Transience in Countable MDPs

Stefan Kiefer
Department of Computer Science, University of Oxford, UK

Richard Mayr
School of Informatics, University of Edinburgh, UK

Mahsa Shirmohammadi
Université de Paris, CNRS, IRIF, F-75013 Paris, France

Patrick Totzke
Department of Computer Science, University of Liverpool, UK

Abstract

The `Transience` objective is not to visit any state infinitely often. While this is not possible in any finite Markov Decision Process (MDP), it can be satisfied in countably infinite ones, e.g., if the transition graph is acyclic.

We prove the following fundamental properties of `Transience` in countably infinite MDPs.

1. There exist uniformly ε-optimal MD strategies (memoryless deterministic) for `Transience`, even in infinitely branching MDPs.
2. Optimal strategies for `Transience` need not exist, even if the MDP is finitely branching. However, if an optimal strategy exists then there is also an optimal MD strategy.
3. If an MDP is universally transient (i.e., almost surely transient under all strategies) then many other objectives have a lower strategy complexity than in general MDPs. E.g., ε-optimal strategies for Safety and co-Büchi and optimal strategies for $\{0,1,2\}$-`Parity` (where they exist) can be chosen MD, even if the MDP is infinitely branching.

2012 ACM Subject Classification Theory of computation → Random walks and Markov chains; Mathematics of computing → Probability and statistics

Keywords and phrases Markov decision processes, Parity, Transience

Related Version This is the full version of a CONCUR 2021 paper [13].

1 Introduction

> Those who cannot remember the past are condemned to repeat it.
>
> *George Santayana (1905) [22]*

The famous aphorism above has often been cited (with small variations), e.g., by Winston Churchill in a 1948 speech to the House of Commons, and carved into several monuments all over the world [22].

We prove that the aphorism is false. In fact, even those who cannot remember anything at all are *not* condemned to repeat the past. With the right strategy they can avoid repeating the past equally well as everyone else. More formally, playing for `Transience` does not require any memory. We show that there always exist ε-optimal memoryless deterministic strategies for `Transience`, and if optimal strategies exist then there also exist optimal memoryless deterministic strategies.[1]

[1] Our result applies to MDPs (also called games against nature). It is an open question whether it generalizes to countable stochastic 2-player games. (However, it is easy to see that the adversary needs infinite memory in general, even if the player is passive [14, 16].)

Background. We study Markov decision processes (MDPs), a standard model for dynamic systems that exhibit both stochastic and controlled behavior [21]. MDPs play a prominent role in many domains, e.g., artificial intelligence and machine learning [26, 24], control theory [5, 1], operations research and finance [25, 12, 6, 23], and formal verification [2, 25, 11, 8, 3, 7].

An MDP is a directed graph where states are either random or controlled. Its observed behavior is described by runs, which are infinite paths that are, in part, determined by the choices of a controller. If the current state is random then the next state is chosen according to a fixed probability distribution. Otherwise, if the current state is controlled, the controller can choose a distribution over all possible successor states. By fixing a strategy for the controller (and initial state), one obtains a probability space of runs of the MDP. The goal of the controller is to optimize the expected value of some objective function on the runs.

The *strategy complexity* of a given objective characterizes the type of strategy necessary to achieve an optimal (resp. ε-optimal) value for the objective. General strategies can take the whole history of the run into account (history-dependent; (H)), while others use only bounded information about it (finite memory; (F)) or base decisions only on the current state (memoryless; (M)). Moreover, the strategy type depends on whether the controller can randomize (R) or is limited to deterministic choices (D). The simplest type, MD, refers to memoryless deterministic strategies.

Acyclicity and Transience. An MDP is called acyclic iff its transition graph is acyclic. While finite MDPs cannot be acyclic (unless they have deadlocks), countable MDPs can. In acyclic countable MDPs, the strategy complexity of Büchi/Parity objectives is lower than in the general case: ε-optimal strategies for Büchi/Parity objectives require only one bit of memory in acyclic MDPs, while they require infinite memory (an unbounded step-counter, plus one bit) in general countable MDPs [14, 15].

The concept of *transience* can be seen as a generalization of acyclicity. In a Markov chain, a state s is called *transient* iff the probability of returning from s to s is < 1 (otherwise the state is called recurrent). This means that a transient state is almost surely visited only finitely often. The concept of transient/recurrent is naturally lifted from Markov chains to MDPs, where they depend on the chosen strategy.

We define the `Transience` objective as the set of runs that do not visit any state infinitely often. We call an MDP *universally transient* iff it almost-surely satisfies `Transience` under every strategy. Thus every acyclic MDP is universally transient, but not vice-versa; cf. Figure 1. In particular, universal transience does not just depend on the structure of the transition graph, but also on the transition probabilities. Universally transient MDPs have interesting properties. Many objectives (e.g., Safety, Büchi, co-Büchi) have a lower strategy complexity than in general MDPs; see below.

We also study the strategy complexity of the `Transience` objective itself, and how it interacts with other objectives, e.g., how to attain a Büchi objective in a transient way.

Our contributions.

1. We show that there exist uniformly ε-optimal MD strategies (memoryless deterministic) for `Transience`, even in infinitely branching MDPs. This is unusual, since (apart from reachability objectives) most other objectives require infinite memory if the MDP is infinitely branching, e.g., all objectives generalizing Safety [17].
   Our result is shown in several steps. First we show that there exist ε-optimal deterministic 1-bit strategies for `Transience`. Then we show how to dispense with the 1-bit memory and obtain ε-optimal MD strategies for `Transience`. Finally, we make these MD strategies uniform, i.e., independent of the start state.
2. We show that optimal strategies for `Transience` need not exist, even if the MDP is

finitely branching. If they do exist then there are also MD optimal strategies. More generally, there exists a single MD strategy that is optimal from every state that allows optimal strategies for `Transience`.

3. If an MDP is universally transient (i.e., almost surely transient under all strategies) then many other objectives have a lower strategy complexity than in general MDPs, e.g., ε-optimal strategies for Safety and co-Büchi and optimal strategies for $\{0,1,2\}$-`Parity` (where they exist) can be chosen MD, even if the MDP is infinitely branching.

For our proofs we develop some technical results that are of independent interest. We generalize Ornstein's plastering construction [20] from reachability to tail objectives and thus obtain a general tool to infer uniformly ε-optimal MD strategies from non-uniform ones (cf. Theorem 7). Secondly, in Section 6 we develop the notion of the *conditioned MDP* (cf. [17]). For tail objectives, this allows to obtain uniformly ε-optimal MD strategies wrt. *multiplicative errors* from those with merely additive errors.

2 Preliminaries

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

Markov Decision Processes. We define Markov decision processes (MDPs for short) over countably infinite state spaces as tuples $\mathcal{M} = (S, S_\Box, S_\bigcirc, \longrightarrow, P)$ where S is the countable set of states partitioned into a set $S_\Box$ of *controlled states* and a set $S_\bigcirc$ of *random states*. The *transition relation* is $\longrightarrow \subseteq S \times S$, and $P : S_\bigcirc \to \mathcal{D}(S)$ is a *probability function*. We write $s \longrightarrow s'$ if $(s,s') \in \longrightarrow$, and refer to s' as a *successor* of s. We assume that every state has at least one successor. The probability function P assigns to each random state $s \in S_\bigcirc$ a probability distribution $P(s)$ over its set of successors. A *sink* is a subset $T \subseteq S$ closed under the $\longrightarrow$ relation.

An MDP is *acyclic* if the underlying graph $(S, \longrightarrow)$ is acyclic. It is *finitely branching* if every state has finitely many successors and *infinitely branching* otherwise. An MDP without controlled states ($S_\Box = \emptyset$) is a *Markov chain*.

Strategies and Probability Measures. A *run* ρ is an infinite sequence $s_0 s_1 \cdots$ of states such that $s_i \longrightarrow s_{i+1}$ for all $i \in \mathbb{N}$; a *partial run* is a finite prefix of a run. We write $\rho(i) = s_i$ and say that (partial) run $s_0 s_1 \cdots$ *visits* s if $s = s_i$ for some i. It *starts in* s if $s = s_0$.

A *strategy* is a function $\sigma : S^* S_\Box \to \mathcal{D}(S)$ that assigns to partial runs $\rho s \in S^* S_\Box$ a distribution over the successors of s. We write $\Sigma_\mathcal{M}$ for the set of all strategies in $\mathcal{M}$. A strategy σ and an initial state $s_0 \in S$ induce a standard probability measure on sets of infinite runs. We write $\mathcal{P}_{\mathcal{M},s_0,\sigma}(\mathfrak{R})$ for the probability of a measurable set $\mathfrak{R} \subseteq s_0 S^\omega$ of runs starting from s_0. It is defined for the cylinders $s_0 s_1 \dots s_n S^\omega \in S^\omega$ as $\mathcal{P}_{\mathcal{M},s_0,\sigma}(s_0 s_1 \dots s_n S^\omega) \stackrel{\text{def}}{=} \prod_{i=0}^{n-1} \bar{\sigma}(s_0 s_1 \dots s_i)(s_{i+1})$, where $\bar{\sigma}$ is the map that extends σ by $\bar{\sigma}(ws) = P(s)$ for all $ws \in S^* S_\bigcirc$. By Carathéodory's theorem [4], the measure for cylinders extends uniquely to a probability measure $\mathcal{P}_{\mathcal{M},s_0,\sigma}$ on all measurable subsets of $s_0 S^\omega$. We will write $\mathcal{E}_{\mathcal{M},s_0,\sigma}$ for the expectation w.r.t. $\mathcal{P}_{\mathcal{M},s_0,\sigma}$.

Strategy Classes. Strategies $\sigma : S^* S_\Box \to \mathcal{D}(S)$ are in general *randomized* (R) in the sense that they take values in $\mathcal{D}(S)$. A strategy σ is *deterministic* (D) if $\sigma(\rho)$ is a Dirac distribution for all partial runs $\rho \in S^* S_\Box$.

We formalize the amount of *memory* needed to implement strategies in Appendix A. The two classes of *memoryless* and *1-bit* strategies are central to this paper. A strategy σ is *memoryless* (M) if σ bases its decision only on the last state of the run: $\sigma(\rho s) = \sigma(\rho' s)$ for all $\rho, \rho' \in S^*$. We may view M-strategies as functions $\sigma : S_\Box \to \mathcal{D}(S)$. A 1-bit strategy σ may

base its decision also on a memory mode $\mathsf{m} \in \{0,1\}$. Formally, a 1-bit strategy σ is given as a tuple (u, m_0) where $\mathsf{m}_0 \in \{0,1\}$ is the initial memory mode and $u : \{0,1\} \times S \to \mathcal{D}(\{0,1\} \times S)$ is an update function such that

- for all controlled states $s \in S_\Box$, the distribution $u((\mathsf{m}, s))$ is over $\{0,1\} \times \{s' \mid s \longrightarrow s'\}$.
- for all random states $s \in S_\bigcirc$, we have that $\sum_{\mathsf{m}' \in \{0,1\}} u((\mathsf{m}, s))(\mathsf{m}', s') = P(s)(s')$.

Note that this definition allows for updating the memory mode upon visiting random states. We write $\sigma[\mathsf{m}_0]$ for the strategy obtained from σ by setting the initial memory mode to m_0.

MD strategies are both memoryless and deterministic; and *deterministic 1-bit* strategies are both deterministic and 1-bit.

Objectives. The objective of the controller is determined by a predicate on infinite runs. We assume familiarity with the syntax and semantics of the temporal logic LTL [9]. Formulas are interpreted on the underlying structure $(S, \longrightarrow)$ of the MDP $\mathcal{M}$. We use $[\![\varphi]\!]^{\mathcal{M},s} \subseteq sS^\omega$ to denote the set of runs starting from s that satisfy the LTL formula φ, which is a measurable set [27]. We also write $[\![\varphi]\!]^{\mathcal{M}}$ for $\bigcup_{s \in S} [\![\varphi]\!]^{\mathcal{M},s}$. Where it does not cause confusion we will identify φ and $[\![\varphi]\!]$ and just write $\mathcal{P}_{\mathcal{M},s,\sigma}(\varphi)$ instead of $\mathcal{P}_{\mathcal{M},s,\sigma}([\![\varphi]\!]^{\mathcal{M},s})$.

Given a set $T \subseteq S$ of states, the *reachability* objective $\texttt{Reach}(T) \stackrel{\text{def}}{=} \mathsf{F}T$ is the set of runs that visit T at least once. The *safety* objective $\texttt{Safety}(T) \stackrel{\text{def}}{=} \mathsf{G}\neg T$ is the set of runs that never visit T.

Let $\mathcal{C} \subseteq \mathbb{N}$ be a finite set of colors. A *color function* $Col : S \to \mathcal{C}$ assigns to each state s its color $Col(s)$. The parity objective, written as $\texttt{Parity}(Col)$, is the set of infinite runs such that the largest color that occurs infinitely often along the run is even. To define this formally, let $even(\mathcal{C}) = \{i \in \mathcal{C} \mid i \equiv 0 \mod 2\}$. For $\triangleright \in \{<, \leq, =, \geq, >\}$, $n \in \mathbb{N}$, and $Q \subseteq S$, let $[Q]^{Col \triangleright n} \stackrel{\text{def}}{=} \{s \in Q \mid Col(s) \triangleright n\}$ be the set of states in Q with color $\triangleright n$. Then

$$\texttt{Parity}(Col) \stackrel{\text{def}}{=} \bigvee_{i \in even(\mathcal{C})} \left(\mathsf{GF}[S]^{Col=i} \wedge \mathsf{FG}[S]^{Col \leq i}\right).$$

We write $\mathcal{C}$-$\texttt{Parity}$ for the parity objectives with the set of colors $\mathcal{C} \subseteq \mathbb{N}$. The classical Büchi and co-Büchi objectives correspond to $\{1,2\}$-$\texttt{Parity}$ and $\{0,1\}$-$\texttt{Parity}$, respectively.

An objective φ is called a *tail objective* (in $\mathcal{M}$) iff for every run $\rho'\rho$ with some finite prefix ρ' we have $\rho'\rho \in \varphi \Leftrightarrow \rho \in \varphi$. For every coloring Col, $\texttt{Parity}(Col)$ is tail. Reachability objectives are not always tail but in MDPs where the target set T is a sink $\texttt{Reach}(T)$ is tail.

Optimal and ε-optimal Strategies. Given an objective φ, the *value* of state s in an MDP $\mathcal{M}$, denoted by $\texttt{val}_{\mathcal{M},\varphi}(s)$, is the supremum probability of achieving φ. Formally, we have $\texttt{val}_{\mathcal{M},\varphi}(s) \stackrel{\text{def}}{=} \sup_{\sigma \in \Sigma} \mathcal{P}_{\mathcal{M},s,\sigma}(\varphi)$ where Σ is the set of all strategies. For $\varepsilon \geq 0$ and state $s \in S$, we say that a strategy is *ε-optimal* from s iff $\mathcal{P}_{\mathcal{M},s,\sigma}(\varphi) \geq \texttt{val}_{\mathcal{M},\varphi}(s) - \varepsilon$. A 0-optimal strategy is called *optimal*. An optimal strategy is *almost-surely winning* iff $\texttt{val}_{\mathcal{M},\varphi}(s) = 1$.

Considering an MD strategy as a function $\sigma : S_\Box \to S$ and $\varepsilon \geq 0$, σ is *uniformly* ε-optimal (resp. uniformly optimal) if it is ε-optimal (resp. optimal) from every $s \in S$.

Throughout the paper, we may drop the subscripts and superscripts from notations, if it is understood from the context. The missing proofs can be found in the appendix.

3 Transience and Universally Transient MDPs

In this section we define the transience property for MDPs, a natural generalization of the well-understood concept of transient Markov chains. We enumerate crucial characteristics of this objective and define the notion of universally transient MDPs.

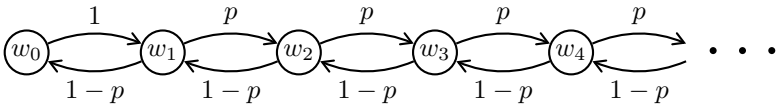

Figure 1 Gambler's Ruin with restart: The state w_i illustrates that the controller's wealth is i, and the coin tosses are in the controller's favor with probability p. For all i, $\mathcal{P}_{w_i}(\texttt{Transience}) = 0$ if $p \leq \frac{1}{2}$; and $\mathcal{P}_{w_i}(\texttt{Transience}) = 1$ otherwise.

Fix a countable MDP $\mathcal{M} = (S, S_\Box, S_\bigcirc, \longrightarrow, P)$. Define the transience objective, denoted by `Transience`, to be the set of runs that do not visit any state of $\mathcal{M}$ infinitely often, i.e.,

$$\texttt{Transience} \stackrel{\text{def}}{=} \bigwedge_{s \in S} \mathsf{FG}\, \neg s.$$

The `Transience` objective is tail, as it is closed under removing finite prefixes of runs. Also note that `Transience` cannot be encoded in a parity objective.

We call $\mathcal{M}$ *universally transient* iff for all states s_0, for all strategies σ, the `Transience` property holds almost-surely from s_0, i.e.,

$$\forall s_0 \in S \;\; \forall \sigma \in \Sigma \;\; \mathcal{P}_{\mathcal{M},s_0,\sigma}(\texttt{Transience}) = 1.$$

The MDP in Figure 1 models the classical Gambler's Ruin Problem with restart; see [10, Chapter 14]. It is well-known that if the controller starts with wealth i and if $p \leq \frac{1}{2}$, the probability of ruin (visiting the state w_0) is $\mathcal{P}_{w_i}(\mathsf{F}\, w_0) = 1$. Consequently, the probability of re-visiting w_0 infinitely often is 1, implying that $\mathcal{P}_{w_i}(\texttt{Transience}) = 0$. In contrast, for the case with $p > \frac{1}{2}$, for all states w_i, the probability of re-visiting w_i is strictly below 1. Hence, the `Transience` property holds almost-surely. This example indicates that the transience property depends on the probability values of the transitions and not just on the underlying transition graph, and thus may require arithmetic reasoning. In particular, the MDP in Figure 1 is universally transient iff $p > \frac{1}{2}$.

In general, optimal strategies for `Transience` need not exist:

▶ **Lemma 1.** *There exists a finitely branching countable MDP with initial state s_0 such that*
- $\texttt{val}_{\texttt{Transience}}(s) = 1$ *for all controlled states s,*
- *there does not exist any optimal strategy σ such that $\mathcal{P}_{s_0,\sigma}(\texttt{Transience}) = 1$.*

Proof. Consider a countable MDP $\mathcal{M}$ with set $S = \{\ell_i, \ell'_i, r_i, x_i \mid i \geq 1\} \cup \{\ell_0, \bot\}$ of states; see Figure 2. For all $i \geq 1$ the state x_{i+1} is the unique successor of x_i so that $(x_i)_{i \geq 1}$ form an acyclic ladder; the value of `Transience` is 1 for all x_i. The state $\bot$ is sink, and its value is 0. The states $(r_i)_{i \geq 1}$ are all random, and $r_i \xrightarrow{1-2^{-i}} x_i$ and $r_i \xrightarrow{2^{-i}} \bot$. Observe that the value of `Transience` is $1 - 2^{-i}$ for the r_i.

The states $(\ell_i)_{i \in \mathbb{N}}$ are controlled whereas the states $(\ell'_i)_{i \geq 1}$ are random. By interleaving of these states, we construct a "recurrent ladder" of decisions: $\ell_0 \to \ell_1$ and for all $i \geq 1$, state ℓ_i has two successors ℓ'_i and r_i. In random states ℓ'_i, as in Gambler's Ruin with a fair coin, the successors are ℓ_{i-1} or ℓ_{i+1}, each with equal probability. In each state $(\ell_i)_{i \geq 1}$, the controller decides to either stay on the ladder by going to ℓ'_i or leaves the ladder to r_i. As in Figure 1, if the controller stays on the ladder forever, the probability of `Transience` is 0.

Starting in ℓ_0, for all $i > 0$, strategy σ_i that stays on the ladder until visiting ℓ_i (which happens eventually almost surely) and then leaves the ladder to r_i achieves `Transience` with probability $1 - 2^i$. Hence, $\texttt{val}_{\texttt{Transience}}(\ell_0) = 1$.

Recall that transience cannot be achieved with a positive probability by staying on the acyclic ladder forever. But any strategy that leaves the ladder with a positive probability

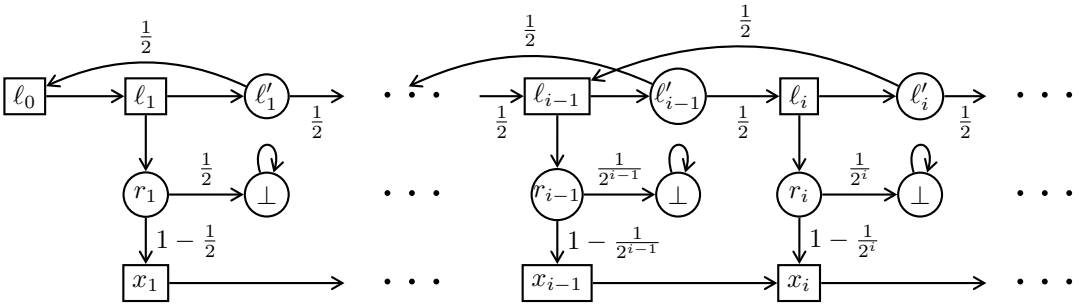

Figure 2 A partial illustration of the MDP in Lemma 1, in which there is no optimal strategy for `Transience`, starting from states ℓ_i. For readability, we have three copies of the state $\perp$. We call the ladder consisting of the interleaved controlled states ℓ_i and random states ℓ'_i a "recurrent ladder": if the controller stays on this ladder forever, it faithfully simulates a Gambler's Ruin with a fair coin, and the probability of `Transience` will be 0.

comes with a positive probability of falling into $\perp$, thus is not optimal either. Thus there is no optimal strategy for `Transience`. ◀

Reduction to Finitely Branching MDPs. In our main results, we will prove that for the `Transience` property there always exist ε-optimal MD strategies in finitely branching countable MDPs; and if an optimal strategy exists, there will exist an optimal MD strategy. We generalize these results to infinitely branching countable MDPs by the following reduction:

▶ **Lemma 2.** *Given an infinitely branching countable MDP $\mathcal{M}$ with an initial state s_0, there exists a finitely branching countable $\mathcal{M}'$ with a set S' of states such that $s_0 \in S'$ and*

1. *each strategy α_1 in $\mathcal{M}$ is mapped to a unique strategy β_1 in $\mathcal{M}'$ where*

$$\mathcal{P}_{s_0,\alpha_1}(\texttt{Transience}) = \mathcal{P}_{s_0,\beta_1}(\texttt{Transience}),$$

2. *and conversely, every MD strategy β_2 in $\mathcal{M}'$ is mapped to an MD strategy α_2 in $\mathcal{M}$ where*

$$\mathcal{P}_{s_0,\alpha_2}(\texttt{Transience}) \geq \mathcal{P}_{s_0,\beta_2}(\texttt{Transience}).$$

Proof sketch. See Appendix B for the complete construction. In order to construct $\mathcal{M}'$ from $\mathcal{M}$, for each controlled state $s \in S$ in $\mathcal{M}$ that has infinitely many successors $(s_i)_{i\geq 1}$, a "recurrent ladder" is introduced; see Figure 3. Since the probability of `Transience` is 0 for all those runs that eventually stay forever on a recurrent ladder, the controller should exit such ladders to play optimally for `Transience`. Infinitely branching random states can be dealt with in an easier way. ◀

Properties of Universally Transient MDPs.

Notice that acyclicity implies universal transience, but not vice-versa.

▶ **Lemma 3.** *For every countable MDP $\mathcal{M} = (S, S_\Box, S_\bigcirc, \longrightarrow, P)$, the following conditions are equivalent.*

1. *$\mathcal{M}$ is universally transient, i.e., $\forall s_0, \forall \sigma. \mathcal{P}_{\mathcal{M},s_0,\sigma}(\texttt{Transience}) = 1$.*
2. *For every initial state s_0 and state s, the objective of re-visiting s infinitely often has value zero, i.e., $\forall s_0, s \sup_\sigma \mathcal{P}_{\mathcal{M},s_0,\sigma}(\mathsf{GF}(s)) = 0$.*
3. *For every state s the value of the objective to re-visit s is strictly below 1, i.e., $Re(s) \stackrel{def}{=} \sup_\sigma \mathcal{P}_{\mathcal{M},s,\sigma}(\mathsf{XF}(s)) < 1$.*

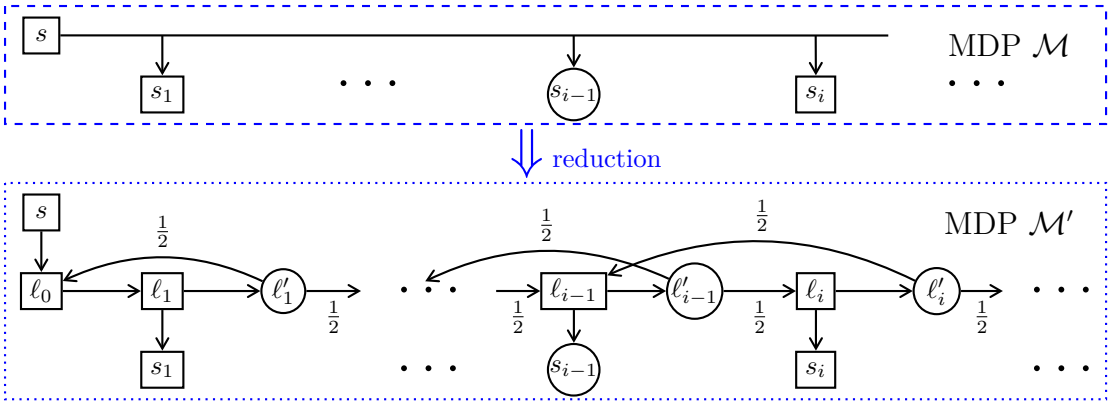

Figure 3 A partial illustration of the reduction in Lemma 2.

4. *For every state s there exists a finite bound $B(s)$ such that for every state s_0 and strategy σ from s_0 the expected number of visits to s is $\leq B(s)$.*
5. *For all states s_0, s, under every strategy σ from s_0 the expected number of visits to s is finite.*

Proof. Towards (1) $\Rightarrow$ (2), consider an arbitrary strategy σ from the initial state s_0 and some state s. By (1) we have $\forall\sigma.\mathcal{P}_{\mathcal{M},s_0,\sigma}(\texttt{Transience}) = 1$ and thus $0 = \mathcal{P}_{\mathcal{M},s_0,\sigma}(\neg\texttt{Transience}) = \mathcal{P}_{\mathcal{M},s_0,\sigma}(\bigcup_{s'\in S}\mathsf{GF}(s')) \geq \mathcal{P}_{\mathcal{M},s_0,\sigma}(\mathsf{GF}(s))$ which implies (2).

Towards (2) $\Rightarrow$ (1), consider an arbitrary strategy σ from the initial state s_0. By (2) we have $0 = \sum_{s\in S}\mathcal{P}_{\mathcal{M},s_0,\sigma}(\mathsf{GF}(s)) \geq \mathcal{P}_{\mathcal{M},s_0,\sigma}(\bigcup_{s\in S}\mathsf{GF}(s)) = \mathcal{P}_{\mathcal{M},s_0,\sigma}(\neg\texttt{Transience})$ and thus $\mathcal{P}_{\mathcal{M},s_0,\sigma}(\texttt{Transience}) = 1$.

We now show the implications (2) $\Rightarrow$ (3) $\Rightarrow$ (4) $\Rightarrow$ (5) $\Rightarrow$ (2).

Towards $\neg(3) \Rightarrow \neg(2)$, $\neg(3)$ implies $\exists s.Re(s) = 1$ and thus $\forall\varepsilon > 0.\exists\sigma_\varepsilon\, \mathcal{P}_{\mathcal{M},s,\sigma_\varepsilon}(\mathsf{XF}(s)) \geq 1-\varepsilon$. Let $\varepsilon_i \stackrel{\text{def}}{=} 2^{-(i+1)}$. We define the strategy σ to play like σ_{ε_i} between the i-th and $(i+1)$th visit to s. Since $\sum_{i=1}^{\infty}\varepsilon_i < \infty$, we have $\prod_{i=1}^{\infty}(1-\varepsilon_i) > 0$. Therefore $\mathcal{P}_{\mathcal{M},s,\sigma}(\mathsf{GF}(s)) \geq \prod_{i=1}^{\infty}(1-\varepsilon_i) > 0$, which implies $\neg(2)$, where $s_0 = s$.

Towards (3) $\Rightarrow$ (4), regardless of s_0 and the chosen strategy, the expected number of visits to s is upper-bounded by $B(s) \stackrel{\text{def}}{=} \sum_{n=0}^{\infty}(n+1)\cdot(Re(s))^n < \infty$.

The implication (4) $\Rightarrow$ (5) holds trivially.

Towards $\neg(2) \Rightarrow \neg(5)$, by $\neg(2)$ there exist states s_0, s and a strategy σ such that $\mathcal{P}_{\mathcal{M},s_0,\sigma}(\mathsf{GF}(s)) > 0$. Thus the expected number of visits to s is infinite, which implies $\neg(5)$. ◀

We remark that if an MDP is *not* universally transient (unlike in Lemma 3(5)), for a strategy σ, the expected number of visits to some state can be infinite, even if σ attains $\texttt{Transience}$ almost surely.

Consider the MDP $\mathcal{M}$ with controlled states $\{s_0, s_1, \dots\}$, initial state s_0 and transitions $s_0 \to s_0$ and $s_k \to s_{k+1}$ for every $k \geq 0$. We define a strategy σ that, while in state s_0, proceeds in rounds $i = 1, 2, \dots$. In the i-th round it tosses a fair coin. If Heads then it goes to s_1. If Tails then it loops around s_0 exactly 2^i times and then goes to round $i+1$. In every round the probability of going to s_1 is $1/2$ and therefore the probability of staying in s_0 forever is $(1/2)^\infty = 0$. Thus $\mathcal{P}_{\mathcal{M},s_0,\sigma}(\texttt{Transience}) = 1$. However, the expected number of visits to s_0 is $\geq \sum_{i=1}^{\infty}\left(\frac{1}{2}\right)^i \cdot 2^i = \infty$.

4 MD Strategies for Transience

We show that there exist uniformly ε-optimal MD strategies for $\texttt{Transience}$ and that optimal strategies, where they exist, can also be chosen MD.

First we show that there exist ε-optimal deterministic 1-bit strategies for $\texttt{Transience}$ (in Corollary 5) and then we show how to dispense with the 1-bit memory (in Lemma 6).

It was shown in [14] that there exist ε-optimal deterministic 1-bit strategies for Büchi objectives in *acyclic* countable MDPs (though not in general MDPs). These 1-bit strategies will be similar to the 1-bit strategies for $\texttt{Transience}$ that we aim for in (not necessarily acyclic) countable MDPs. In Lemma 4 below we first strengthen the result from [14] and construct ε-optimal deterministic 1-bit strategies for objectives $\texttt{Büchi}(F) \cap \texttt{Transience}$. From this we obtain deterministic 1-bit strategies for $\texttt{Transience}$ (Corollary 5).

▶ **Lemma 4.** *Let $\mathcal{M}$ be a countable MDP, I a finite set of initial states, F a set of states and $\varepsilon > 0$. Then there exists a deterministic 1-bit strategy for $\texttt{Büchi}(F) \cap \texttt{Transience}$ that is ε-optimal from every $s \in I$.*

Proof sketch. The full proof can be found in Appendix C. It follows the proof of [14, Theorem 5], which considers $\texttt{Büchi}(F)$ conditions for *acyclic* (and hence universally transient) MDPs. The only part of that proof that requires modification is [14, Lemma 10], which is replaced here by Lemma 18 to deal with general MDPs.

In short, from every $s \in I$ there exists an ε-optimal strategy σ_s for $\varphi \stackrel{\text{def}}{=} \texttt{Büchi}(F) \cap \texttt{Transience}$. We observe the behavior of the finitely many σ_s for $s \in I$ on an infinite, increasing sequence of finite subsets of S. Based on Lemma 18, we can define a second stronger objective $\varphi' \subseteq \varphi$ and show $\forall_{s \in I} \mathcal{P}_{\mathcal{M},s,\sigma_s}(\varphi') \geq \texttt{val}_{\mathcal{M},\varphi}(s) - 2\varepsilon$. We then construct a deterministic 1-bit strategy σ' that is optimal for φ' from all $s \in I$ and thus 2ε-optimal for φ. Since ε can be chosen arbitrarily small, the result follows. ◀

Unlike for the $\texttt{Transience}$ objective alone (see below), the 1-bit memory is strictly necessary for the $\texttt{Büchi}(F) \cap \texttt{Transience}$ objective in Lemma 4. The 1-bit lower bound for $\texttt{Büchi}(F)$ objectives in [14] holds even for acyclic MDPs where $\texttt{Transience}$ is trivially true.

▶ **Corollary 5.** *Let $\mathcal{M}$ be a countable MDP, I a finite set of initial states, F a set of states and $\varepsilon > 0$.*

1. *If $\forall s \in I\ \texttt{val}_{\mathcal{M},\texttt{Büchi}(F)}(s) = \texttt{val}_{\mathcal{M},\texttt{Büchi}(F)\cap\texttt{Transience}}(s)$ then there exists a deterministic 1-bit strategy for $\texttt{Büchi}(F)$ that is ε-optimal from every $s \in I$.*
2. *If $\mathcal{M}$ is universally transient then there exists a deterministic 1-bit strategy for $\texttt{Büchi}(F)$ that is ε-optimal from every $s \in I$.*
3. *There exists a deterministic 1-bit strategy for $\texttt{Transience}$ that is ε-optimal from every $s \in I$.*

Proof. Towards (1), since $\forall s \in I\ \texttt{val}_{\mathcal{M},\texttt{Büchi}(F)}(s) = \texttt{val}_{\mathcal{M},\texttt{Büchi}(F)\cap\texttt{Transience}}(s)$, strategies that are ε-optimal for $\texttt{Büchi}(F) \cap \texttt{Transience}$ are also ε-optimal for $\texttt{Büchi}(F)$. Thus the result follows from Lemma 4.

Item (2) follows directly from (1), since the precondition always holds in universally transient MDPs.

Towards (3), let $F \stackrel{\text{def}}{=} S$. Then we have $\texttt{Büchi}(F) \cap \texttt{Transience} = \texttt{Transience}$ and we obtain from Lemma 4 that there exists a deterministic 1-bit strategy for $\texttt{Transience}$ that is ε-optimal from every $s \in I$. ◀

Note that every acyclic MDP is universally transient and thus Corollary 5(2) implies the upper bound on the strategy complexity of $\texttt{Büchi}(F)$ from [14] (but not vice-versa).

In the next step we show how to dispense with the 1-bit memory and obtain non-uniform ε-optimal MD strategies for $\texttt{Transience}$.

▶ **Lemma 6.** *Let $\mathcal{M} = (S, S_\Box, S_\bigcirc, \longrightarrow, P)$ be a countable MDP with initial state s_0, and $\varepsilon > 0$. There exists an MD strategy σ that is ε-optimal for $\texttt{Transience}$ from s_0, i.e., $\mathcal{P}_{\mathcal{M},s_0,\sigma}(\texttt{Transience}) \geq \texttt{val}_{\mathcal{M},\texttt{Transience}}(s_0) - \varepsilon$.*

Proof. By Lemma 2 it suffices to prove the property for finitely branching MDPs. Thus without restriction in the rest of the proof we assume that $\mathcal{M}$ is finitely branching.

Let $\varepsilon' \stackrel{\text{def}}{=} \varepsilon/2$. We instantiate Corollary 5(3) with $I \stackrel{\text{def}}{=} \{s_0\}$ and obtain that there exists an ε'-optimal deterministic 1-bit strategy $\hat{\sigma}$ for $\texttt{Transience}$ from s_0.

We now construct a slightly modified MDP $\mathcal{M}'$ as follows. Let $S_{bad} \subseteq S$ be the subset of states where $\hat{\sigma}$ attains zero for $\texttt{Transience}$ in *both* memory modes, i.e., $S_{bad} \stackrel{\text{def}}{=} \{s \in S \mid \mathcal{P}_{\mathcal{M},s,\sigma[0]}(\texttt{Transience}) = \mathcal{P}_{\mathcal{M},s,\sigma[1]}(\texttt{Transience}) = 0\}$. Let $S_{good} \stackrel{\text{def}}{=} S \setminus S_{bad}$. We obtain $\mathcal{M}'$ from $\mathcal{M}$ by making all states in S_{bad} losing sinks (for $\texttt{Transience}$), by deleting all outgoing edges and adding a self-loop instead. It follows that

$$\mathcal{P}_{\mathcal{M},s_0,\hat{\sigma}}(\texttt{Transience}) = \mathcal{P}_{\mathcal{M}',s_0,\hat{\sigma}}(\texttt{Transience}) \tag{1}$$

$$\forall \sigma.\ \mathcal{P}_{\mathcal{M},s_0,\sigma}(\texttt{Transience}) \geq \mathcal{P}_{\mathcal{M}',s_0,\sigma}(\texttt{Transience}) \tag{2}$$

In the following we show that it is possible to play in such a way that, for every $s \in S_{good}$, the expected number of visits to s is *finite*. We obtain the deterministic 1-bit strategy σ' in $\mathcal{M}'$ by modifying $\hat{\sigma}$ as follows. In every state s and memory mode $x \in \{0,1\}$ where $\hat{\sigma}[x]$ attains 0 for $\texttt{Transience}$ and $\hat{\sigma}[1-x]$ attains > 0 the strategy σ' sets the memory bit to $1-x$. (Note that only states $s \in S_{good}$ can be affected by this change.) It follows that

$$\forall s \in S.\ \mathcal{P}_{\mathcal{M}',s,\sigma'}(\texttt{Transience}) \geq \mathcal{P}_{\mathcal{M}',s,\hat{\sigma}}(\texttt{Transience}) \tag{3}$$

Moreover, from all states in S_{good} in $\mathcal{M}'$ the strategy σ' attains a strictly positive probability of $\texttt{Transience}$ in *both* memory modes, i.e., for all $s \in S_{good}$ we have

$$t(s,\sigma') \stackrel{\text{def}}{=} \min_{x \in \{0,1\}} \mathcal{P}_{\mathcal{M}',s,\sigma'[i]}(\texttt{Transience}) > 0.$$

Let $r(s,\sigma',x)$ be the probability, when playing $\sigma'[x]$ from state s, of reaching s again in the *same* memory mode x. For every $s \in S_{good}$ we have $r(s,\sigma',x) < 1$, since $t(s,\sigma') > 0$.

Let $R(s)$ be the expected number of visits to state s when playing σ' from s_0 in $\mathcal{M}'$, and $R_x(s)$ the expected number of visits to s in memory mode $x \in \{0,1\}$. For all $s \in S_{good}$ we have that

$$R(s) = R_0(s) + R_1(s) \leq \sum_{n=1}^{\infty} n \cdot r(s,\sigma',0)^{n-1} + \sum_{n=1}^{\infty} n \cdot r(s,\sigma',1)^{n-1} < \infty \tag{4}$$

where the first equality holds by linearity of expectations. Thus the expected number of visits to s is *finite*.

Now we upper-bound the probability of visiting S_{bad}. We have $\mathcal{P}_{\mathcal{M}',s_0,\sigma'}(\texttt{Transience}) \geq \mathcal{P}_{\mathcal{M}',s_0,\hat{\sigma}}(\texttt{Transience}) = \mathcal{P}_{\mathcal{M},s_0,\hat{\sigma}}(\texttt{Transience}) \geq \texttt{val}_{\mathcal{M},\texttt{Transience}}(s_0) - \varepsilon'$ by (3), (1) and the ε'-optimality of $\hat{\sigma}$. Since states in S_{bad} are losing sinks in $\mathcal{M}'$, it follows that

$$\mathcal{P}_{\mathcal{M}',s_0,\sigma'}(\mathsf{F} S_{bad}) \leq 1 - \mathcal{P}_{\mathcal{M}',s_0,\sigma'}(\texttt{Transience}) \leq 1 - \texttt{val}_{\mathcal{M},\texttt{Transience}}(s_0) + \varepsilon' \tag{5}$$

We now augment the MDP $\mathcal{M}'$ by assigning costs to transitions as follows. Let $i : S \to \mathbb{N}$ be an enumeration of the state space, i.e., a bijection. Let $S'_{good} \stackrel{\text{def}}{=} \{s \in S_{good} \mid R(s) > 0\}$ be the subset of states in S_{good} that are visited with non-zero probability when playing σ' from s_0. Each transition $s' \to s$ is assigned a cost:

- If $s' \in S_{bad}$ then $s \in S_{bad}$ by def. of $\mathcal{M}'$. We assign cost 0.
- If $s' \in S_{good}$ and $s \in S_{bad}$ we assign cost $K/(1 - \mathtt{val}_{\mathcal{M},\mathtt{Transience}}(s_0) + \varepsilon')$ for $K \stackrel{\text{def}}{=} (1 + \varepsilon')/\varepsilon'$.
- If $s' \in S_{good}$ and $s \in S'_{good}$ we assign cost $2^{-i(s)}/R(s)$. This is well defined, since $R(s) > 0$.
- $s' \in S_{good}$ and $s \in S_{good} \setminus S'_{good}$ we assign cost 1.

Note that all transitions leading to states in S_{good} are assigned a non-zero cost, since $R(s)$ is finite by (4).

When playing σ' from s_0 in $\mathcal{M}'$, the expected total cost is upper-bounded by

$$\mathcal{P}_{\mathcal{M}',s_0,\sigma'}(\mathsf{F}S_{bad}) \cdot K/(1 - \mathtt{val}_{\mathcal{M},\mathtt{Transience}}(s_0) + \varepsilon') + \sum_{s \in S'_{good}} R(s) \cdot 2^{-i(s)}/R(s)$$

The first part is $\leq K$ by (5) and the second part is ≤ 1, since $R(s) < \infty$ by (4). Therefore the expected total cost is $\leq K + 1$, i.e., σ' witnesses that it is possible to attain a finite expected cost that is upper-bounded by $K + 1$.

Now we define our MD strategy σ. Let σ be an optimal MD strategy on $\mathcal{M}'$ (from s_0) that minimizes the expected cost. It exists, as a finite expected cost is attainable and $\mathcal{M}'$ is finitely branching; see [21, Theorem 7.3.6].

We now show that σ attains $\mathtt{Transience}$ with high probability in $\mathcal{M}'$ (and in $\mathcal{M}$). Since σ is cost-optimal, its attained cost from s_0 is upper-bounded by that of σ', i.e., $\leq K + 1$. Since the cost of entering S_{bad} is $K/(1 - \mathtt{val}_{\mathcal{M},\mathtt{Transience}}(s_0) + \varepsilon')$, we have $\mathcal{P}_{\mathcal{M}',s_0,\sigma}(\mathsf{F}S_{bad}) \cdot K/(1 - \mathtt{val}_{\mathcal{M},\mathtt{Transience}}(s_0) + \varepsilon') \leq K + 1$ and thus

$$\mathcal{P}_{\mathcal{M}',s_0,\sigma}(\mathsf{F}S_{bad}) \leq \frac{K+1}{K}(1 - \mathtt{val}_{\mathcal{M},\mathtt{Transience}}(s_0) + \varepsilon') \tag{6}$$

For every state $s \in S_{good}$, all transitions into s have the same fixed non-zero cost. Thus every run that visits some state $s \in S_{good}$ infinitely often has infinite cost. Since the expected cost of playing σ from s_0 is $\leq K + 1$, such runs must be a null-set, i.e.,

$$\mathcal{P}_{\mathcal{M}',s_0,\sigma}(\neg\mathtt{Transience} \wedge \mathsf{G}S_{good}) = 0 \tag{7}$$

Thus

$$\begin{aligned}
&\mathcal{P}_{\mathcal{M},s_0,\sigma}(\mathtt{Transience}) \\
&\geq \mathcal{P}_{\mathcal{M}',s_0,\sigma}(\mathtt{Transience}) && \text{by (2)} \\
&= 1 - \mathcal{P}_{\mathcal{M}',s_0,\sigma}(\mathsf{F}S_{bad}) && \text{by (7)} \\
&\geq 1 - \frac{K+1}{K}(1 - \mathtt{val}_{\mathcal{M},\mathtt{Transience}}(s_0) + \varepsilon') && \text{by (6)} \\
&= \mathtt{val}_{\mathcal{M},\mathtt{Transience}}(s_0) - \varepsilon' - (1/K)(1 - \mathtt{val}_{\mathcal{M},\mathtt{Transience}}(s_0) + \varepsilon') \\
&\geq \mathtt{val}_{\mathcal{M},\mathtt{Transience}}(s_0) - \varepsilon' - (1/K)(1 + \varepsilon') \\
&= \mathtt{val}_{\mathcal{M},\mathtt{Transience}}(s_0) - 2\varepsilon' && \text{def. of } K \\
&= \mathtt{val}_{\mathcal{M},\mathtt{Transience}}(s_0) - \varepsilon && \text{def. of } \varepsilon'
\end{aligned}$$

◀

Now we lift the result of Lemma 6 from non-uniform to uniform strategies (and to optimal strategies) and obtain the following theorem. The proof is a generalization of a "plastering" construction by Ornstein [20] (see also [16]) from reachability to tail objectives, which works by fixing MD strategies on ever expanding subsets of the state space.

▶ **Theorem 7.** *Let $\mathcal{M} = (S, S_\Box, S_\bigcirc, \longrightarrow, P)$ be a countable MDP, and let φ be an objective that is tail in $\mathcal{M}$. Suppose for every $s \in S$ there exist ε-optimal MD strategies for φ. Then:*

1. *There exist uniform ε-optimal MD strategies for φ.*
2. *There exists a single MD strategy that is optimal from every state that has an optimal strategy.*

▶ **Theorem 8.** *In every countable MDP there exist uniform ε-optimal MD strategies for* `Transience`*. Moreover, there exists a single MD strategy that is optimal for* `Transience` *from every state that has an optimal strategy.*

Proof. Immediate from Lemma 6 and Theorem 7, since `Transience` is a tail objective. ◀

5 Strategy Complexity in Universally Transient MDPs

The strategy complexity of parity objectives in general MDPs is known [15]. Here we show that some parity objectives have a lower strategy complexity in universally transient MDPs. It is known [14] that there are acyclic (and hence universally transient) MDPs where ε-optimal strategies for $\{1,2\}$-`Parity` (and optimal strategies for $\{1,2,3\}$-`Parity`, resp.) require 1 bit.

We show that, for all simpler parity objectives in the Mostowski hierarchy [19], universally transient MDPs admit uniformly (ε-)optimal MD strategies (unlike general MDPs [15]). These results (Theorems 10 and 11) ultimately rely on the existence of uniformly ε-optimal strategies for safety objectives. While such strategies always exist for finitely branching MDPs – simply pick a value-maximal successor – this is not the case for infinitely branching MDPs [17]. However, we show that universal transience implies the existence of uniformly ε-optimal strategies for safety objectives even for *infinitely branching* MDPs.

▶ **Theorem 9.** *For every universally transient countable MDP, safety objective and $\varepsilon > 0$ there exists a uniformly ϵ-optimal MD strategy.*

Proof. Let $\mathcal{M} = (S, S_\Box, S_\bigcirc, \longrightarrow, P)$ be a universally transient MDP and $\varepsilon > 0$. Assume w.l.o.g. that the target $T \subseteq S$ of the objective $\varphi = \texttt{Safety}(T)$ is a (losing) sink and let $\iota : S \to \mathbb{N}$ be an enumeration of the state space S.

By Lemma 3(3), for every state s we have $Re(s) \stackrel{\text{def}}{=} \sup_\sigma \mathcal{P}_{\mathcal{M},s,\sigma}(\mathsf{XF}(s)) < 1$ and thus $R(s) \stackrel{\text{def}}{=} \sum_{i=0}^{\infty} Re(s)^i < \infty$. This means that, independent of the chosen strategy, $Re(s)$ upper-bounds the chance to return to s, and $R(s)$ bounds the expected number of visits to s.

Suppose that σ is an MD strategy which, at any state $s \in S_\Box$, picks a successor s' with

$$\texttt{val}(s') \quad \geq \quad \texttt{val}(s) - \frac{\varepsilon}{2^{\iota(s)+1} \cdot R(s)}.$$

This is possible even if $\mathcal{M}$ is infinitely branching, by the definition of value and the fact that $R(s) < \infty$. We show that $\mathcal{P}_{\mathcal{M},s_0,\sigma}(\texttt{Safety}(T)) \geq \texttt{val}(s_0) - \varepsilon$ holds for every initial state s_0, which implies the claim of the theorem.

Towards this, we define a function `cost` that labels each transition in the MDP with a real-valued cost: For every controlled transition $s \longrightarrow s'$ let $\texttt{cost}((s,s')) \stackrel{\text{def}}{=} \texttt{val}(s) - \texttt{val}(s') \geq 0$. Random transitions have cost zero. We will argue that when playing σ from any start state

s_0, its attainment w.r.t. the objective $\texttt{Safety}(T)$ equals the value of s_0 minus the expected total cost, and that this cost is bounded by ε.

For any $i \in \mathbb{N}$ let us write s_i for the random variable denoting the state just after step i, and $\texttt{Cost}(i) \stackrel{\text{def}}{=} \texttt{cost}(s_i, s_{i+1})$ for the cost of step i in a random run. We observe that under σ the expected total cost is bounded in the limit, i.e.,

$$\lim_{n\to\infty} \mathcal{E}\left(\sum_{i=0}^{n-1} \texttt{Cost}(i)\right) \leq \varepsilon. \tag{8}$$

We moreover note that for every n,

$$\mathcal{E}(\texttt{val}(s_n)) = \mathcal{E}(\texttt{val}(s_0)) - \mathcal{E}\left(\sum_{i=0}^{n-1} \texttt{Cost}(i)\right). \tag{9}$$

Full proofs of the above two equations can be found in Appendix E. Together they imply

$$\liminf_{n\to\infty} \mathcal{E}(\texttt{val}(s_n)) = \texttt{val}(s_0) - \lim_{n\to\infty} \mathcal{E}\left(\sum_{i=0}^{n-1} \texttt{cost}(i)\right) \geq \texttt{val}(s_0) - \varepsilon. \tag{10}$$

Finally, to show the claim let $[s_n \notin T] : S^\omega \to \{0,1\}$ be the random variable that indicates that the n-th state is not in the target set T. Note that $[s_n \notin T] \geq \texttt{val}(s_n)$ because target states have value 0. We have:

$$\begin{aligned}
\mathcal{P}_{\mathcal{M},s_0,\sigma}(\texttt{Safety}(T)) &= \mathcal{P}_{\mathcal{M},s_0,\sigma}\left(\bigwedge_{i=0}^{\infty} \mathsf{X}^i \neg T\right) && \text{semantics of } \texttt{Safety}(T) = \mathsf{G}\neg T \\
&= \lim_{n\to\infty} \mathcal{P}_{\mathcal{M},s_0,\sigma}\left(\bigwedge_{i=0}^{n} \mathsf{X}^i \neg T\right) && \text{continuity of measures} \\
&= \lim_{n\to\infty} \mathcal{P}_{\mathcal{M},s_0,\sigma}(\mathsf{X}^n \neg T) && T \text{ is a sink} \\
&= \lim_{n\to\infty} \mathcal{E}([s_n \notin T]) && \text{definition of } [s_n \notin T] \\
&\geq \liminf_{n\to\infty} \mathcal{E}(\texttt{val}(s_n)) && \text{as } [s_n \notin T] \geq \texttt{val}(s_n) \\
&\geq \texttt{val}(s_0) - \varepsilon && \text{Equation (10).}
\end{aligned}$$

◀

We can now combine Theorem 9 with the results from [15] to show the existence of MD strategies assuming universal transience.

▶ **Theorem 10.** *For universally transient MDPs optimal strategies for* $\{0,1,2\}$-$\texttt{Parity}$*, where they exist, can be chosen uniformly MD.*

Formally, let $\mathcal{M}$ *be a universally transient MDP with states* S*,* $Col : S \to \{0,1,2\}$*, and* $\varphi = \texttt{Parity}(Col)$*. There exists an MD strategy* σ' *that is optimal for all states* s *that have an optimal strategy:* $\left(\exists \sigma \in \Sigma.\, \mathcal{P}_{\mathcal{M},s,\sigma}(\varphi) = \texttt{val}_{\mathcal{M}}(s)\right) \implies \mathcal{P}_{\mathcal{M},s,\sigma'}(\varphi) = \texttt{val}_{\mathcal{M}}(s)$.

Proof. Let $\mathcal{M}_+$ be the conditioned version of $\mathcal{M}$ w.r.t. φ (see [15, Def. 19] for a precise definition). By Lemma 17, $\mathcal{M}_+$ is still a universally transient MDP and therefore by Theorem 9, there exist uniformly ε-optimal MD strategies for every safety objective and every $\varepsilon > 0$. The claim now follows from [15, Theorem 22]. ◀

▶ **Theorem 11.** *For every universally transient countable MDP* $\mathcal{M}$*, co-Büchi objective and* $\varepsilon > 0$ *there exists a uniformly* ε*-optimal MD strategy.*

Formally, let $\mathcal{M}$ *be a universally transient countable MDP with states* S*,* $Col : S \to \{0,1\}$ *be a coloring,* $\varphi = \texttt{Parity}(Col)$ *and* $\varepsilon > 0$*.*

There exists an MD strategy σ' *s.t. for every state* s*,* $\mathcal{P}_{\mathcal{M},s,\sigma'}(\varphi) \geq \texttt{val}_{\mathcal{M}}(s) - \varepsilon$*.*

Proof. This directly follows from Theorem 9 and [15, Theorem 25]. ◀

6 The Conditioned MDP

Given an MDP $\mathcal{M}$ and an objective φ that is tail in $\mathcal{M}$, a construction of a *conditioned* MDP $\mathcal{M}_+$ was provided in [17, Lemma 6] that, very loosely speaking, "scales up" the probability of φ so that any strategy σ is optimal in $\mathcal{M}$ if it is almost surely winning in $\mathcal{M}_+$. For certain tail objectives, this construction was used in [17] to reduce the sufficiency of MD strategies for *optimal* strategies to the sufficiency of MD strategies for *almost surely winning* strategies, which is a special case that may be easier to handle.

However, the construction was restricted to states that *have* an optimal strategy. In fact, states in $\mathcal{M}$ that do not have an optimal strategy do not appear in $\mathcal{M}_+$. In the following, we lift this restriction by constructing a more general version of the conditioned MDP, called $\mathcal{M}_*$. The MDP $\mathcal{M}_*$ will contain all states from $\mathcal{M}$ that have a positive value w.r.t. φ in $\mathcal{M}$. Moreover, all these states will have value 1 in $\mathcal{M}_*$. It will then follow from Lemma 13(3) below that an ε-optimal strategy in $\mathcal{M}_*$ is $\varepsilon\mathtt{val}_{\mathcal{M}}(s_0)$-optimal in $\mathcal{M}$. This allows us to reduce the sufficiency of MD strategies for ε-optimal strategies to the sufficiency of MD strategies for ε-optimal strategies for states with value 1. In fact, it also follows that if an MD strategy σ is uniform ε-optimal in $\mathcal{M}_*$, it is *multiplicatively* uniform ε-optimal in $\mathcal{M}$, i.e., $\mathcal{P}_{\mathcal{M},s,\sigma}(\varphi) \geq (1-\varepsilon)\cdot\mathtt{val}_{\mathcal{M}}(s)$ holds for all states s.

▶ **Definition 12.** *For an MDP $\mathcal{M} = (S, S_\Box, S_\bigcirc, \longrightarrow, P)$ and an objective φ that is tail in $\mathcal{M}$, define the* conditioned version *of $\mathcal{M}$ w.r.t. φ to be the MDP $\mathcal{M}_* = (S_*, S_{*\Box}, S_{*\bigcirc}, \longrightarrow_*, P_*)$ with*

$$
\begin{aligned}
S_{*\Box} &= \{s \in S_\Box \mid \mathtt{val}_{\mathcal{M}}(s) > 0\}\\
S_{*\bigcirc} &= \{s \in S_\bigcirc \mid \mathtt{val}_{\mathcal{M}}(s) > 0\} \cup \{s_\bot\} \cup \{(s,t) \in \longrightarrow \mid s \in S_\Box,\ \mathtt{val}_{\mathcal{M}}(s) > 0\}\\
\longrightarrow_* &= \{(s,(s,t)) \in (S_\Box \times \longrightarrow) \mid \mathtt{val}_{\mathcal{M}}(s) > 0,\ s \longrightarrow t\} \cup\\
&\quad \{(s,t) \in S_\bigcirc \times S \mid \mathtt{val}_{\mathcal{M}}(s) > 0,\ \mathtt{val}_{\mathcal{M}}(t) > 0\} \cup\\
&\quad \{((s,t),t) \in (\longrightarrow \times S) \mid \mathtt{val}_{\mathcal{M}}(s) > 0,\ \mathtt{val}_{\mathcal{M}}(t) > 0\} \cup\\
&\quad \{((s,t),s_\bot) \in (\longrightarrow \times \{s_\bot\}) \mid \mathtt{val}_{\mathcal{M}}(s) > \mathtt{val}_{\mathcal{M}}(t)\} \cup\\
&\quad \{(s_\bot, s_\bot)\}
\end{aligned}
$$

$$
\begin{aligned}
P_*(s,t) &= P(s,t)\cdot\frac{\mathtt{val}_{\mathcal{M}}(t)}{\mathtt{val}_{\mathcal{M}}(s)} & P_*((s,t),t) &= \frac{\mathtt{val}_{\mathcal{M}}(t)}{\mathtt{val}_{\mathcal{M}}(s)}\\
P_*((s,t),s_\bot) &= 1 - \frac{\mathtt{val}_{\mathcal{M}}(t)}{\mathtt{val}_{\mathcal{M}}(s)} & P_*(s_\bot,s_\bot) &= 1
\end{aligned}
$$

for a fresh state $s_\bot$.

The conditioned MDP is well-defined. Indeed, as φ is tail in $\mathcal{M}$, for any $s \in S_\bigcirc$ we have $\mathtt{val}_{\mathcal{M}}(s) = \sum_{s\longrightarrow t} P(s,t)\mathtt{val}_{\mathcal{M}}(t)$, and so if $\mathtt{val}_{\mathcal{M}}(s) > 0$ then $\sum_{s\longrightarrow t} P_*(s,t) = 1$.

▶ **Lemma 13.** *Let $\mathcal{M} = (S, S_\Box, S_\bigcirc, \longrightarrow, P)$ be an MDP, and let φ be an objective that is tail in $\mathcal{M}$. Let $\mathcal{M}_* = (S_*, S_{*\Box}, S_{*\bigcirc}, \longrightarrow_*, P_*)$ be the conditioned version of $\mathcal{M}$ w.r.t. φ. Let $s_0 \in S_* \cap S$. Let $\sigma \in \Sigma_{\mathcal{M}_*}$, and note that σ can be transformed to a strategy in $\mathcal{M}$ in a natural way. Then:*

1. *For all $n \geq 0$ and all partial runs $s_0 s_1 \cdots s_n \in s_0 S_*^*$ in $\mathcal{M}_*$ with $s_n \in S$:*

$$
\mathtt{val}_{\mathcal{M}}(s_0)\cdot\mathcal{P}_{\mathcal{M}_*,s_0,\sigma}(s_0 s_1\cdots s_n S_*^\omega) \;=\; \mathcal{P}_{\mathcal{M},s_0,\sigma}(\overline{s_0 s_1\cdots s_n} S^\omega)\cdot\mathtt{val}_{\mathcal{M}}(s_n)\,,
$$

where $\overline{w}$ for a partial run w in $\mathcal{M}_*$ refers to its natural contraction to a partial run in $\mathcal{M}$; i.e., $\overline{w}$ is obtained from w by deleting all states of the form (s,t).

2. *For all measurable $\mathfrak{R} \subseteq s_0(S_* \setminus \{s_\perp\})^\omega$ we have*

$$\mathcal{P}_{\mathcal{M},s_0,\sigma}(\overline{\mathfrak{R}}) \;\geq\; \mathtt{val}_{\mathcal{M}}(s_0) \cdot \mathcal{P}_{\mathcal{M}_*,s_0,\sigma}(\mathfrak{R}) \;\geq\; \mathcal{P}_{\mathcal{M},s_0,\sigma}(\overline{\mathfrak{R}} \cap [\![\varphi]\!]^{s_0}),$$

where $\overline{\mathfrak{R}}$ is obtained from $\mathfrak{R}$ by deleting, in all runs, all states of the form (s,t).

3. *We have $\mathtt{val}_{\mathcal{M}}(s_0) \cdot \mathcal{P}_{\mathcal{M}_*,s_0,\sigma}(\varphi) = \mathcal{P}_{\mathcal{M},s_0,\sigma}(\varphi)$. In particular, $\mathtt{val}_{\mathcal{M}_*}(s_0) = 1$, and, for any $\varepsilon \geq 0$, strategy σ is ε-optimal in $\mathcal{M}_*$ if and only if it is $\varepsilon\mathtt{val}_{\mathcal{M}}(s_0)$-optimal in $\mathcal{M}$.*

Lemma 13.3 provides a way of proving the existence of MD strategies that attain, for each state s, a fixed *fraction* (arbitrarily close to 1) of the value of s:

▶ **Theorem 14.** *Let $\mathcal{M} = (S, S_\Box, S_\bigcirc, \longrightarrow, P)$ be an MDP, and let φ be an objective that is tail in $\mathcal{M}$. Let $\mathcal{M}_* = (S_*, S_{*\Box}, S_{*\bigcirc}, \longrightarrow_*, P_*)$ be the conditioned version of $\mathcal{M}$ w.r.t. φ. Let $\varepsilon \geq 0$. Any MD strategy σ that is uniformly ε-optimal in $\mathcal{M}_*$ (i.e., $\mathcal{P}_{\mathcal{M}_*,s,\sigma}(\varphi) \geq \mathtt{val}_{\mathcal{M}_*}(s) - \varepsilon$ holds for all $s \in S_*$) is* multiplicatively ε-optimal *in $\mathcal{M}$ (i.e., $\mathcal{P}_{\mathcal{M},s,\sigma}(\varphi) \geq (1-\varepsilon)\mathtt{val}_{\mathcal{M}}(s)$ holds for all $s \in S$).*

Proof. Immediate from Lemma 13.3. ◀

As an application of Theorem 14, we can strengthen the first statement of Theorem 8 towards *multiplicatively* (see Theorem 14) uniform ε-optimal MD strategies for `Transience`.

▶ **Corollary 15.** *In every countable MDP there exist multiplicatively uniform ε-optimal MD strategies for* `Transience`.

Proof. Let $\mathcal{M}$ be a countable MDP, and $\mathcal{M}_*$ its conditioned version w.r.t. `Transience`. Let $\varepsilon > 0$. By Theorem 8, there is a uniform ε-optimal MD strategy σ for `Transience` in $\mathcal{M}_*$. By Theorem 14, strategy σ is multiplicatively uniform ε-optimal in $\mathcal{M}$. ◀

The following lemma, stating that universal transience is closed under "conditioning", is needed for the proof of Lemma 17 below.

▶ **Lemma 16.** *Let $\mathcal{M} = (S, S_\Box, S_\bigcirc, \longrightarrow, P)$ be an MDP, and let φ be an objective that is tail in $\mathcal{M}$. Let $\mathcal{M}_* = (S_*, S_{*\Box}, S_{*\bigcirc}, \longrightarrow_*, P_*)$ be the conditioned version of $\mathcal{M}$ w.r.t. φ, where $s_\perp$ is replaced by an infinite chain $s_\perp^1 \longrightarrow s_\perp^2 \longrightarrow \cdots$. If $\mathcal{M}$ is universally transient, then so is $\mathcal{M}_*$.*

In [17, Lemma 6] a variant, say $\mathcal{M}_+$, of the conditioned MDP $\mathcal{M}_*$ from Definition 12 was proposed. This variant $\mathcal{M}_+$ differs from $\mathcal{M}_*$ in that $\mathcal{M}_+$ has only those states s from $\mathcal{M}$ that have an optimal strategy, i.e., a strategy σ with $\mathcal{P}_{\mathcal{M},s,\sigma}(\varphi) = \mathtt{val}_{\mathcal{M}}(s)$. Further, for any transition $s \longrightarrow t$ in $\mathcal{M}_+$ where s is a controlled state, we have $\mathtt{val}_{\mathcal{M}}(s) = \mathtt{val}_{\mathcal{M}}(t)$, i.e., $\mathcal{M}_+$ does not have value-decreasing transitions emanating from controlled states. The following lemma was used in the proof of Theorem 10:

▶ **Lemma 17.** *Let $\mathcal{M}$ be an MDP, and let φ be an objective that is tail in $\mathcal{M}$. Let $\mathcal{M}_+$ be the conditioned version w.r.t. φ in the sense of [17, Lemma 6]. If $\mathcal{M}$ is universally transient, then so is $\mathcal{M}_+$.*

7 Conclusion

The `Transience` objective admits ε-optimal (resp. optimal) MD strategies even in *infinitely* branching MDPs. This is unusual, since ε-optimal strategies for most other objectives require infinite memory if the MDP is infinitely branching (in particular all objectives generalizing Safety [17]).

`Transience` encodes a notion of continuous progress, which can be used as a tool to reason about the strategy complexity of other objectives in countable MDPs. E.g., our result on `Transience` is used in [18] as a building block to show upper bounds on the strategy complexity of certain threshold objectives w.r.t. mean payoff, total payoff and point payoff.

References

1 Pieter Abbeel and Andrew Y. Ng. Learning first-order Markov models for control. In *Advances in Neural Information Processing Systems 17*. MIT Press, 2004. URL: http://papers.nips.cc/paper/2569-learning-first-order-markov-models-for-control.

2 Galit Ashkenazi-Golan, János Flesch, Arkadi Predtetchinski, and Eilon Solan. Reachability and safety objectives in Markov decision processes on long but finite horizons. *Journal of Optimization Theory and Applications*, 2020.

3 Christel Baier and Joost-Pieter Katoen. *Principles of Model Checking*. MIT Press, 2008.

4 Patrick Billingsley. *Probability and Measure*. Wiley, 1995. Third Edition.

5 Vincent D. Blondel and John N. Tsitsiklis. A survey of computational complexity results in systems and control. *Automatica*, 2000.

6 Nicole Bäuerle and Ulrich Rieder. *Markov Decision Processes with Applications to Finance*. Springer-Verlag Berlin Heidelberg, 2011.

7 K. Chatterjee and T. Henzinger. A survey of stochastic ω-regular games. *Journal of Computer and System Sciences*, 2012.

8 Edmund M. Clarke, Thomas A. Henzinger, Helmut Veith, and Roderick Bloem, editors. *Handbook of Model Checking*. Springer, 2018. doi:10.1007/978-3-319-10575-8.

9 E.M. Clarke, O. Grumberg, and D. Peled. *Model Checking*. MIT Press, Dec. 1999.

10 William Feller. *An Introduction to Probability Theory and Its Applications*. Wiley & Sons, second edition, 1966.

11 János Flesch, Arkadi Predtetchinski, and William Sudderth. Simplifying optimal strategies in limsup and liminf stochastic games. *Discrete Applied Mathematics*, 2018.

12 T.P. Hill and V.C. Pestien. The existence of good Markov strategies for decision processes with general payoffs. *Stoch. Processes and Appl.*, 1987.

13 S. Kiefer, R. Mayr, M. Shirmohammadi, and P. Totzke. Transience in countable MDPs. In *International Conference on Concurrency Theory*, LIPIcs, 2021. Full version at https://arxiv.org/abs/2012.13739.

14 Stefan Kiefer, Richard Mayr, Mahsa Shirmohammadi, and Patrick Totzke. Büchi objectives in countable MDPs. In *International Colloquium on Automata, Languages and Programming*, LIPIcs, 2019. Full version at https://arxiv.org/abs/1904.11573. doi:10.4230/LIPIcs.ICALP.2019.119.

15 Stefan Kiefer, Richard Mayr, Mahsa Shirmohammadi, and Patrick Totzke. Strategy Complexity of Parity Objectives in Countable MDPs. In *International Conference on Concurrency Theory*, 2020. doi:10.4230/LIPIcs.CONCUR.2020.7.

16 Stefan Kiefer, Richard Mayr, Mahsa Shirmohammadi, Patrick Totzke, and Dominik Wojtczak. How to play in infinite MDPs (invited talk). In *International Colloquium on Automata, Languages and Programming*, 2020. doi:10.4230/LIPIcs.ICALP.2020.3.

17 Stefan Kiefer, Richard Mayr, Mahsa Shirmohammadi, and Dominik Wojtczak. Parity Objectives in Countable MDPs. In *Annual IEEE Symposium on Logic in Computer Science*, 2017. doi:10.1109/LICS.2017.8005100.

18 Richard Mayr and Eric Munday. Strategy Complexity of Mean Payoff, Total Payoff and Point Payoff Objectives in Countable MDPs. In *International Conference on Concurrency Theory*, LIPIcs, 2021. The full version is available on arXiv.

19 A. Mostowski. Regular expressions for infinite trees and a standard form of automata. In *Computation Theory*, LNCS, 1984.

20 Donald Ornstein. On the existence of stationary optimal strategies. *Proceedings of the American Mathematical Society*, 1969. doi:10.2307/2035700.

21 Martin L. Puterman. *Markov Decision Processes: Discrete Stochastic Dynamic Programming*. John Wiley & Sons, Inc., 1st edition, 1994.

22 George Santayana. Reason in common sense. In *Volume 1 of The Life of Reason*. 1905. URL: https://en.wikipedia.org/wiki/George_Santayana.

23 Manfred Schäl. Markov decision processes in finance and dynamic options. In *Handbook of Markov Decision Processes*. Springer, 2002.

24 Olivier Sigaud and Olivier Buffet. *Markov Decision Processes in Artificial Intelligence*. John Wiley & Sons, 2013.

25 William D. Sudderth. Optimal Markov strategies. *Decisions in Economics and Finance*, 2020.

26 R.S. Sutton and A.G Barto. *Reinforcement Learning: An Introduction*. Adaptive Computation and Machine Learning. MIT Press, 2018.

27 Moshe Y. Vardi. Automatic verification of probabilistic concurrent finite-state programs. In *Annual Symposium on Foundations of Computer Science*. IEEE Computer Society, 1985. doi:10.1109/SFCS.1985.12.

A Strategy Classes

We formalize the amount of *memory* needed to implement strategies. Let M be a countable set of memory modes. An *update function* is a function $u : \mathsf{M} \times S \to \mathcal{D}(\mathsf{M} \times S)$ that meets the following two conditions, for all modes $\mathsf{m} \in \mathsf{M}$:

- for all controlled states $s \in S_\Box$, the distribution $u((\mathsf{m}, s))$ is over $\mathsf{M} \times \{s' \mid s \longrightarrow s'\}$.
- for all random states $s \in S_\bigcirc$, we have that $\sum_{\mathsf{m}' \in \mathsf{M}} u((\mathsf{m}, s))(\mathsf{m}', s') = P(s)(s')$.

An update function u together with an initial memory m_0 induce a strategy $u[\mathsf{m}_0] : S^* S_\Box \to \mathcal{D}(S)$ as follows. Consider the Markov chain with states set $\mathsf{M} \times S$, transition relation $(\mathsf{M} \times S)^2$ and probability function u. Any partial run $\rho = s_0 \cdots s_i$ in $\mathcal{M}$ gives rise to a set $H(\rho) = \{(\mathsf{m}_0, s_0) \cdots (\mathsf{m}_i, s_i) \mid \mathsf{m}_0, \ldots, \mathsf{m}_i \in \mathsf{M}\}$ of partial runs in this Markov chain. Each $\rho s \in s_0 S^* S_\Box$ induces a probability distribution $\mu_{\rho s} \in \mathcal{D}(\mathsf{M})$, the probability $\mu_{\rho s}(\mathsf{m})$ is the probability of being in state (m, s) conditioned on having taken some partial run from $H(\rho s)$. We define $u[\mathsf{m}_0]$ such that $u[\mathsf{m}_0](\rho s)(s') \stackrel{\text{def}}{=} \sum_{\mathsf{m}, \mathsf{m}' \in \mathsf{M}} \mu_{\rho s}(\mathsf{m}) u((\mathsf{m}, s))(\mathsf{m}', s')$ for all $\rho s \in S^* S_\Box$ and $s' \in S$.

We say that a strategy σ can be *implemented* with memory M (and initial memory m_0) if there exists an update function u such that $\sigma = u[\mathsf{m}_0]$. In this case we may also write $\sigma[\mathsf{m}_0]$ to explicitly specify the initial memory mode m_0. Based on this, we can define several classes of strategies:

A strategy σ is *memoryless* (M) (also called *positional*) if it can be implemented with a memory of size 1. We may view M-strategies as functions $\sigma : S_\Box \to \mathcal{D}(S)$. A strategy σ is *finite memory* (F) if there exists a finite memory M implementing σ. More specifically, a strategy is 1-*bit* if it can be implemented with a memory of size 2. Such a strategy is then determined by a function $u : \{0,1\} \times S \to \mathcal{D}(\{0,1\} \times S)$. *Deterministic 1-bit* strategies are are both deterministic and 1-bit.

B Missing Proofs from Section 3

In this section, we prove Lemma 2 from the main body.

▶ **Lemma 2.** *Given an infinitely branching countable MDP $\mathcal{M}$ with an initial state s_0, there exists a finitely branching countable $\mathcal{M}'$ with a set S' of states such that $s_0 \in S'$ and*

1. *each strategy α_1 in $\mathcal{M}$ is mapped to a unique strategy β_1 in $\mathcal{M}'$ where*

$$\mathcal{P}_{s_0,\alpha_1}(\texttt{Transience}) = \mathcal{P}_{s_0,\beta_1}(\texttt{Transience}),$$

2. *and conversely, every MD strategy β_2 in $\mathcal{M}'$ is mapped to an MD strategy α_2 in $\mathcal{M}$ where*

$$\mathcal{P}_{s_0,\alpha_2}(\texttt{Transience}) \geq \mathcal{P}_{s_0,\beta_2}(\texttt{Transience}).$$

Proof. Given an infinitely branching MDP $\mathcal{M} = (S, S_\Box, S_\bigcirc, \longrightarrow, P)$ with set S of states and an initial state $s_0 \in S$, we construct a finitely branching $\mathcal{M}'$ with set S' of states such that $s_0 \in S'$. The reduction uses the concept of "recurrent ladders"; see Figure 2.

The reduction is as follows.

- For all controlled state s in $\mathcal{M}$ with infinite branching $s \to s_i$ for all $i \geq 1$, we introduce a recurrent ladder in $\mathcal{M}'$, consisting the controlled states $(\ell_{s,i})_{i \in \mathbb{N}}$ and random states $(\ell'_{s,i})_{i \geq 1}$. The set of transitions includes $s \to \ell_{s,0}$ and $\ell_{s,0} \to \ell_{s,1}$, and for all $i \geq 1$ two transitions $\ell_{s,i} \to \ell'_{s,i}$, and $\ell_{s,i} \to s_i$. Moreover, $\ell'_{s,i} \xrightarrow{\frac{1}{2}} \ell_{s,i+1}$ and $\ell'_{s,i} \xrightarrow{\frac{1}{2}} \ell_{s,i-1}$. Here, all states of the recurrent ladder are fresh states.

- For all random states s in $\mathcal{M}$ with infinite branching $s \xrightarrow{p_i} s_i$ for all $i \geq 1$, we use a gadget $s \xrightarrow{1} z_1, z_i \xrightarrow{1-p'_i} z_{i+1}, z_i \xrightarrow{p'_i} s_i$ for all $i \geq 1$, with fresh random states z_i and suitably adjusted probabilities p'_i to ensure that the gadget is left at state s_i with exact probability p_i, i.e., $p'_i = p_i/(\prod_{j=1}^{i-1}(1-p'_j))$.

  See Figure 3 for a partial illustration.
  Given $\rho = q_0 q_1 \cdots q_n \in S^+$ denote by $\mathrm{last}(\rho) = q_n$ the last state of ρ.

For the first item, let α_1 be a general strategy $\alpha_1 : S^* S_\Box \to \mathcal{D}(S)$ in $\mathcal{M}$. We define β_1 in $\mathcal{M}'$ with the use of memory $\mathsf{M} = S^* \times \{\bot\} \cup \{i \in \mathbb{N} \mid i \leq 1\})$ and an update function u; see Appendix A. The definition of $u : \mathsf{M} \times S' \to \mathcal{D}(\mathsf{M} \times S')$ is as follows. For all $q, q' \in S'$ and $\rho \in S^*$,

- for all $\mathsf{m} = (\rho, \bot)$ and $\mathsf{m}' = (\rho q, \bot)$,

$$u(\mathsf{m}, q)(\mathsf{m}', q') = \begin{cases} P(q)(q') & \text{if } q \in S_\bigcirc; \\ \\ \alpha_1(\rho q)(q') & \text{if } q \in S_\Box \text{ is finitely branching in } \mathcal{M}; \end{cases}$$

- for all $\mathsf{m} = (\rho, \bot)$ and $\mathsf{m}' = (\rho q, j)$ with $j \geq 1$,

$$u(\mathsf{m}, q)(\mathsf{m}', q') = \begin{cases} \alpha_1(\rho q)(q_j) & \text{if } q \in S_\Box \text{ is infinitely branching in } \mathcal{M} \text{ with } q \to q_i \text{ for all } i \geq 1, \text{ and } q' = \ell_{q,0}; \end{cases}$$

- for all $\mathsf{m}, \mathsf{m}' = (\rho, j)$ with j,

$$u(\mathsf{m}, q)(\mathsf{m}, q') = \begin{cases} 1 & \text{if } s = \mathrm{last}(\rho) \text{ was infinitely branching in } \mathcal{M}, \text{ and if } q = \ell_{s,i},\ q' = \ell'_{s,i} \text{ and } i < j; \end{cases}$$

- for all $\mathsf{m} = (\rho, j)$ and $\mathsf{m}' = (\rho, \bot)$ with $j \geq 1$,

$$u(\mathsf{m}, q)(\mathsf{m}, q') = \begin{cases} 1 & \text{if } s = \mathrm{last}(\rho) \text{ was infinitely branching in } \mathcal{M} \text{ with } s \to s_i \text{ for all } i \geq 1, \text{ and if } q = \ell_{s,j} \text{ and } q' = s_j; \end{cases}$$

- and $u(\mathsf{m}, q)(\mathsf{m}', q') = 0$ otherwise.

The strategy β_1 consists of the above update function u and initial memory $\mathsf{m}_0 = (\epsilon, \bot)$ where ϵ is the empty run. Intuitively speaking, in every step β_1 considers the memory (ρ, x) and the current state q to simulate what α_1 would have played in $\mathcal{M}$. The memory (ρ, x) is such that ρ invariantly demonstrates the history of run projected into the state space S of $\mathcal{M}$ (omitting the introduced states due to the reduction). The second component x in the memory is $\bot$ if the current state is in S, and otherwise it is a natural number $j \geq 1$. Such a natural number j indicates that the controller is currently on a recurrent ladder and must leaves the ladder at the j-th controlled state on the ladder. Subsequently, β_1 starts with memory $(\epsilon, \bot)$ and $q = s_0$,

- when q is a random state in $\mathcal{M}$, β_1 only append q to ρ to keep track of the history;
- when q is a finitely branching state in $\mathcal{M}$, β_1 plays as $\alpha_1(\rho q)$ and append q to ρ;
- when q is an infinitely branching state in $\mathcal{M}$ with successors $(q_j)_{j \geq 1}$, for every $j \geq 1$, the strategy β_1 chooses the first state $\ell_{q,0}$ of the recurrent ladder for q while flipping the memory from $(\rho, \bot)$ to (ρ, j) with probability $\sigma(\rho q)(q_j)$. This requires the ladder to be traversed to state $\ell_{q,j}$ and left from there to q_j, the j-th successor of q in $\mathcal{M}$. Furthermore, β_1 append q to ρ;

- when q is $\ell_{s,i}$ and memory is (ρ, j) with $\text{last}(\rho) = s$, if $i \leq j$ then β_1 continues to stay on the recurrent ladder by picking $\ell'_{s,i}$;
- when q is $\ell_{s,j}$ and memory is (ρ, j) with $\text{last}(\rho) = s$, β_1 leaves the ladder from $\ell_{s,j}$ to q_j which is the j-th successor of state s in $\mathcal{M}$;. In addition, the memory (ρ, j) is flipped back to $(\rho, \bot)$.

By the construction of $\mathcal{M}'$ and β_1, it follows that β_1 in $\mathcal{M}'$ faithfully simulates α_1 in $\mathcal{M}$ and thus $\mathcal{P}_{\mathcal{M},s_0,\alpha_1}(\texttt{Transience}) = \mathcal{P}_{\mathcal{M}',s_0,\beta_1}(\texttt{Transience})$.

For the second item, let $\beta_2 : S'_\Box \to S'_\Box$ be an MD strategy in $\mathcal{M}'$ where $S'_\Box \subseteq S$ is the set of controlled states in $\mathcal{M}'$. We define an MD strategy $\alpha_2 : S \to S$ in $\mathcal{M}$ as follows. For all controlled states $s \in S'$,

$$\alpha_2(s) = \begin{cases} \beta_2(q) & \text{if } s \in S_\Box \text{ is finitely branching in } \mathcal{M}; \\ s_j & \text{if } s \in S_\Box \text{ is infinitely branching in } \mathcal{M} \text{ with the successors } (s_i)_{i\geq 1}, \\ & \text{and if there exists } j \in \mathbb{N} \text{ such that } \beta_2(s) = \ell_{s,0},\ \beta_2(\ell_{s,0}) = \ell_{s,1} \text{ and} \\ & \beta_2(\ell_{s,i}) = \ell'_{s,i} \text{ for all } 0 < i < j, \text{ and } \beta_2(\ell_{s,j}) = s_j; \\ s_1 & \text{if } s \in S_\Box \text{ is infinitely branching in } \mathcal{M} \text{ with the successors } (s_i)_{i\geq 1}, \text{ and} \\ & \text{if } \beta_2(s) = \ell_{s,0},\ \beta_2(\ell_{s,0}) = \ell_{s,1} \text{ and } \beta_2(\ell_{s,i}) = \ell'_{s,i} \text{ for all } i > 0. \end{cases}$$

Note that the above strategy is well-defined, as in every recurrent ladder in $\mathcal{M}'$, either there exists some j such that β_2 exits the ladder at its j-th controller state, or β_2 choose to stay on the ladder forever. In the latter case, by a Gambler's Ruin argument, the probability of `Transience` for those runs staying on the ladder forever is 0. By the construction of $\mathcal{M}'$, α_2 faithfully simulates β_2 unless when β_2 stays on a ladder forever and the prospect of `Transience` becomes 0. In those cases, α_2 continues playing what β_2 would have played if it exited the s-ladder at $\ell_{s,1}$.

It follows that $\mathcal{P}_{\mathcal{M},s_0,\alpha_2}(\texttt{Transience}) \geq \mathcal{P}_{\mathcal{M}',s_0,\beta_2}(\texttt{Transience})$. ◀

C 1-Bit Strategy for $\texttt{Büchi}(F) \cap \texttt{Transience}$

▶ **Lemma 4.** *Let $\mathcal{M}$ be a countable MDP, I a finite set of initial states, F a set of states and $\varepsilon > 0$. Then there exists a deterministic 1-bit strategy for $\texttt{Büchi}(F) \cap \texttt{Transience}$ that is ε-optimal from every $s \in I$.*

Proof. We prove the claim for finitely branching $\mathcal{M}$ first and transfer the result to general MDPs at the end.

Let $\mathcal{M} = (S, S_\Box, S_\bigcirc, \longrightarrow, P)$ be a finitely branching countable MDP, $I \subseteq S$ a finite set of initial states and $F \subseteq S$ a set of goal states and $\varphi \stackrel{\text{def}}{=} \texttt{Büchi}(F) \cap \texttt{Transience}$ the objective.

For every $\varepsilon > 0$ and every $s \in I$ there exists an ε-optimal strategy σ_s such that

$$\mathcal{P}_{\mathcal{M},s,\sigma_s}(\varphi) \geq \texttt{val}_{\mathcal{M},\varphi}(s) - \varepsilon. \tag{11}$$

However, the strategies σ_s might differ from each other and might use randomization and a large (or even infinite) amount of memory. We will construct a single deterministic strategy σ' that uses only 1 bit of memory such that $\forall_{s\in I}\, \mathcal{P}_{\mathcal{M},s,\sigma'}(\varphi) \geq \texttt{val}_{\mathcal{M},\varphi}(s) - 2\varepsilon$. This proves the claim as ε can be chosen arbitrarily small.

In order to construct σ', we first observe the behavior of the finitely many σ_s for $s \in I$ on an infinite, increasing sequence of finite subsets of S. Based on this, we define a second

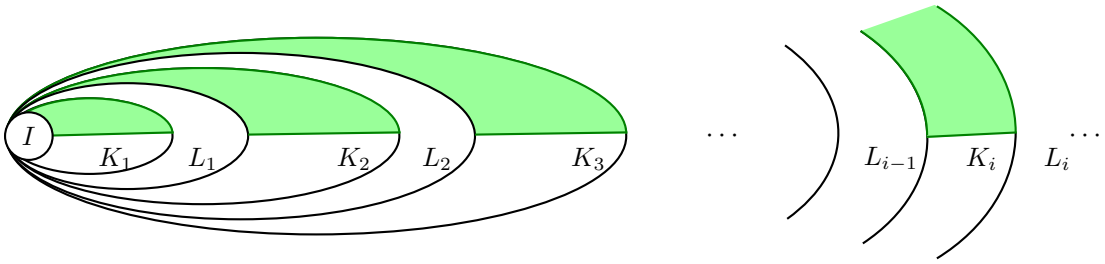

Figure 4 To show the bubble construction. The green region in K_1 is F_1, and for all $i \geq 2$, the green region in $K_i \setminus L_{i-1}$ is F_i.

stronger objective φ' with

$$\varphi' \subseteq \varphi, \tag{12}$$

and show that all σ_s attain at least $\mathtt{val}_{\mathcal{M},\varphi}(s) - 2\varepsilon$ w.r.t. φ', i.e.,

$$\forall_{s\in I}\, \mathcal{P}_{\mathcal{M},s,\sigma_s}(\varphi') \geq \mathtt{val}_{\mathcal{M},\varphi}(s) - 2\varepsilon. \tag{13}$$

We construct σ' as a deterministic 1-bit *optimal* strategy w.r.t. φ' from all $s \in I$ and obtain

$$\begin{aligned}
\mathcal{P}_{\mathcal{M},s,\sigma'}(\varphi) &\geq \mathcal{P}_{\mathcal{M},s,\sigma'}(\varphi') && \text{by Equation (12)}\\
&\geq \mathcal{P}_{\mathcal{M},s,\sigma_s}(\varphi') && \text{by optimality of } \sigma' \text{ for } \varphi'\\
&\geq \mathtt{val}_{\mathcal{M},\varphi}(s) - 2\varepsilon && \text{by Equation (13).}
\end{aligned}$$

Behavior of σ, objective φ' and properties Equation (12) and Equation (13). We start with some notation. Let $\mathsf{bubble}_k(X)$ be the set of states that can be reached from some state in the set X within at most k steps. Since $\mathcal{M}$ is finitely branching, $\mathsf{bubble}_k(X)$ is finite if X is finite. Let $\mathsf{F}^{\leq k}(X) \stackrel{\text{def}}{=} \{\rho \in S^\omega \mid \exists t \leq k.\, \rho(t) \in X\}$ and $\mathsf{F}^{\geq k}(X) \stackrel{\text{def}}{=} \{\rho \in S^\omega \mid \exists t \geq k.\, \rho(t) \in X\}$ denote the property of visiting the set X (at least once) within at most (resp. at least) k steps. Moreover, let $\varepsilon_i \stackrel{\text{def}}{=} \varepsilon \cdot 2^{-(i+1)}$.

▶ **Lemma 18.** *Assume the setup of Lemma 4, $\varphi \stackrel{def}{=} \mathtt{Büchi}(F) \cap \mathtt{Transience}$ and a strategy σ_s from each $s \in I$. Let $X \subseteq S$ be a finite set of states and $\varepsilon' > 0$.*

1. *There is $k \in \mathbb{N}$ such that $\forall_{s\in I}\, \mathcal{P}_{\mathcal{M},s,\sigma_s}(\varphi \cap \neg(\mathsf{F}^{\leq k}(F \setminus X))) \leq \varepsilon'$.*
2. *There is $l \in \mathbb{N}$ such that $\forall_{s\in I}\, \mathcal{P}_{\mathcal{M},s,\sigma_s}(\varphi \cap \mathsf{F}^{\geq l}(X)) \leq \varepsilon'$.*

Proof. It suffices to show the properties for a single s, σ_s since one can take the maximal k, l over the finitely many $s \in I$.

We observe that $\varphi \subseteq \mathtt{Transience} = \bigcap_{s\in S} \mathsf{FG}\neg(s) \subseteq \bigcap_{s\in X} \mathsf{FG}\neg(s) = \mathsf{FG}\neg(X)$, where the last equivalence is due to the finiteness of X.

Towards 1, we have $\varphi = \mathsf{GF}F \cap \mathtt{Transience} \subseteq \mathsf{GF}F \cap \mathsf{FG}\neg(X) \subseteq \mathsf{GF}(F \setminus X) \subseteq \mathsf{F}(F \setminus X) = \bigcup_{k\in\mathbb{N}} \mathsf{F}^{\leq k}(F \setminus X)$ and therefore that $\varphi \cap \bigcap_{k\in\mathbb{N}} \neg(\mathsf{F}^{\leq k}(F \setminus X)) = \emptyset$. It follows from the continuity of measures that $\lim_{k\to\infty} \mathcal{P}_{\mathcal{M},s,\sigma_s}(\varphi \cap \neg(\mathsf{F}^{\leq k}(F \setminus X))) = 0$.

Towards 2, we have $\varphi \cap \cap_l \mathsf{F}^{\geq l}(X) \subseteq \mathsf{FG}\neg(X) \cap \cap_l \mathsf{F}^{\geq l}(X) = \emptyset$. By continuity of measures we obtain $\lim_{l\to\infty} \mathcal{P}_{\mathcal{M},s,\sigma_s}(\varphi \cap \mathsf{F}^{\geq l}(X)) = 0$. ◀

In the following, let us write $\overline{X}$ to denote the complement of a set $X \subseteq S^\omega$ of runs.

By Lemma 18(1) there is a k_1 such that for $K_1 \stackrel{\text{def}}{=} \mathsf{bubble}_{k_1}(I)$ and $F_1 \stackrel{\text{def}}{=} F \cap K_1$ we have $\forall_{s\in I}\, \mathcal{P}_{\mathcal{M},s,\sigma_s}(\varphi \cap \overline{K_1^* F_1 S^\omega}) \leq \varepsilon_1$. We define the pattern

$$R_1 \stackrel{\text{def}}{=} (K_1 \setminus F_1)^* F_1$$

and obtain $\forall_{s\in I}\, \mathcal{P}_{\mathcal{M},s,\sigma_s}(\varphi \cap \overline{R_1 S^\omega}) \leq \varepsilon_1$. By Lemma 18(2) there is an $l_1 > k_1$ such that $\forall_{s\in I}\, \mathcal{P}_{\mathcal{M},s,\sigma_s}(\mathsf{F}^{\geq l_1}(K_1)) \leq \varepsilon_1$. Define $L_1 \stackrel{\text{def}}{=} \mathsf{bubble}_{l_1}(I)$. By Lemma 18(1) there is a $k_2 > l_1$ such that for $K_2 \stackrel{\text{def}}{=} \mathsf{bubble}_{k_2}(I)$ and $F_2 \stackrel{\text{def}}{=} F \cap K_2 \setminus L_1$ we have $\forall_{s\in I}\, \mathcal{P}_{\mathcal{M},s,\sigma_s}(\varphi \cap \overline{K_2^* F_2 S^\omega}) \leq \varepsilon_2$. We define the pattern

$$R_2 \stackrel{\text{def}}{=} (K_2 \setminus F_2)^* F_2$$

and obtain $\forall_{s\in I}\, \mathcal{P}_{\mathcal{M},s,\sigma_s}(\varphi \cap \overline{R_2 S^\omega}) \leq \varepsilon_2$ and, via a union bound, $\forall_{s\in I}\, \mathcal{P}_{\mathcal{M},s,\sigma_s}(\varphi \cap \overline{R_2 (S \setminus K_1)^\omega}) \leq \varepsilon_1 + \varepsilon_2$. By another union bound it follows that $\forall_{s\in I}\, \mathcal{P}_{\mathcal{M},s,\sigma_s}(\varphi \cap \overline{R_1 R_2 (S \setminus K_1)^\omega}) \leq 2\varepsilon_1 + \varepsilon_2$.

Proceed inductively for $i = 2, 3, \ldots$ as follows (see Figure 4 for an illustration). By Lemma 18(2) there is an $l_i > k_i$ such that $\forall_{s\in I}\, \mathcal{P}_{\mathcal{M},s,\sigma_s}(\mathsf{F}^{\geq l_i}(K_i)) \leq \varepsilon_i$. Define $L_i \stackrel{\text{def}}{=} \mathsf{bubble}_{l_i}(I)$. By Lemma 18(1) there is $k_{i+1} > l_i$ such that for $K_{i+1} \stackrel{\text{def}}{=} \mathsf{bubble}_{k_{i+1}}(I)$ and $F_{i+1} \stackrel{\text{def}}{=} F \cap K_{i+1} \setminus L_i$ we have $\forall_{s\in I}\, \mathcal{P}_{\mathcal{M},s,\sigma_s}(\varphi \cap \overline{(K_{i+1} \setminus F_{i+1})^* F_{i+1} S^\omega}) \leq \varepsilon_{i+1}$. By a union bound, $\forall_{s\in I}\, \mathcal{P}_{\mathcal{M},s,\sigma_s}(\varphi \cap \overline{(K_{i+1} \setminus F_{i+1})^* F_{i+1} (S \setminus K_i)^\omega}) \leq \varepsilon_i + \varepsilon_{i+1}$. By an induction hypothesis we have $\forall_{s\in I}\, \mathcal{P}_{\mathcal{M},s,\sigma_s}(\varphi \cap \overline{R_1 R_2 \ldots R_i (S \setminus K_{i-1})^\omega}) \leq 2\varepsilon_1 + \cdots + 2\varepsilon_{i-1} + \varepsilon_i$. We define the pattern

$$R_{i+1} \stackrel{\text{def}}{=} (K_{i+1} \setminus (F_{i+1} \cup K_{i-1}))^* F_{i+1}.$$

Using that $(K_{i+1} \setminus F_{i+1})^* F_{i+1} (S \setminus K_i)^\omega \cap R_1 R_2 \ldots R_i (S \setminus K_{i-1})^\omega \subseteq R_1 R_2 \ldots R_{i+1} (S \setminus K_i)^\omega$, we get

$$\forall_{s\in I}\, \mathcal{P}_{\mathcal{M},s,\sigma_s}(\varphi \cap \overline{R_1 R_2 \ldots R_{i+1} (S \setminus K_i)^\omega}) \quad \leq \quad 2\varepsilon_1 + \cdots + 2\varepsilon_i + \varepsilon_{i+1} \quad \leq \quad \varepsilon. \tag{14}$$

We now define the Borel objectives $R_{\leq i} \stackrel{\text{def}}{=} R_1 R_2 \ldots R_i S^\omega$ and

$$\varphi' \stackrel{\text{def}}{=} \bigcap_{i\in\mathbb{N}} R_{\leq i}.$$

Since $F_i \cap F_k = \emptyset$ for $i \neq k$ and φ' implies a visit to the set F_i for all $i \in \mathbb{N}$, we have $\varphi' \subseteq \texttt{Büchi}(F)$. Now we show that $\varphi' \subseteq \texttt{Transience}$. Let s be an arbitrary state and ρ a run from some state in I that satisfies φ'. If s is not reachable from I then ρ never visits s. Otherwise, there exists some minimal j such that $s \in K_j$. The run ρ must eventually visit F_{j+1} and after visiting F_{j+1} it cannot visit K_j (and thus s) any more. Therefore ρ visits s only finitely often. Thus $\varphi' \subseteq \texttt{Transience}$. Together we have $\varphi' \subseteq \texttt{Büchi}(F) \cap \texttt{Transience} = \varphi$ and obtain Equation (12).

Moreover, $R_{\leq 1} \supseteq R_{\leq 2} \supseteq R_{\leq 3} \ldots$ is an infinite decreasing sequence of Borel objectives.

For every $s \in I$ we have

$$\begin{aligned}
\mathcal{P}_{\mathcal{M},s,\sigma_s}(\varphi') &= \mathcal{P}_{\mathcal{M},s,\sigma_s}(\cap_{i=1}^{\infty} R_{\leq i}) && \text{by def. of } \varphi' \\
&= \lim_{i\to\infty} \mathcal{P}_{\mathcal{M},s,\sigma_s}(R_{\leq i}) && \text{by cont. of measures} \\
&= \lim_{i\to\infty} 1 - \mathcal{P}_{\mathcal{M},s,\sigma_s}(\overline{R_{\leq i}}) && \text{by duality} \\
&= \lim_{i\to\infty} 1 - (\mathcal{P}_{\mathcal{M},s,\sigma_s}(\overline{R_{\leq i}} \cap \varphi) + \mathcal{P}_{\mathcal{M},s,\sigma_s}(\overline{R_{\leq i}} \cap \overline{\varphi})) && \text{case split} \\
&\geq \lim_{i\to\infty} 1 - (\varepsilon + \mathcal{P}_{\mathcal{M},s,\sigma_s}(\overline{R_{\leq i}} \cap \overline{\varphi})) && \text{by Equation (14)} \\
&\geq \lim_{i\to\infty} 1 - (\varepsilon + \mathcal{P}_{\mathcal{M},s,\sigma_s}(\overline{\varphi'} \cap \overline{\varphi})) && \text{since } \varphi' \subseteq R_{\leq i} \\
&= 1 - (\varepsilon + 1 - \mathcal{P}_{\mathcal{M},s,\sigma_s}(\varphi' \cup \varphi)) && \text{by duality} \\
&= \mathcal{P}_{\mathcal{M},s,\sigma_s}(\varphi) - \varepsilon && \text{by Equation (12)} \\
&\geq \mathtt{val}_{\mathcal{M},\varphi}(s) - 2\varepsilon && \text{by Equation (11)}
\end{aligned}$$

Thus we obtain property Equation (13).

Definition of the 1-bit strategy σ'. We now define our deterministic 1-bit strategy σ' that is optimal for objective φ' from *every* $s \in I$. First we define certain "suffix" objectives of φ'. Recall that $R_i = (K_i \setminus (F_i \cup K_{i-2}))^* F_i$. Let $R_{i,j} \stackrel{\text{def}}{=} R_i R_{i+1} \dots R_j S^\omega$ and $R_{\geq i} \stackrel{\text{def}}{=} \bigcap_{j \geq i} R_{i,j}$. In particular, this means that $\varphi' = R_{\geq 1}$. Every run w from some state $s \in I$ that satisfies φ' can be split into parts before and after the first visit to set F_i, i.e., $w = w_1 s' w_2$ where $w_1 s' \in R_{\leq i}$, $s' \in F_i$ and $s' w_2 \in R_{\geq i+1}$. (Note also that w_2 cannot visit any states in K_{i-1}.) Thus it will be useful to consider the objectives $R_{\geq i+1}$ for runs that start in states $s' \in F_i$. For every state $s' \in F_i$ we consider its value w.r.t. the objective $R_{\geq i+1}$, i.e., $\mathtt{val}_{\mathcal{M},R_{\geq i+1}}(s') \stackrel{\text{def}}{=} \sup_{\hat{\sigma}} \mathcal{P}_{\mathcal{M},s',\hat{\sigma}}(R_{\geq i+1})$.

For every $i \geq 1$ we consider the finite subspace $K_i \setminus K_{i-2}$. In particular, it contains the sets F_{i-1} and F_i. (For completeness let $K_0 \stackrel{\text{def}}{=} F_0 \stackrel{\text{def}}{=} I$ and $K_{-1} \stackrel{\text{def}}{=} \emptyset$.) It is not enough to maximize the probability of reaching the set F_i in each K_i individually. One also needs to maximize the potential of visiting further sets $F_{i+1}, F_{i+2}, \dots$ in the indefinite future. Thus we define the bounded total reward objective B_i for runs starting in F_{i-1} as follows. Runs that exit the subspace (either by leaving K_i or by visiting K_{i-2}) before visiting F_i get reward 0. When some run reaches the set F_i *for the first time* in some state s' then this run gets the reward of $\mathtt{val}_{\mathcal{M},R_{\geq i+1}}(s')$. We can consider an induced finite MDP $\hat{\mathcal{M}}$ with state space $K_i \setminus K_{i-2}$, plus a sink state (with reward 0) that is reached immediately after visiting any state in F_i and whenever one exits the set $K_i \setminus K_{i-2}$. In $\hat{\mathcal{M}}$ one gets a reward of $\mathtt{val}_{\mathcal{M},R_{\geq i+1}}(s')$ for visiting $s' \in F_i$ as above. By [21, Theorem 7.1.9], there exists a uniform optimal MD strategy σ_i for this bounded total reward objective on the induced finite MDP $\hat{\mathcal{M}}$, which can be directly applied for objective B_i on the subspace $K_i \setminus K_{i-2}$ in $\mathcal{M}$. (The strategy σ_i is not necessarily unique, but our results hold regardless of which of them is picked.)

We now define σ' by combining different MD strategies σ_i, depending on the current state and on the value of the 1-bit memory. The intuition is that the strategy σ' has two modes: normal-mode and next-mode. In a state $s' \in K_i \setminus K_{i-1}$, if the memory is $i \pmod 2$ then the strategy is in normal-mode and plays towards reaching F_i. Otherwise, the strategy is in next-mode and plays towards reaching F_{i+1} (normally this happens because F_i has already been seen).

Initially σ' starts in a state $s \in I$ with the 1-bit memory set to 1. We define the behavior of σ' in a state $s' \in K_i \setminus K_{i-1}$ for every $i \geq 1$.

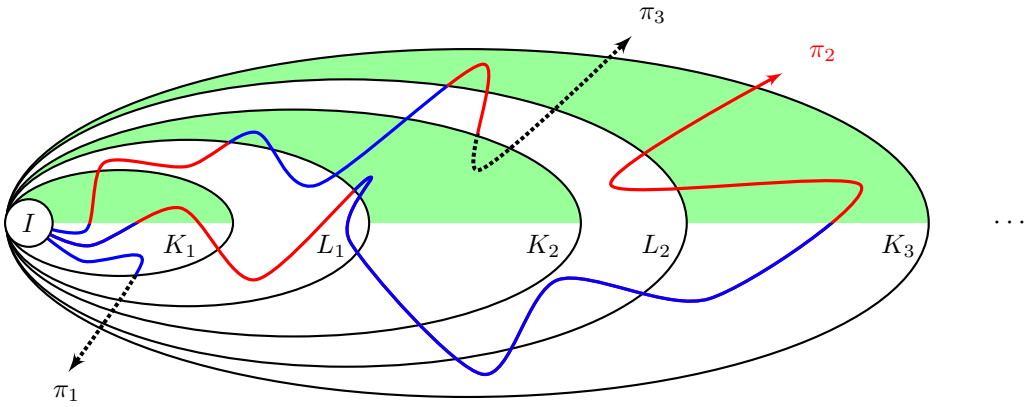

Figure 5 Memory updates along runs π_1, π_2, π_3, drawn in blue while the memory-bit is one and in red while the bit is zero. Both π_1 and π_3 violate φ' and are drawn as dotted lines once they do.

- If the 1-bit memory is $i \pmod 2$ and $s' \notin F_i$ then σ' plays like σ_i. (Intuitively, one plays towards F_i, since one has not yet visited it.)
- If the 1-bit memory is $i \pmod 2$ and $s' \in F_i$ then the 1-bit memory is set to $(i+1) \pmod 2$, and σ' plays like σ_{i+1}. (Intuitively, one records the fact that one has already seen F_i and then targets the next set F_{i+1}.)
- If the 1-bit memory is $(i+1) \pmod 2$ then σ' plays like σ_{i+1}. (Intuitively, one plays towards F_{i+1}, since one has already visited F_i.)

Observe that if a run according to σ' exits some set K_i (and thus enters $K_{i+1} \setminus K_i$) with the bit still set to $i \pmod 2$ (normal-mode) then this run has not visited F_i and thus does not satisfy the objective φ'. (Or the same has happened earlier for some $j < i$, in which case also the objective φ' is violated.) An example is the run π_1 in Figure 5.

However, if a run according to σ' exits some set K_i (and thus enters $K_{i+1} \setminus K_i$) with the bit set to $(i+1) \pmod 2$ (thus σ_{i+1} in next-mode) then in the new set $K_{i'} \setminus K_{i'-1}$ with $i' = i+1$ the bit is set to $i' \pmod 2$ and σ' continues to play like σ_{i+1} in normal-mode. Even if this run returns (temporarily) to K_i (but not to K_{i-1}) the strategy σ' continues to play like σ_{i+1} in next-mode. An example is the run π_2 in Figure 5.

Finally, if a run returns to K_{i-1} after having visited F_i then it fails the objective φ'. An example is the run π_3 in Figure 5.

The 1-bit strategy σ' is optimal for φ' from every $s \in I$. In the following let $s \in I$ be an arbitrary initial state in I. For any run from s, let $\mathsf{firstin}(F_i)$ be the first state s' in F_i that is visited (if any). We define a bounded reward objective B'_i for runs starting at s as follows. Every run that does not satisfy the objective $R_{\leq i}$ gets assigned reward 0. Otherwise, consider a run from s that satisfies $R_{\leq i}$. When this run reaches the set F_i for the first time in some state s' then this run gets a reward of $\mathtt{val}_{\mathcal{M},R_{\geq i+1}}(s')$. Note that this reward is ≤ 1.

We show that for all $i \in \mathbb{N}$

$$\mathtt{val}_{\mathcal{M},\varphi'}(s) = \mathtt{val}_{\mathcal{M},B'_i}(s) \tag{15}$$

Towards the $\geq$ inequality, let $\hat{\sigma}$ be an $\hat{\varepsilon}$-optimal strategy for B'_i from s. We define the strategy $\hat{\sigma}'$ to play like $\hat{\sigma}$ until a state $s' \in F_i$ is reached and then to switch to some $\hat{\varepsilon}$-optimal strategy for objective $R_{\geq i+1}$ from s'. Every run from s that satisfies φ' can be split into parts, before and after the first visit to the set F_i, i.e., $\varphi' = \{w_1 s' w_2 \mid w_1 s' \in R_{\leq i}, s' \in F_i, s' w_2 \in R_{\geq i+1}\}$.

Therefore we obtain that $\mathcal{P}_{\mathcal{M},s,\hat{\sigma}'}(\varphi') \geq \mathcal{E}_{\mathcal{M},s,\hat{\sigma}}(B'_i) - \hat{\varepsilon} \geq \mathtt{val}_{\mathcal{M},B'_i}(s) - 2\hat{\varepsilon}$. Since this holds for every $\hat{\varepsilon} > 0$, we obtain $\mathtt{val}_{\mathcal{M},\varphi'}(s) \geq \mathtt{val}_{\mathcal{M},B'_i}(s)$.

Towards the $\leq$ inequality, let $\hat{\sigma}$ be any strategy for φ' from s. We have $\mathcal{P}_{\mathcal{M},s,\hat{\sigma}}(\varphi') \leq \sum_{s' \in F_i} \mathcal{P}_{\mathcal{M},s,\hat{\sigma}}(R_{\leq i} \cap \mathsf{firstin}(F_i) = s') \cdot \mathtt{val}_{\mathcal{M},R_{\geq i+1}}(s') = \mathcal{E}_{\mathcal{M},s,\hat{\sigma}}(B'_i)$. Thus $\mathtt{val}_{\mathcal{M},\varphi'}(s) \leq \mathtt{val}_{\mathcal{M},B'_i}(s)$. Together we obtain Equation (15).

For all $i \in \mathbb{N}$ and every state $s' \in F_i$ we show that

$$\mathtt{val}_{\mathcal{M},R_{\geq i+1}}(s') = \mathtt{val}_{\mathcal{M},B_{i+1}}(s') \tag{16}$$

Towards the $\geq$ inequality, let $\hat{\sigma}$ be an $\hat{\varepsilon}$-optimal strategy for B_{i+1} from $s' \in F_i$. We define the strategy $\hat{\sigma}'$ to play like $\hat{\sigma}$ until a state $s'' \in F_{i+1}$ is reached and then to switch to some $\hat{\varepsilon}$-optimal strategy for objective $R_{\geq i+2}$ from s''. We have that $\mathcal{P}_{\mathcal{M},s',\hat{\sigma}'}(R_{\geq i+1}) \geq \mathcal{E}_{\mathcal{M},s',\hat{\sigma}}(B_{i+1}) - \hat{\varepsilon} \geq \mathtt{val}_{\mathcal{M},B_{i+1}}(s) - 2\hat{\varepsilon}$. Since this holds for every $\hat{\varepsilon} > 0$, we obtain $\mathtt{val}_{\mathcal{M},R_{\geq i+1}}(s') \geq \mathtt{val}_{\mathcal{M},B_{i+1}}(s')$.

Towards the $\leq$ inequality, let $\hat{\sigma}$ be any strategy for $R_{\geq i+1}$ from $s' \in F_i$. We have

$$\begin{aligned}\mathcal{P}_{\mathcal{M},s',\hat{\sigma}}(R_{\geq i+1}) &\leq \sum_{s'' \in F_{i+1}} \mathcal{P}_{\mathcal{M},s',\hat{\sigma}}(R_{i+1}S^\omega \cap \mathsf{firstin}(F_{i+1}) = s'') \cdot \mathtt{val}_{\mathcal{M},R_{\geq i+2}}(s'') \\ &= \mathcal{E}_{\mathcal{M},s',\hat{\sigma}}(B_{i+1}).\end{aligned}$$

Thus $\mathtt{val}_{\mathcal{M},R_{\geq i+1}}(s') \leq \mathtt{val}_{\mathcal{M},B_{i+1}}(s')$. Together we obtain Equation (16).

We show, by induction on i, that σ' is optimal for B'_i for all $i \in \mathbb{N}$ from start state s, i.e.,

$$\mathcal{E}_{\mathcal{M},s,\sigma'}(B'_i) = \mathtt{val}_{\mathcal{M},B'_i}(s) \tag{17}$$

In the base case of $i = 1$ we have that $B'_1 = B_1$. The strategy σ' plays σ_1 until reaching F_1, which is optimal for objective B_1 and thus optimal for B'_1. For the induction step we assume (IH) that σ' is optimal for B'_i.

$$\begin{aligned}
\mathtt{val}_{\mathcal{M},B'_{i+1}}(s) &= \mathtt{val}_{\mathcal{M},B'_i}(s) && \text{by Equation (15)} \\
&= \mathcal{E}_{\mathcal{M},s,\sigma'}(B'_i) && \text{by (IH)} \\
&= \sum_{s' \in F_i} \mathcal{P}_{\mathcal{M},s,\sigma'}(R_{\leq i} \cap \mathsf{firstin}(F_i) = s') \cdot \mathtt{val}_{\mathcal{M},R_{\geq i+1}}(s') && \text{by def. of } B'_i \\
&= \sum_{s' \in F_i} \mathcal{P}_{\mathcal{M},s,\sigma'}(R_{\leq i} \cap \mathsf{firstin}(F_i) = s') \cdot \mathtt{val}_{\mathcal{M},B_{i+1}}(s') && \text{by Equation (16)} \\
&= \sum_{s' \in F_i} \mathcal{P}_{\mathcal{M},s,\sigma'}(R_{\leq i} \cap \mathsf{firstin}(F_i) = s') \cdot \mathcal{E}_{\mathcal{M},s',\sigma_{i+1}}(B_{i+1}) && \text{opt. of } \sigma_{i+1} \text{ for } B_{i+1} \\
&= \mathcal{E}_{\mathcal{M},s,\sigma'}(B'_{i+1}) && \text{by def. of } \sigma' \text{ and } B'_{i+1}
\end{aligned}$$

So σ' attains the value $\mathtt{val}_{\mathcal{M},B'_{i+1}}(s)$ of the objective B'_{i+1} from s and is optimal. Thus Equation (17).

Now we show that σ' performs well on the objectives $R_{\leq i}$ for all $i \in \mathbb{N}$.

$$\mathcal{P}_{\mathcal{M},s,\sigma'}(R_{\leq i}) \geq \mathtt{val}_{\mathcal{M},\varphi'}(s) \tag{18}$$

We have

$$\begin{aligned}
\mathcal{P}_{\mathcal{M},s,\sigma'}(R_{\leq i}) &\geq \mathcal{E}_{\mathcal{M},s,\sigma'}(B'_i) && \text{since } B'_i \text{ gives rewards 0 for runs } \notin R_{\leq i} \text{ and } \leq 1 \text{ otherwise} \\
&= \mathtt{val}_{\mathcal{M},B'_i}(s) && \text{by Equation (17)} \\
&= \mathtt{val}_{\mathcal{M},\varphi'}(s) && \text{by Equation (15)}
\end{aligned}$$

So we get Equation (18). Now we are ready to prove the optimality of σ' for φ' from s.

$$\begin{aligned}
\mathcal{P}_{\mathcal{M},s,\sigma'}(\varphi') &= \mathcal{P}_{\mathcal{M},s,\sigma'}(\cap_{i\in\mathbb{N}} R_{\leq i}) && \text{by def. of } \varphi' \\
&= \lim_{i\to\infty} \mathcal{P}_{\mathcal{M},s,\sigma'}(R_{\leq i}) && \text{by continuity of measures from above} \\
&\geq \lim_{i\to\infty} \mathtt{val}_{\mathcal{M},\varphi'}(s) && \text{by Equation (18)} \\
&= \mathtt{val}_{\mathcal{M},\varphi'}(s)
\end{aligned}$$

This concludes the proof that σ' is optimal for φ' and hence 2ε-optimal for φ for every initial state $s \in I$.

From finitely to infinitely branching MDPs. Let $\mathcal{M}$ be an infinitely branching MDP with a finite set of initial states I and $\varepsilon > 0$. We derive a finitely branching MDP $\mathcal{M}'$ with sufficiently similar behavior wrt. our objective $\varphi = \mathtt{Büchi}(F) \cap \mathtt{Transience}$. Every controlled state x with infinite branching $x \to y_i$ for all $i \in \mathbb{N}$ is replaced by a gadget $x \to z_1, z_i \to z_{i+1}, z_i \to y_i$ for all $i \in \mathbb{N}$ with fresh controlled states z_i. Infinitely branching random states with $x \xrightarrow{p_i} y_i$ for all $i \in \mathbb{N}$ are replaced by a gadget $x \xrightarrow{1} z_1, z_i \xrightarrow{1-p'_i} z_{i+1}, z_i \xrightarrow{p'_i} y_i$ for all $i \in \mathbb{N}$, with fresh random states z_i and suitably adjusted probabilities p'_i to ensure that the gadget is left at state y_i with probability p_i, i.e., $p'_i = p_i/(\prod_{j=1}^{i-1}(1-p'_j))$.

We apply the above result for finitely branching MDPs to $\mathcal{M}'$ and obtain a 1-bit deterministic ε-optimal strategy σ' for our objective $\varphi = \mathtt{Büchi}(F) \cap \mathtt{Transience}$ from all states $s \in I$. We construct a 1-bit deterministic ε-optimal strategy σ'' for $\mathcal{M}$ as follows. Consider some state x that is infinitely branching in $\mathcal{M}$ and its associated gadget in $\mathcal{M}'$. Whenever a run in $\mathcal{M}'$ according to σ' reaches x with some memory value $\alpha \in \{0,1\}$ there exist values p_i for the probability that the gadget is left at state y_i. Let $p \stackrel{\text{def}}{=} 1 - \sum_{i\in\mathbb{N}} p_i$ be the probability that the gadget is never left. (If x is controlled then only one p_i (or p) is nonzero, since σ' is deterministic. If x is random then $p = 0$.) Since σ' is deterministic, the memory updates are deterministic, and thus there are values $\alpha'_i \in \{0,1\}$ such that whenever the gadget is left at state y_i the memory will be α'_i. We now define the behavior of the 1-bit deterministic strategy σ'' at state x with memory α in $\mathcal{M}$.

If x is controlled and $p \neq 1$ then σ'' picks the successor state y_i where $p_i = 1$ and sets the memory to α'_i. If $p = 1$ then any run according to σ' that enters the gadget does not satisfy the objective $\varphi = \mathtt{Büchi}(F) \cap \mathtt{Transience}$, since the states in the gadget are disjoint from F. I.e., every run that eventually stays in some gadget forever does not even satisfy $\mathtt{Büchi}(F)$, and thus does not satisfy φ. Thus σ'' performs at least as well in $\mathcal{M}$ regardless of its choice, e.g., pick successor y_1 and $\alpha' = \alpha$.

If x is random then $p = 0$ and the successor is chosen according to the defined distribution (which is the same in $\mathcal{M}$ and $\mathcal{M}'$) and σ'' can only update its memory. Whenever the successor y_i is chosen, σ'' updates the memory to α'_i.

In states that are not infinitely branching in $\mathcal{M}$, σ'' does exactly the same in $\mathcal{M}$ as σ' in $\mathcal{M}'$.

Since the gadgets do not intersect F, σ'' performs at least as well in $\mathcal{M}$ as σ' in $\mathcal{M}'$ and is thus ε-optimal from every $s \in I$. ◀

▶ Remark 19. Note that the last step in the proof of Lemma 4, lifting the result from finitely branching MDPs to infinitely branching MDPs, does require this particular construction. It cannot be shown by applying Lemma 2. The construction used for Lemma 2 (i.e., Figure 3) can only lift MD strategies, but not deterministic 1-bit strategies. The problem is that the construction in Figure 3 introduces extra randomness and multiple paths to the same

exit from the ladder. While an MD strategy on the finitely branching MDP $\mathcal{M}'$ induces a corresponding MD strategy on the infinitely branching MDP $\mathcal{M}$, the same does not hold for deterministic 1-bit strategies. In contrast, the different construction in the last part of the proof of Lemma 4 preserves deterministic 1-bit strategies, but works only for the $\texttt{Büchi}(F) \cap \texttt{Transience}$ objective, not for $\texttt{Transience}$ alone.

D Missing Proofs from Section 4

We prove Theorem 7 from the main body:

▶ **Theorem 7.** *Let $\mathcal{M} = (S, S_\Box, S_\bigcirc, \longrightarrow, P)$ be a countable MDP, and let φ be an objective that is tail in $\mathcal{M}$. Suppose for every $s \in S$ there exist ε-optimal MD strategies for φ. Then:*

1. *There exist uniform ε-optimal MD strategies for φ.*
2. *There exists a single MD strategy that is optimal from every state that has an optimal strategy.*

D.1 Proof of Item 1 of Theorem 7

Proof. We follow Ornstein's proof [20] as presented in [16]. Recall that an MD strategy σ can be viewed as a function $\sigma : S_\Box \to S$ such that for all $s \in S_\Box$, the state $\sigma(s)$ is a successor state of s. Starting from the original MDP $\mathcal{M}$ we successively *fix* more and more controlled states, by which we mean select an outgoing transition and remove all others. While this is in general an infinite (but countable) process, it defines an MD strategy in the limit. Visually, we "plaster" the whole state space by the fixings.

Put the states in some order, i.e., $s_1, s_2, \ldots$ with $S = \{s_1, s_2, \ldots\}$. The plastering proceeds in *rounds*, one round for every state. Let $\mathcal{M}_i$ be the MDP obtained from $\mathcal{M}$ after the fixings of the first $i-1$ rounds (with $\mathcal{M}_1 = \mathcal{M}$). In round i we fix controlled states in such a way that

(A) the probability, starting from s_i, of φ using only random and *fixed* controlled states is not much less than the value $\texttt{val}_{\mathcal{M}_i}(s_i)$; and
(B) for all states s, the value $\texttt{val}_{\mathcal{M}_{i+1}}(s)$ is almost as high as $\texttt{val}_{\mathcal{M}_i}(s)$.

The purpose of goal (A) is to guarantee good progress towards φ when starting from s_i. The purpose of goal (B) is to avoid fixings that would cause damage to the values of other states.

Now we describe round i. Consider the MDP $\mathcal{M}_i$ after the fixings from the first $i-1$ rounds, and let $\varepsilon_i > 0$. Recall that we wish to fix a part of the state space so that s_i has a high probability of φ using only random and fixed controlled states. By assumption there is an MD strategy σ such that $\mathcal{P}_{\mathcal{M}_i, s_i, \sigma}(\varphi) \geq \texttt{val}_{\mathcal{M}_i}(s_i) - \varepsilon_i^2$. Fixing σ everywhere would accomplish goal (A), but potentially compromise goal (B). So instead we are going to fix σ only for states where σ does well: define

$$G \stackrel{\text{def}}{=} \{s \in S \mid \mathcal{P}_{\mathcal{M}_i, s, \sigma}(\varphi) \geq \texttt{val}_{\mathcal{M}_i}(s) - \varepsilon_i\}$$

and obtain $\mathcal{M}_{i+1}$ from $\mathcal{M}_i$ by fixing σ on G. (Note that σ does not "contradict" earlier fixings, because in the MDP $\mathcal{M}_i$ the previously fixed states have only one outgoing transition left.)

We have to check that with this fixing we accomplish the two goals above. Indeed, we accomplish goal (A): by its definition strategy σ is ε_i^2-optimal from s_i, so the probability of ever entering $S \setminus G$ (where σ is less than ε_i-optimal) cannot be large:

$$\mathcal{P}_{\mathcal{M}_i, s_i, \sigma}(\texttt{Reach}(S \setminus G)) \;\leq\; \varepsilon_i \tag{19}$$

In slightly more detail, this inequality holds because the probability that the ε_i^2-optimal strategy σ enters a state whose value is underachieved by σ by at least ε_i can be at most ε_i. We give a detailed proof of (19) in Lemma 20 below. It follows from the ε_i^2-optimality of σ and from (19) that we have $\mathcal{P}_{\mathcal{M}_i,s_i,\sigma}(\varphi \wedge \neg \mathtt{Reach}(S \setminus G)) \geq \mathtt{val}_{\mathcal{M}_i}(s_i) - \varepsilon_i - \varepsilon_i^2$. So in $\mathcal{M}_{i+1}$ we obtain for *all* strategies σ':

$$\mathcal{P}_{\mathcal{M}_{i+1},s_i,\sigma'}(\varphi) \;\geq\; \mathtt{val}_{\mathcal{M}_i}(s_i) - \varepsilon_i - \varepsilon_i^2 \tag{20}$$

We also accomplish goal (B): the difference between $\mathcal{M}_i$ and $\mathcal{M}_{i+1}$ is that σ is fixed on G, but σ performs well from G on. So we obtain for *all* states s:

$$\mathtt{val}_{\mathcal{M}_{i+1}}(s) \;\geq\; \mathtt{val}_{\mathcal{M}_i}(s) - \varepsilon_i \tag{21}$$

In slightly more detail, this inequality holds because any strategy in $\mathcal{M}_i$ can be transformed into a strategy in $\mathcal{M}_{i+1}$, with the difference that once the newly fixed part G is entered, the strategy switches to the strategy σ, which (by the definition of $\mathcal{M}_{i+1}$) is consistent with the fixing and (by the definition of G) is ε_i-optimal from there. We give a detailed proof of (21) in Lemma 21 below. This completes the description of round i.

Let $\varepsilon \in (0,1)$, and for all $i \geq 1$, choose $\varepsilon_i \stackrel{\text{def}}{=} \frac{\varepsilon}{2} \cdot 2^{-i}$. Let σ be an arbitrary MD strategy that is compatible with all fixings. (This strategy σ is actually unique.) It follows that σ is playable in all $\mathcal{M}_i$. We have for all $i \geq 1$:

$$\begin{aligned}
\mathcal{P}_{\mathcal{M},s_i,\sigma}(\varphi) \;&\geq\; \mathtt{val}_{\mathcal{M}_i}(s_i) - \varepsilon_i - \varepsilon_i^2 && \text{by (20)} \\
&\geq\; \mathtt{val}_{\mathcal{M}_i}(s_i) - 2\varepsilon_i && \text{as } \varepsilon_i < 1 \\
&\geq\; \mathtt{val}_{\mathcal{M}_i}(s_i) - \frac{\varepsilon}{2} && \text{choice of } \varepsilon_i \\
&\geq\; \mathtt{val}_{\mathcal{M}}(s_i) - \sum_{j=1}^{i-1} \varepsilon_j - \frac{\varepsilon}{2} && \text{by (21)} \\
&\geq\; \mathtt{val}_{\mathcal{M}}(s_i) - \varepsilon && \text{choice of } \varepsilon_j
\end{aligned}$$

Thus, the MD strategy σ is ε-optimal for all states. ◀

▶ **Lemma 20.** *Equation* (19) *holds.*

Proof. For a state $s \in S \setminus G$, define the event L_s as the set of runs that leave G such that s is the first visited state in $S \setminus G$. Then we have:

$$\mathcal{P}_{\mathcal{M}_i,s_i,\sigma}(\mathtt{Reach}(S \setminus G)) \;=\; \sum_{s \in S \setminus G} \mathcal{P}_{\mathcal{M}_i,s_i,\sigma}(L_s)$$

Since φ is tail and using the Markov property:

$$\begin{aligned}
\mathcal{P}_{\mathcal{M}_i,s_i,\sigma}(\varphi) \;=\; &\mathcal{P}_{\mathcal{M}_i,s_i,\sigma}(\neg \mathtt{Reach}(S \setminus G) \wedge \varphi) \;+ \\
&\sum_{s \in S \setminus G} \mathcal{P}_{\mathcal{M}_i,s_i,\sigma}(L_s) \cdot \mathcal{P}_{\mathcal{M}_i,s,\sigma}(\varphi)
\end{aligned}$$

By the definition of G it follows:

$$\begin{aligned}
\mathcal{P}_{\mathcal{M}_i,s_i,\sigma}(\varphi) \;\leq\; &\mathcal{P}_{\mathcal{M}_i,s_i,\sigma}(\neg \mathtt{Reach}(S \setminus G) \wedge \varphi) \;+ \\
&\sum_{s \in S \setminus G} \mathcal{P}_{\mathcal{M}_i,s_i,\sigma}(L_s) \cdot (\mathtt{val}_{\mathcal{M}_i}(s) - \varepsilon_i)
\end{aligned} \tag{22}$$

On the other hand, σ is ε_i^2-optimal for s_i, hence:

$$\begin{aligned}\mathcal{P}_{\mathcal{M}_i,s_i,\sigma}(\varphi) &\geq -\varepsilon_i^2 + \mathtt{val}_{\mathcal{M}_i}(s)\\ &\geq -\varepsilon_i^2 + \mathcal{P}_{\mathcal{M}_i,s_i,\sigma}(\varphi \wedge \neg\mathtt{Reach}(S\setminus G)) + \\ &\qquad \sum_{s\in S\setminus G} \mathcal{P}_{\mathcal{M}_i,s_i,\sigma}(L_s)\cdot \mathtt{val}_{\mathcal{M}_i}(s)\end{aligned} \tag{23}$$

By combining (22) and (23) we obtain:

$$\varepsilon_i^2 \;\geq\; \varepsilon_i \cdot \sum_{s\in S\setminus G} \mathcal{P}_{\mathcal{M}_i,s_i,\sigma}(L_s) \;=\; \varepsilon_i \cdot \mathcal{P}_{\mathcal{M}_i,s_i,\sigma}(\mathtt{Reach}(S\setminus G)) \qquad \blacktriangleleft$$

▶ **Lemma 21.** *Equation (21) holds.*

Proof. For a state $s' \in G$, define the event $E_{s'}$ as the set of runs that enter G such that s' is the first visited state in G. Fix any state $s \in S$ and any strategy σ_i in $\mathcal{M}_i$. We transform σ_i into a strategy σ_{i+1} in $\mathcal{M}_{i+1}$ such that σ_{i+1} behaves like σ_i until G is entered, at which point σ_{i+1} switches to the MD strategy σ, which we recall is compatible with $\mathcal{M}_{i+1}$ and is ε_i-optimal from G in $\mathcal{M}_i$. To show (21) it suffices to show that $\mathcal{P}_{\mathcal{M}_{i+1},s,\sigma_{i+1}}(\varphi) \geq \mathcal{P}_{\mathcal{M}_i,s,\sigma_i}(\varphi) - \varepsilon_i$. We have:

$$\begin{aligned}\mathcal{P}_{\mathcal{M}_{i+1},s,\sigma_{i+1}}(\varphi) &= \mathcal{P}_{\mathcal{M}_{i+1},s,\sigma_{i+1}}(\neg\mathtt{Reach}(G)\wedge\varphi) + && \varphi \text{ is tail}\\ &\qquad \sum_{s'\in G} \mathcal{P}_{\mathcal{M}_{i+1},s,\sigma_{i+1}}(E_{s'})\cdot \mathcal{P}_{\mathcal{M}_{i+1},s',\sigma_{i+1}}(\varphi) && \text{Markov property}\\ &= \mathcal{P}_{\mathcal{M}_i,s,\sigma_i}(\neg\mathtt{Reach}(G)\wedge\varphi) + && \text{using def. of } \sigma_{i+1}\\ &\qquad \sum_{s'\in G} \mathcal{P}_{\mathcal{M}_i,s,\sigma_i}(E_{s'})\cdot \mathcal{P}_{\mathcal{M}_i,s',\sigma}(\varphi)\end{aligned}$$

Further we have for all $s' \in G$:

$$\begin{aligned}\mathcal{P}_{\mathcal{M}_i,s',\sigma}(\varphi) &\geq \mathtt{val}_{\mathcal{M}_i}(s') - \varepsilon_i && \text{as } s' \in G\\ &\geq \mathcal{P}_{\mathcal{M}_i,s',\sigma_i}(\varphi) - \varepsilon_i\end{aligned}$$

Plugging this in above, we obtain:

$$\begin{aligned}\mathcal{P}_{\mathcal{M}_{i+1},s,\sigma_{i+1}}(\varphi) &\geq \mathcal{P}_{\mathcal{M}_i,s,\sigma_i}(\neg\mathtt{Reach}(G)\wedge\varphi) + \\ &\qquad \sum_{s'\in G} \mathcal{P}_{\mathcal{M}_i,s,\sigma_i}(E_{s'})\cdot(\mathcal{P}_{\mathcal{M}_i,s',\sigma_i}(\varphi) - \varepsilon_i)\\ &\geq \mathcal{P}_{\mathcal{M}_i,s,\sigma_i}(\neg\mathtt{Reach}(G)\wedge\varphi) + \\ &\qquad \Big(\sum_{s'\in G} \mathcal{P}_{\mathcal{M}_i,s,\sigma_i}(E_{s'})\cdot \mathcal{P}_{\mathcal{M}_i,s',\sigma_i}(\varphi)\Big) - \varepsilon_i\\ &= \mathcal{P}_{\mathcal{M}_i,s,\sigma_i}(\varphi) - \varepsilon_i\end{aligned} \qquad \blacktriangleleft$$

D.2 Proof of Item 2 of Theorem 7

Proof. As discussed in Section 6, in [17, Lemma 6] there is a construction of a certain *conditioned version* of $\mathcal{M}$ (similar to $\mathcal{M}_*$ from Definition 12), say $\mathcal{M}_+$. The construction is such that φ is tail also in $\mathcal{M}_+$. By [17, Lemma 6, item 2] it suffices to exhibit a single MD strategy in $\mathcal{M}_+$ that is *almost surely winning* from all states that have an *almost surely winning* strategy.

Obtain from $\mathcal{M}_+$ an MDP $\mathcal{M}'$ by restricting the state space to those states that have an almost surely winning strategy, and eliminating all transitions leaving these states. In $\mathcal{M}'$ all states have an almost surely winning strategy, as an almost surely winning strategy may never enter a state that does not have an almost surely winning strategy (using the fact that φ is tail). Let σ be a uniform $\frac{1}{2}$-optimal MD strategy (in $\mathcal{M}'$), which exists by item 1. It suffices to show that σ is (in $\mathcal{M}'$) almost surely winning from all states that have an almost surely winning strategy.

We follow the argument from [16, Theorem 6]. We have $\mathcal{P}_{\mathcal{M}',s,\sigma}(\varphi) \geq \frac{1}{2}$ for all states s. Thus, for any run $s_0 s_1 \cdots$ in $\mathcal{M}'$ we have $\mathcal{P}_{\mathcal{M}',s_i,\sigma}(\neg\varphi) \leq \frac{1}{2}$ for all i; in particular, the sequence $(\mathcal{P}_{\mathcal{M}',s_i,\sigma}(\neg\varphi))_i$ does not converge to 1. As a consequence of Lévy's zero-one law, since $\neg\varphi$ is tail, the events $\neg\varphi$ and $\{s_0 s_1 \cdots \mid \lim_{i\to\infty} \mathcal{P}_{\mathcal{M}',s_i,\sigma}(\neg\varphi) = 1\}$ are equal up to a null set. Thus, for all states s we have $\mathcal{P}_{\mathcal{M}',s,\sigma}(\neg\varphi) = 0$; hence, $\mathcal{P}_{\mathcal{M}',s,\sigma}(\varphi) = 1$. ◀

E Missing Proofs from Section 5

▶ **Theorem 9.** *For every universally transient countable MDP, safety objective and $\varepsilon > 0$ there exists a uniformly ϵ-optimal MD strategy.*

Proof. Let $\mathcal{M} = (S, S_\Box, S_\bigcirc, \longrightarrow, P)$ be a universally transient MDP and $\varepsilon > 0$. Assume w.l.o.g. that the target $T \subseteq S$ of the objective $\varphi = \texttt{Safety}(T)$ is a (losing) sink and let $\iota : S \to \mathbb{N}$ be an enumeration of the state space S.

By Lemma 3(3), for every state s we have $Re(s) \stackrel{\text{def}}{=} \sup_\sigma \mathcal{P}_{\mathcal{M},s,\sigma}(\mathsf{XF}(s)) < 1$ and thus $R(s) \stackrel{\text{def}}{=} \sum_{i=0}^{\infty} Re(s)^i < \infty$. This means that, independent of the chosen strategy, $Re(s)$ upper-bounds the chance to return to s, and $R(s)$ bounds the expected number of visits to s.

Suppose that σ is an MD strategy which, at any state $s \in S_\Box$, picks a successor s' with

$$\texttt{val}(s') \quad \geq \quad \texttt{val}(s) - \frac{\varepsilon}{2^{\iota(s)+1} \cdot R(s)}.$$

This is possible even if $\mathcal{M}$ is infinitely branching, by the definition of value and the fact that $R(s) < \infty$. We show that $\mathcal{P}_{\mathcal{M},s_0,\sigma}(\texttt{Safety}(T)) \geq \texttt{val}(s_0) - \varepsilon$ holds for every initial state s_0, which implies the claim of the theorem.

Towards this, we define a function $\texttt{cost}$ that labels each transition in the MDP with a real-valued cost: For every controlled transition $s \longrightarrow s'$ let $\texttt{cost}((s,s')) \stackrel{\text{def}}{=} \texttt{val}(s) - \texttt{val}(s') \geq 0$. Random transitions have cost zero. We will argue that when playing σ from any start state s_0, its attainment w.r.t. the objective $\texttt{Safety}(T)$ equals the value of s_0 minus the expected total cost, and that this cost is bounded by ε.

For any $i \in \mathbb{N}$ let us write s_i for the random variable denoting the state just after step i, and $\texttt{Cost}(i) \stackrel{\text{def}}{=} \texttt{cost}(s_i, s_{i+1})$ for the cost of step i in a random run. We now show that under σ the expected total cost is bounded in the limit, i.e.,

$$\lim_{n\to\infty} \mathcal{E}\left(\sum_{i=0}^{n-1} \texttt{Cost}(i)\right) \leq \varepsilon. \tag{24}$$

To show this, let us decompose the cost function as $\texttt{cost} = \sum_s \texttt{cost}_s$ where $\texttt{cost}_s$ is a local cost function for state s that assigns $\texttt{val}(s) - \texttt{val}(s')$ to all controlled transitions $s \longrightarrow s'$ *starting in* s and zero otherwise. Similarly, we let $\texttt{Cost}(n) \stackrel{\text{def}}{=} \sum_s \texttt{Cost}_s(n)$, where $\texttt{Cost}_s(n)$ is the random variable denoting the cost incurred on in step n from s. We thus have

$$\lim_{n\to\infty} \mathcal{E}\left(\sum_{i=0}^{n-1} \texttt{Cost}(i)\right) = \lim_{n\to\infty} \mathcal{E}\left(\sum_{s\in S}\sum_{i=0}^{n-1} \texttt{Cost}_s(i)\right) = \sum_{s\in S} \lim_{n\to\infty} \mathcal{E}\left(\sum_{i=0}^{n-1} \texttt{Cost}_s(i)\right)$$

where the last equality holds by convergence of monotone series.

We now show an upper bound on $\lim_{n\to\infty} \mathcal{E}\left(\sum_{i=0}^{n-1} \mathtt{Cost}_s(i)\right)$ for some fixed state s. Costs are only incurred at state s, and each time they are upper-bounded by $\frac{\varepsilon}{2^{\iota(s)+1}\cdot R(s)}$. Moreover, the probability of returning from s to s is upper-bounded by $Re(s)$. This means that $\lim_{n\to\infty} \mathcal{E}\left(\sum_{i=0}^{n-1} \mathtt{Cost}_s(i)\right) \leq \sum_{i=0}^{\infty} Re(s)^i \frac{\varepsilon}{2^{\iota(s)+1}\cdot R(s)} = \frac{\varepsilon}{2^{\iota(s)+1}}$, which in turn implies Eq. (24) as then $\lim_{n\to\infty} \mathcal{E}\left(\sum_{i=0}^{n-1} \mathtt{Cost}(i)\right) = \sum_{s\in S} \lim_{n\to\infty} \mathcal{E}\left(\sum_{i=0}^{n-1} \mathtt{Cost}_s(i)\right) \leq \sum_s \frac{\varepsilon}{2^{\iota(s)+1}} = \varepsilon$.

Next, we show that for every n,

$$\mathcal{E}(\mathtt{val}(s_n)) = \mathcal{E}(\mathtt{val}(s_0)) - \mathcal{E}\left(\sum_{i=0}^{n-1} \mathtt{Cost}(i)\right). \tag{25}$$

By induction on n where the base case $n = 0$ trivially holds. For the induction step,

$$\begin{aligned}
\mathcal{E}(\mathtt{val}(s_{n+1})) &= \mathcal{E}(\mathtt{val}(s_n) + \mathtt{val}(s_{n+1}) - \mathtt{val}(s_n)) \\
&= \mathcal{E}(\mathtt{val}(s_n)) + \mathcal{E}(\mathtt{val}(s_{n+1}) - \mathtt{val}(s_n)) \\
&= \mathcal{E}(\mathtt{val}(s_n)) + \mathcal{P}(s_n \in S_{\bigcirc})\mathcal{E}(\mathtt{val}(s_{n+1}) - \mathtt{val}(s_n) \mid s_n \in S_{\bigcirc}) \\
&\qquad + \mathcal{P}(s_n \in S_{\Box})\mathcal{E}(\mathtt{val}(s_{n+1}) - \mathtt{val}(s_n) \mid s_n \in S_{\Box}) \\
&= \mathcal{E}(\mathtt{val}(s_n)) + 0 - \mathcal{P}(s_n \in S_{\Box})\mathcal{E}(\mathtt{Cost}(n) \mid s_n \in S_{\Box}) \\
&= \mathcal{E}(\mathtt{val}(s_n)) - \mathcal{P}(s_n \in S_{\bigcirc})\mathcal{E}(\mathtt{Cost}(n) \mid s_n \in S_{\bigcirc}) \\
&\qquad - \mathcal{P}(s_n \in S_{\Box})\mathcal{E}(\mathtt{Cost}(n) \mid s_n \in S_{\Box}) \\
&= \mathcal{E}(\mathtt{val}(s_n)) - \mathcal{E}(\mathtt{Cost}(n)) \\
&= \mathcal{E}(\mathtt{val}(s_0)) - \sum_{i=0}^{n-1} \mathcal{E}(\mathtt{Cost}(i)) - \mathcal{E}(\mathtt{Cost}(n)) \\
&= \mathcal{E}(\mathtt{val}(s_0)) - \mathcal{E}\left(\sum_{i=0}^{n} \mathtt{Cost}(i)\right).
\end{aligned}$$

From Equations (24) and (25) we get

$$\liminf_{n\to\infty} \mathcal{E}(\mathtt{val}(s_n)) = \mathtt{val}(s_0) - \lim_{n\to\infty} \mathcal{E}\left(\sum_{i=0}^{n-1} \mathtt{cost}(i)\right) \geq \mathtt{val}(s_0) - \varepsilon. \tag{26}$$

Finally, to show the claim let $[s_n \notin T] : S^\omega \to \{0, 1\}$ be the random variable that indicates that the n-th state is not in the target set T. Note that $[s_n \notin T] \geq \mathtt{val}(s_n)$ because target states have value 0. We have:

$$\begin{aligned}
\mathcal{P}_{\mathcal{M},s_0,\sigma}(\mathtt{Safety}(T)) &= \mathcal{P}_{\mathcal{M},s_0,\sigma}\left(\bigwedge_{i=0}^{\infty} \mathsf{X}^i \neg T\right) && \text{semantics of } \mathtt{Safety}(T) = \mathsf{G}\neg T \\
&= \lim_{n\to\infty} \mathcal{P}_{\mathcal{M},s_0,\sigma}\left(\bigwedge_{i=0}^{n} \mathsf{X}^i \neg T\right) && \text{continuity of measures} \\
&= \lim_{n\to\infty} \mathcal{P}_{\mathcal{M},s_0,\sigma}(\mathsf{X}^n \neg T) && T \text{ is a sink} \\
&= \lim_{n\to\infty} \mathcal{E}([s_n \notin T]) && \text{definition of } [s_n \notin T] \\
&\geq \liminf_{n\to\infty} \mathcal{E}(\mathtt{val}(s_n)) && \text{as } [s_n \notin T] \geq \mathtt{val}(s_n) \\
&\geq \mathtt{val}(s_0) - \varepsilon && \text{Equation (26).}
\end{aligned}$$

◀

F Missing Proofs from Section 6

We prove Lemma 13 from the main body:

▶ **Lemma 13.** *Let $\mathcal{M} = (S, S_\Box, S_\bigcirc, \longrightarrow, P)$ be an MDP, and let φ be an objective that is tail in $\mathcal{M}$. Let $\mathcal{M}_* = (S_*, S_{*\Box}, S_{*\bigcirc}, \longrightarrow_*, P_*)$ be the conditioned version of $\mathcal{M}$ w.r.t. φ. Let $s_0 \in S_* \cap S$. Let $\sigma \in \Sigma_{\mathcal{M}_*}$, and note that σ can be transformed to a strategy in $\mathcal{M}$ in a natural way. Then:*

1. *For all $n \geq 0$ and all partial runs $s_0 s_1 \cdots s_n \in s_0 S_*^*$ in $\mathcal{M}_*$ with $s_n \in S$:*

$$\mathtt{val}_{\mathcal{M}}(s_0) \cdot \mathcal{P}_{\mathcal{M}_*,s_0,\sigma}(s_0 s_1 \cdots s_n S_*^\omega) \;=\; \mathcal{P}_{\mathcal{M},s_0,\sigma}(\overline{s_0 s_1 \cdots s_n} S^\omega) \cdot \mathtt{val}_{\mathcal{M}}(s_n)\,,$$

   where $\overline{w}$ for a partial run w in $\mathcal{M}_*$ refers to its natural contraction to a partial run in $\mathcal{M}$; i.e., $\overline{w}$ is obtained from w by deleting all states of the form (s,t).
2. *For all measurable $\mathfrak{R} \subseteq s_0(S_* \setminus \{s_\bot\})^\omega$ we have*

$$\mathcal{P}_{\mathcal{M},s_0,\sigma}(\overline{\mathfrak{R}}) \;\geq\; \mathtt{val}_{\mathcal{M}}(s_0) \cdot \mathcal{P}_{\mathcal{M}_*,s_0,\sigma}(\mathfrak{R}) \;\geq\; \mathcal{P}_{\mathcal{M},s_0,\sigma}(\overline{\mathfrak{R}} \cap [\![\varphi]\!]^{s_0})\,,$$

   where $\overline{\mathfrak{R}}$ is obtained from $\mathfrak{R}$ by deleting, in all runs, all states of the form (s,t).
3. *We have $\mathtt{val}_{\mathcal{M}}(s_0) \cdot \mathcal{P}_{\mathcal{M}_*,s_0,\sigma}(\varphi) = \mathcal{P}_{\mathcal{M},s_0,\sigma}(\varphi)$. In particular, $\mathtt{val}_{\mathcal{M}_*}(s_0) = 1$, and, for any $\varepsilon \geq 0$, strategy σ is ε-optimal in $\mathcal{M}_*$ if and only if it is $\varepsilon\mathtt{val}_{\mathcal{M}}(s_0)$-optimal in $\mathcal{M}$.*

Proof. We prove the equality in item 1 by induction on n. For $n = 0$ it is trivial. For the step, suppose the equality holds for some n. Let $s_0 s_1 \cdots s_n \in s_0 S_*^*$ be a partial run in $\mathcal{M}_*$ with $s_n \in S$.

Let $s_n \in S_\Box$ and $s_{n+1} \in S_* \cap S$. We have:

$$\begin{aligned}
&\mathtt{val}_{\mathcal{M}}(s_0) \cdot \mathcal{P}_{\mathcal{M}_*,s_0,\sigma}(s_0 s_1 \cdots s_n (s_n, s_{n+1}) s_{n+1} S_*^\omega) \\
&= \mathtt{val}_{\mathcal{M}}(s_0) \cdot \mathcal{P}_{\mathcal{M}_*,s_0,\sigma}(s_0 s_1 \cdots s_n S_*^\omega) \cdot \sigma(s_0 s_1 \ldots s_n)((s_n, s_{n+1})) \cdot \frac{\mathtt{val}_{\mathcal{M}}(s_{n+1})}{\mathtt{val}_{\mathcal{M}}(s_n)} && \text{def. of } P_* \\
&= \mathcal{P}_{\mathcal{M},s_0,\sigma}(\overline{s_0 s_1 \cdots s_n} S^\omega) \cdot \sigma(s_0 s_1 \ldots s_n)((s_n, s_{n+1})) \cdot \mathtt{val}_{\mathcal{M}}(s_{n+1}) && \text{ind. hyp.} \\
&= \mathcal{P}_{\mathcal{M},s_0,\sigma}(\overline{s_0 s_1 \cdots s_n} S^\omega) \cdot \sigma(\overline{s_0 s_1 \ldots s_n})(s_{n+1}) \cdot \mathtt{val}_{\mathcal{M}}(s_{n+1}) && \sigma \text{ in } \mathcal{M} \\
&= \mathcal{P}_{\mathcal{M},s_0,\sigma}(\overline{s_0 s_1 \cdots s_n (s_n s_{n+1}) s_{n+1}} S^\omega) \cdot \mathtt{val}_{\mathcal{M}}(s_{n+1})
\end{aligned}$$

Let $s_n \in S_\bigcirc$ and $s_{n+1} \in S_* \cap S$. We have:

$$\begin{aligned}
&\mathtt{val}_{\mathcal{M}}(s_0) \cdot \mathcal{P}_{\mathcal{M}_*,s_0,\sigma}(s_0 s_1 \cdots s_n s_{n+1} S_*^\omega) \\
&= \mathtt{val}_{\mathcal{M}}(s_0) \cdot \mathcal{P}_{\mathcal{M}_*,s_0,\sigma}(s_0 s_1 \cdots s_n S_*^\omega) \cdot P_*(s_n)(s_{n+1}) \\
&= \mathcal{P}_{\mathcal{M},s_0,\sigma}(\overline{s_0 s_1 \cdots s_n} S^\omega) \cdot P_*(s_n)(s_{n+1}) \cdot \mathtt{val}_{\mathcal{M}}(s_n) && \text{ind. hyp.} \\
&= \mathcal{P}_{\mathcal{M},s_0,\sigma}(\overline{s_0 s_1 \cdots s_n} S^\omega) \cdot P(s_n)(s_{n+1}) \cdot \mathtt{val}_{\mathcal{M}}(s_{n+1}) && \text{def. of } P_* \\
&= \mathcal{P}_{\mathcal{M},s_0,\sigma}(\overline{s_0 s_1 \cdots s_n s_{n+1}} S^\omega) \cdot \mathtt{val}_{\mathcal{M}}(s_{n+1})
\end{aligned}$$

This completes the inductive step, and we have proved item 1.

Towards item 2, define an MDP $\mathcal{M}'_* = (S'_*, S'_{*\Box}, S'_{*\bigcirc}, \longrightarrow'_*, P'_*)$ with "intermediate" states like (s,t) in $\mathcal{M}_*$, but with transition probabilities as in $\mathcal{M}$; more precisely:

$$\begin{aligned}
S'_{*\Box} &= S_\Box \\
S'_{*\bigcirc} &= S_\bigcirc \cup \{(s,t) \in \longrightarrow \mid s \in S_\Box\} \\
\longrightarrow'_* &= \{(s,(s,t)) \in (S_\Box \times \longrightarrow) \mid s \longrightarrow t\} \cup (S_\bigcirc \times S) \cup \\
&\qquad \{((s,t),t) \in (\longrightarrow \times S) \mid s \in S_\Box\} \\
P'_*(s,t) &= P(s,t) \\
P'_*((s,t),t) &= 1
\end{aligned}$$

Then we have

$$\mathcal{P}_{\mathcal{M}'_*,s_0,\sigma}(\mathfrak{R}) \;=\; \mathcal{P}_{\mathcal{M},s_o,\sigma}(\overline{\mathfrak{R}}) \qquad \text{for all measurable } \mathfrak{R} \subseteq s_0(S'_*)^\omega. \tag{27}$$

Let $s_0 s_1 \cdots s_n \in s_0(S'_*)^*$. If $s_0 s_1 \cdots s_n$ is a partial run in $\mathcal{M}_*$, then we have:

$$\begin{aligned}
&\mathcal{P}_{\mathcal{M}'_*,s_0,\sigma}(s_0 s_1 \cdots s_n (S'_*)^\omega) \\
&\;=\; \mathcal{P}_{\mathcal{M},s_0,\sigma}(\overline{s_0 s_1 \cdots s_n} S^\omega) && \text{Equation (27)} \\
&\;\geq\; \mathcal{P}_{\mathcal{M},s_0,\sigma}(\overline{s_0 s_1 \cdots s_n} S^\omega) \cdot \mathtt{val}_{\mathcal{M}}(s_n) \\
&\;=\; \mathtt{val}_{\mathcal{M}}(s_0) \cdot \mathcal{P}_{\mathcal{M}_*,s_0,\sigma}(s_0 s_1 \cdots s_n S_*^\omega) && \text{item 1} \\
&\;=\; \mathtt{val}_{\mathcal{M}}(s_0) \cdot \mathcal{P}_{\mathcal{M}_*,s_0,\sigma}(s_0 s_1 \cdots s_n (S'_*)^\omega \cap S_*^\omega)
\end{aligned}$$

Otherwise (i.e., $s_0 s_1 \cdots s_n$ is not a partial run in $\mathcal{M}_*$), the same inequality holds trivially. Invoking Lemma 22 below with $S := S'_*$ and $s := s_0$ and $\mu(\mathfrak{R}) := \mathcal{P}_{\mathcal{M}_*,s_0,\sigma}(\mathfrak{R} \cap S_*^\omega)$ and $\mu'(\mathfrak{R}) := \mathcal{P}_{\mathcal{M}'_*,s_0,\sigma}(\mathfrak{R})$ and $x := \mathtt{val}_{\mathcal{M}}(s_0)$ yields

$$\mathcal{P}_{\mathcal{M}'_*,s_0,\sigma}(\mathfrak{R}) \;\geq\; \mathtt{val}_{\mathcal{M}}(s_0) \cdot \mathcal{P}_{\mathcal{M}_*,s_0,\sigma}(\mathfrak{R} \cap S_*^\omega) \quad \text{for all measurable } \mathfrak{R} \subseteq s_0(S'_*)^\omega\,.$$

By Equation (27), the first inequality of item 2 follows.

Towards the second inequality of item 2, define $[\![\varphi]\!]^{s_0}_- \stackrel{\text{def}}{=} \{\rho \in s_0(S'_*)^\omega \mid \overline{\rho} \in [\![\varphi]\!]^{s_0}\}$. If $\mathcal{P}_{\mathcal{M}'_*,s_0,\sigma}(s_0 s_1 \cdots s_n (S'_*)^\omega \cap [\![\varphi]\!]^{s_0}_-) > 0$, then $s_0 s_1 \cdots s_n$ is a partial run in $\mathcal{M}_*$ and we have:

$$\begin{aligned}
&\mathtt{val}_{\mathcal{M}}(s_0) \cdot \mathcal{P}_{\mathcal{M}_*,s_0,\sigma}(s_0 s_1 \cdots s_n (S'_*)^\omega \cap S_*^\omega) \\
&\;=\; \mathtt{val}_{\mathcal{M}}(s_0) \cdot \mathcal{P}_{\mathcal{M}_*,s_0,\sigma}(s_0 s_1 \cdots s_n S_*^\omega) \\
&\;=\; \mathcal{P}_{\mathcal{M},s_0,\sigma}(\overline{s_0 s_1 \cdots s_n} S^\omega) \cdot \mathtt{val}_{\mathcal{M}}(s_n) && \text{item 1} \\
&\;\geq\; \mathcal{P}_{\mathcal{M},s_0,\sigma}(\overline{s_0 s_1 \cdots s_n} S^\omega) \cdot \mathcal{P}_{\mathcal{M},s_0,\sigma}([\![\varphi]\!]^{s_0} \mid \overline{s_0 s_1 \cdots s_n} S^\omega) && \varphi \text{ is tail} \\
&\;=\; \mathcal{P}_{\mathcal{M},s_0,\sigma}(\overline{s_0 s_1 \cdots s_n} S^\omega \cap [\![\varphi]\!]^{s_0}) \\
&\;=\; \mathcal{P}_{\mathcal{M}'_*,s_0,\sigma}(s_0 s_1 \cdots s_n (S'_*)^\omega \cap [\![\varphi]\!]^{s_0}_-) && \text{Equation (27)}
\end{aligned}$$

Otherwise (i.e., $\mathcal{P}_{\mathcal{M}'_*,s_0,\sigma}(s_0 s_1 \cdots s_n (S'_*)^\omega \cap [\![\varphi]\!]^{s_0}_-) = 0$), the same inequality holds trivially. Invoking Lemma 22 with $S := S'_*$ and $s := s_0$ and $\mu(\mathfrak{R}) := \mathcal{P}_{\mathcal{M}'_*,s_0,\sigma}(\mathfrak{R} \cap [\![\varphi]\!]^{s_0}_-)$ and $\mu'(\mathfrak{R}) := \mathcal{P}_{\mathcal{M}_*,s_0,\sigma}(\mathfrak{R} \cap S_*^\omega)$ and $x := 1/\mathtt{val}_{\mathcal{M}}(s_0)$ yields

$$\mathtt{val}_{\mathcal{M}}(s_0) \cdot \mathcal{P}_{\mathcal{M}_*,s_0,\sigma}(\mathfrak{R} \cap S_*^\omega) \;\geq\; \mathcal{P}_{\mathcal{M}'_*,s_0,\sigma}(\mathfrak{R} \cap [\![\varphi]\!]^{s_0}_-) \quad \text{for all measurable } \mathfrak{R} \subseteq s_0(S'_*)^\omega\,.$$

By Equation (27), the second inequality of item 2 follows.

Item 3 follows from item 2, with $\overline{\mathfrak{R}} = [\![\varphi]\!]^{s_0}$. ◀

The following lemma was used in the preceding proof.

▶ **Lemma 22.** *Let S be countable and $s \in S$. Call a set of the form swS^ω for $w \in S^*$ a* cylinder. *Let μ, μ' be measures on sS^ω defined in the standard way, i.e., first on cylinders and then extended to all measurable sets $\mathfrak{R} \subseteq sS^\omega$. Suppose there is $x \geq 0$ such that $x \cdot \mu(\mathfrak{C}) \leq \mu'(\mathfrak{C})$ for all cylinders $\mathfrak{C}$. Then $x \cdot \mu(\mathfrak{R}) \leq \mu'(\mathfrak{R})$ holds for all measurable $\mathfrak{R} \subseteq sS^\omega$.*

Proof. Let $\mathcal{C} = \{\mathfrak{C} \subseteq sS^\omega \mid \mathfrak{C} \text{ cylinder}\}$ denote the class of cylinders. This class generates an algebra $\mathcal{C}_* \supseteq \mathcal{C}$, which is the closure of $\mathcal{C}$ under finite union and complement. The classes $\mathcal{C}$ and $\mathcal{C}_*$ generate the same σ-algebra $\sigma(\mathcal{C})$. The class $\mathcal{C}_*$ is a set of countable disjoint unions of cylinders [4, Section 2]. Hence $x \cdot \mu(\mathfrak{R}) \leq \mu'(\mathfrak{R})$ for all $\mathfrak{R} \in \mathcal{C}_*$.

Define

$$\mathcal{Q} = \{\mathfrak{R} \in \sigma(\mathcal{C}) \mid x \cdot \mu(\mathfrak{R}) \leq \mu'(\mathfrak{R})\}\,.$$

We have $\mathcal{C} \subseteq \mathcal{C}_* \subseteq \mathcal{Q} \subseteq \sigma(\mathcal{C})$. We show that $\mathcal{Q}$ is a *monotone* class, i.e., if $\mathfrak{R}_1, \mathfrak{R}_2, \ldots \in \mathcal{Q}$, then $\mathfrak{R}_1 \subseteq \mathfrak{R}_2 \subseteq \cdots$ implies $\bigcup_i \mathfrak{R}_i \in \mathcal{Q}$, and $\mathfrak{R}_1 \supseteq \mathfrak{R}_2 \supseteq \cdots$ implies $\bigcap_i \mathfrak{R}_i \in \mathcal{Q}$. Suppose $\mathfrak{R}_1, \mathfrak{R}_2, \ldots \in \mathcal{Q}$ and $\mathfrak{R}_1 \subseteq \mathfrak{R}_2 \subseteq \cdots$. Then:

$$\begin{aligned} x \cdot \mu\Big(\bigcup_i \mathfrak{R}_i\Big) &= \sup_i x \cdot \mu(\mathfrak{R}_i) && \text{measures are continuous from below} \\ &\leq \sup_i \mu'(\mathfrak{R}_i) && \text{definition of } \mathcal{Q} \\ &= \mu'\Big(\bigcup_i \mathfrak{R}_i\Big) && \text{measures are continuous from below} \end{aligned}$$

So $\bigcup_i \mathfrak{R}_i \in \mathcal{Q}$. Using the fact that measures are continuous from above, one can similarly show that if $\mathfrak{R}_1, \mathfrak{R}_2, \ldots \in \mathcal{Q}$ and $\mathfrak{R}_1 \supseteq \mathfrak{R}_2 \supseteq \cdots$ then $\bigcap_i \mathfrak{R}_i \in \mathcal{Q}$. Hence $\mathcal{Q}$ is a monotone class.

Now the *monotone class theorem* (see, e.g., [4, Theorem 3.4]) implies that $\sigma(\mathcal{C}) \subseteq \mathcal{Q}$, thus $\mathcal{Q} = \sigma(\mathcal{C})$. Hence $x \cdot \mu(\mathfrak{R}) \leq \mu'(\mathfrak{R})$ for all $\mathfrak{R} \in \sigma(\mathcal{C})$. ◀

We prove Lemma 16 from the main body:

▶ **Lemma 16.** *Let $\mathcal{M} = (S, S_\Box, S_\bigcirc, \longrightarrow, P)$ be an MDP, and let φ be an objective that is tail in $\mathcal{M}$. Let $\mathcal{M}_* = (S_*, S_{*\Box}, S_{*\bigcirc}, \longrightarrow_*, P_*)$ be the conditioned version of $\mathcal{M}$ w.r.t. φ, where $s_\perp$ is replaced by an infinite chain $s_\perp^1 \longrightarrow s_\perp^2 \longrightarrow \cdots$. If $\mathcal{M}$ is universally transient, then so is $\mathcal{M}_*$.*

Proof. For any state $s_0 \in S_* \cap S$, let

$$\mathfrak{R}_{s_0} \stackrel{\text{def}}{=} \{s_0 s_1 \cdots \in s_0 S_*^\omega \mid \exists\, i \geq 1 : s_0 = s_i\}$$

denote the event of returning to s_0. Suppose $\mathcal{M}_*$ is not universally transient. By Lemma 3(3) there exists $s_0 \in S_* \cap S$ such that $\mathtt{val}_{\mathcal{M}_*, \mathfrak{R}_{s_0}}(s_0) = 1$. We show that, in $\mathcal{M}$, for any $C > 0$ there exists a strategy under which the expected number of returns to s_0 is at least C. By Lemma 3(4) this implies that $\mathcal{M}$ is not universally transient.

Let $C > 0$. Let $\mathfrak{R}$ be the event, in $\mathcal{M}_*$, starting in s_0, of returning to s_0 at least $2C/\mathtt{val}_{\mathcal{M},\varphi}(s_0)$ times, and denote by X the random variable counting the number of returns to s_0. Since $\mathtt{val}_{\mathcal{M}_*, \mathfrak{R}_{s_0}}(s_0) = 1$, we also have $\mathtt{val}_{\mathcal{M}_*, \mathfrak{R}}(s_0) = 1$, and so there exists a strategy σ with $\mathcal{P}_{\mathcal{M}_*, s_0, \sigma}(\mathfrak{R}) \geq \frac{1}{2}$. By the first inequality of Lemma 13.2 we have $\mathcal{P}_{\mathcal{M}, s_0, \sigma}(\overline{\mathfrak{R}}) \geq \mathtt{val}_{\mathcal{M},\varphi}(s_0) \cdot \frac{1}{2}$. It follows:

$$\mathcal{E}_{\mathcal{M}, s_0, \sigma}(X) \;\geq\; \mathcal{P}_{\mathcal{M}, s_0, \sigma}(\overline{\mathfrak{R}}) \cdot 2C/\mathtt{val}_{\mathcal{M},\varphi}(s_0) \;\geq\; C$$ ◀

In [17, Lemma 6] a variant, say $\mathcal{M}_+$, of the conditioned MDP $\mathcal{M}_*$ from Definition 12 was proposed. This variant $\mathcal{M}_+$ differs from $\mathcal{M}_*$ in that $\mathcal{M}_+$ has only those states s from $\mathcal{M}$ that have an optimal strategy, i.e., a strategy σ with $\mathcal{P}_{\mathcal{M}, s, \sigma}(\varphi) = \mathtt{val}_{\mathcal{M}}(s)$. Further, for any transition $s \longrightarrow t$ in $\mathcal{M}_+$ where s is a controlled state, we have $\mathtt{val}_{\mathcal{M}}(s) = \mathtt{val}_{\mathcal{M}}(t)$, i.e., $\mathcal{M}_+$ does not have value-decreasing transitions emanating from controlled states.

As a consequence, in contrast to $\mathcal{M}_*$, in $\mathcal{M}_+$ there is no need for intermediate states of the form (s, t): Since $\mathtt{val}_{\mathcal{M}}(s) = \mathtt{val}_{\mathcal{M}}(t)$, an intermediate state (s, t) would transition to t with probability 1. Therefore, such intermediate states do not appear in $\mathcal{M}_+$. Instead, in $\mathcal{M}_+$ there is a direct transition from s to t like in the original MDP $\mathcal{M}$. As a further consequence, the state $s_\perp$ does not appear in $\mathcal{M}_+$ (it would not be reachable).

Any strategy σ in $\mathcal{M}_+$ can be naturally applied also in $\mathcal{M}_*$: whenever σ moves from a controlled state s to a state t (hence s and t have the same value), in $\mathcal{M}_*$ strategy σ moves instead to the random state (s, t) (from which $\mathcal{M}_*$ transitions to t with probability 1).

This correspondence is exploited in the proof of the following lemma from the main body:

▶ **Lemma 17.** *Let $\mathcal{M}$ be an MDP, and let φ be an objective that is tail in $\mathcal{M}$. Let $\mathcal{M}_+$ be the conditioned version w.r.t. φ in the sense of [17, Lemma 6]. If $\mathcal{M}$ is universally transient, then so is $\mathcal{M}_+$.*

Proof. Suppose $\mathcal{M}$ is universally transient. We show that $\mathcal{M}_+$ is universally transient. Indeed, let s_0 be any state in $\mathcal{M}_+$, and let σ be any strategy in $\mathcal{M}_+$. Write $\mathfrak{R}$ for the event of returning to s_0 in $\mathcal{M}_*$, and $\overline{\mathfrak{R}}$ for the event of returning to s_0 in $\mathcal{M}_+$. We have $\mathcal{P}_{\mathcal{M}_+,s_0,\sigma}(\overline{\mathfrak{R}}) = \mathcal{P}_{\mathcal{M}_*,s_0,\sigma}(\mathfrak{R})$. Since $\mathcal{M}_*$ is universally transient by Lemma 16, by Lemma 3(3) this probability is less than 1. Applying Lemma 3(3) again, it follows that $\mathcal{M}_+$ is universally transient. ◀