem 7, $\mathtt{val}_{\mathcal{G}}(c_x) \geq 1/2$ for every x. Moreover, the game from c_x is finitely branching unless we return to u. Thus there exists a strategy σ_x and a finite horizon h_x such that $\inf_\pi \mathcal{P}_{\mathcal{G},c_x,\sigma_x,\pi}(\mathtt{Reach}_{h_x}(\{c_0\})) \geq 1/4$. Then σ plays from u as follows. If Minimizer moves $u \longrightarrow c_x$ then first play σ_x for h_x steps, unless c_0 or u are reached first. Then play to reach u again, i.e., the next time that the play reaches a state c_i play $c_i \longrightarrow c_{i,1} \longrightarrow d_{i,1}$ (thus either Maximizer wins or the play returns to u). So after every visit to u the Maximizer strategy σ wins with probability $\geq 1/4$ before seeing u again, and otherwise the play returns to u. Thus $\inf_\pi \mathcal{P}_{\mathcal{G},s,\sigma,\pi}(\mathtt{Reach}(\{c_0\})) \geq 1 - (1/4)^\infty = 1$. This proves item 1.

Claim 10 *Suppose that for each MR Maximizer strategy σ and every $\varepsilon > 0$ there is a Minimizer strategy $\pi(\sigma, \varepsilon)$ such that, from u, the probability of visiting c_0 before revisiting u is at most ε. Then item 2 is true.*

Proof of the claim Suppose the precondition of the claim. Let σ be an MR Maximizer strategy. Let π be the Minimizer strategy that, after the i-th visit to u, continues to play $\pi(\sigma, \varepsilon \cdot 2^{-i})$ until the next visit to u. It follows that $\mathcal{P}_{\mathcal{G},u,\sigma,\pi}(\mathtt{Reach}(\{c_0\})) \leq \sum_{i \geq 1} \varepsilon \cdot 2^{-i} = \varepsilon$. ☐

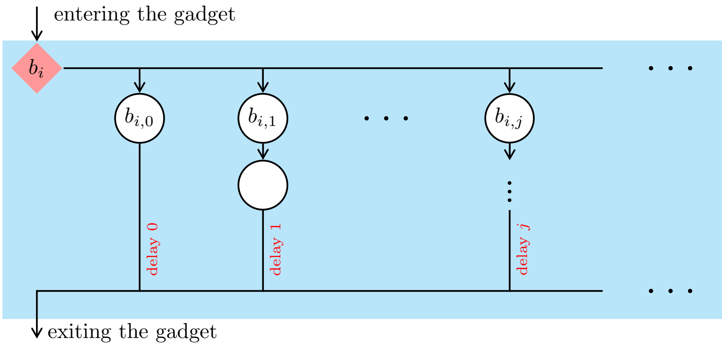

Fig. 7 Delay Gadget D_i, used in Figure 6.

It remains to prove the precondition of the claim. Let σ be an MR Maximizer strategy, and let $\varepsilon > 0$. By Theorem 7(2), there are $i \in \mathbb{N}$ and a Minimizer strategy π such that $\mathcal{P}_{\mathcal{G}',c_i,\sigma,\pi}(\texttt{Reach}(\{c_0\})) \leq \varepsilon$, where $\mathcal{G}'$ is the finitely branching subgame of $\mathcal{G}$ from Definition 3. The Minimizer strategy $\pi(\sigma,\varepsilon)$ that, in $\mathcal{G}$, from u goes to c_i and then plays π has the required property. $\square$

6 No Markov Strategies for Reachability in Infinitely Branching Games

We strengthen the result from the previous section by modifying the game so that even Markov strategies are useless for Maximizer. The modification of the game allows Minimizer to cause an arbitrary but finite delay before any state c_i is entered.

Definition 6 *Consider the infinitely branching turn-based reachability game from Definition 5. We modify it as follows. For each $i \in \mathbb{N}$ we add a Minimizer-controlled state b_i and redirect all transitions going into c_i to go into b_i instead. Each b_i is infinitely branching: for each $j \in \mathbb{N}$ we add a random state $b_{i,j}$ and a transition $b_i \longrightarrow b_{i,j}$. We add further states so that the game moves (deterministically, via a chain of random states) from $b_{i,j}$ to c_i in exactly j steps. See Figures 6 and 7 for a depiction of this game. The objective is still $\texttt{Reach}(\{c_0\})$.*

Theorem 11 *Let $\mathcal{G}$ be the infinitely branching turn-based reachability game from Definition 6.*

1. *All states in $\mathcal{G}$ are almost sure winning. I.e., for every state s there exists a Maximizer strategy σ such that $\inf_\pi \mathcal{P}_{\mathcal{G},s,\sigma,\pi}(\texttt{Reach}(\{c_0\})) = 1$.*
2. *For every Markov Maximizer strategy σ it holds that $\inf_\pi \mathcal{P}_{\mathcal{G},u,\sigma,\pi}(\texttt{Reach}(\{c_0\})) = 0$. I.e., no Markov Maximizer strategy is ε-optimal from state u for any $\varepsilon < 1$.*

Proof Item 1 follows from Theorem 9(1), as the modification in Definition 6 only allows Minimizer to cause finite delays.

Towards item 2, the idea of the proof is that for every Markov Maximizer strategy σ, Minimizer can cause delays that make σ behave in the way it would after a long time. This way, Minimizer turns σ approximately to an MR-strategy, which is useless by Theorem 9(2).

In more detail, fix any Markov Maximizer strategy σ. As in the proof of Theorem 7(2), we can assume that whenever the game is in c_i, the strategy σ almost surely moves eventually to some $d_{i,j}$. Let $p_{i,j,t}$ be the probability that strategy σ, when it is in c_i at time t, moves to $d_{i,j}$. Thus, a corresponding Maximizer strategy in the (concurrent) Big Match, when it is in c_i at time t, picks action 1 with probability $f(i,t) \stackrel{\text{def}}{=} \sum_j p_{i,j,t}(1/j)$; cf. the proof of Theorem 7(2). For each $i \in \mathbb{N}$, let $f(i)$ be an accumulation point of $f(i,1), f(i,2), \ldots$; e.g., take $f(i) \stackrel{\text{def}}{=} \liminf_t f(i,t)$. We have that

$$\forall_{i\in\mathbb{N}}\ \forall_{t_0\in\mathbb{N}}\ \forall_{\varepsilon>0}\ \exists_{t\geq t_0}:\quad f(i,t) \leq f(i)+\varepsilon \tag{6}$$

$$\forall_{i\in\mathbb{N}}\ \forall_{t_0\in\mathbb{N}}\ \forall_{\varepsilon>0}\ \exists_{t\geq t_0}:\quad f(i,t) \geq f(i)-\varepsilon \tag{7}$$

Similarly to the proof of Theorem 9(2) (see Claim 10 therein), it suffices to show that after each visit to u, Minimizer can make the probability of visiting c_0 before seeing u again arbitrarily small. Let $\varepsilon > 0$. We show that Minimizer has a strategy π to make this probability at most ε.

Consider the first case where $\sum_{i\geq 1} f(i) < \infty$. Then there is $i_0 \in \mathbb{N}$ such that $\sum_{i\geq i_0} f(i) \leq \varepsilon/2$. In u, strategy π moves to b_{i_0}. Whenever the game is in some b_i, strategy π moves to some $b_{i,j}$ so that the game will arrive in c_i at a time t that satisfies $f(i,t) \leq f(i) + 2^{-i}\cdot\varepsilon/2$; such t exists due to (6). In c_i Maximizer (using σ) moves (eventually) to some $d_{i,j}$. Then π always chooses "action 1"; i.e., π moves to $r^1_{i,j}$. In this way, the play, restricted to states c_i, is either of the form $c_{i_0}, c_{i_0+1}, \ldots$ (Maximizer loses) or of the form $c_{i_0}, c_{i_0+1}, \ldots, c_{i_0+k}, c_0$ (Maximizer wins). The probability of the latter can be bounded similarly to the proof of Theorem 4(3); i.e., we have

$$\mathcal{P}_{\mathcal{G},u,\sigma,\pi}(\texttt{Reach}(\{c_0\})) \;\leq\; \sum_{i=i_0}^{\infty} f(i) + 2^{-i}\cdot\frac{\varepsilon}{2} \;\leq\; \frac{\varepsilon}{2}+\frac{\varepsilon}{2} \;=\; \varepsilon.$$

Now consider the second case where $\sum_{i\geq 1} f(i) = \infty$. Then there is $i_0 \in \mathbb{N}$ such that $\sum_{i=1}^{i_0} f(i) \geq \frac{1}{\varepsilon}$. In u, strategy π moves to b_{i_0}. Whenever the game is in some b_i, strategy π moves to some $b_{i,j}$ so that the game will arrive in c_i at a time t that satisfies $f(i,t) \geq f(i) - 2^{-i}$; such t exists due to (7). In c_i Maximizer (using σ) moves (eventually) to some $d_{i,j}$. Then π always chooses "action 0"; i.e., π moves to $r^0_{i,j}$. In this way, the play, restricted to states c_i, is either of the form $c_{i_0}, c_{i_0-1}, \ldots, c_0$ (Maximizer wins) or of the form $c_{i_0}, c_{i_0-1}, \ldots, c_{i_0-k}$ (for some $k < i_0$), followed by lose, u (Maximizer does not reach c_0 before revisiting u). The probability of the former can be bounded similarly to the proof of Theorem 4(3); i.e., the probability that the play reaches c_0 before u is upper-bounded by

$$\begin{aligned}\prod_{i=1}^{i_0}(1-\max\{f(i)-2^{-i},0\}) \quad &\leq \quad \frac{1}{1+\sum_{i=1}^{i_0}(f(i)-2^{-i})} \qquad \text{by Proposition 29}\\ &\leq \quad \frac{1}{\sum_{i=1}^{i_0} f(i)} \;\leq\; \varepsilon.\end{aligned}$$

□

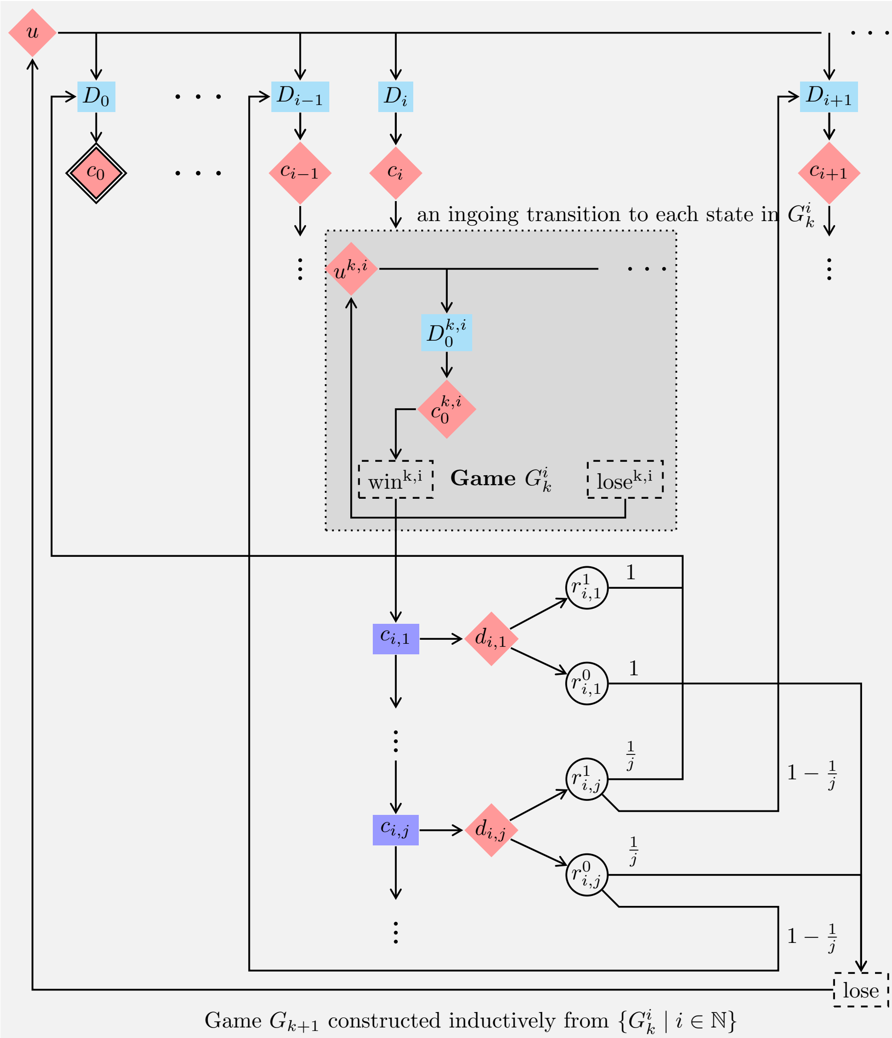

Fig. 8 The scheme of the nested construction in Definition 7.

7 Good Strategies for Reachability Require Infinite Memory

We show that even finite private memory, in addition to a step counter, is useless for Maximizer in infinitely branching reachability games. To this end, we define a nested version of the game of Definition 6, where the memory requirements increase unboundedly with the nesting depth.

Definition 7 *Let $\mathcal{G}$ be the game from Definition 6. We inductively define the* k-nested *game $\mathcal{G}_k$ as follows (see Figure 8). For the base case, let $\mathcal{G}_1 \stackrel{\text{def}}{=} \mathcal{G}$.*

For every $i \geq 1$ let $\mathcal{G}_k^i$ be a fresh copy of $\mathcal{G}_k$ and let $u^{k,i}$ (resp. $c_0^{k,i}$) be the initial state u (resp. the target state c_0) in $\mathcal{G}_k^i$. For every $k \geq 1$ we construct $\mathcal{G}_{k+1}$ by modifying $\mathcal{G}$ as follows. The idea is that at every state c_i Maximizer first needs to win the subgame $\mathcal{G}_k^i$ before continuing in the game $\mathcal{G}_{k+1}$, but Minimizer can choose at which state s in $\mathcal{G}_k^i$ the subgame is entered.

We make the state c_i Minimizer-controlled and replace the previous Maximizer transition $c_i{\longrightarrow}c_{i,1}$ by Minimizer transitions $c_i{\longrightarrow}s$ for every state s in $\mathcal{G}_k^i$. Moreover, we add the transitions $c_0^{k,i}{\longrightarrow}\mathrm{win}^{k,i}{\longrightarrow}c_{i,1}$. (The new state $\mathrm{win}^{k,i}$ is not strictly needed. It just indicates that Maximizer has won and exited the subgame $\mathcal{G}_k^i$.) Note that also in $\mathcal{G}_{k+1}$ Minimizer can introduce arbitrary delays between states b_i and c_i.

The objective in $\mathcal{G}_{k+1}$ is still $\mathtt{Reach}(\{c_0\})$.

Lemma 12 *For any $k \geq 1$ let $\mathcal{G}_k$ be the infinitely branching turn-based reachability game from Definition 7.*

1. *All states in $\mathcal{G}_k$ are almost sure winning. I.e., for every state s there exists a Maximizer strategy σ such that*

$$\inf_{\pi} \mathcal{P}_{\mathcal{G}_k,s,\sigma,\pi}(\mathtt{Reach}(\{c_0\})) = 1.$$

2. *For each Maximizer strategy σ with a step counter plus a private finite memory with $\leq k$ modes*

$$\inf_{\pi} \mathcal{P}_{\mathcal{G}_k,u,\sigma,\pi}(\mathtt{Reach}(\{c_0\})) = 0.$$

I.e., for any $\varepsilon < 1$ there does not exist any ε-optimal step counter plus k memory mode Maximizer strategy σ from state u in $\mathcal{G}_k$.

Proof We show Item 1 by induction on k.

In the base case of $k = 1$ we have $\mathcal{G}_1 = \mathcal{G}$ from Definition 6, and thus the result holds by Theorem 11(Item 1).

Induction step $k{\longrightarrow}k+1$. For every state s in $\mathcal{G}_{k+1}$ outside of any subgame, let $\sigma'(s)$ be the almost sure winning Maximizer strategy from s in the non-nested game $\mathcal{G}_1$, obtained from above. By the induction hypothesis, for any state s in a subgame $\mathcal{G}_k^i$ there exists a Maximizer strategy $\sigma_k^i(s)$ from s that almost surely wins this subgame $\mathcal{G}_k^i$.

We now construct a Maximizer strategy $\sigma(s)$ from any state s in $\mathcal{G}_{k+1}$. If s is not in any strict subgame then $\sigma(s)$ plays like $\sigma'(s)$ outside of the subgames. Whenever a subgame $\mathcal{G}_k^i$ is entered at some state x then it plays like $\sigma_k^i(x)$ until $\mathrm{win}^{k,i}$ is reached (which happens eventually almost surely by the definition of $\sigma_k^i(x)$) and the play exits the subgame, and then it continues with the outer strategy $\sigma'(s)$. Similarly, if the start state s is in some subgame $\mathcal{G}_k^i$ then it first plays $\sigma_k^i(s)$ until $\mathrm{win}^{k,i}$ is reached (which happens eventually almost surely by the definition of $\sigma_k^i(s)$) and the play exits the subgame, and then it continues like the strategy $\sigma'(c_{i,1})$ described above. Then $\sigma(s)$ wins almost surely, since the strategies σ' and σ_k^i win almost surely.

Towards Item 2, we show, by induction on k, a slightly stronger property. For each Maximizer strategy σ in $\mathcal{G}_k$ with a step counter plus a private finite memory with $\leq k$ modes, for every $\delta > 0$ there exists a Minimizer strategy π that upper-bounds Maximizer's attainment by δ regardless of Maximizer's initial memory mode and the starting time. Formally,

$$\forall \delta > 0\, \exists \pi\, \forall \mathsf{m}\, \forall t\ \mathcal{P}_{\mathcal{G}_k,u,\sigma[\mathsf{m}](t),\pi}(\mathtt{Reach}(\{c_0\})) \leq \delta \tag{8}$$

For the base case $k = 1$ we have $\mathcal{G}_1 = \mathcal{G}$ from Definition 6. Since $k = 1$, Maximizer has only one memory mode. Moreover, Minimizer's strategy from the proof of Theorem 11(2) works regardless of the starting time t (since it just chooses delays in the states b_i to satisfy (6) and (7)). Thus we obtain (8).

Induction step $k \longrightarrow k+1$. Consider the game $\mathcal{G}_{k+1}$ and a fixed Maximizer strategy σ with a step counter plus $(k+1)$ private memory modes from state u. Let $\{0, 1, \ldots, k\}$ denote the $k + 1$ private memory modes of σ.

From every state c_i in $\mathcal{G}_{k+1}$ we enter the subgame $\mathcal{G}_k^i$, at some state chosen by Minimizer. When (and if) Maximizer wins this subgame then we are in state $\mathrm{win}^{k,i}$ and the game $\mathcal{G}_{k+1}$ continues with Maximizer's choice at $c_{i,1}$, etc.

Consider the state c_i, visited at some time t with some Maximizer memory mode m. Let π' be some Minimizer strategy. Then let $\alpha(i, \mathsf{m}, t, \pi')$ be the probability that Maximizer will play action "1" (wrt. the encoded concurrent game) in the next round in $\mathcal{G}_{k+1}$ after winning the subgame $\mathcal{G}_k^i$ (i.e., after reaching $\mathrm{win}^{k,i}$), or loses the subgame (never reaches $\mathrm{win}^{k,i}$). So $\alpha(i, \mathsf{m}, t, \pi')$ is the probability of losing the subgame $\mathcal{G}_k^i$ plus $\sum_j (1/j) \cdot p_j$, where p_j is the probability of winning the subgame and then directly going to $d_{i,j}$ (i.e., in the same round, without seeing any other state c_i before). Formally, let $C_{-i} \overset{\text{def}}{=} \{c_j \mid j \neq i\}$ and define

$$\begin{aligned}\alpha(i, \mathsf{m}, t, \pi') \ \overset{\text{def}}{=}\ & \mathcal{P}_{\mathcal{G}_{k+1},c_i,\sigma[\mathsf{m}](t),\pi'}(\neg\mathtt{Reach}(\{\mathrm{win}^{k,i}\})) \\ & + \sum_j (1/j) \mathcal{P}_{\mathcal{G}_{k+1},c_i,\sigma[\mathsf{m}](t),\pi'}(d_{i,j} \ \mathsf{before}\ C_{-i})\end{aligned}$$

(where $(d_{i,j}\ \mathsf{before}\ C_{-i})$ denotes the set of plays that visit state $d_{i,j}$ before visiting any state in C_{-i}). The probability $\alpha(i, \mathsf{m}, t, \pi')$ depends on i (since we are looking at state c_i), on Maximizer's private memory mode $\mathsf{m} \in \{0, 1, \ldots, k\}$ at state c_i, at the time $t \in \mathbb{N}$ when we are at c_i and on Minimizer's strategy π'. Let $\alpha(i, \mathsf{m}, t) \overset{\text{def}}{=} \sup_{\pi'} \alpha(i, \mathsf{m}, t, \pi')$ be the supremum over all Minimizer strategies. Let $\alpha(i, t) \overset{\text{def}}{=} \min_{\mathsf{m} \in \{0,\ldots,k\}} \alpha(i, \mathsf{m}, t)$ the minimum over all memory modes. Intuitively, when entering the subgame $\mathcal{G}_k^i$ from c_i at time t, for each Maximizer memory mode m, for each $\varepsilon > 0$, Minimizer has a strategy to make the probability of Maximizer playing action "1" after winning the subgame (or else losing the subgame) at least $\alpha(i, t) - \varepsilon$. Let $\alpha(i)$ be the maximal accumulation point of the infinite sequence $\alpha(i, 1), \alpha(i, 2), \ldots$, i.e., $\alpha(i) \overset{\text{def}}{=} \limsup_t \alpha(i, t)$. We have:

$$\forall i \in \mathbb{N}\, \forall \varepsilon > 0\, \forall t_0 \in \mathbb{N}\, \exists t \geq t_0 : \quad \alpha(i, t) \geq \alpha(i) - \varepsilon \tag{9}$$

$$\forall i \in \mathbb{N}\, \forall \varepsilon > 0\, \exists t_0 \in \mathbb{N}\, \forall t \geq t_0 : \quad \alpha(i, t) \leq \alpha(i) + \varepsilon \tag{10}$$

Now there are two cases.

In the first case, $\sum_i \alpha(i)$ diverges. Intuitively, since we have $\alpha(i) = \limsup_t \alpha(i, t)$, Minimizer chooses the delays at the states b_i in order to make $\alpha(i, t)$ "large", i.e., close to $\alpha(i)$, by using (9).

Analogously to Claim 10, it suffices, for every $\varepsilon > 0$ to construct a Minimizer strategy π in $\mathcal{G}_{k+1}$ from state u that makes the probability of visiting c_0 before revisiting u (denoted as the event "c_0 before u") at most ε.

We construct a Minimizer strategy π that plays as follows. First π picks the transition $u \longrightarrow b_{i_0}$ for a sufficiently high $i_0 \in \mathbb{N}$, to be determined. At every state $d_{i,j}$, outside of the subgames, π always plays action "0", i.e., $d_{i,j} \longrightarrow r^0_{i,j}$. This implies that the state c_i is not visited again, unless the state u is re-visited first. (If Maximizer plays action "1" then he loses this round and the game goes back to u. If Maximizer plays "0" then the game goes down to c_{i-1}.)

At every state b_i at each time t' the strategy π picks a delay such that the game arrives at c_i at a time t such that

$$\alpha(i,t) \geq \alpha(i) - \frac{1}{4}2^{-i}. \tag{11}$$

This is possible by Equation (9). Let T_i be the set of the times t that satisfy (11). Minimizer's strategy ensures that the states c_i are only reached at times $t \in T_i$.

From state c_i, at each time t, for each memory mode m there exists a Minimizer strategy $\pi(i,\mathsf{m},t)$ such that we have $\alpha(i,\mathsf{m},t,\pi(i,\mathsf{m},t)) \geq \alpha(i,\mathsf{m},t) - \frac{1}{4}2^{-i}$, by the definition of $\alpha(i,\mathsf{m},t) \stackrel{\text{def}}{=} \sup_{\pi'} \alpha(i,\mathsf{m},t,\pi')$.

Since the Maximizer memory mode m is private, Minimizer does not know it. So our Minimizer strategy π hedges her bets over all possible $\mathsf{m} \in \{0,\ldots,k\}$ and plays each strategy $\pi(i,\mathsf{m},t)$ with equal probability $\frac{1}{k+1}$. It follows that for every i and time $t \in T_i$ chosen according to Equation (11), after (and if) winning the subgame $\mathcal{G}^i_k$, Maximizer plays action "1" with probability

$$\begin{aligned}\alpha(i,\mathsf{m},t,\pi) \;&\geq\; \frac{1}{k+1}\left(\alpha(i,t) - \frac{1}{4}2^{-i}\right)\\ &\geq\; \frac{1}{k+1}\left(\alpha(i) - \frac{1}{4}2^{-i} - \frac{1}{4}2^{-i}\right) = \frac{1}{k+1}\left(\alpha(i) - \frac{1}{2}2^{-i}\right).\end{aligned} \tag{12}$$

Since, outside of subgames, Minimizer always plays action "0" and each state c_i is visited at most once (unless state u is re-visited), we get for every starting memory mode m' and starting time $t' \in T_{i_0}$ that the probability of visiting c_0 before revisiting u is upper-bounded, i.e.,

$$\mathcal{P}_{\mathcal{G}_{k+1},c_{i_0},\sigma[\mathsf{m}'](t'),\pi}(c_0 \text{ } \mathsf{before} \text{ } u) \;\leq\; \prod_{i=1}^{i_0}(1 - \min_{\mathsf{m}} \inf_{t \in T_i} \alpha(i,\mathsf{m},t,\pi)).$$

Since $\sum_i \alpha(i)$ diverges, it follows from (12) that

$$\sum_{i=1}^{\infty} \min_{\mathsf{m}} \inf_{t \in T_i} \alpha(i,\mathsf{m},t,\pi) \geq \left(\frac{1}{k+1}\sum_i \alpha(i)\right) - \frac{1}{2k+2}$$

also diverges. From Proposition 30 we obtain

$$\prod_{i=1}^{\infty}(1 - \min_{\mathsf{m}} \inf_{t \in T_i} \alpha(i,\mathsf{m},t,\pi)) = 0$$

and hence

$$\lim_{i_0 \to \infty} \prod_{i=1}^{i_0}(1 - \min_{\mathsf{m}} \inf_{t \in T_i} \alpha(i,\mathsf{m},t,\pi)) = 0.$$

Thus, for every $\varepsilon > 0$, there exists a sufficiently large i_0 such that for all $t' \in T_{i_0}$ and m'

$$\mathcal{P}_{\mathcal{G}_{k+1}, c_{i_0}, \sigma[\mathsf{m}'](t'), \pi}(c_0 \text{ before } u) \leq \varepsilon.$$

Hence, for all $t \in \mathbb{N}$ and all m we have

$$\mathcal{P}_{\mathcal{G}_{k+1}, u, \sigma[\mathsf{m}](t), \pi}(c_0 \text{ before } u) \leq \varepsilon.$$

Analogously to Claim 10, we can pick i_0^l and $\varepsilon^l = \delta \cdot 2^{-l}$ after the l-th visit to u such that for all $t \in \mathbb{N}$ and all m we have

$$\mathcal{P}_{\mathcal{G}_{k+1}, u, \sigma[\mathsf{m}](t), \pi}(\mathtt{Reach}(\{c_0\})) \leq \sum_{l=1}^{\infty} \varepsilon_l = \delta$$

as required.

Now we consider the second case where $\sum_i \alpha(i)$ converges. We construct a Minimizer strategy π in $\mathcal{G}_{k+1}$ that plays as follows. First π picks the transition $u{\longrightarrow}b_{i_0}$ for a sufficiently high $i_0 \in \mathbb{N}$, to be determined. At every state $d_{i,j}$, outside of the subgames, π always plays action "1", i.e., $d_{i,j}{\longrightarrow}r^1_{i,j}$. This implies that each state c_i with $i \geq i_0$ is visited at most once, and states c_i with $i < i_0$ are never visited. (If Maximizer plays action "1" then he wins and if he plays "0" then the games goes up to c_{i+1}.)

By Equation (10), for every $i \geq 1$ there exists a time $t_i \in \mathbb{N}$ such that

$$\forall t \geq t_i.\ \alpha(i,t) \leq \alpha(i) + \frac{1}{4} 2^{-i}. \tag{13}$$

Let $T_i = \{t \in \mathbb{N} \mid t \geq t_i\}$ be the set of the times t that satisfy (13). At each state b_i Minimizer's strategy π delays sufficiently long such that c_i is reached at a time $t \in T_i$. This is possible for every arrival time t' at b_i.

Consider a state s in the subgame $\mathcal{G}^i_k$ that is reached at some time t when Maximizer's strategy is in some memory mode m, and let π' be a Minimizer strategy. Let $\beta(s, i, \mathsf{m}, t, \pi')$ be the probability that Maximizer will play action "1" (wrt. the encoded concurrent game) in the next round in $\mathcal{G}_{k+1}$ after winning the subgame $\mathcal{G}^i_k$ (i.e., after reaching $\mathrm{win}^{k,i}$), or loses the subgame (never reaches $\mathrm{win}^{k,i}$). So $\beta(s, i, \mathsf{m}, t, \pi')$ is the probability, from state s, of losing the subgame $\mathcal{G}^i_k$ plus $\sum_j (1/j) \cdot p_j$, where p_j is the probability of winning the subgame and then directly going to $d_{i,j}$ (without visiting any other state c_i in between). To define this formally, let $C_{-i} \stackrel{\text{def}}{=} \{c_j \mid j \neq i\}$. We let

$$\begin{aligned} \beta(s, i, \mathsf{m}, t, \pi') \ &\stackrel{\text{def}}{=}\ \mathcal{P}_{\mathcal{G}_{k+1}, s, \sigma[\mathsf{m}](t), \pi'}(\neg\mathtt{Reach}(\{\mathrm{win}^{k,i}\})) \\ &\qquad + \sum_j (1/j) \mathcal{P}_{\mathcal{G}_{k+1}, s, \sigma[\mathsf{m}](t), \pi'}(d_{i,j} \text{ before } C_{-i}) \end{aligned}$$

Let $\beta(s, i, \mathsf{m}, t) \stackrel{\text{def}}{=} \sup_{\pi'} \beta(s, i, \mathsf{m}, t, \pi')$ be the supremum over all Minimizer strategies. Let

$$\beta(s, i, t) \stackrel{\text{def}}{=} \min_{\mathsf{m} \in \{0, \ldots, k\}} \beta(s, i, \mathsf{m}, t)$$

be the minimum over all memory modes.

Claim 13 *For all states s in $\mathcal{G}^i_k$ and times t we have $\beta(s, i, t+1) \leq (k+1)\alpha(i, t)$.*

Proof Consider the situation where we are at state c_i at time t and the Maximizer strategy σ is in some memory mode m (unknown to Minimizer). Then some particular Minimizer strategy $\hat{\pi}$ could play $c_i \longrightarrow s$ to arrive at s in one step at time $t+1$. Meanwhile, σ can update its memory to some other mode m' (or a distribution over memory modes), still unknown to Minimizer. Then $\hat{\pi}$ hedges her bets by guessing Maximizer's memory mode m'. For each of the $k+1$ possible modes m', the Minimizer strategy $\hat{\pi}$ plays, with probability $\frac{1}{k+1}$, an ε-optimal strategy to maximize the probability that Maximizer plays action "1" after winning $\mathcal{G}_k^i$ (or loses the subgame). Thus

$$\begin{aligned}
\alpha(i,t) &= \min_{\mathsf{m}} \sup_{\pi'} \alpha(i,\mathsf{m},t,\pi') \\
&\geq \min_{\mathsf{m}} \alpha(i,\mathsf{m},t,\hat{\pi}) \\
&\geq \min_{\mathsf{m}'} \frac{1}{k+1} \beta(s,i,\mathsf{m}',t+1,\hat{\pi}) \\
&\geq \frac{1}{k+1} (\min_{\mathsf{m}'} \sup_{\pi'} \beta(s,i,\mathsf{m}',t+1,\pi') - \varepsilon) \\
&= \frac{1}{k+1} (\beta(s,i,t+1) - \varepsilon).
\end{aligned}$$

Since this holds for every $\varepsilon > 0$, the claim follows. □

By Claim 13 for every s in $\mathcal{G}_k^i$ and time $t+1$ there exists at least one memory mode $\mathsf{m}(s,t+1)$ (a mode m where the minimum $\beta(s,i,t+1) = \min_{\mathsf{m} \in \{0,\ldots,k\}} \beta(s,i,\mathsf{m},t+1)$ is realized) such that if σ enters s at time $t+1$ in mode $\mathsf{m}(s,t+1)$ then after winning $\mathcal{G}_k^i$ (if at all) Maximizer plays action "1" with a "small" probability $\leq (k+1)\alpha(i,t)$. Crucially, this property holds for the sup over the Minimizer strategies and thus for *every* Minimizer strategy inside the subgame $\mathcal{G}_k^i$. In particular it holds for the Minimizer strategy π that we will construct. We call $\mathsf{m}(s,t+1)$ the *forbidden* memory mode for state s at time $t+1$.

Above we have defined our Minimizer strategy π so that it adds sufficient delays in the states b_i such that c_i is only visited at times $t \geq t_i$. This implies that states s in $\mathcal{G}_k^i$ are only visited at times $t+1$ where $t \geq t_i$. Since for these times Equation (13) is satisfied, we obtain

$$\forall t \geq t_i.\ \beta(s,i,\mathsf{m}(s,t+1),t+1) \leq (k+1)(\alpha(i) + \frac{1}{4}2^{-i}). \tag{14}$$

Let σ' be a restriction of σ that, inside the subgame $\mathcal{G}_k^i$, is never in the forbidden memory mode $\mathsf{m}(s,t+1)$ at state s at time $t+1$, or else concedes defeat.

Claim 14 *Consider a step counter plus $(k+1)$ mode Maximizer strategy σ' in $\mathcal{G}_k^i$ that is never in the forbidden memory mode $\mathsf{m}(s,t+1)$ at state s at time $t+1$. Then there exists a step counter plus k mode Maximizer strategy σ'' in $\mathcal{G}_k^i$ that performs equally well as σ' against any Minimizer strategy.*

Proof While σ' has memory modes $\{0,\ldots,k\}$, the strategy σ'' only has memory modes $\{0,\ldots,k-1\}$ (in addition to the step counter). However, σ'' can directly imitate the behavior of σ' as follows. Whenever σ' enters memory mode k at some state s and time $t+1$ then σ'' enters memory mode $\mathsf{m}(s,t+1)$ instead. Whenever,

σ'' is in $\mathsf{m}(s, t+1)$ at time state s and time $t+1$ then it plays like σ' at state s in memory mode k. By the condition on the behavior of σ' there is no confusion. $\square$

By Claim 14, the Maximizer strategy σ' is equivalent to a strategy with just a step counter and k memory modes. By the induction hypothesis (8), for this restricted Maximizer strategy σ' there exists a Minimizer strategy π_i in $\mathcal{G}_k^i$ such that

$$\forall \mathsf{m}\ \forall t\ \mathcal{P}_{\mathcal{G}_k^i, u^{k,i}, \sigma'[\mathsf{m}](t), \pi_i}(\mathtt{Reach}(\{c_0^{k,i}\})) \leq \delta \cdot 2^{-(i+1)}. \tag{15}$$

We are now ready to construct Minimizer's strategy π in $\mathcal{G}_{k+1}$. At every state $d_{i,j}$, outside of the subgames, π always plays action "1", i.e., $d_{i,j} \longrightarrow r_{i,j}^1$. This implies that each state c_i is visited at most once. At the states b_i, Minimizer chooses the delays such that c_i is reached at a time $t \geq t_i$, as described above. From each c_i Minimizer goes to state $u^{k,i}$ of the subgame $\mathcal{G}_k^i$. Inside each subgame $\mathcal{G}_k^i$ Minimizer plays like π_i. By (15), π_i performs well in $\mathcal{G}_k^i$ (regardless of the initial memory mode and time) if Maximizer limits himself to σ'.

Now we show that π performs well in $\mathcal{G}_{k+1}$. Since π first picks the transition $u \longrightarrow b_{i_0}$ and then always plays action "1" outside of the subgames, it follows that each subgame $\mathcal{G}_k^i$ with $i \geq i_0$ is played at most once, and subgames $\mathcal{G}_k^i$ with $i < i_0$ are never played. For each subgame $\mathcal{G}_k^i$, let Forb_i be the set of plays where Maximizer enters a forbidden memory mode (for the current state and time) at least once. [3]

From Equation (15) we obtain that Maximizer loses $\mathcal{G}_k^i$ (and thus $\mathcal{G}_{k+1}$) with high probability if he *never* enters a forbidden memory mode.

$$\begin{aligned}
&\max_{\mathsf{m}} \sup_t \mathcal{P}_{\mathcal{G}_{k+1}, c_i, \sigma[\mathsf{m}](t), \pi}(\mathtt{Reach}(\{c_0\}) \cap \overline{\mathrm{Forb}_i}) \\
&\qquad \leq \max_{\mathsf{m}} \sup_t \mathcal{P}_{\mathcal{G}_k^i, u^{k,i}, \sigma'[\mathsf{m}](t), \pi_i}(\mathtt{Reach}(\{c_0^{k,i}\})) \\
&\qquad \leq \delta \cdot 2^{-(i+1)}.
\end{aligned} \tag{16}$$

On the other hand, we can show that if Maximizer does enter a forbidden memory mode (for the current state and time) in $\mathcal{G}_k^i$ then his chance of playing action "1" (and thus winning $\mathcal{G}_{k+1}$ in that round) after (and if) winning the subgame $\mathcal{G}_k^i$ is small. This holds for *every* Minimizer's strategy inside $\mathcal{G}_k^i$ and thus in particular this holds for our chosen Minimizer strategy π_i.

Recall that π ensures that states in $\mathcal{G}_k^i$ are only reached at times $t+1$ where $t \geq t_i$, and thus Equation (14) applies. Hence, at c_i, Maximizer's chance of satisfying Forb_i and still winning the game in this round (without going to c_{i+1} and the next subgame G_k^{i+1}) are upper bounded by $(k+1)(\alpha(i) + \frac{1}{4}2^{-i})$. For all m and all $t \geq t_i$ we have

$$\begin{aligned}
&\mathcal{P}_{\mathcal{G}_{k+1}, c_i, \sigma[\mathsf{m}](t), \pi}(\mathtt{Reach}(\{c_0\}) \cap \mathrm{Forb}_i \cap \neg\mathtt{Reach}(\{c_{i+1}\})) \\
&\qquad \leq (k+1)(\alpha(i) + \frac{1}{4}2^{-i})
\end{aligned} \tag{17}$$

[3] Strictly speaking, Forb_i is not an event in $\mathcal{G}_k^i$, since it refers to the memory mode of Maximizer's strategy σ. However, since we fix σ first, we can consider the MDP that is induced by fixing σ in $\mathcal{G}_k^i$. Then Maximizer's memory mode m and the step counter t are encoded into the states, which are of the form (s, m, t). In this MDP, Forb_i is a measurable event, actually an open set. However, since Maximizer's memory is private, Minimizer has only partial observation in this MDP, i.e., she cannot distinguish between states (s, m, t) and (s, m', t). Indeed the Minimizer strategy that we construct does not assume any knowledge of the memory mode, but instead hedges her bets.

Since in our current case $\sum_i \alpha(i)$ converges, it follows that $\sum_i (k+1)(\alpha(i) + \frac{1}{4}2^{-i})$ also converges, and thus there exists a sufficiently large $i_0 \in \mathbb{N}$ such that

$$\sum_{i \geq i_0} (k+1)(\alpha(i) + \frac{1}{4}2^{-i}) \leq \delta/2 \tag{18}$$

Let

$$nFo(i, \mathsf{m}', t') \stackrel{\text{def}}{=} \mathcal{P}_{\mathcal{G}_{k+1}, c_i, \sigma[\mathsf{m}'](t'), \pi}(\mathtt{Reach}(\{c_0\}) \cap \overline{\text{Forb}_i})$$

and

$$Fo(i, \mathsf{m}', t') \stackrel{\text{def}}{=} \mathcal{P}_{\mathcal{G}_{k+1}, c_i, \sigma[\mathsf{m}'](t'), \pi}(\mathtt{Reach}(\{c_0\}) \cap \text{Forb}_i \cap \neg\mathtt{Reach}(\{c_{i+1}\}))$$

Then from (16), (17) and (18) we obtain that for every initial memory mode m and time t,

$$\begin{aligned}
\mathcal{P}_{\mathcal{G}_{k+1}, u, \sigma[\mathsf{m}](t), \pi}(\mathtt{Reach}(\{c_0\}) &\leq \sum_{i \geq i_0} \max_{\mathsf{m}'} \sup_{t' \geq t_i} nFo(i, \mathsf{m}', t') + \sum_{i \geq i_0} \max_{\mathsf{m}'} \sup_{t' \geq t_i} Fo(i, \mathsf{m}', t') \\
&\leq \sum_{i \geq i_0} \delta \cdot 2^{-(i+1)} + \sum_{i \geq i_0} (k+1)(\alpha(i) + \frac{1}{4}2^{-i}) \\
&\leq \delta/2 + \delta/2 = \delta
\end{aligned}$$

□

In order to show that Maximizer needs infinite memory, in addition to a step counter, we combine all the nested games $\mathcal{G}_k$ into a single game.

Definition 8 *For all $k \geq 1$ consider the nested games $\mathcal{G}_k$ from Definition 7 with initial state u^k and target state c_0^k, respectively. We construct a game $\mathcal{G}$ with initial state s_0, target state f, Minimizer-controlled transitions $s_0 \longrightarrow u_k$ for all k, and Maximizer controlled transitions $c_0^k \longrightarrow f$ for all k. The objective in $\mathcal{G}$ is $\mathtt{Reach}(\{f\})$.*

Theorem 15 *Let $\mathcal{G}$ be the infinitely branching turn-based reachability game from Definition 8.*

1. *All states in $\mathcal{G}$ are almost sure winning. I.e., for every state s there exists a Maximizer strategy σ such that $\inf_\pi \mathcal{P}_{\mathcal{G}, s, \sigma, \pi}(\mathtt{Reach}(\{f\})) = 1$.*
2. *For each Maximizer strategy σ with a step counter plus a private finite memory we have*
$$\inf_\pi \mathcal{P}_{\mathcal{G}, s_0, \sigma, \pi}(\mathtt{Reach}(\{f\})) = 0.$$
I.e., for any $\varepsilon < 1$ there does not exist any ε-optimal step counter plus finite private memory Maximizer strategy σ from state s_0 in $\mathcal{G}$.

Proof Towards Item 1, every state in $\mathcal{G}_k$ is almost sure winning by Lemma 12(1). Thus, after the first step $s_0 \longrightarrow u^k$ into some game $\mathcal{G}_k$, Maximizer just needs to play the respective almost sure winning strategy in $\mathcal{G}_k$.

Towards Item 2, consider a Maximizer strategy σ with a step counter and a finite memory with some number of modes $k \in \mathbb{N}$. Then, by Lemma 12(2), for every $\delta > 0$, Minimizer can choose a first step $s_0 \longrightarrow u^k$ into $\mathcal{G}_k$ and a strategy in $\mathcal{G}_k$ that upper-bounds Maximizer's attainment to $\leq \delta$. □

Remark 1 Theorem 15 has implications even for games with finite action sets (resp. finitely branching turn-based games).

Consider a finitely-branching game where the states are labeled with rewards in $\{-1, 1\}$. The objective of Maximizer is to ensure that the lim inf of the seen rewards is ≥ 0. (Equivalently that states with reward -1 are visited only finitely often. This is also called a co-Büchi objective Kiefer et al (2017b).) The infinitely branching reachability game of Theorem 15 can be encoded into a finitely branching lim inf game, and thus the lower bound of Theorem 15 carries over. One just replaces every infinite Minimizer branching $s \to s_i$ for $i \in \mathbb{N}$ by a Minimizer-controlled gadget $s \to s_1' \to s_2' \dots$ and $s_i' \to s_i$ with new states s_i' that have reward 1. Minimizer cannot stay in states s_i' forever, since their rewards of 1 makes this winning for Maximizer. Thus the finitely branching gadgets faithfully encode Minimizer's original infinitely branching choice. Finally, the target state c_0 is given reward 1 and a self-loop, and all other states are given reward -1. Thus the $\liminf \geq 0$ objective in the new game corresponds to the reachability objective to reach state c_0 in the original game. The only problem with this construction is that the new gadgets incur extra steps, i.e., the step counters in the two games do not coincide. Thus the property that a step counter does not help Maximizer does not follow immediately from Theorem 15 if taken as a black box. However, the delay gadgets D_i in Definition 7 (and their finitely branching encoding in the new game) still ensure that the step counter does not help Maximizer. I.e., the proof of the lower bound for the finitely branching $\liminf \geq 0$ game is nearly identical to the proof of Theorem 15.

In the new finitely branching game above, the objective to reach c_0 also coincides with the objective to attain a high *expected* lim inf of the rewards. Thus ε-optimal (resp. optimal) strategies to maximize the expected lim inf also require infinite memory.

Finally, if one flips the signs of the rewards of all transitions in the game above, then the objective to reach c_0 coincides with the objective to minimize the expected lim sup of the rewards. Thus ε-optimal (resp. optimal) strategies to minimize the expected lim sup also require infinite memory. This solves the open question of Secchi in (Secchi, 1998, Section 5).

8 Special Cases with Restrictions on Minimizer's Actions

By Theorem 15, ε-optimal strategies for Maximizer require infinite memory (even in turn-based reachability games). The reason for this are Minimizer's infinite action sets (resp. infinite branching). If Minimizer has only finite action sets then Maximizer's strategies can be simpler (MR, resp. MD for turn-based games; cf. Table 1). However, the connection between Minimizer's infinite action sets and Maximizer's need for infinite memory is not as direct as it might seem.

In this section we consider the restricted case where Minimizer can use an infinite action set at most finitely often in any play. We show that in this case Maximizer has uniformly ε-optimal 1-bit strategies.

However, first we show a matching lower bound. Even if Minimizer can use an infinite action set only once, still 1 bit of memory is required for ε-optimal

Maximizer strategies. The following theorem shows that, even in turn-based reachability games, if Minimizer can use infinite branching just once, MR strategies cannot be ε-optimal for Maximizer for any $\varepsilon < 1/2$.

Definition 9 *Consider the finitely branching turn-based game from Definition 3, and add a new infinitely branching Minimizer-controlled initial state u and Minimizer-transitions $u \longrightarrow c_x$ for all $x \in \mathbb{N}$.*

Theorem 16 *There exists a turn-based game $\mathcal{G}$ as in Definition 9 with initial state u and reachability objective* $\mathtt{Reach}(\{c_0\})$ *such that*

1. *All states except u are finitely branching, and all plays from u use infinite branching exactly once.*
2. $\mathtt{val}_{\mathcal{G}}(u) \geq 1/2$.
3. $\inf_\pi \mathcal{P}_{\mathcal{G},u,\sigma,\pi}(\mathtt{Reach}(\{c_0\})) = 0$ *holds for every Maximizer MR strategy σ*

Proof Item 1 holds by the construction in Definition 9, since u is the only state with infinite branching and plays cannot return to u.

Towards Item 2., Theorem 7(Item 1) yields $\mathtt{val}_{\mathcal{G}}(c_x) \geq 1/2$ for every $x \in \mathbb{N}$, and thus $\mathtt{val}_{\mathcal{G}}(u) \geq 1/2$.

Towards Item 3., by Theorem 7(Item 2), for every MR Maximizer strategy σ we have $\limsup_{x\to\infty} \inf_\pi \mathcal{P}_{\mathcal{G},c_x,\sigma,\pi}(\mathtt{Reach}(\{c_0\})) = 0$. Since in our game the Minimizer strategy π gets to pick the transition $u \to c_x$ for an arbitrary $x \in \mathbb{N}$, $\inf_\pi \mathcal{P}_{\mathcal{G},u,\sigma,\pi}(\mathtt{Reach}(\{c_0\})) = 0$. $\square$

Towards the upper bound, we first define the case where Minimizer can use infinite action sets only finitely often in any play.

Definition 10 *Let $\mathcal{G}$ be a concurrent game on a countable set of states S and $S^\infty \stackrel{\text{def}}{=} \{s \in S \mid |B(s)| = \infty\}$ the states with an infinite Minimizer action set. Let $S' \subseteq S$ be the subset of states s such that every play from s (under any strategies) visits S^∞ only finitely often.*

Different plays from the same start state can have different numbers of visits to S^∞. Even if this number is finite for every play, there is no uniform finite upper bound. Thus the condition of Definition 10 on plays does not imply a finite bound for the start state. However, we show that an ordinal bound exists in this case.

First we recall a classic result on well-founded relations.

Theorem 17 ((Jech, 2002, Theorem 2.27)) *If E is a well-founded relation on V, then there exists a unique function $\rho : V \to \mathbb{O}rd$ such that for all $x \in V$*

$$\rho(x) = \sup\{\rho(y) + 1 \mid yEx\}.$$

In particular, yEx implies $\rho(y) < \rho(x)$. Moreover, if V is countable then $\sup \rho(V)$ is a countable ordinal.

Lemma 18 *Let $\mathcal{G}$ be a concurrent reachability game on a countable set of states S and $S', S^\infty \subseteq S$ as in Definition 10. Let $\rightarrow \subseteq S \times S$ be the induced game graph, i.e., $x \rightarrow y \iff \exists a, b.\, y \in \mathtt{supp}(p(x,a,b))$.*

There exists a unique function $I : S' \to \mathbb{O}rd$ such that

$$x \rightarrow y \text{ implies } I(x) \geq I(y) \tag{19}$$

$$x \in S^\infty \cap S' \ \wedge\ x \rightarrow y \text{ implies } I(x) > I(y) \tag{20}$$

and $\gamma(\mathcal{G}) \stackrel{\text{def}}{=} \sup I(S')$ is a countable ordinal.

Proof Let $V \stackrel{\text{def}}{=} S' \cap S^\infty$ and $E \subseteq V \times V$ be the reversal of the closure of $\rightarrow$ over states in $S' \setminus S^\infty$, i.e., $(y,x) \in E \iff \exists k \geq 0, z_1, \ldots, z_k \in S' \setminus S^\infty \ : x \rightarrow z_1 \rightarrow \cdots \rightarrow z_k \rightarrow y$.

From the definition of S' we obtain that E is a well-founded relation on V. By Theorem 17, there exists a unique function $\rho : V \to \mathbb{O}rd$ such that for all $x \in V$

$$\rho(x) = \sup\{\rho(y) + 1 \mid yEx\}. \tag{21}$$

By convention, $\sup \emptyset = 0$. We define our ranking function $I : V \to \mathbb{O}rd$ by $I(x) \stackrel{\text{def}}{=} \rho(x)+1$. Intuitively, $I(x)$ is the upper bound on the number of visits to S^∞, including the current state x. We now extend the function I from V to S'. For every $x \in S' \setminus V$ let

$$I(x) \stackrel{\text{def}}{=} \sup\{\rho(y) \mid y \in V \wedge x \rightarrow^+ y\}.$$

Again, by convention, $\sup \emptyset = 0$. So the states x that cannot reach S^∞ have $I(x) = 0$. Then Equation (19) and Equation (20) follow directly from (21) and the definition of I. Since S and S' are countable, $\sup I(S')$ is a countable ordinal by Theorem 17. □

It follows from Lemma 18 that states in S' can be part of cycles, but not part of any cycle that contains a state from S^∞. E.g., in the game in Figure 8 we have $S' = \{c_0\}$, i.e., none of the states are in S' except for the target.

Now we show that 1 bit of memory is sufficient for Maximizer, provided that Minimizer can use infinite action sets only finitely often in any play. I.e., for every $\varepsilon > 0$, Maximizer has a public 1-bit strategy for reachability that is uniformly ε-optimal from S'.

First we need a slight generalization of the reachability objective.

Definition 11 (Weighted reachability) *Let $\mathcal{G}$ be a concurrent game on S and $T \subseteq S$. Let $f : T \to [0,1]$ be a reward function. We lift f to plays $f : Z^\omega \to [0,1]$ as follows. If a play $h \in Z^\omega$ never visits T then $f(h) \stackrel{\text{def}}{=} 0$. Otherwise, let $f(h) \stackrel{\text{def}}{=} f(s)$ where s is the first state in T that is visited by h. Let $\mathcal{W}_f$ denote the weighted reachability objective, i.e., to maximize the expected payoff wrt. function f.*

Weighted reachability generalizes reachability (just let $f(s) = 1$ for all $s \in T$). Now we generalize Theorem 6 to weighted reachability.

Theorem 19 *For any concurrent game with finite action sets and weighted reachability objective, for any $\varepsilon > 0$, Maximizer has a uniformly ε-optimal public 1-bit strategy. If the game is turn-based and finitely branching, Maximizer has a deterministic such strategy.*

Proof We can encode weighted reachability into ordinary reachability. Given a concurrent game $\mathcal{G}$ with target set T and weighted reachability objective $\mathcal{W}_f$, we construct a modified game $\mathcal{G}'$ with target set $\{t\}$, where t is a new state, as follows. From every state $s \in T$, regardless of the chosen actions, the game goes to t with probability $f(s)$ and to a special new sink state $\perp$ with probability $1 - f(s)$. Then $\mathcal{W}_f$ in $\mathcal{G}$ coincides with $\mathtt{Reach}(\{t\})$ in $\mathcal{G}'$, i.e., $\mathcal{E}_{\mathcal{G},s,\sigma,\pi}(f) = \mathcal{P}_{\mathcal{G}',s,\sigma,\pi}(\mathtt{Reach}(\{t\}))$. The result follows from Theorem 6, since the 1-bit strategy can be carried from $\mathcal{G}'$ to $\mathcal{G}$. □

Theorem 20 *Let $\mathcal{G}$ be a concurrent game on a countable set of states S with reachability objective $\mathtt{Reach}(T)$ and S' as in Definition 10.*

For every $\varepsilon > 0$, Maximizer has a public 1-bit strategy that is uniformly ε-optimal from every state in S'. If the game is turn-based then Maximizer has a deterministic such strategy.

Proof Let $I : S' \to \mathbb{O}rd$ be the function from Lemma 18. For every ordinal $\alpha \in \mathbb{O}rd$ let $S_\alpha \stackrel{\text{def}}{=} \{s \in S' \mid I(s) = \alpha\}$. We have $S' = \bigcup_{\alpha \le \gamma(\mathcal{G})} S_\alpha$ for the countable ordinal $\gamma(\mathcal{G})$ by Lemma 18. Let $S_{<\alpha} \stackrel{\text{def}}{=} \bigcup_{\beta<\alpha} S_\beta$ and $S_{\le\alpha} \stackrel{\text{def}}{=} \bigcup_{\beta\le\alpha} S_\beta$. We can assume without restriction that the states in T are absorbing and thus $T \subseteq S_0$.

Let the function $f : T \to [0, 1]$ be defined by $f(s) \stackrel{\text{def}}{=} \mathtt{val}_{\mathcal{G},\mathtt{Reach}(T)}(s)$. In particular $f(s) = 1$ for every $s \in T$. The function f is lifted to plays as in Definition 11. Since $\gamma(\mathcal{G})$ is a countable ordinal, the set $\{\alpha \in \mathbb{O}rd \mid \alpha \le \gamma(\mathcal{G})\}$ is countable and thus we can pick an injection $g : \{\alpha \in \mathbb{O}rd \mid \alpha \le \gamma(\mathcal{G})\} \to \mathbb{N}$.

For every ordinal $\alpha \le \gamma(\mathcal{G})$ we consider a restricted subgame $\mathcal{G}_\alpha$ of $\mathcal{G}$ that is played on the subspace $S_{\le\alpha}$. Define $\varepsilon_\alpha \stackrel{\text{def}}{=} \varepsilon \cdot 2^{-g(\alpha)}$ and let $T_\alpha \stackrel{\text{def}}{=} S_{<\alpha} \cup T$ be the target set. We consider the weighted reachability objective $\mathcal{W}_{f_\alpha}$ where $f_\alpha : T_\alpha \to [0, 1]$ with $f_\alpha(s) \stackrel{\text{def}}{=} f(s)$.

For every $s \in S_\alpha$ we show that

$$\mathtt{val}_{\mathcal{G}_\alpha,\mathcal{W}_{f_\alpha}}(s) = \mathtt{val}_{\mathcal{G},\mathtt{Reach}(T)}(s) \tag{22}$$

If $\alpha = 0$ then the equality holds trivially, since $T_0 = T$ and $f(s) = 1$ for every $s \in T$. For $\alpha > 0$, first assume towards a contradiction that $\mathtt{val}_{\mathcal{G}_\alpha,\mathcal{W}_{f_\alpha}}(s) > \mathtt{val}_{\mathcal{G},\mathtt{Reach}(T)}(s)$ for some state $s \in S_\alpha$. Let $\varepsilon \stackrel{\text{def}}{=} (\mathtt{val}_{\mathcal{G}_\alpha,\mathcal{W}_{f_\alpha}}(s) - \mathtt{val}_{\mathcal{G},\mathtt{Reach}(T)}(s))/3 > 0$ and σ an ε-optimal Maximizer strategy from s for $\mathcal{W}_{f_\alpha}$ in $\mathcal{G}_\alpha$. We construct a Maximizer strategy σ' in $\mathcal{G}$ from s as follows. Initially, σ' plays like σ. Then upon reaching some state s' in $S_{<\alpha}$ it switches to an ε-optimal strategy for $\mathtt{Reach}(T)$ from s'. Hence, we get that $\mathcal{P}_{\mathcal{G},s,\sigma',\pi}(\mathtt{Reach}(T)) \ge \mathtt{val}_{\mathcal{G}_\alpha,\mathcal{W}_{f_\alpha}}(s) - 2\varepsilon > \mathtt{val}_{\mathcal{G},\mathtt{Reach}(T)}(s)$, a contradiction. Therefore $\mathtt{val}_{\mathcal{G}_\alpha,\mathcal{W}_{f_\alpha}}(s) \le \mathtt{val}_{\mathcal{G},\mathtt{Reach}(T)}(s)$.

Towards the other inequality, consider an ε-optimal strategy σ from s for $\mathtt{Reach}(T)$ in $\mathcal{G}$ and apply it in $\mathcal{G}_\alpha$. Let π be an arbitrary Minimizer strategy. Let $\mathfrak{R}$

be the set of induced plays from s. Let $\mathfrak{R}_{s'} \subseteq \mathfrak{R}$ be the subset of plays where s' is the first visited state with $I(s') < \alpha$. Then

$$\begin{aligned}\mathtt{val}_{\mathcal{G}_\alpha,\mathcal{W}_{f_\alpha}}(s) \geq \mathcal{E}_{\mathcal{G}_\alpha,s,\sigma,\pi}(f_\alpha) &\geq \sum_{s' \in S_{<\alpha}} \mathcal{P}_{\mathcal{G},s,\sigma,\pi}(\mathfrak{R}_{s'}) \cdot \mathtt{val}_{\mathcal{G},\mathtt{Reach}(T)}(s') \\ &\geq \sum_{s' \in S_{<\alpha}} \mathcal{P}_{\mathcal{G},s,\sigma,\pi}(\mathfrak{R}_{s'}) \cdot \mathcal{P}_{\mathcal{G},s',\sigma,\pi}(\mathtt{Reach}(T)) \\ &= \mathcal{P}_{\mathcal{G},s,\sigma,\pi}(\mathtt{Reach}(T)) \\ &\geq \mathtt{val}_{\mathcal{G},\mathtt{Reach}(T)}(s) - \varepsilon.\end{aligned}$$

Since this holds for every $\varepsilon > 0$, it follows that $\mathtt{val}_{\mathcal{G}_\alpha,\mathcal{W}_{f_\alpha}}(s) \geq \mathtt{val}_{\mathcal{G},\mathtt{Reach}(T)}(s)$.

We now define Maximizer's public 1-bit strategy σ on S' in $\mathcal{G}$. For every $\alpha \leq \gamma(\mathcal{G})$, by Theorem 19, there exists a uniformly ϵ_α-optimal public 1-bit strategy σ_α for Maximizer on $\mathcal{G}_\alpha$ with objective $\mathcal{W}_{f_\alpha}$.

The strategy σ starts with the memory bit set to 0. In every state s with $I(s) = \alpha$ the strategy σ plays like σ_α. Whenever we make a step $s \to s'$ with $I(s') < I(s)$ then set the memory bit to 0. (It is impossible that $I(s') > I(s)$ by the definition of I.) Since all the σ_α are public 1-bit, so is σ. In the special case of turn-based games, the σ_α are also deterministic by Theorem 19 and thus also σ is deterministic.

We now show by induction on α (where $\alpha \leq \gamma(\mathcal{G})$) that σ is uniformly ε'_α-optimal for objective $\mathtt{Reach}(T)$ from every state $s \in S_{\leq\alpha}$, where $\varepsilon'_\alpha \stackrel{\text{def}}{=} \sum_{\beta\leq\alpha} \varepsilon_\beta$.

In the base case of $\alpha = 0$ we have $\varepsilon'_0 = \varepsilon_0$ and $\sigma = \sigma_0$ on $\mathcal{G}_0$. Since $T_0 = T \subseteq S_0$, the objectives $\mathcal{W}_{f_0}$ and $\mathtt{Reach}(T)$ coincide (formally, $\mathcal{E}_{\mathcal{G}_0,s,\sigma,\pi}(f_0) = \mathcal{P}_{\mathcal{G}_0,s,\sigma,\pi}(\mathtt{Reach}(T))$) and the result follows from the properties of σ_0.

For the induction step let $\alpha > 0$. If $s \in S_{<\alpha}$ then the claim holds by induction hypothesis. Now let $s \in S_\alpha$ and π be an arbitrary Minimizer strategy. Let $\mathfrak{R}$ be the set of induced plays from s under σ and π and $\mathfrak{R}_{s'} \subseteq \mathfrak{R}$ be the subset of plays where s' is the first visited state with $I(s') < \alpha$.

$$\begin{aligned}&\mathcal{P}_{\mathcal{G},s,\sigma,\pi}(\mathtt{Reach}(T)) \\ &\geq \sum_{s' \in S_{<\alpha}} \mathcal{P}_{\mathcal{G},s,\sigma,\pi}(\mathfrak{R}_{s'}) \cdot \inf_{\pi'} \mathcal{P}_{\mathcal{G},s',\sigma,\pi'}(\mathtt{Reach}(T)) && \text{since } T \subseteq S_0 \\ &\geq \sum_{s' \in S_{<\alpha}} \mathcal{P}_{\mathcal{G},s,\sigma,\pi}(\mathfrak{R}_{s'}) \cdot (\mathtt{val}_{\mathcal{G},\mathtt{Reach}(T)}(s') - \varepsilon'_{I(s')}) && \text{ind. hyp.} \\ &= \sum_{s' \in S_{<\alpha}} \mathcal{P}_{\mathcal{G},s,\sigma,\pi}(\mathfrak{R}_{s'}) \cdot f_\alpha(s') - \sum_{s' \in S_{<\alpha}} \mathcal{P}_{\mathcal{G},s,\sigma,\pi}(\mathfrak{R}_{s'}) \cdot \varepsilon'_{I(s')} \\ &\geq \mathcal{E}_{\mathcal{G}_\alpha,s,\sigma_\alpha,\pi}(f_\alpha) - \sup_{s' \in S_{<\alpha}} \varepsilon'_{I(s')} \\ &\geq \mathtt{val}_{\mathcal{G}_\alpha,\mathcal{W}_{f_\alpha}}(s) - \varepsilon_\alpha - \sup_{s' \in S_{<\alpha}} \varepsilon'_{I(s')} && \sigma_\alpha \text{ is } \varepsilon_\alpha\text{-optimal} \\ &= \mathtt{val}_{\mathcal{G},\mathtt{Reach}(T)}(s) - \varepsilon_\alpha - \sup_{s' \in S_{<\alpha}} \varepsilon'_{I(s')} && \text{by (22)} \\ &= \mathtt{val}_{\mathcal{G},\mathtt{Reach}(T)}(s) - (\varepsilon_\alpha + \sup_{s' \in S_{<\alpha}} \varepsilon'_{I(s')}) \\ &\geq \mathtt{val}_{\mathcal{G},\mathtt{Reach}(T)}(s) - \varepsilon'_\alpha\end{aligned}$$

Therefore, for every $s \in S' = S_{\leq\gamma(\mathcal{G})}$ our strategy σ is $\varepsilon'_{\gamma(\mathcal{G})}$-optimal. Moreover, $\varepsilon'_{\gamma(\mathcal{G})} = \sum_{\beta\leq\gamma(\mathcal{G})} \varepsilon_\beta = \sum_{\beta\leq\gamma(\mathcal{G})} \varepsilon \cdot 2^{-g(\beta)} \leq \varepsilon$, since $g : \{\alpha \in \mathbb{O}rd \mid \alpha \leq \gamma(\mathcal{G})\} \to \mathbb{N}$ is injective. Thus σ is uniformly ε-optimal from S' in $\mathcal{G}$. $\square$

9 Optimal Maximizer Strategies

In countably infinite reachability games, optimal strategies for Maximizer need not exist in general even if the game is turn-based, in fact not even in countably infinite finitely-branching MDPs Kiefer et al (2017b).

In this section we study the memory requirements of optimal Maximizer strategies under the condition that such an optimal strategy exists.

If we allow infinite action sets for Minimizer (resp. infinite Minimizer branching in turn-based games) then optimal (and even almost surely winning) Maximizer strategies require infinite memory by Theorem 15. Thus, in the rest of this section, we consider games with finite action sets (resp. turn-based finitely-branching games).

9.1 Concurrent Games

We show that optimal Maximizer strategies require infinite memory, even with a step counter.

Definition 12 *We define a concurrent reachability game $\mathcal{G}$ with finite action sets, initial state s_0 and target state t. At state s_0 both players have only one action. The distribution $p(s_0)$ over successor states is defined as $p(s_0)(s_i') = \frac{1}{2^i}$ for all $i \geq 1$. At all states s_i' (where $i \geq 1$), Maximizer has only one action, so s_i' is effectively Minimizer-controlled. At s_i' she chooses between an action, say a_i, that leads with probability $1 - \frac{1}{2^i}$ to the target t and with probability $\frac{1}{2^i}$ to a losing sink state, and an action that leads to a state s_1''. There are also states s_i'' for all $i \geq 2$. At all states s_i'' (where $i \geq 1$), Minimizer has only one action, so s_i'' is effectively Maximizer-controlled. At s_i'' he chooses between an action that leads to s_{i+1}'' and an action that leads with probability $1 - \frac{1}{2^i}$ to t and with probability $\frac{1}{2^i}$ to a losing sink state. The latter action is also called a_i, without risk of confusion, because it is equivalent to Minimizer's action a_i at s_i'.*

Proposition 21 *There exists a concurrent reachability game $\mathcal{G}$ with finite action sets, initial state s_0 and objective $\texttt{Reach}(\{t\})$ as in Definition 12, such that*

1. *Maximizer has an optimal strategy from s_0.*
2. *Every randomized Maximizer strategy from s_0 that uses only a step counter and finite private memory is not optimal.*

Proof We have $\texttt{val}_{\mathcal{G}}(s_i'') = 1$ for all i, and so $\texttt{val}_{\mathcal{G}}(s_i') = 1 - \frac{1}{2^i}$ for all i, and so $\texttt{val}_{\mathcal{G}}(s_0) = \sum_{i=1}^{\infty} \frac{1}{2^i} \cdot (1 - \frac{1}{2^i})$. (The latter series equals $\frac{2}{3}$, but that will not be needed.) It follows that the only optimal Minimizer strategy chooses at s_i' (where $i \geq 1$) action a_i. Note that Maximizer does not make any nontrivial choices if Minimizer plays her optimal strategy.

Towards Item 1, Maximizer's optimal strategy σ from s_0 is defined as follows. In plays where the state s_1'' is not reached, Maximizer does not make any decisions. If s_1'' is reached, Maximizer considers the history of this play: If Minimizer chose the action that leads from s'_i to s''_1 for some $i \geq 1$, then Maximizer chooses actions that lead, from s''_1 via states $s''_2, \ldots, s''_{i-1}$ to state s''_i for the same i, and at state s''_i he chooses action a_i. Note that in this way Maximizer takes the action that Minimizer refused to take (although it would have been optimal for her) at s'_i. With this Maximizer strategy σ, for every Minimizer strategy π, the probability to reach t equals $\mathcal{P}_{\mathcal{G},s_0,\sigma,\pi}(\mathtt{Reach}(\{t\}) = \sum_{i=1}^{\infty} \frac{1}{2^i} \cdot (1 - \frac{1}{2^i}) = \mathtt{val}_{\mathcal{G}}(s_0)$; i.e., σ is optimal.

Towards Item 2, we note that the step counter from s_0 is implicit in the states of $\mathcal{G}$ (except in the target t and the losing sink state), and thus superfluous for Maximizer strategies. Hence it suffices to prove the property for Maximizer strategies with finite memory. Let σ be an FR Maximizer strategy with finitely many memory modes $\{1, \ldots, k\}$. At state s''_1 this strategy σ can base its decision only on the current memory mode $\mathsf{m} \in \{1, \ldots, k\}$. Let $X(\mathsf{m}) \stackrel{\text{def}}{=} \inf_\pi \mathcal{P}_{\mathcal{G},s''_1,\sigma[\mathsf{m}],\pi}(\mathtt{Reach}(\{t\})$ be the probability of reaching the target if σ is in mode m at state s''_1. (From state s''_1 only Maximizer plays, thus Minimizer has no influence.) Since $X(\mathsf{m}) < 1$ and the memory is finite, we have $Y \stackrel{\text{def}}{=} \max_{\mathsf{m} \in \{1,\ldots,k\}} X(\mathsf{m}) < 1$. There exists a number i sufficiently large such that $Y < 1 - \frac{1}{2^i}$. Let π be a Minimizer strategy from s_0 that takes the action that leads from s'_i to s''_1, but chooses action a_j at all states s'_j with $j \neq i$. Then we have

$$\mathcal{P}_{\mathcal{G},s_0,\sigma,\pi}(\mathtt{Reach}(\{t\}) \;\leq\; \frac{1}{2^i} Y + \sum_{j \neq i} \frac{1}{2^j} \cdot \left(1 - \frac{1}{2^j}\right) \;<\; \sum_{j=1}^{\infty} \frac{1}{2^j} \cdot \left(1 - \frac{1}{2^j}\right) \;=\; \mathtt{val}_{\mathcal{G}}(s_0)$$

and thus σ is not optimal. □

In the special case of almost surely winning Maximizer strategies, no memory is needed, not even for uniform strategies.

Theorem 22 *Given a concurrent game with finite action sets and a reachability objective, there is a single randomized memoryless Maximizer strategy that is almost surely winning from every state that admits an almost surely winning strategy.*

Proof Let $\mathcal{G}$ be a concurrent game with state space S, and let $\mathtt{Reach}(T)$ be a reachability objective.

Without restriction, we can assume that all states in S admit an almost surely winning strategy. Otherwise, we consider the subgame $\mathcal{G}'$ obtained by restricting $\mathcal{G}$ to S', where $S' \subseteq S$ is the subset of states that admit an almost surely winning strategy in $\mathcal{G}$. Then all states in S' admit an almost surely winning strategy in $\mathcal{G}'$. (Note that this construction of $\mathcal{G}'$ would not work if we replaced the "almost surely winning" condition by the weaker condition of "having value 1".)

In order to construct a memoryless Maximizer strategy $\hat{\sigma}$ that wins almost surely from every state, we inductively define a sequence of modified games $\mathcal{G}_i$ in which the strategy of Maximizer is already fixed on a finite subset of the state space, and where all states in $\mathcal{G}_i$ still admit an almost surely winning strategy. Fix an enumeration $s_1, s_2, \ldots$ of S in which very state s appears *infinitely often*.

For the base case we have $\mathcal{G}_0 \stackrel{\text{def}}{=} \mathcal{G}$ and the property holds by our assumption on $\mathcal{G}$.

Given $\mathcal{G}_i$ we construct $\mathcal{G}_{i+1}$ as follows. We use Lemma 5 to get a memoryless strategy σ_i and a finite subset of states R_i s.t. $\inf_\pi \mathcal{P}_{\mathcal{G}_i,s_i,\sigma_i,\pi}(\texttt{Reach}_{R_i}(T)) \geq \texttt{val}_{\mathcal{G}_i}(s_i) - 2^{-i} = 1 - 2^{-i}$.

Let $\mathcal{G}'_i$ be the subgame of $\mathcal{G}_i$ that is restricted to R_i and further let

$$R'_i \stackrel{\text{def}}{=} \{s \in R_i \mid \inf_\pi \mathcal{P}_{\mathcal{G}'_i,s,\sigma_i,\pi}(\texttt{Reach}_{R_i}(T)) > 0\}$$

be the subset of states in R_i where σ_i has strictly positive attainment in $\mathcal{G}'_i$. In particular, we have $s_i \in R'_i$ for all $i \geq 1$. Since R'_i is finite, we have

$$\lambda_i \stackrel{\text{def}}{=} \min_{s \in R'_i} \inf_\pi \mathcal{P}_{\mathcal{G}'_i,s,\sigma_i,\pi}(\texttt{Reach}_{R_i}(T)) > 0.$$

We now construct $\mathcal{G}_{i+1}$ by modifying $\mathcal{G}_i$ as follows. For every state $s \in R'_i$ we fix Maximizer's (randomized) action according to σ_i. Then $\inf_\pi \mathcal{P}_{\mathcal{G}_{i+1},s_i,\sigma,\pi}(\texttt{Reach}(T)) \geq 1 - 2^{-i}$ and $\inf_\pi \mathcal{P}_{\mathcal{G}_{i+1},s,\sigma,\pi}(\texttt{Reach}_{R'_i}(T)) \geq \lambda_i$ for all $s \in R'_i$ and all $\sigma \in \Sigma_{\mathcal{G}_{i+1}}$ (and thus in particular for the strategy $\hat{\sigma}$ that we will construct).

Now we show that in $\mathcal{G}_{i+1}$ all states s still have an almost surely winning strategy.

Let σ be an a.s. winning Maximizer strategy from s in $\mathcal{G}_i$, which exists by the induction hypothesis. We now define an a.s. winning Maximizer strategy σ' from s in $\mathcal{G}_{i+1}$.

If the game does not enter R'_i then σ' plays exactly as σ (which is possible since outside R'_i no Maximizer actions have been fixed). If the game enters R'_i then it will reach the target within R'_i with probability $\geq \lambda_i$. Moreover, if the game stays inside R'_i forever then it will almost surely reach the target, since $(1-\lambda_i)^\infty = 0$. Plays that do not stay inside R'_i then exit R'_i at some state $s' \notin R'_i$. Then, from s', σ' plays an a.s. winning strategy w.r.t. $\mathcal{G}_i$ (which exists by the induction hypothesis).

Now we show that σ' wins almost surely from s in $\mathcal{G}_{i+1}$. The set of plays which visit R'_i infinitely often eventually reach the target (except for a nullset), since $(1 - \lambda_i)^\infty = 0$. The set of plays which visit R'_i only finitely often eventually forever play an a.s. winning strategy outside of R'_i and thus (except for a nullset) also eventually reach the target. Hence σ' almost surely wins from s in $\mathcal{G}_{i+1}$.

Finally, we can construct the memoryless Maximizer strategy $\hat{\sigma}$. Since our enumeration of the states $s_1, s_2, \ldots$ contains every state in S and $s_i \in R'_i$, in the limit of the games $\mathcal{G}_\infty$ all Maximizer choices are fixed. The memoryless Maximizer strategy $\hat{\sigma}$ plays according to these fixed choices, i.e., it plays like σ_i at state s_i for all $i \in \mathbb{N}$.

We obtain $\inf_\pi \mathcal{P}_{\mathcal{G},s_i,\hat{\sigma},\pi}(\texttt{Reach}(T)) \geq \inf_\pi \mathcal{P}_{\mathcal{G},s_i,\sigma_i,\pi}(\texttt{Reach}(T)) \geq 1 - 2^{-i}$ for all $i \in \mathbb{N}$. Let $s \in S$. Since $s = s_i$ holds for infinitely many i, we obtain $\inf_\pi \mathcal{P}_{\mathcal{G},s,\hat{\sigma},\pi}(\texttt{Reach}(T)) = 1$ as required. □

9.2 Finitely Branching Turn-Based Games

For finitely branching turn-based reachability games it is shown in (Kučera, 2011, Prop. 5.7.b) that without a step counter, optimal Maximizer strategies require infinite memory. In the following proposition we show that even *with* a step counter, optimal Maximizer strategies require in general at least *some* memory.

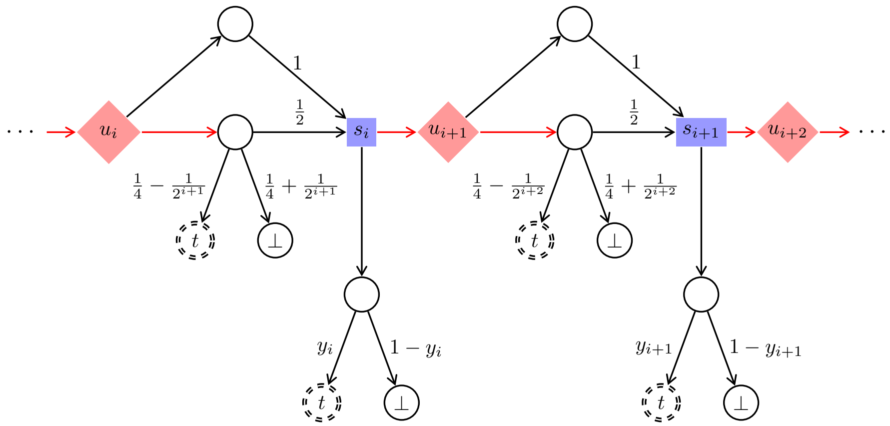

Fig. 9 Finitely branching turn-based reachability game $\mathcal{G}$, where optimal Maximizer strategies cannot be Markov. For the clarity we have drawn several copies of the target state t. The number y_i is defined to be $\frac{1}{2} - \frac{1}{2^{i+1}}$; see Proposition 23.

Proposition 23 *There exist a finitely branching turn-based reachability game $\mathcal{G}$, initial state u_1 and objective* $\mathtt{Reach}(\{t\})$*, as shown in Figure 9, such that*

1. *Maximizer has an optimal strategy from u_1.*
2. *Every randomized Maximizer strategy from u_1 that uses only a step counter and no memory is not optimal.*

Proof Consider first a version of $\mathcal{G}$, say $\mathcal{G}'$, in which the only outgoing transition from the states u_i is the horizontal one, shown in red in Figure 9. I.e., in $\mathcal{G}'$ Minimizer does not have any choice and thus $\mathcal{G}'$ can be regarded as a maximizing MDP. Let σ be the MD Maximizer strategy that chooses at all states s_i the horizontal outgoing transition, shown in red in Figure 9. Then we have

$$\mathcal{P}_{\mathcal{G}',u_i,\sigma}(\mathtt{Reach}(\{t\})) \;=\; \sum_{j=0}^{\infty} \frac{1}{2^j} \cdot \left(\frac{1}{4} - \frac{1}{2^{i+j+1}}\right) \;=\; \frac{1}{2} - \frac{1}{3} \cdot \frac{1}{2^{i-1}}\,.$$

In $\mathcal{G}'$, strategy σ is optimal for Maximizer everywhere; indeed, with the strategy σ' that chooses at state s_i the vertical outgoing transition, we have

$$\begin{aligned} \mathcal{P}_{\mathcal{G}',s_i,\sigma'}(\mathtt{Reach}(\{t\})) \;=\; \frac{1}{2} - \frac{1}{2^{i+1}} \;&<\; \frac{1}{2} - \frac{1}{3} \cdot \frac{1}{2^i} \\ &=\; \inf_{\pi} \mathcal{P}_{\mathcal{G}',u_{i+1},\sigma,\pi}(\mathtt{Reach}(\{t\}))\,. \end{aligned} \tag{23}$$

Consider now the original game $\mathcal{G}$ as shown in Figure 9. Since Minimizer has additional options, the value at each state is not larger than at the corresponding state in $\mathcal{G}'$. However, we show that Maximizer still has an optimal strategy, $\hat{\sigma}$, that, starting from u_1, attains the same value ($\frac{1}{2} - \frac{1}{3}$) as in $\mathcal{G}'$. Namely, define $\hat{\sigma}$ so that as long as Minimizer chooses the horizontal (red) outgoing transitions at u_i, Maximizer chooses the horizontal (red) outgoing transition at s_i; once Minimizer deviates and chooses the non-horizontal outgoing transition at, say, u_i, then Maximizer responds by choosing the vertical outgoing transition at s_i. (The strategy $\hat{\sigma}$ is a deterministic public 1-bit strategy, but we do not need that here.)

Intuitively, for Minimizer a "deviation", i.e., choosing a non-horizontal outgoing transition, is value-increasing and thus suboptimal. But she may try to lay a trap for Maximizer and trick him into visiting all states u_i, s_i. To stop this from happening, Maximizer, using $\hat{\sigma}$, responds to a Minimizer deviation by also deviating, i.e., by choosing a vertical outgoing transition. Such a deviation is suboptimal for him, but the game is constructed so that a Maximizer deviation decreases the value less than Minimizer has previously increased it by her deviation. In effect, with $\hat{\sigma}$, Maximizer attains as much as in $\mathcal{G}'$ if Minimizer never deviates; if Minimizer deviates, Maximizer attains slightly more than in $\mathcal{G}'$. Thus, $\hat{\sigma}$ is optimal.

Formally, let π be any Minimizer strategy. Denote by D_i the event that Minimizer deviates at u_i, i.e., chooses the non-horizontal outgoing transition at u_i. Since the Maximizer strategy $\hat{\sigma}$ responds by choosing the vertical outgoing transition at s_i, we have

$$\mathcal{P}_{\mathcal{G},u_1,\hat{\sigma},\pi}(\mathtt{Reach}(\{t\}) \mid D_i) \;=\; \frac{1}{2} - \frac{1}{2^{i+1}} \;>\; \frac{1}{2} - \frac{1}{3} \cdot \frac{1}{2^{i-1}} \;=\; \mathcal{P}_{\mathcal{G}',u_i,\sigma}(\mathtt{Reach}(\{t\}))\,,$$

i.e., by deviating at u_i, Minimizer increases the probability of reaching t compared to her not deviating at u_i or thereafter (which corresponds to playing in $\mathcal{G}'$). We have already argued that $\mathcal{P}_{\mathcal{G}',u_i,\sigma}(\mathtt{Reach}(\{t\})) = \mathtt{val}_{\mathcal{G}'}(u_i) \geq \mathtt{val}_{\mathcal{G}}(u_i)$. It follows that $\hat{\sigma}$ is optimal, which concludes the proof of Item 1.

Towards Item 2, note that in $\mathcal{G}$ the step counter from u_1 is implicit in the current state. In particular, starting from u_1, if a state s_i is visited then it is visited as the $3i$-th state. It follows that a step counter is not useful for Maximizer strategies. Thus, it suffices to show that no memoryless strategy for Maximizer is optimal. Let σ be any memoryless Maximizer strategy. If σ chooses at every s_i the horizontal outgoing transition, the probability of reaching t is zero if Minimizer never chooses the horizontal outgoing transition at any u_i; thus, σ is not optimal. Hence, we can assume that there is a state s_i at which σ chooses with a positive probability the vertical outgoing transition. Denote by E_i the event that Maximizer chooses the vertical outgoing transition at s_i. Let π be the Minimizer strategy that at all u_j chooses the horizontal outgoing transition. Recall that π is optimal for Minimizer everywhere. Similarly to (23) above, we have

$$\mathcal{P}_{\mathcal{G},u_1,\sigma,\pi}(\mathtt{Reach}(\{t\}) \mid E_i) \;=\; \frac{1}{2} - \frac{1}{2^{i+1}} \;<\; \frac{1}{2} - \frac{1}{3} \cdot \frac{1}{2^{i}} \;=\; \mathtt{val}_{\mathcal{G}}(u_{i+1}) \;=\; \mathtt{val}_{\mathcal{G}}(s_i)\,.$$

Thus, σ is not optimal. As σ was chosen arbitrarily, Maximizer does not have an optimal memoryless strategy. This proves Item 2. ☐

Matching the lower bound from Proposition 23, we show in the following proposition that, with a step counter, 1 bit of public memory suffices for optimal Maximizer strategies.

Theorem 24 *Let $\mathcal{G}$ be a finitely branching turn-based reachability game with initial state s_0 and objective $\mathtt{Reach}(\{t\})$. Suppose that Maximizer has an optimal strategy σ, i.e., $\mathcal{P}_{\mathcal{G},s_0,\sigma,\pi}(\mathtt{Reach}(\{t\})) \geq \mathtt{val}_{\mathcal{G}}(s_0)$ holds for all Minimizer strategies π. Then Maximizer also has a deterministic such strategy that uses 1 bit of public memory and a step counter.*

For the proof we use (Kiefer et al, 2017a, Theorem 5(2)), slightly generalized as the following lemma.

Lemma 25 *Let $\mathcal{G}$ be a finitely branching turn-based reachability game, in which Minimizer does not have value-increasing transitions; i.e., for all transitions $s \longrightarrow s'$ with $s \in S_\Diamond$ we have $\mathtt{val}_{\mathcal{G}}(s) = \mathtt{val}_{\mathcal{G}}(s')$. Then there is a single MD Maximizer strategy that is optimal from every state that admits an optimal strategy.*

Proof The statement follows from (Kiefer et al, 2017a, Thm. 5(2)), but there it is stated only for a single initial state that admits an optimal strategy. Therefore, denote by $S_{opt} \subseteq S$ the set of states that admit an optimal strategy. Add a fresh random state, say s_0, such that the support of $P(s_0)$ equals S_{opt}. This might require infinite branching, but one can easily encode infinite branching of random states into finite branching in the case of reachability objectives, using a "ladder" of fresh intermediate finitely branching random states. Since every state in S_{opt} admits an optimal strategy, the new state s_0 admits an optimal strategy. The mentioned result (Kiefer et al, 2017a, Thm. 5(2)) applied to s_0 gives an MD Maximizer strategy σ that is optimal starting from s_0. But then σ must be optimal from every state in S_{opt}. □

Proof of Theorem 24 Denote by $\bar{\mathcal{G}}$ the game obtained from $\mathcal{G}$ by deleting Minimizer's value-increasing transitions, i.e., those transitions $s \longrightarrow s'$ with $s \in S_\Diamond$ and $\mathtt{val}_{\mathcal{G}}(s) < \mathtt{val}_{\mathcal{G}}(s')$. As $\mathcal{G}$ is finitely branching, each Minimizer state still has at least one outgoing transition, and all states have the same value in $\mathcal{G}$ and $\bar{\mathcal{G}}$, and all states that admit an optimal Maximizer strategy in $\mathcal{G}$ admit an optimal Maximizer strategy in $\bar{\mathcal{G}}$ and vice versa. Denote by $S_{opt} \subseteq S$ the set of states that admit an optimal Maximizer strategy. By Lemma 25 there exists an MD Maximizer strategy $\bar{\sigma}$ in $\bar{\mathcal{G}}$ that is optimal from every state in S_{opt}. Thus, for any $s \in S_\Box \cap S_{opt}$, the transition $s \longrightarrow s'$ that $\bar{\sigma}$ prescribes preserves the value, i.e., $\mathtt{val}(s) = \mathtt{val}(s')$, and $s' \in S_{opt}$. By assumption, $s_0 \in S_{opt}$. Thus, $\bar{\sigma}$ is optimal from s_0 in $\bar{\mathcal{G}}$. Hence, $\bar{\sigma}$ is optimal from s_0 in $\mathcal{G}$ if Minimizer never chooses a value-increasing transition $s \longrightarrow s'$ with $s \in S_\Diamond$ and $\mathtt{val}(s) < \mathtt{val}(s')$. We view such a transition as a *gift* from Minimizer of size $\mathtt{val}(s') - \mathtt{val}(s)$.

Let us now sketch a first draft of a Maximizer strategy that is optimal from s_0 in $\mathcal{G}$.

- Play the strategy $\bar{\sigma}$ until Minimizer gives a gift of, say, ε. Use the 1 bit of public memory to record the fact that a gift has been given.
- Then play an ε-optimal MD strategy, which exists by Lemma 5.

The problem with this draft strategy is that storing the *size* of Minimizer's gift ε appears to require infinite memory, not just 1 bit. I.e., the different ε-optimal MD strategies might prescribe different choices for different ε. Therefore, we use the step counter to deduce a lower bound on any (nonzero) gift that Minimizer may have given up to that point in time. Indeed, for any $i \in \mathbb{N}$ define $\varepsilon_i > 0$ as the smallest nonzero gift that Minimizer can give over all histories $s_0 s_1 \cdots s_i s_{i+1} \in \{s_0\} S^{i+1}$. This is well defined due to the finite branching in $\mathcal{G}$, because only finitely many different gifts can be given in the first i rounds. Another problem with the draft strategy is that the ε-optimal MD strategy from Lemma 5 is ε-optimal only for a finite set of initial states (uniformly ε-optimal positional strategies do not always exist; see Theorem 7).

Therefore, we partition time into infinitely many finite *phases* $\Phi_1 = \{1, \ldots, t_1\}, \Phi_2 = \{t_1 + 1, \ldots, t_2\}, \Phi_3 = \{t_2 + 1, \ldots, t_3\}$, etc, and refer to $\Phi_1, \Phi_3, \ldots$

as *odd* and to $\Phi_2, \Phi_4, \ldots$ as *even* phases. The length of the phases is determined inductively as follows. Denote by S_i the set of states in which the game may be at the beginning of Φ_i (under some pair of strategies). We have $S_1 = \{s_0\}$ and the set S_i is determined by the previous phases $\Phi_1, \ldots, \Phi_{i-1}$. Since $\mathcal{G}$ is finitely branching, S_i is finite.

- Each odd phase Φ_i is long enough so that we have $\inf_\pi \mathcal{P}_{\bar{\mathcal{G}},s,\bar{\sigma},\pi}(\mathtt{Reach}_{\Phi_i}(\{t\})) \geq \frac{\mathtt{val}(s)}{2}$ for all $s \in S_i \cap S_{opt}$, where we write $\mathtt{Reach}_{\Phi_i}(\{t\})$ for the event that t is reached within $t_i - t_{i-1}$ steps, i.e., the length of phase Φ_i. That is, if Φ_i begins at a state $s \in S_i \cap S_{opt}$ and if Minimizer does not give a gift during Φ_i, the Maximizer strategy $\bar{\sigma}$ realizes at least half of the value of s already within Φ_i.
- For each even phase Φ_i, by Lemma 5, there is an MD Maximizer strategy σ_i so that Φ_i can be made long enough so that we have $\inf_\pi \mathcal{P}_{\mathcal{G},s,\sigma_i,\pi}(\mathtt{Reach}_{\Phi_i}(\{t\})) \geq \mathtt{val}(s) - \varepsilon_{\min \Phi_i}$ for all $s \in S_i$. That is, if Minimizer has given a gift in a previous phase and Φ_i begins at a state $s \in S_i$, the Maximizer strategy σ_i "undercuts" Minimizer's gift and realizes most of the value of s already within Φ_i.

We now define a deterministic Maximizer strategy σ from s_0 that uses a step counter and 1 bit of public memory. Later we show that σ is optimal from s_0. Strategy σ uses two modes, m_0 and m_1, where m_0 is the initial mode. Strategy σ *updates* the mode as follows.

(U1) While in m_0 and in an odd phase: if Minimizer gives a gift, switch to m_1. I.e., Maximizer uses the bit to remember that Minimizer has given a gift and will undercut it in the next even phase. The step counter implies a lower bound on the size of the gift.

(U2) While in m_0 and upon entering an odd phase: if the new state does not admit an optimal strategy, switch to m_1. This can only happen if Minimizer has given a gift in some previous *even* phase. (If Minimizer had given a gift in some previous odd phase then the memory mode would already be m_1. If Minimizer has never given a gift then the current state would still admit an optimal strategy, since Maximizer never voluntarily leaves S_{opt}.)

Note that once the mode has been switched to m_1 it is never switched back to m_0. Strategy σ *plays* as follows.

(P1) While in m_0 and in S_{opt}: play $\bar{\sigma}$. This keeps the game in S_{opt}, at least until possibly Minimizer gives a gift.

(P2) While in m_0 and in a state $s \in S_\Box \setminus S_{opt}$: choose a value-preserving transition, i.e., $s{\longrightarrow}s'$ with $\mathtt{val}(s) = \mathtt{val}(s')$. Such a transition must exist, due to the finite branching in $\mathcal{G}$.

(P3) While in m_1 during an odd phase: choose a value-preserving transition, i.e., $s{\longrightarrow}s'$ with $\mathtt{val}(s) = \mathtt{val}(s')$. Such a transition must exist, due to the finite branching in $\mathcal{G}$.

(P4) While in m_1 during an even phase Φ_i: play the MD strategy σ_i from the definition of the even phase Φ_i. It follows from (U1) and (U2) that

σ_i has been played from the beginning of Φ_i. (Here Maximizer undercuts Minimizer's previous gift.)

Note that not all possible gifts by Minimizer are detected, i.e., result in a switch to memory mode m_1. Minimizer could give a gift in an even phase while staying in S_{opt}, or the game might just temporarily leave S_{opt} but move back to S_{opt} before the next odd phase, thus avoiding rule (U2). However, this is not problem for Maximizer. Since the game returns to S_{opt} before the next odd phase, Maximizer is fine to just continue playing $\bar{\sigma}$ by (P1), because he will realize at least half of the value during the next odd phase.

To show that σ is optimal from s_0, fix an arbitrary Minimizer strategy π for the rest of this proof, and assume that the target t is a sink. Let us write $\mathcal{P}$ for $\mathcal{P}_{\mathcal{G},s_0,\sigma,\pi}$ and $\mathcal{E}$ for the associated expectation. We need to show that $\mathtt{val}(s_0) \le \mathcal{P}(\mathtt{Reach}(\{t\}))$.

For a play $s_0 s_1 \cdots \in \{s_0\}S^\omega$, define a random variable τ_1, taking values in $\mathbb{N} \cup \{\infty\}$, such that $\tau_1 = \infty$ if Minimizer never gives a gift, and $\tau_1 = j < \infty$ if $s_j \longrightarrow s_{j+1}$ is the first Minimizer gift. Also define a random variable τ_2, taking values in $\mathbb{N} \cup \{\infty\}$, such that $\tau_2 = \infty$ if no mode switch from m_0 to m_1 ever occurs, and $\tau_2 = k < \infty$ if k is the beginning of the even phase following a mode switch from m_0 to m_1. The random variables $\tau_1 + 1, \tau_2$ are both stopping times. As long as Minimizer does not give any gift, Maximizer plays $\bar{\sigma}$ and, by (P1), keeps the game in S_{opt}, and thus, by (U1) and (U2), the mode remains m_0. Hence, $\tau_1 < \tau_2 \le \infty$ or $\tau_1 = \infty = \tau_2$.

Further, define random variables $V_0, V_1, \ldots$ and $W_0, W_1, \ldots$, taking values in $[0,1]$, by $V_i \stackrel{\text{def}}{=} \mathtt{val}(s_i)$ for all $i \le \tau_1$, and $V_i = V_{\tau_1}$ for all $i \ge \tau_1$, and $W_i \stackrel{\text{def}}{=} \mathtt{val}(s_i)$ for all $i \le \tau_2$, and $W_i = W_{\tau_2}$ for all $i \ge \tau_2$. By (P1), (P2) and (P3), Maximizer preserves the value in each of his transitions, at least until τ_2. Thus, $W_0, W_1, \ldots$ is a submartingale. Moreover, Minimizer preserves the value in each of her transitions, until τ_1. Thus, $V_0, V_1, \ldots$ is a martingale. By Doob's martingale convergence theorem, $V_0, V_1, \ldots$ and $W_0, W_1, \ldots$ converge almost surely to random variables, which we may call, without risk of confusion, V_{τ_1} (which equals V_{τ_1+1}) and W_{τ_2}, respectively. By the optional stopping theorem, we have

$$\mathtt{val}(s_0) \;=\; \mathcal{E}V_0 \;=\; \mathcal{E}V_{\tau_1+1} \;=\; \mathcal{E}V_{\tau_1}\,. \tag{24}$$

On the event $\tau_2 = \infty$, by (U1) and (U2), Minimizer does not give a gift in any odd phase and the state at the beginning of every odd phase admits an optimal strategy for Maximizer. Thus, on $\tau_2 = \infty$, by (P1), Maximizer plays $\bar{\sigma}$ throughout every odd phase. By the definition of odd phases, in every odd phase Maximizer realizes at least half of the value of the state at the beginning of the phase. It follows that

$$\mathcal{P}(\{\tau_2 = \infty\} \cap \{W_\infty > 0\} \setminus \mathtt{Reach}(\{t\})) \;=\; 0\,. \tag{25}$$

Thus,

$$\begin{aligned}\mathcal{E}(V_\infty \mid \tau_1 = \infty) \;&\le\; \mathcal{P}(V_\infty > 0 \mid \tau_1 = \infty)\\ &=\; \mathcal{P}(W_\infty > 0 \mid \tau_1 = \infty) \;\le\; \mathcal{P}(\mathtt{Reach}(\{t\}) \mid \tau_1 = \infty)\,.\end{aligned}$$

Hence, continuing (24),

$$\begin{aligned}\mathtt{val}(s_0) \;&=\; \mathcal{P}(\tau_1 = \infty) \cdot \mathcal{E}(V_\infty \mid \tau_1 = \infty) + \sum_{0 \le j < \infty} \mathcal{P}(\tau_1 = j) \cdot \mathcal{E}(V_j \mid \tau_1 = j)\\ &\le\; \mathcal{P}(\mathtt{Reach}(\{t\}), \tau_1 = \infty) + \sum_{0 \le j < \infty} \mathcal{P}(\tau_1 = j) \cdot \mathcal{E}(V_j \mid \tau_1 = j)\,.\end{aligned} \tag{26}$$

Let $j \in \mathbb{N}$. It follows from the definitions of τ_1 and ε_j that on $\tau_1 = j$ we have $W_{j+1} \geq W_j + \varepsilon_j$. Thus,

$$
\begin{aligned}
&\mathcal{E}(V_j \mid \tau_1 = j) \\
&= \mathcal{E}(W_j \mid \tau_1 = j) && \text{by def. of } V_j, W_j \\
&\leq -\varepsilon_j + \mathcal{E}(W_{j+1} \mid \tau_1 = j) && \text{as explained above} \\
&\leq -\varepsilon_j + \mathcal{E}(W_{\tau_2} \mid \tau_1 = j) && \text{opt. stopping theorem} \\
&= -\varepsilon_j + \mathcal{P}(\tau_2 = \infty \mid \tau_1 = j) \cdot \mathcal{E}(W_\infty \mid \tau_1 = j, \tau_2 = \infty) + \\
&\quad + \sum_{j+1 \leq k < \infty} \mathcal{P}(\tau_2 = k \mid \tau_1 = j) \cdot \mathcal{E}(W_k \mid \tau_1 = j, \tau_2 = k) .
\end{aligned}
\tag{27}
$$

Concerning the first expectation, we have

$$
\begin{aligned}
\mathcal{E}(W_\infty \mid \tau_1 = j, \tau_2 = \infty) &\leq \mathcal{P}(W_\infty > 0 \mid \tau_1 = j, \tau_2 = \infty) \\
&\leq \mathcal{P}(\mathtt{Reach}(\{t\}) \mid \tau_1 = j, \tau_2 = \infty) \quad \text{by (25).}
\end{aligned}
\tag{28}
$$

Concerning the second expectation, let $k > j$, and denote by $H(j,k)$ the set of histories $s_0 \cdots s_k \in \{s_0\}S^k$ such that for some (hence, all) extension(s) $r = s_0 \cdots s_k s_{k+1} \cdots$ we have $\tau_1(r) = j$ and $\tau_2(r) = k$. Then we have

$$
\begin{aligned}
&\mathcal{P}(\tau_1 = j, \tau_2 = k) \cdot \mathcal{E}(W_k \mid \tau_1 = j, \tau_2 = k) \\
&\qquad = \sum_{h = s_0 \cdots s_k \in H(j,k)} \mathcal{P}(\{h\}S^\omega) \cdot \mathtt{val}(s_k) && \text{by the definitions} \\
&\qquad \leq \sum_{h \in H(j,k)} \mathcal{P}(\{h\}S^\omega) \cdot \big(\mathcal{P}(\mathtt{Reach}(\{t\}) \mid \{h\}S^\omega) + \varepsilon_k\big) && \text{by (P4)} \\
&\qquad \leq \sum_{h \in H(j,k)} \mathcal{P}(\{h\}S^\omega) \cdot \big(\mathcal{P}(\mathtt{Reach}(\{t\}) \mid \{h\}S^\omega) + \varepsilon_j\big) && \text{as } \varepsilon_k \leq \varepsilon_j \\
&\qquad = \mathcal{P}(\tau_1 = j, \tau_2 = k) \cdot \big(\mathcal{P}(\mathtt{Reach}(\{t\}) \mid \tau_1 = j, \tau_2 = k) + \varepsilon_j\big) .
\end{aligned}
$$

Thus,

$$
\begin{aligned}
&\sum_{j+1 \leq k < \infty} \mathcal{P}(\tau_2 = k \mid \tau_1 = j) \cdot \mathcal{E}(W_k \mid \tau_1 = j, \tau_2 = k) \\
&\qquad \leq \sum_{j+1 \leq k < \infty} \mathcal{P}(\tau_2 = k \mid \tau_1 = j) \cdot \big(\mathcal{P}(\mathtt{Reach}(\{t\}) \mid \tau_1 = j, \tau_2 = k) + \varepsilon_j\big) \\
&\qquad \leq \mathcal{P}(\mathtt{Reach}(\{t\}) \mid \tau_2 < \infty, \tau_1 = j) + \varepsilon_j .
\end{aligned}
$$

Combined with Equations (27) and (28) this gives

$$
\begin{aligned}
\mathcal{E}(V_j \mid \tau_1 = j) &\leq \mathcal{P}(\mathtt{Reach}(\{t\}) \mid \tau_2 = \infty, \tau_1 = j) + \mathcal{P}(\mathtt{Reach}(\{t\}) \mid \tau_2 < \infty, \tau_1 = j) \\
&\leq \mathcal{P}(\mathtt{Reach}(\{t\}) \mid \tau_1 = j) .
\end{aligned}
$$

Combined with (26), we obtain

$$
\mathtt{val}(s_0) \leq \mathcal{P}(\mathtt{Reach}(\{t\}), \tau_1 = \infty) + \mathcal{P}(\mathtt{Reach}(\{t\}), \tau_1 < \infty) = \mathcal{P}(\mathtt{Reach}(\{t\})) ,
$$

as required. $\square$

10 Minimizer Strategies

In the previous sections we have considered the strategy complexity of Maximizer's strategies. In this section we complete the picture of the strategy complexity of Minimizer; cf. Table 2.

We begin by considering games on acyclic graphs. Positional strategies in acyclic games yield Markov strategies in general games, since an encoded step counter makes the graph acyclic. In fact, the following result about acyclic games is slightly more general, since not all acyclic graphs yield an implicit step counter, i.e., the same state might be reached via paths of different lengths.

Lemma 26 *Acyclic turn-based reachability games admit uniformly ε-optimal MD strategies for Minimizer. That is, for every such game $\mathcal{G} = (S, (S_\Box, S_\Diamond, S_\bigcirc), \longrightarrow, P)$, reachability target $T \subseteq S$ and every $0 < \varepsilon < 1$ there exists an MD Minimizer strategy π which satisfies, for every state $s_0 \in S$ and every Maximizer strategy σ, that $\mathcal{P}_{\mathcal{G},s_0,\sigma,\pi}(\mathtt{Reach}(T)) \le \mathtt{val}_{\mathcal{G},\mathtt{Reach}(T)}(s_0)(1+\varepsilon)$.*

Proof Let us shortly write $\mathtt{val}(s) = \mathtt{val}_{\mathcal{G},\mathtt{Reach}(T)}(s)$ for the value of a state s and let $\iota : S \to \mathbb{N} \setminus \{0\}$ be an enumeration of the state space starting at 1. Define π as the MD-strategy that, at any state $s \in S_\Diamond$, picks a successor s' such that

$$\mathtt{val}(s') \quad \le \quad \mathtt{val}(s)(1 + \ln(1+\varepsilon)2^{-\iota(s)}).$$

To show that this strategy π satisfies the claim we (over)estimate the error by $L(s) \stackrel{\text{def}}{=} \prod_{s' \in Post^*(s)}(1 + \ln(1+\varepsilon)2^{-\iota(s')})$ where $Post^*(s) \subseteq S$ is the set of states reachable from state $s \in S$ (under any pair of strategies). Notice that this guarantees that

$$\begin{aligned} 1 < L(s) &\le \prod_{i>0}\left(1 + \ln(1+\varepsilon)2^{-i}\right) \\ &\le \exp\left(\sum_{i>0} \ln(1+\varepsilon)2^{-i}\right) \\ &\le \exp(\ln(1+\varepsilon)) = 1+\varepsilon \end{aligned} \tag{29}$$

where the third inequality uses that $1 + x \le \exp(x)$.

Let σ be an arbitrary Maximizer strategy. For this pair σ, π of strategies let's consider plays $(X_i)_{i\ge 0}$ that start in $s_0 \in S$ and proceed according to σ, π, and let $\mathcal{E}_{\mathcal{G},s_0,\sigma,\pi}$ be the expectation with respect to $\mathcal{P}_{\mathcal{G},s_0,\sigma,\pi}$.

An induction on n using our choice of strategy gives, for every initial state $s_0 \in S$, that

$$\mathcal{E}_{\mathcal{G},s_0,\sigma,\pi}(\mathtt{val}(X_n)) \le \mathtt{val}(s_0)L(s_0). \tag{30}$$

Indeed, this trivially holds for $n = 0$ as $\mathcal{E}_{\mathcal{G},s_0,\sigma,\pi}(\mathtt{val}(X_0)) = \mathtt{val}(s_0)$ and $L(s_0) > 1$. For the induction step there are three cases.

Case 1: $s_0 \in S_\Diamond$ and $\pi(s_0) = s$. Let $\sigma[s_0 \to s]$ denote the Maximizer strategy from s that behaves just like σ does after observing the first step, i.e., satisfies $\sigma[s_0 \to s](sh) = \sigma(s_0 sh)$ for all suffix histories $h \in S^*$. Then

$$\mathcal{E}_{\mathcal{G},s_0,\sigma,\pi}(\mathtt{val}(X_{n+1})) = \mathcal{E}_{\mathcal{G},s,\sigma[s_0\to s],\pi}(\mathtt{val}(X_n))$$

$$\begin{aligned}
&\leq \mathtt{val}(s)L(s) && \text{ind. hyp.}\\
&\leq \mathtt{val}(s_0)\left(1+\ln(1+\varepsilon)2^{-\iota(s_0)}\right)L(s) && \text{def. of } \pi\\
&\leq \mathtt{val}(s_0)L(s_0) && \text{acyclicity; def. of } L(s_0).
\end{aligned}$$

Case 2: $s_0 \in S_\Box$. Again, for any state s let $\sigma[s_0 \to s]$ denote the suffix strategy consistent with σ after the first step. Then

$$\begin{aligned}
\mathcal{E}_{\mathcal{G},s_0,\sigma,\pi}(\mathtt{val}(X_{n+1})) &= \sum_{s\in S}\sigma(s_0)(s)\cdot\mathcal{E}_{\mathcal{G},s,\sigma[s_0\to s],\pi}(\mathtt{val}(X_n))\\
&\leq \sum_{s\in S}\sigma(s_0)(s)\cdot\mathtt{val}(s)L(s)\\
&\leq \sum_{s\in S}\sigma(s_0)(s)\cdot\mathtt{val}(s)\left(1+\ln(1+\varepsilon)2^{-\iota(s_0)}\right)L(s)\\
&\leq \sum_{s\in S}\sigma(s_0)(s)\cdot\mathtt{val}(s)L(s_0)\\
&\leq \mathtt{val}(s_0)L(s_0),
\end{aligned}$$

where the first inequality holds by induction hypothesis, the second holds because $1 < (1+\ln(1+\varepsilon)2^{-\iota(s_0)})$, and the third is acyclicity and the definition of $L(s_0)$.

Case 3: $s_0 \in S_\bigcirc$ is analogous to case 2, with the only difference that the initial successor distribution is $P(s_0)$, the one fixed by the game, instead of $\sigma(s_0)$ and the last inequality becomes an equality.

Together with the observation (Equation (29)) that $L(s_0) \leq (1+\varepsilon)$ for every s_0, we derive that

$$\liminf_{n\to\infty}\mathcal{E}_{\mathcal{G},s_0,\sigma,\pi}(\mathtt{val}(X_n)) \leq \mathtt{val}(s_0)(1+\varepsilon). \tag{31}$$

Finally, to show the claim, let $[X_n \in T] : S^\omega \to \{0,1\}$ be the random variable that indicates that the nth state is in T. Note that $[X_n \in T] \leq \mathtt{val}(X_n)$ because target states have value 1. We conclude that

$$\begin{aligned}
\mathcal{P}_{\mathcal{G},s_0,\sigma,\pi}(\mathtt{Reach}(T)) &= \mathcal{P}_{\mathcal{G},s_0,\sigma,\pi}\left(\bigcup_{i=0}^{\infty}\mathtt{Reach}_i(T)\right) && \text{semantics of } \mathtt{Reach}(T)\\
&= \lim_{n\to\infty}\mathcal{P}_{\mathcal{G},s_0,\sigma,\pi}\left(\bigcup_{i=0}^{n}\mathtt{Reach}_i(T)\right) && \text{continuity of measures}\\
&= \lim_{n\to\infty}\mathcal{P}_{\mathcal{G},s_0,\sigma,\pi}\left(\mathtt{Reach}_n(T)\right) && T \text{ is a sink}\\
&= \lim_{n\to\infty}\mathcal{E}_{\mathcal{G},s_0,\sigma,\pi}([X_n \in T]) && \text{definition of } [X_n \in T]\\
&\leq \liminf_{n\to\infty}\mathcal{E}(\mathtt{val}(X_n)) && \text{as } [X_n \in T] \leq \mathtt{val}(X_n)\\
&\leq \mathtt{val}(s_0)(1+\varepsilon) && \text{by Equation (31).}
\end{aligned}$$

□

Theorem 27 *Turn-based games, even infinitely branching ones, with reachability objective admit uniformly ε-optimal strategies for Minimizer that are deterministic and Markov.*

Proof For a given game $\mathcal{G} = (S, (S_\Box, S_\Diamond, S_\bigcirc), \longrightarrow, P)$ and reachability target $T \subseteq S$, one can construct the acyclic game that encodes the stage (clock value) into the states: $\mathcal{G}' = (S, (S'_\Box, S'_\Diamond, S'_\bigcirc), \longrightarrow', P')$ where $S' = S \times \mathbb{N}$, $S'_\Box = S_\Box \times \mathbb{N}$, $S'_\Diamond = S_\Diamond \times \mathbb{N}$ $S'_\bigcirc = S_\bigcirc \times \mathbb{N}$, and for all $i \in \mathbb{N}$, $(s,i) {\longrightarrow'} (t, i{+}1) \iff s{\longrightarrow}t$ and $P((s,i))((t,i{+}1)) = P(s)(t)$.

Every Markov strategy in $\mathcal{G}$ uniquely gives rise to a positional strategy in $\mathcal{G}'$ and vice versa. The claim now follows from Lemma 26. $\Box$

In infinitely branching turn-based reachability games, optimal Minimizer strategies need not exist Kiefer et al (2017b). When they do exist, they may need infinite memory. We slightly improve the result of (Kučera, 2011, Proposition 5.6.(a)) by showing that even a step counter does not help Minimizer.

Definition 13 *We define an infinitely branching turn-based reachability game $\mathcal{G}$ with initial state s_0 and target state t. Let s_0 be Maximizer-controlled. We have transitions $s_0 \to s'_i$ for all $i \geq 1$. All states s'_i are random states with $P(s'_i)(t) = 1/2 - 2^{-i}$ and and $P(s'_i)(u) = 1/2 + 2^{-i}$. The state u is Minimizer-controlled with transitions $u \to s''_i$ for all $i \geq 1$. All states s''_i are random states with $P(s''_i)(t) = 2^{-i}$ and $P(s''_i)(b) = 1 - 2^{-i}$ for a losing sink state b.*

Proposition 28 *There exists an infinitely branching turn-based reachability game $\mathcal{G}$ with initial state s_0 and objective* $\texttt{Reach}(\{t\})$ *as in Definition 13, such that*

1. *Minimizer has an optimal strategy from s_0.*
2. *Every randomized Minimizer strategy from s_0 that uses only a step counter and finite private memory is not optimal.*

Proof Towards Item 1, we note that $\texttt{val}_\mathcal{G}(u) = 0$ and thus $\texttt{val}_\mathcal{G}(s_0) = 1/2$. Minimizer's optimal strategy π from s_0 is defined as follows. In plays where the state u is not reached, Minimizer does not make any decisions. If state u is reached, Minimizer considers the history of this play: If Maximizer made the step $s_0 \to s'_i$ for some $i \geq 1$, then Minimizer plays $u \to s''_i$ for the same i. Now we show that π is optimal for Minimizer from s_0. Let σ be an arbitrary Maximizer strategy from s_0 and let x_i be the probability that σ chooses the step $s_0 \to s'_i$. This must be a distribution, i.e., $\sum_{i\geq1} x_i = 1$. Then we have

$$\begin{aligned}\mathcal{P}_{\mathcal{G},s_0,\sigma,\pi}(\texttt{Reach}(\{t\}) &= \sum_{i\geq1} x_i((1/2 - 2^{-i}) + (1/2 + 2^{-i})2^{-i}) \\ &\leq \sum_{i\geq1} x_i(1/2) = 1/2 = \texttt{val}_\mathcal{G}(s_0)\end{aligned} \tag{32}$$

as required.

Towards Item 2, we note that the step counter from s_0 is implicit in the states of $\mathcal{G}$, and thus superfluous for Minimizer strategies. Hence it suffices to prove the property for Minimizer strategies with finite memory. Let π be an FR Minimizer strategy with finitely many memory modes $\{1, \ldots, k\}$. In state u this strategy π can base its decision only on the current memory mode $\mathsf{m} \in \{1, \ldots, k\}$. Let $X(\mathsf{m}) \stackrel{\text{def}}{=}$

$\mathcal{P}_{\mathcal{G},u,\sigma,\pi[\mathsf{m}]}(\mathtt{Reach}(\{t\})$ be the probability of reaching the target if π is in mode m at state u. (From state u only Minimizer plays, thus Maximizer has no influence.) Since $X(\mathsf{m}) > 0$ and the memory is finite, we have $Y \stackrel{\text{def}}{=} \min_{\mathsf{m}\in\{1,\ldots,k\}} X(\mathsf{m}) > 0$. There exists a number i sufficiently large such that $2^{-i} < Y/2$. Let σ be a Maximizer strategy from s_0 that chooses the transition $s_0 \to s_i'$. Then we have

$$\mathcal{P}_{\mathcal{G},s_0,\sigma,\pi}(\mathtt{Reach}(\{t\}) \geq (1/2 - 2^{-i}) + (1/2 + 2^{-i})Y > 1/2 = \mathtt{val}_{\mathcal{G}}(s_0)$$

and thus π is not optimal. □

11 Conclusion

Our results show that infinite action sets for Minimizer make stochastic reachability games much more difficult for Maximizer. In particular, Proposition 1, the existence of positional ε-optimal Maximizer strategies in concurrent games, no longer holds if Minimizer is allowed infinite action sets. Even in turn-based reachability games with infinitely branching Minimizer states, Maximizer strategies based on a step counter plus arbitrary finite private memory are useless.

Moreover, our results highlight the differences between MDPs and 2-player stochastic games w.r.t. the reachability objective. While Ornstein's construction Ornstein (1969) can turn ε-optimal positional strategies into uniform ones in MDPs, this is not possible even in finitely branching turn-based stochastic games. On the other hand, just 1 bit of public memory is enough for uniformly ε-optimal strategies, even in concurrent reachability games with finite action sets.

Statements and Declarations

- **Funding.** This work has been supported by the EPSRC, project No. EP/V025848/1, and by the Royal Society, grant IES\R3\213110.
- **Competing interests.** The authors declare they have no financial, or non-financial interests, and have no potential conflicts of interest to declare.

Appendix A Technical Lemmas

The following inequality is due to Weierstrass.

Proposition 29 ((Bromwich, 1955, p. 104-105)) *Given an infinite sequence of real numbers a_n with $0 \leq a_n \leq 1$,*

$$\prod_{k=1}^{n}(1-a_k) \leq \frac{1}{1+\sum_{k=1}^{n} a_k}$$

Proof By induction on n. In case $n=1$ we have $(1-a_1)(1+a_1) = (1-a_1^2) \leq 1$ as required. For the induction hypothesis we assume

$$\prod_{k=1}^{n}(1-a_k)\left(1+\sum_{k=1}^{n} a_k\right) \leq 1$$

For the induction step we have

$$\begin{aligned}
\prod_{k=1}^{n+1}(1-a_k)\left(1+\sum_{k=1}^{n+1} a_k\right) &= (1-a_{n+1})\prod_{k=1}^{n}(1-a_k)\left(\left(1+\sum_{k=1}^{n} a_k\right)+a_{n+1}\right)\\
&\leq (1-a_{n+1})\left(1+a_{n+1}\prod_{k=1}^{n}(1-a_k)\right)\\
&\leq (1-a_{n+1})(1+a_{n+1})\\
&= (1-a_{n+1}^2) \leq 1
\end{aligned}$$

□

Proposition 30 *Given an infinite sequence of real numbers a_n with $0 \leq a_n < 1$, we have*

$$\prod_{n=1}^{\infty}(1-a_n) > 0 \quad \Leftrightarrow \quad \sum_{n=1}^{\infty} a_n < \infty.$$

and the "⇒" implication holds even for the weaker assumption $0 \leq a_n \leq 1$.

Proof If $a_n = 1$ for any n then the "⇒" implication is vacuously true, but the "⇐" implication does not hold in general. In the following we assume $0 \leq a_n < 1$.

In the case where a_n does not converge to zero, the property is trivial. In the case where $a_n \to 0$, it is shown by taking the logarithm of the product and using the limit comparison test as follows.

Taking the logarithm of the product gives the series

$$\sum_{n=1}^{\infty} \ln(1-a_n)$$

whose convergence (to a finite number ≤ 0) is equivalent to the positivity of the product. It is also equivalent to the convergence (to a number ≥ 0) of its negation $\sum_{n=1}^{\infty} -\ln(1-a_n)$. But observe that (by L'Hôpital's rule)

$$\lim_{x \to 0} \frac{-\ln(1-x)}{x} = 1.$$

Since $a_n \to 0$ we have

$$\lim_{n\to\infty} \frac{-\ln(1-a_n)}{a_n} = 1.$$

By the limit comparison test, the series $\sum_{n=1}^{\infty} -\ln(1-a_n)$ converges if and only if the series $\sum_{n=1}^{\infty} a_n$ converges. $\square$

Proposition 31 *Given an infinite sequence of real numbers a_n with $0 \leq a_n \leq 1$,*

$$\prod_{n=1}^{\infty} a_n > 0 \quad \Rightarrow \quad \forall \varepsilon > 0\, \exists N.\ \prod_{n=N}^{\infty} a_n \geq (1-\varepsilon).$$

Proof If $a_n = 0$ for any n then the property is vacuously true. In the following we assume $a_n > 0$. Since $\prod_{n=1}^{\infty} a_n > 0$, by taking the logarithm we obtain $\sum_{n=1}^{\infty} \ln(a_n) > -\infty$. Thus for every $\delta > 0$ there exists an N s.t. $\sum_{n=N}^{\infty} \ln(a_n) \geq -\delta$. By exponentiation we obtain $\prod_{n=N}^{\infty} a_n \geq \exp(-\delta)$. By picking $\delta = -\ln(1-\varepsilon)$ the result follows. $\square$

References

de Alfaro L, Henzinger TA (2001) Interface automata. In: International Symposium on Foundations of Software Engineering. ACM, pp 109–120, https://doi.org/10.1145/503209, URL https://doi.org/10.1145/503209

de Alfaro L, Henzinger TA, Kupferman O (1998) Concurrent reachability games. In: 39th Annual Symposium on Foundations of Computer Science, FOCS '98, November 8-11, 1998, Palo Alto, California, USA. IEEE Computer Society, pp 564–575, https://doi.org/10.1109/SFCS.1998.743507

Altman E, Avrachenkov K, Marquez R, Miller GB (2005) Zero-sum constrained stochastic games with independent state processes. Math Methods Oper Res 62(3):375–386. https://doi.org/10.1007/s00186-005-0034-4, URL https://doi.org/10.1007/s00186-005-0034-4

Altman E, Avrachenkov K, Bonneau N, et al (2007) Constrained stochastic games in wireless networks. In: IEEE (ed) IEEE global communications conference (GLOBECOM), Washington

Aumann RJ (1981) Survey of repeated games. Essays in game theory and mathematical economics in honor of Oskar Morgenstern

Bacharach M (2019) Economics and the Theory of Games. CRC Press

Bertrand N, Genest B, Gimbert H (2017) Qualitative determinacy and decidability of stochastic games with signals. Journal of the ACM 64(5):33:1–33:48. URL https://doi.org/10.1145/3107926

Billingsley P (1995) Probability and Measure. Wiley, New York, NY, third Edition

Blackwell D, Ferguson TS (1968) The Big Match. The Annals of Mathematical Statistics 39(1):159 – 163. https://doi.org/10.1214/aoms/1177698513, URL https://doi.org/10.1214/aoms/1177698513

Bouyer P, Markey N, Randour M, et al (2016) Reachability in Networks of Register Protocols under Stochastic Schedulers. In: International Colloquium on Automata, Languages and Programming (ICALP), Leibniz International Proceedings in Informatics (LIPIcs), vol 55. Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik, Dagstuhl, Germany, pp 106:1–106:14

Brázdil T, Brožek V, Kučera A, Obdrzálek J (2011) Qualitative reachability in stochastic BPA games. Information and Computation 209. URL http://arxiv.org/abs/1003.0118

Brázdil T, Kiefer S, Kučera A, et al (2014) Zero-reachability in probabilistic multi-counter automata. In: ACM/IEEE Symposium on Logic in Computer Science (LICS). Association for Computing Machinery, New York, NY, USA, CSL-LICS '14

Bromwich T (1955) An Introduction to the Theory of Infinite Series. McMillan and Company, London

Chen T, Forejt V, Kwiatkowska M, et al (2013) Prism-games: A model checker for stochastic multi-player games. In: International Conference on TOOLS and Algorithms for the Construction and Analysis of Systems, Springer, pp 185–191

Condon A (1992) The complexity of stochastic games. Information and Computation 96(2):203–224. https://doi.org/https://doi.org/10.1016/0890-5401(92)90048-K

Filar JA (1980) Algorithms for solving some undiscounted stochastic games. PhD thesis, University of Illinois at Chicago

Filar JA (1981) Ordered field property for stochastic games when the player who controls transitions changes from state to state. Journal of Optimization Theory and Applications 34(4):503–515

Flesch J, Predtetchinski A, Sudderth WD (2018) Characterization and simplification of optimal strategies in positive stochastic games. Journal of Applied Probability 55(3):728–741

Flesch J, Predtetchinski A, Sudderth W (2020) Positive zero-sum stochastic games with countable state and action spaces. Applied Mathematics and Optimization 82:499–516

Flesch J, Herings PJ, Maes J, Predtetchinski A (2021) Subgame maxmin strategies in zero-sum stochastic games with tolerance levels. Dynamic Games and Applications 11(4):704–737

Fristedt B, Lapic SK, Sudderth WD (1995) The big match on the integers. In: G.J. O (ed) New Trends in Dynamic Games and Applications. Birkhäuser Boston, Annals of the International Society of Dynamic Games

Gillette D (1958) Stochastic games with zero stop probabilities. In: Contributions to the Theory of Games (AM-39), Volume III. Princeton University Press, p 179–188, https://doi.org/10.1515/9781400882151-011

Gimbert H, Horn F (2010) Solving simple stochastic tail games. In: Annual ACM-SIAM Symposium on Discrete Algorithms, pp 847–862, https://doi.org/10.1137/1.9781611973075, URL https://doi.org/10.1137/1.9781611973075

Hansen K, Ibsen-Jensen R, Neyman A (2018) The big match with a clock and a bit of memory. In: Economics and Computation (EC). Association for Computing Machinery, New York, NY, USA, pp 149–50, https://doi.org/10.1145/3219166.3219198, URL https://doi.org/10.1145/3219166.3219198

Jaskiewicz A, Nowak AS (2011) Stochastic games with unbounded payoffs: Applications to robust control in economics. Dynamic Games and Applications 1(2):253–279. https://doi.org/10.1007/s13235-011-0013-8, URL https://doi.org/10.1007/s13235-011-0013-8

Jech T (2002) Set Theory: The Third Millennium Edition, 3rd edn. Springer Monographs in Mathematics, https://doi.org/10.1007/3-540-44761-X

Kalai E (1990) Bounded rationality and strategic complexity in repeated games. In: Ichiishi T, Neyman A, Tauman Y (eds) Game Theory and Applications. Economic Theory, Econometrics, and Mathematical Economics, Academic Press, San Diego, p 131–157, https://doi.org/https://doi.org/10.1016/B978-0-12-370182-4.50010-6

Kiefer S, Mayr R, Shirmohammadi M, Wojtczak D (2017a) On strong determinacy of countable stochastic games. In: ACM/IEEE Symposium on Logic in Computer Science (LICS). IEEE Computer Society, pp 1–12, https://doi.org/10.1109/LICS.2017.8005134

Kiefer S, Mayr R, Shirmohammadi M, Wojtczak D (2017b) Parity Objectives in Countable MDPs. In: ACM/IEEE Symposium on Logic in Computer

Science (LICS), https://doi.org/10.1109/LICS.2017.8005100

Kiefer S, Mayr R, Shirmohammadi M, Totzke P (2019) Büchi objectives in countable MDPs. In: International Colloquium on Automata, Languages and Programming (ICALP), pp 119:1–119:14, https://doi.org/10.4230/LIPIcs.ICALP.2019.119, full version at https://arxiv.org/abs/1904.11573

Kiefer S, Mayr R, Shirmohammadi M, Totzke P (2020a) Strategy Complexity of Parity Objectives in Countable MDPs. In: International Conference on Concurrency Theory (CONCUR), pp 7:1–:17, https://doi.org/10.4230/LIPIcs.CONCUR.2020.39

Kiefer S, Mayr R, Shirmohammadi M, et al (2020b) How to play in infinite MDPs (invited talk). In: International Colloquium on Automata, Languages and Programming (ICALP), pp 3:1–3:18, https://doi.org/10.4230/LIPIcs.ICALP.2020.3

Kumar PR, Shiau TH (1981) Existence of value and randomized strategies in zero-sum discrete-time stochastic dynamic games. SIAM Journal on Control and Optimization 19(5):617–634. https://doi.org/10.1137/0319039

Kučera A (2011) Turn-based stochastic games. In: Apt KR, Grädel E (eds) Lectures in Game Theory for Computer Scientists. Cambridge University Press, p 146–184, https://doi.org/10.1017/CBO9780511973468.006

Laraki R, Maitra AP, Sudderth WD (2013) Two-person zero-sum stochastic games with semicontinuous payoff. Dynamic Games and Applications 3(2):162–171

Maitra A, Sudderth W (1998) Finitely additive stochastic games with borel measurable payoffs. International Journal of Game Theory 27(2):257–267. https://doi.org/10.1007/s001820050071, URL https://doi.org/10.1007/s001820050071

Maitra AP, Sudderth WD (1996) Discrete Gambling and Stochastic Games. Springer-Verlag

Maitra AP, Sudderth WD (2007) Subgame-perfect equilibria for stochastic games. Mathematics of Operations Research 32(3):711–722. https://doi.org/10.1287/moor.1070.0264

Martin DA (1998) The determinacy of blackwell games. Journal of Symbolic Logic 63(4):1565–1581. https://doi.org/10.2307/2586667

Neyman A, Sorin S (2003) Stochastic Games and Applications. Springer, Dordrecht, https://doi.org/10.1007/978-94-010-0189-2, URL https://doi.org/10.1007/978-94-010-0189-2

Nowak A, Raghavan T (1991) Positive stochastic games and a theorem of Ornstein. In: Raghavan T, Ferguson T, Parthasarathy T, Vrieze O (eds) Stochastic Games And Related Topics, Theory and Decision Library (Game Theory, Mathematical Programming and Operations Research), vol 7. Springer, https://doi.org/10.1007/978-94-011-3760-7_11

Nowak A, Szajowski K (2005) Advances in Dynamic Games: Applications to Economics, Finance, Optimization, and Stochastic Control. Birkhaeuser, Boston, https://doi.org/10.1007/b138903

Ornstein D (1969) On the existence of stationary optimal strategies. Proceedings of the American Mathematical Society 20(2):563–569. https://doi.org/10.2307/2035700

Puterman ML (1994) Markov Decision Processes: Discrete Stochastic Dynamic Programming, 1st edn. John Wiley & Sons, Inc., New York, NY, USA, https://doi.org/10.1002/9780470316887

Raghavan TES, Ferguson T, Parthasarathy T, Vrieze O (2012) Stochastic Games and Related Topics: in honor of Professor LS Shapley, vol 7. Springer Science & Business Media, https://doi.org/10.1007/978-94-011-3760-7

Secchi P (1997) Stationary strategies for recursive games. Math Oper Res 22(2):494–512. https://doi.org/10.1287/moor.22.2.494, URL https://doi.org/10.1287/moor.22.2.494

Secchi P (1998) On the existence of good stationary strategies for nonleavable stochastic games. International Journal of Game Theory 27:61–81. https://doi.org/10.1007/BF01243195

Shapley LS (1953) Stochastic games. Proceedings of the National Academy of Sciences 39(10):1095–1100. https://doi.org/10.1073/pnas.39.10.1095

Solan E, Vieille N (2015) Stochastic games. Proceedings of the National Academy of Sciences 112(45):13,743–13,746. https://doi.org/10.1073/pnas.1513508112, URL https://www.pnas.org/content/112/45/13743, https://arxiv.org/abs/https://www.pnas.org/content/112/45/13743.full.pdf

Sorin S (1992) Repeated games with complete information. In: Aumann R, Hart S (eds) Handbook of Game Theory with Economic Applications, Handbook of Game Theory with Economic Applications, vol 1. Elsevier, chap 4, p 71–107, URL https://ideas.repec.org/h/eee/gamchp/1-04.html

Svoreňová M, Kwiatkowska M (2016) Quantitative verification and strategy synthesis for stochastic games. European Journal of Control 30:15–30

Vrieze O (1987) Stochastic Games with Finite State and Action Spaces. CWI Tract - Centrum voor Wiskunde en Informatica, Centrum voor Wiskunde en Informatica, URL https://books.google.fr/books?id=59tbPwAACAAJ

Vrieze O, Tijs SH, Raghavan TES, Filar JA (1983) A finite algorithm for the switching control stochastic game. Operations-Research-Spektrum 5(1):15–24. URL https://doi.org/10.1007/BF01720283