Strategy Complexity of Parity Objectives in Countable MDPs

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

Richard Mayr
School of Informatics, University of Edinburgh, UK

Mahsa Shirmohammadi
CNRS & IRIF, Université de Paris, FR

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

──── **Abstract** ────

We study countably infinite MDPs with parity objectives. Unlike in finite MDPs, optimal strategies need not exist, and may require infinite memory if they do. We provide a complete picture of the exact strategy complexity of ε-optimal strategies (and optimal strategies, where they exist) for all subclasses of parity objectives in the Mostowski hierarchy. Either MD-strategies, Markov strategies, or 1-bit Markov strategies are necessary and sufficient, depending on the number of colors, the branching degree of the MDP, and whether one considers ε-optimal or optimal strategies. In particular, 1-bit Markov strategies are necessary and sufficient for ε-optimal (resp. optimal) strategies for general parity objectives.

1 Introduction

Background. Markov decision processes (MDPs) are a standard model for dynamic systems that exhibit both stochastic and controlled behavior [17]. MDPs play a prominent role in numerous domains, including artificial intelligence and machine learning [20, 19], control theory [4, 1], operations research and finance [5, 18], and formal verification [7, 2].

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 type of strategy necessary to achieve an optimal (resp. ε-optimal) value for a given objective is called its *strategy complexity*. There are different types of strategies, depending on whether one can take the whole history of the run into account (history-dependent; (H)), or whether one is limited to a finite amount of memory (finite memory; (F)) or whether decisions are based 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. *Markov strategies* are strategies that base their decisions only on the current state and the number of steps in the history of the run. Thus they do use infinite memory, but only in a very
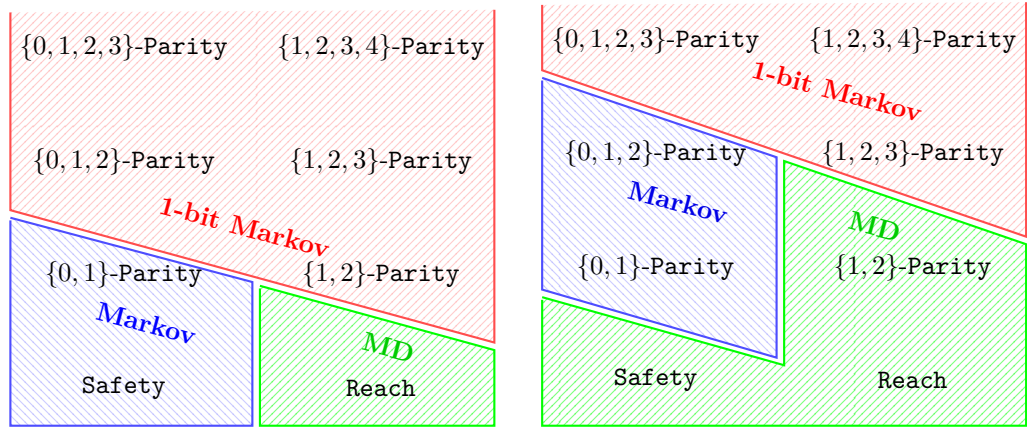
(a) ε-optimal strategies for infinitely branching MDPs.

(b) Optimal strategies for infinitely branching MDPs.

■ **Figure 1** These diagrams show the strategy complexity of ε-optimal strategies and optimal strategies (where they exist) for parity objectives. Depending on the position in the Mostowski hierarchy, either MD-strategies (green), deterministic Markov-strategies (blue) or deterministic 1-bit Markov strategies (red) are necessary and sufficient (and randomization does not help [12]). If the MDPs are finitely branching then the Markov strategies can be replaced by MD-strategies (i.e., the blue parts turn green), but the deterministic 1-bit Markov part (red) remains unchanged.

restricted form by maintaining an unbounded step-counter. Slightly more general are *1-bit Markov strategies* that use 1 bit of extra memory in addition to a step-counter.

Parity objectives. We study countably infinite MDPs with parity objectives. Parity conditions are widely used in temporal logic and formal verification, e.g., they can express ω-regular languages and modal μ-calculus [9]. Every state has a *color*, out of a finite set of colors encoded as natural numbers. A run is winning iff the highest color that is seen infinitely often is even. The controller wants to maximize the probability of winning runs. The Mostowski hierarchy [15] is a classification of parity conditions based on restricting the set of allowed colors. For instance, $\{1,2,3\}$-Parity objectives only use colors 1, 2, and 3. This includes Büchi ($\{1,2\}$-Parity) and co-Büchi objectives ($\{0,1\}$-Parity), both of which further subsume reachability and safety objectives.

Related work. In *finite* MDPs, there always exist optimal MD-strategies for parity objectives. In fact, this holds even for finite turn-based 2-player stochastic parity games [6, 23]. Similarly, there always exist optimal MD-strategies in countably infinite *non-stochastic* turn-based 2-player parity games [22].

The picture is more complex for countably infinite MDPs. Optimal strategies need not exist (not even for reachability objectives [17, 16]), and ε-optimal strategies for Büchi objectives [10] and optimal strategies for parity objectives [13] require infinite memory.

The paper [13] gave a complete classification whether MD-strategies suffice or whether infinite memory is required for ε-optimal (resp. optimal) strategies for all subclasses of parity objectives in the Mostowski-hierarchy.

However, the mere fact that infinite memory is required for (a subclass of) parity does not establish the precise strategy complexity. E.g., are Markov strategies (or Markov strategies with finite extra memory) sufficient?

In [12] we showed that deterministic 1-bit Markov strategies are both necessary and sufficient for ε-optimal strategies for Büchi objectives. I.e., deterministic 1-bit Markov strategies

are sufficient, but neither randomized Markov strategies nor randomized finite-memory strategies are sufficient. This solved a 40-year old problem in gambling theory from [10, 11]. The same paper [12] showed that even for finitely branching MDPs with $\{1, 2, 3\}$-`Parity` objectives, optimal strategies (where they exist) need to be *at least* deterministic 1-bit Markov in general, i.e., neither randomized Markov nor randomized finite-memory strategies are sufficient.

While the lower bounds for ε-optimal strategies for Büchi objectives (resp. for optimal strategies for $\{1, 2, 3\}$-`Parity` objectives) carry over to general parity objectives, the upper bounds on the strategy complexity of ε-optimal (resp. optimal) parity remained open.

A basic upper bound and related conjecture. A basic upper bound on the complexity of ε-optimal strategies for parity can be obtained by using a combination of the results of [12] on Büchi objectives (1-bit Markov) and Lévy's zero-one law as follows. (However, note that the following argument does not work directly for optimal strategies.)

Informally speaking, Lévy's zero-one law implies that, for a tail objective (like parity) and any strategy, the level of attainment from the current state almost surely converges to either zero or one. I.e., the runs that always stay in states where the strategy attains something in $(0, 1)$ is a null-set (cf. Appendix B). A consequence for parity is that almost all winning runs must eventually, with ever higher probability, commit to winning by some particular color. Thus, with minimal losses (e.g., $\varepsilon/2$), after a sufficiently long finite prefix (depending on ε), one can switch to a strategy that aims to visit some *particular* color x infinitely often. The latter objective is like a Büchi objective where the states of color x are accepting and states of color $> x$ are considered losing sinks. By [12], an $\varepsilon/2$-optimal strategy for such a Büchi objective can be chosen 1-bit Markov. However, one would also need to remember which color x one is supposed to win by and *stick to that color*. The latter is critical, since strategies that switch focus between winning colors infinitely often (e.g., if they follow some local criteria based on the value of the current state wrt. various colors) can end up losing. Overall, the memory needed for such an ε-optimal strategy for parity is: $\lceil \log_2(c) \rceil$ bits for c even colors to remember which color x one is supposed to win by and Markov plus 1 bit for the Büchi strategy (see above), where the Markov step-counter also determines whether one still plays in the prefix. Thus Markov plus $(1 + \lceil \log_2(c) \rceil)$ bits are sufficient. This argument would suggest that more memory is required for more colors. However, our result shows that this is *not* the case.

Our contributions. We show *tight* upper bounds on the strategy complexity of ε-optimal (resp. optimal) strategies for parity objectives: They can be chosen as deterministic 1-bit Markov, regardless of the number of colors. I.e., we provide matching upper bounds to the lower bounds from [12].

In Section 3 we prove Theorem 1. An iterative plastering construction (i.e., fixing player choices on larger and larger subspaces) builds an ε-optimal 1-bit Markov strategy where the probability of never switching between winning even colors is $\geq 1 - \varepsilon$. Its correctness relies heavily on Lévy's zero-one law. The number of iterations is finite and proportional to the number of even colors. It eliminates the need to remember the winning color x and the $\lceil \log_2(c) \rceil$ part of the memory.

▶ **Theorem 1.** *Consider an MDP $\mathcal{M}$, a parity objective and a finite set S_0 of initial states. For every $\varepsilon > 0$ there exists a deterministic 1-bit Markov strategy that is ε-optimal from every state $s \in S_0$.*

In Section 4 we prove Theorem 2. If an optimal strategy exists, then an optimal 1-bit Markov strategy can be constructed by the so-called *sea urchin* construction. It is a very

complex plastering construction with infinitely many iterations that uses the results of Theorem 1 and Lévy's zero-one law as building blocks. Its name comes from the shape of the subspace in which player choices get fixed: a growing finite body (around a start set S_0) with a finite, but increasing, number of spikes, where each spike is of infinite size; cf. Figure 4. E.g., if the initial states are almost surely winning then, at the stage with i spikes, this strategy attains parity with some probability $\geq 1 - 2^{-i}$ already *inside* this subspace, and in the limit of $i \to \infty$ it attains parity almost surely. A further step even yields a single deterministic 1-bit Markov strategy that is optimal from every state that has an optimal strategy.

▶ **Theorem 2.** *Consider an MDP $\mathcal{M}$ with a parity objective and let S_{opt} be the subset of states that have an optimal strategy.*

     There exists a deterministic 1-bit Markov strategy that is optimal from every $s \in S_{opt}$.

In Theorem 1 and Theorem 2 the initial content of the 1-bit memory is irrelevant (cf. Lemma 9, Lemma 18 and Remark 8).

Moreover, we show in Section 5 and Section 6 that in certain subcases deterministic Markov strategies are necessary and sufficient (i.e., these require a Markov step-counter, but not the extra bit): optimal strategies for co-Büchi and $\{0, 1, 2\}$-`Parity`, and ε-optimal strategies for safety and co-Büchi. In the special case of finitely branching MDPs, these Markov strategies (but not the 1-bit Markov strategies) can be replaced by MD-strategies.

Together with the previously established lower bounds, this yields a complete picture of the *exact* strategy complexity of parity objectives at all levels of the Mostowski hierarchy, for countable MDPs. Figure 1 gives a complete overview.

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.

We study *Markov decision processes* (MDPs) over countably infinite state spaces. Formally, an MDP $\mathcal{M} = (S, S_\square, S_\bigcirc, \longrightarrow, P)$ consists of a countable set S of *states*, which is partitioned into a set $S_\square$ of *controlled states* and a set $S_\bigcirc$ of *random states*, a *transition relation* $\longrightarrow \subseteq S \times S$, and a *probability function* $P : S_\bigcirc \to \mathcal{D}(S)$. 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 ($S_\square = \emptyset$) 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 $i \in \mathbb{N}$; write $\rho(i) \stackrel{\text{def}}{=} s_i$ for the i-th state along ρ. A *partial run* is a finite prefix of a run. We say that (partial) run ρ *visits* s if $s = \rho(i)$ for some i, and that ρ *starts in* s if $s = \rho(0)$.

A *strategy* is a function $\sigma : S^* S_\square \to \mathcal{D}(S)$ that assigns to partial runs $\rho s \in S^* S_\square$ a distribution over the successors of s. A (partial) run $s_0 s_1 \cdots$ is *induced by* strategy σ if for all i either $s_i \in S_\square$ and $\sigma(s_0 s_1 \cdots s_i)(s_{i+1}) > 0$, or $s_i \in S_\bigcirc$ and $P(s_i)(s_{i+1}) > 0$.

A strategy σ and an initial state $s_0 \in S$ induce a standard probability measure on sets of infinite plays. We write $\mathbb{P}_{\mathcal{M}, s_0, \sigma}(\mathcal{R})$ for the probability of a measurable set $\mathcal{R} \subseteq s_0 S^\omega$ of runs starting from s_0. As usual, it is first defined on the *cylinders* $s_0 s_1 \ldots s_n S^\omega$, where

$s_1, \ldots, s_n \in S$: if $s_0 s_1 \ldots s_n$ is not a partial run induced by σ then $\mathbb{P}_{\mathcal{M}, s_0, \sigma}(s_0 s_1 \ldots s_n S^\omega) \stackrel{\text{def}}{=} 0$. Otherwise, $\mathbb{P}_{\mathcal{M}, s_0, \sigma}(s_0 s_1 \ldots s_n S^\omega) \stackrel{\text{def}}{=} \prod_{i=0}^{n-1} \bar{\sigma}(s_0 s_1 \ldots s_i)(s_{i+1})$, 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 [3], this extends uniquely to a probability measure $\mathbb{P}_{\mathcal{M}, s_0, \sigma}$ on measurable subsets of $s_0 S^\omega$. We will write $\mathbb{E}_{\mathcal{M}, s_0, \sigma}$ for the expectation w.r.t. $\mathbb{P}_{\mathcal{M}, s_0, \sigma}$. We may drop the subscripts from notations, if it is understood.

Objectives. The objective of the player is determined by a predicate on infinite plays. We assume familiarity with the syntax and semantics of the temporal logic LTL [8]. Formulas are interpreted on the structure $(S, \longrightarrow)$. We use $\llbracket \varphi \rrbracket^s \subseteq sS^\omega$ to denote the set of runs starting from s that satisfy the LTL formula φ, which is a measurable set [21]. We also write $\llbracket \varphi \rrbracket$ for $\bigcup_{s \in S} \llbracket \varphi \rrbracket^s$. Where it does not cause confusion we will identify φ and $\llbracket \varphi \rrbracket$ and just write $\mathbb{P}_{\mathcal{M}, s, \sigma}(\varphi)$ instead of $\mathbb{P}_{\mathcal{M}, s, \sigma}(\llbracket \varphi \rrbracket^s)$.

Given a set $T \subseteq S$ of states, the *reachability* objective $\mathtt{Reach}(T)$ is the set of runs that visit T at least once; and the *safety objective* $\mathtt{Safety}(T)$ 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 $\mathtt{Parity}(Col)$, 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 $\rhd \in \{<, \leq, =, \geq, >\}$, $n \in \mathbb{N}$, and $Q \subseteq S$, let $[Q]^{Col \rhd n} \stackrel{\text{def}}{=} \{s \in Q \mid Col(s) \rhd n\}$ be the set of states in Q with color $\rhd n$. Then

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

The Mostowski hierarchy [15] classifies parity objectives by restricting the range of Col to a set of colors $\mathcal{C} \subseteq \mathbb{N}$. We write $\mathcal{C}$-$\mathtt{Parity}$ for such restricted parity objectives. In particular, the classical Büchi and co-Büchi objectives correspond to $\{1, 2\}$-$\mathtt{Parity}$ and $\{0, 1\}$-$\mathtt{Parity}$, respectively. These two classes are incomparable but both subsume the reachability and safety objectives. Assuming that T is a sink, $\mathtt{Reach}(T) = \mathtt{Parity}(Col)$ for the coloring with $Col(s) = 1 \iff s \notin T$ and $\mathtt{Safety}(T) = \mathtt{Parity}(Col)$ for the coloring with $Col(s) = 1 \iff s \in T$. Similarly, $\{0, 1, 2\}$-$\mathtt{Parity}$ and $\{1, 2, 3\}$-$\mathtt{Parity}$ are incomparable, but they both subsume (modulo renaming of colors) Büchi and co-Büchi objectives.

An objective φ is called a *tail objective* (resp. *suffix-closed*) iff for every run $\rho' \rho$ with some finite prefix ρ' we have $\rho' \rho \in \varphi \Leftrightarrow \rho \in \varphi$ (resp. $\rho' \rho \in \varphi \Rightarrow \rho \in \varphi$). In particular, $\mathtt{Parity}(Col)$ is tail for every coloring Col. Moreover, if φ is suffix-closed then $\mathsf{F}\varphi$ is tail.

Strategy Classes. Strategies $\sigma : S^* S_\square \to \mathcal{D}(S)$ 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 $\rho \in S^* S_\square$.

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 $s \in S_\square$, 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 $u[\mathsf{m}_0] : S^* S_\square \to \mathcal{D}(S)$ 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 $\rho s \in s_0 S^* S_\square$ induces a probability distribution $\mu_{\rho s} \in \mathcal{D}(\mathsf{M})$, 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 $\rho s \in S^* S_\square$ 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 $\sigma : S_\square \to \mathcal{D}(S)$.
- A strategy σ is *finite memory* (F) if there exists a finite memory M implementing σ. More specifically, a strategy is *k-bit* if it can be implemented with a memory of size 2^k. Such a strategy is then determined by a function $u : \{0, 1\}^k \times S \to \mathcal{D}(\{0, 1\}^k \times S)$.
- A strategy σ is *Markov* if it can be implemented with the natural numbers $\mathsf{M} = \mathbb{N}$ as the memory, initial memory mode $\mathsf{m}_0 = 0$ and a function u such that the distribution $u(\mathsf{m}, s)$ is over $\{\mathsf{m} + 1\} \times S$ for all $\mathsf{m} \in \mathsf{M}$ and $s \in S$. Intuitively, such a strategy depends only on the current state and the number of steps taken so far.
- A strategy σ is *k-bit Markov* if it can be implemented with memory $\mathsf{M} = \mathbb{N} \times \{0, 1\}^k$, $\mathsf{m}_0 \in \{0\} \times \{0, 1\}^k$ and a function u such that the distribution $u((n, b, s))$ is over $\{n + 1\} \times \{0, 1\}^k \times S$ for all $(n, b) \in \mathsf{M}$ and $s \in S$.

Deterministic 1-bit strategies are central in this paper; by this we mean strategies that are both deterministic and 1-bit.

Optimal and ε-optimal Strategies. Given an objective φ, the value of state s in an MDP $\mathcal{M}$, denoted by $\mathtt{val}_{\mathcal{M}}(s)$, is the supremum probability of achieving φ. Formally, we have $\mathtt{val}_{\mathcal{M}}(s) \stackrel{\text{def}}{=} \sup_{\sigma \in \Sigma} \mathbb{P}_{\mathcal{M}, s, \sigma}(\varphi)$ 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 $\mathbb{P}_{\mathcal{M}, s, \sigma}(\varphi) \geq \mathtt{val}_{\mathcal{M}}(s) - \varepsilon$. A 0-optimal strategy is called optimal. An optimal strategy is almost-surely winning if $\mathtt{val}_{\mathcal{M}}(s) = 1$.

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

Fixing and Safe Sets. Let σ be an MD strategy. Given a set $S' \subseteq S$ of states, write $\mathcal{M}[\sigma, S']$ for the MDP obtained from $\mathcal{M}$ by fixing the strategy σ for all states in S', that is, $\mathcal{M}[\sigma, S'] \stackrel{\text{def}}{=} (S, S_\square \setminus S', S_\bigcirc \cup S', \longrightarrow, P')$ where $P'(s) \stackrel{\text{def}}{=} \sigma(s)$ for all $s \in S'$.

   For an objective φ and a threshold $\beta \in [0, 1]$, denote by $Safe_{\mathcal{M}, \sigma, \varphi}(\beta)$ the set of all states s starting from which σ attains at least probability β; and denote by $Safe_{\mathcal{M}, \varphi}(\beta)$ the set of states whose value for φ is at least β. Formally,

$$Safe_{\mathcal{M}, \sigma, \varphi}(\beta) \stackrel{\text{def}}{=} \{s \in S \mid \mathbb{P}_{\mathcal{M}, s, \sigma}(\varphi) \geq \beta\}, \quad Safe_{\mathcal{M}, \varphi}(\beta) \stackrel{\text{def}}{=} \{s \in S \mid \mathtt{val}_{\mathcal{M}, \varphi}(s) \geq \beta\}. \quad (1)$$

3 ε-Optimal Strategies for Parity

In this section we prove Theorem 1, stating that ε-optimal strategies for parity objectives can be chosen 1-bit Markov. Given an MDP we convert it by three successive reductions to a structurally simpler MDP where strategies require less sophistication to achieve parity.

First reduction (Finitely Branching). This reduction converts an infinitely branching MDP $\mathcal{M}$ to a finitely branching one $\mathcal{M}'$, with a clear bijection between the strategies in $\mathcal{M}$ and $\mathcal{M}'$. The construction, first presented in our previous work [12], replaces each controlled state s, that has infinitely many successors $(s_i)_{i \in \mathbb{N}}$, with a "ladder" of controlled states $(q_i)_{i \in \mathbb{N}}$, where each q_i has only two successors: q_{i+1} and s_i. Roughly speaking, the controller

choice of successor s_n at s in $\mathcal{M}$, is simulated by a series of choices q_{i+1} at q_i, $0 \leq i < n$, followed by a choice of successor s_n in state q_n in $\mathcal{M}'$, and vice versa.

To prevent scenarios when the controller in $\mathcal{M}'$ stays on a ladder and never commits to a decision, we assign color 1 to all states $(q_i)_{i \geq 1}$ on the ladder (q_0 inherits the color of s). Hence, a hesitant run on the ladder is losing for parity. So w.l.o.g. we can assume that the given $\mathcal{M}$ is finitely branching.

▶ **Lemma 3.**
1. *Suppose that for every finitely branching acyclic MDP with a finite set S_0 of initial states, and a parity objective, there exist ε-optimal deterministic 1-bit strategies from S_0.*
   Then even for every infinitely branching acyclic MDP with a finite set S_0 of initial states and a parity objective, there exist ε-optimal deterministic 1-bit strategies from S_0.
2. *Suppose that for every finitely branching acyclic MDP with a parity objective, there exists a deterministic 1-bit strategy that is optimal from all states that have an optimal strategy. Then even for every infinitely branching acyclic MDP with a parity objective, there exists a deterministic 1-bit strategy that is optimal from all states that have an optimal strategy.*

Second reduction (Acyclicity). A deterministic 1-bit Markov strategy can be seen as a function $\sigma : \mathbb{N} \times \{0, 1\} \times S \to \{0, 1\} \times S$, where σ has access to an internal bit $b \in \{0, 1\}$, which can be updated freely, and a step counter $k \in \mathbb{N}$, which increments by one in each step. Having b and k, σ produces a decision based on the current state of the MDP.

Following [12], we encode the step-counter from strategies into MDPs s.t. the current state of the system uniquely determines the length of the path taken so far. This translation allows us to focus on acyclic MDPs.

▶ **Lemma 4.** *Consider MDPs with a parity objective and $k \in \mathbb{N}$.*
1. *Suppose that for every acyclic MDP $\mathcal{M}'$ and every finite set of initial states S_0' and $\varepsilon > 0$, there exists a deterministic k-bit strategy that is ε-optimal from all states $s \in S_0'$.*
   Then for every MDP $\mathcal{M}$ and every finite set of initial states S_0 and $\varepsilon > 0$, there exists a deterministic k-bit Markov strategy that is ε-optimal from all states $s \in S_0$.
2. *Suppose that for every acyclic MDP $\mathcal{M}'$ and $\varepsilon > 0$, there exists a deterministic k-bit strategy that is ε-optimal from all states. Then for every MDP $\mathcal{M}$ and $\varepsilon > 0$, there exists a deterministic k-bit Markov strategy that is ε-optimal from all states.*
3. *Suppose that for every acyclic MDP $\mathcal{M}'$, where S_{opt}' is the subset of states that have an optimal strategy, there exists a deterministic k-bit strategy that is optimal from all states $s \in S_{opt}'$. Then for every MDP $\mathcal{M}$, where S_{opt} is the subset of states that have an optimal strategy, there exists a deterministic k-bit Markov strategy that is optimal from all states $s \in S_{opt}$.*

By Lemma 4, the sufficiency of deterministic 1-bit strategies in acyclic MDPs implies the sufficiency of deterministic 1-bit Markov strategies in general MDPs. Thus to prove Theorem 1, it suffices to prove the following:

▶ **Theorem 5.** *Consider an acyclic MDP $\mathcal{M}$, a parity objective and a finite set S_0 of states. For every $\varepsilon > 0$ there exists a deterministic 1-bit strategy that is ε-optimal from every $s \in S_0$.*

Third reduction (Layered MDP). This reduction is in the same spirit of the previous one, in which the bit $b \in \{0, 1\}$ is transferred from strategies to MDPs. Given an MDP $\mathcal{M}$, the corresponding *layered* MDP $\mathcal{L}(\mathcal{M})$ has two copies of each state $s \in S$ and each transition $t \in \longrightarrow_1$ of $\mathcal{M}$, one augmented with bit 0 and another with bit 1: (s, i) and (t, j) with $i, j \in \{0, 1\}$. The states (s, i) are random if $s \in S_\bigcirc$ and controlled if $s \in S_\square$. All the (t, j) are controlled.

If there is a transition $t = (a, b)$ from state a to b in $\mathcal{M}$, there will be two transitions from (a, i) to (t, i), and four transitions from (t, i) to (b, j) in $\mathcal{L}(\mathcal{M})$; see Figure 2.

A 1-bit deterministic strategy in $\mathcal{M}$ at a state a picks a single successor b and may flip the bit from i to j; this is simulated in $\mathcal{L}(\mathcal{M})$ with an MD strategy σ within two consecutive steps: σ first chooses the transition $t = (a, b)$ by $\sigma(a, i) = (t, i)$ and then updates the bit by $\sigma(t, i) = (b, j)$ thereby moving from layer i to layer j. The controlled states (t, i) are essential for a correct simulation, since otherwise the controller cannot freely flip the bit (switch between layers) after it observes the successor chosen randomly at a random state.

▶ **Definition 6** (Layered MDP). *Given an MDP $\mathcal{M} = (S, S_\square, S_\bigcirc, \longrightarrow_1, P_1)$ with coloring $Col_1 : S \to \mathcal{C}$, we define the corresponding* layered MDP $\mathcal{L}(\mathcal{M}) = (L, L_\square, L_\bigcirc, \longrightarrow_2, P_2)$ *with coloring $Col_2 : L \to \mathcal{C}$ as follows.*

- *$L \stackrel{def}{=} (S \cup \longrightarrow_1) \times \{0, 1\}$ where the set of controlled states is $L_\square \stackrel{def}{=} (S_\square \cup \longrightarrow_1) \times \{0, 1\}$.*
- *For all $t \in \longrightarrow_1$ such that $t = (s, s')$ and for all $i, j \in \{0, 1\}$, we have:*
  1. *$(s, i) \longrightarrow_2 (t, i)$ and $(t, i) \longrightarrow_2 (s', j)$,*
  2. *$P(s, i)((t, i)) \stackrel{def}{=} P(s)(s')$ iff $s \in S_\bigcirc$, and*
  3. *$Col_2((s, i)) \stackrel{def}{=} Col_1(s)$ and $Col_2((t, i)) \stackrel{def}{=} Col_1(s')$.*

The layered MDP of an acyclic MDP is acyclic. For $q \in S \cup \longrightarrow_1$, we refer to the copies of q in layer 0 and layer 1 as *siblings*: $(q, 0)$ and $(q, 1)$. A set $B \subseteq L$ is *closed* if for each state $(q, i) \in B$ its sibling is also in B. Denote by $Cl(B)$ the minimal closed superset of B.

▶ **Lemma 7.** *Consider an acyclic MDP $\mathcal{M} = (S, S_\square, S_\bigcirc, \longrightarrow, P)$ with a parity objective $\varphi = \mathtt{Parity}(Col)$ and let $\mathcal{L}(\mathcal{M})$ be the corresponding layered MDP.*

For every deterministic 1-bit strategy $u[\mathsf{m}_0]$ in $\mathcal{M}$ there is a corresponding MD strategy τ in $\mathcal{L}(\mathcal{M})$, and vice-versa, such that for every $s_0 \in S$, $\mathbb{P}_{\mathcal{L}(\mathcal{M}),(s_0,\mathsf{m}_0),\tau}(\varphi) = \mathbb{P}_{\mathcal{M},s_0,u[\mathsf{m}_0]}(\varphi)$.

▶ Remark 8. We note that, in a layered system $\mathcal{L}(\mathcal{M})$, any two siblings have the same value w.r.t. a parity objective φ. Moreover, any state s in $\mathcal{M}$ has an optimal strategy iff $(s, 0) \in \mathcal{L}(\mathcal{M})$ has an optimal strategy iff its sibling $(s, 1)$ has an optimal strategy.

Suppose τ is an MD strategy in $\mathcal{L}(\mathcal{M})$ that is optimal for all states that have an optimal strategy. Let u be the update function of a corresponding 1-bit strategy in $\mathcal{M}$, derived as described in Lemma 7. Then for every state s in $\mathcal{M}$ that has an optimal strategy we have $\mathbb{P}_{\mathcal{M},s,u[0]}(\varphi) = \mathbb{P}_{\mathcal{L}(\mathcal{M}),(s,0),\tau}(\varphi) = \mathbb{P}_{\mathcal{L}(\mathcal{M}),(s,1),\tau}(\varphi) = \mathbb{P}_{\mathcal{M},s,u[1]}(\varphi)$. That is, both $u[0]$ and $u[1]$ are optimal from s, so the initial memory mode is irrelevant. ◀

To prove Theorem 5, given an acyclic MDP, a set of initial states S_0 and $\varepsilon > 0$, we consider the layered MDP $\mathcal{L}(\mathcal{M})$ and set $L_0 = S_0 \times \{0\}$ of initial states. In the following lemma, we prove that there exists a single MD strategy that is ε-optimal starting from every state $\ell_0 \in L_0$ in $\mathcal{L}(\mathcal{M})$. This and Lemma 7 will directly lead to Theorem 5.

▶ **Lemma 9.** *Consider an acyclic MDP $\mathcal{M}$ and parity objective $\varphi = \mathtt{Parity}(Col)$. Let $\mathcal{L}(\mathcal{M})$ be the layered MDP of $\mathcal{M}$ and Col. For all finite sets L_0 of states in $\mathcal{L}(\mathcal{M})$ and all $\varepsilon > 0$ there exists a single MD strategy that is ε-optimal for φ from every state $\ell_0 \in L_0$.*

In the rest of this section, we prove Lemma 9. We fix a layered MDP $\mathcal{L}(\mathcal{M})$ (or simply $\mathcal{L}$) obtained from a given acyclic and finitely branching MDP $\mathcal{M}$ and a coloring $Col : S \to \mathcal{C}$, where the set of states is L and the finite set of initial states is $L_0 \subseteq L$. Let φ be the resulting parity objective in $\mathcal{L}$.

Recall that $even(\mathcal{C}) = 2\mathbb{N} \cap \mathcal{C}$ denotes the set of even colors. We denote by $e_{\max}$ the largest even color in $even(\mathcal{C})$ and assume w.l.o.g., that $even(\mathcal{C})$ contains all even numbers
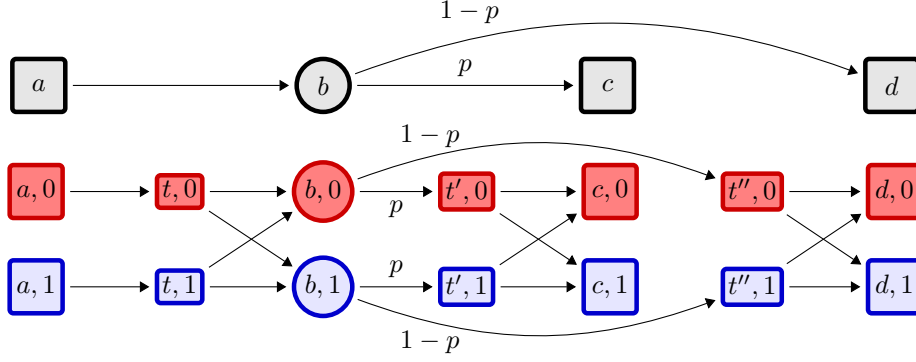
Figure 2 An MDP $\mathcal{M}$ (in grey) and the corresponding layered MDP $\mathcal{L}(\mathcal{M})$ with states of layer 0 and 1 in red and blue, respectively. Here, $t = (a, b)$, $t' = (b, c)$ and $t'' = (b, d)$ are transitions of $\mathcal{M}$.

from 2 to $e_{\max}$ inclusive. We have:

$$
\begin{aligned}
\varphi &\stackrel{\text{def}}{=} \bigvee_{e \in even(\mathcal{C})} \left(\mathsf{GF}[L]^{Col=e} \wedge \mathsf{FG}[L]^{Col \leq e} \right) \\
&= \bigvee_{e \in even(\mathcal{C})} \left(\mathsf{FGF}[L]^{Col=e} \wedge \mathsf{FG}[L]^{Col \leq e} \right) && \text{since } \mathsf{GF}[L]^{Col=e} \text{ is a tail objective} \\
&= \bigvee_{e \in even(\mathcal{C})} \mathsf{F} \left(\mathsf{GF}[L]^{Col=e} \wedge \mathsf{G}[L]^{Col \leq e} \right) && \text{since } \mathsf{FG}A \wedge \mathsf{FG}B = \mathsf{F}(\mathsf{G}A \wedge \mathsf{G}B) \\
&= \bigvee_{e \in even(\mathcal{C})} \mathsf{F}\varphi_e \,,
\end{aligned}
$$

where $\varphi_e \stackrel{\text{def}}{=} \left(\mathsf{GF}[L]^{Col=e} \wedge \mathsf{G}[L]^{Col \leq e} \right)$. Indeed, φ_e is the set of runs that win through color e (i.e., by visiting color e infinitely often and never visiting larger colors). Since the $\mathsf{F}\varphi_e$ are disjoint, for all states ℓ and strategies σ, we have:

$$
\mathbb{P}_{\mathcal{L}, \ell, \sigma}(\varphi) = \sum_{e \in even(\mathcal{C})} \mathbb{P}_{\mathcal{L}, \ell, \sigma}(\mathsf{F}\varphi_e). \tag{2}
$$

Fix $\varepsilon > 0$ and define $\gamma \stackrel{\text{def}}{=} \frac{\varepsilon}{e_{\max}+2}$. To construct an MD strategy $\hat{\sigma}$ that is ε-optimal starting from every state in L_0 we have an iterative procedure. In each iteration, we define $\hat{\sigma}$ at states in some carefully chosen region; and continuing in this fashion, we gradually fix all choices of $\hat{\sigma}$. In an iteration, in order to fix "good" choices in the "right" region we need to carefully observe the behavior of finitely many $\frac{\gamma}{2}$-optimal strategies σ_{ℓ_0}, one for each $\ell_0 \in L_0$, which must respect the choices already fixed in previous iterations. We thus view these strategies σ_{ℓ_0} to be $\frac{\gamma}{2}$-optimal not in $\mathcal{L}$ but in another layered MDP that is derived from $\mathcal{L}$ after fixing the choices of partially defined $\hat{\sigma}$.

In more detail, the proof consists of exactly $\frac{e_{\max}}{2} + 1$ iterations: one iteration for each even color e and a final "reach" iteration. Starting from color 2 and $\mathcal{L}_0 \stackrel{\text{def}}{=} \mathcal{L}$, in the iteration $e \in \{2, \cdots, e_{\max}\}$, we obtain a layered MDP $\mathcal{L}_e$ from $\mathcal{L}_{e-2}$ by fixing a single choice for each controlled state in a set fix_e. Roughly speaking, a run that falls in the set fix_e is likely going to win through φ_e (win through color e). We identify a certain subspace of fix_e, referred to as core_e, such that the following crucial fact holds: Once core_e is visited the run remains in fix_e with probability at least $1 - \gamma$. At the final iteration, we fix the choices of all remaining states to maximize the probability of falling into the union of core_e sets. As

mentioned, the majority of such runs that visit $core_e$, for some color e, will stay in fix_e forever and thus win parity through color e. After all the iterations, all choices of all controlled states are fixed, and this prescribes the MD strategy $\hat{\sigma}$ from L_0 in $\mathcal{L}$.

In order to define the sets fix_e we heavily use Lévy's zero-one law and follow an inductive transformation on objectives. Lévy's zero-one states that, for a given set of (infinite) runs of a Markov chain, if we gradually observe a random run of the chain, we will become more and more certain whether the random run belongs to that set. This law has a strong implications for tail objectives. It asserts that on almost all runs $s_0 s_1 s_2 \cdots$ the limit of the value of s_i w.r.t. a tail objective tends to either 0 or 1.

In each iteration $e \in \{2, \cdots, e_{\max}\}$, we transform an objective ψ_{e-2} to a next objective ψ_e where $\psi_0 \overset{\text{def}}{=} \varphi$ is the parity objective and the result of the last transformation is $\psi_{e_{\max}} = \bigvee_{e \in even(\mathcal{C})} \mathsf{F} core_e$. We will also move from the MDP $\mathcal{L}_{e-2}$ to $\mathcal{L}_e$ after the fixings so as to maintain the following **invariant**: For all $\ell_0 \in L_0$, the value of ℓ_0 for ψ_e in $\mathcal{L}_e$ is almost as high as its value for φ in $\mathcal{L}$, that is

$$\mathtt{val}_{\mathcal{L}_e, \psi_e}(\ell_0) \geq \mathtt{val}_{\mathcal{L}, \varphi}(\ell_0) - e \cdot \gamma. \tag{3}$$

Recall that $\varphi = \bigvee_{e \in even(\mathcal{C})} \mathsf{F}\varphi_e$. Let $\mathsf{Fix}_0 \overset{\text{def}}{=} \emptyset$ and write $\mathsf{Fix}_e \overset{\text{def}}{=} \bigcup_{e' \leq e} Cl(fix_{e'})$ for $e \in \{2, 4, \cdots, e_{\max}\}$. We define:

$$\psi_0 \overset{\text{def}}{=} \bigvee_{e' > 0} \mathsf{F}\varphi_{e'} \wedge \mathsf{G} \neg \mathsf{Fix}_0 = \varphi \qquad\qquad \psi_e \overset{\text{def}}{=} \bigvee_{e' \leq e} \mathsf{F} core_{e'} \vee \bigvee_{e' > e} (\mathsf{F}\varphi_{e'} \wedge \mathsf{G} \neg \mathsf{Fix}_e). \tag{4}$$

At each transformation, we examine the disjunct $\chi_e \overset{\text{def}}{=} \mathsf{F}\varphi_e \wedge \mathsf{G} \neg \mathsf{Fix}_{e-2}$ in ψ_{e-2}. The set of runs satisfying this objective χ_e not only win through color e but also avoid the previously fixed regions. Roughly speaking, the aim is to transform χ_e to $\mathsf{F} core_e$, to move from ψ_{e-2} to ψ_e. We apply Lévy's zero-one law to deduce that the runs satisfying the χ_e are likely to enter a region that has a high value for a slightly simpler objective, namely

$$\theta_e \overset{\text{def}}{=} \varphi_e \wedge \mathsf{G} \neg \mathsf{Fix}_{e-2}. \tag{5}$$

To do so, we observe in $\mathcal{L}_{e-2}$ the behavior of several arbitrary $\frac{\gamma}{2}$-optimal strategies σ_{ℓ_0} for ψ_{e-2}, one for each $\ell_0 \in L_0$. Then, for each σ_{ℓ_0}, we apply Lévy's zero-one law separately; this provides that there exists a finite set R_e of states that have a high value for θ_e, and is reached by one of the σ_{ℓ_0} with probability as high as the probability of satisfying the disjunct χ_e. Now we use our previous results [12] on the strategy complexity of Büchi objectives and prove the existence of an MD strategy τ_e that is almost optimal for θ_e (error less than γ), starting from every state in R_e. We define sets fix_e and $core_e$ to be the set of states from which τ_e attains a high probability for θ_e in $\mathcal{L}_{e-2}$; see Figure 3. Define $\beta \overset{\text{def}}{=} 1 - \gamma$ and $\alpha \overset{\text{def}}{=} 1 - \gamma^2$, and

$$fix_e \overset{\text{def}}{=} Safe_{\mathcal{L}_{e-2}, \tau_e, \theta_e}(\beta) \qquad\qquad core_e \overset{\text{def}}{=} Safe_{\mathcal{L}_{e-2}, \tau_e, \theta_e}(\alpha). \tag{6}$$

We fix the strategy τ_e in the fix_e-region to derive the MDP $\mathcal{L}_e$ from $\mathcal{L}_{e-2}$. Formally,

$$\mathcal{L}_e \overset{\text{def}}{=} \mathcal{L}_{e-2}[\tau_e, fix_e]. \tag{7}$$

Iteration $e \in \{2, \cdots, e_{\max}\}$: For all states $\ell_0 \in L_0$, let σ_{ℓ_0} be a general (not necessarily MD) $\frac{\gamma}{2}$-optimal strategy w.r.t. ψ_{e-2} in the layered MDP $\mathcal{L}_{e-2}$. Consider the Markov chain $\mathcal{C}_{\ell_0}$ induced by $\mathcal{L}_{e-2}$, the fixed initial state ℓ_0 and strategy σ_{ℓ_0}.
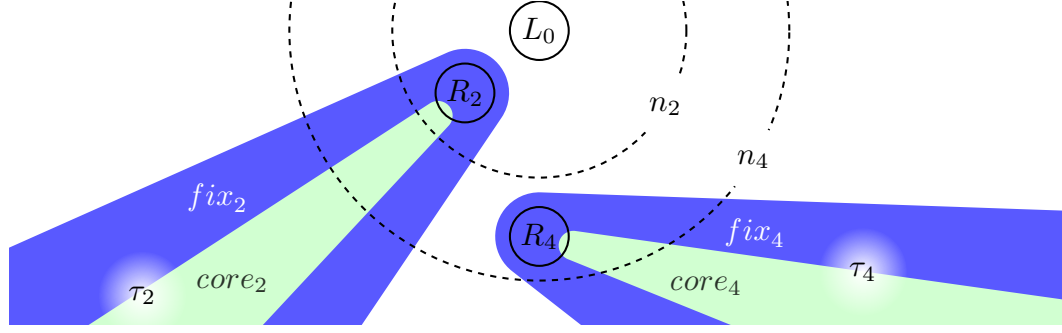
■ **Figure 3** The construction for Lemma 9. In the first iteration, for color 2, we fix the MD strategy τ_2 in the fix_2-region. In the second iteration, for color 4, we fix τ_4 in fix_4, and so on for all even colors. Everywhere else we fix an γ-optimal reachability strategy towards $\bigcup_{e=2}^{e_{\max}} core_e$ (in green).

By definition (Equation 5), θ_e is suffix-closed and $\mathsf{F}\theta_e$ is tail. The strategy σ_{ℓ_0} attains $\mathsf{F}\theta_e$ with probability at least as large as it achieves disjunct χ_e in ψ_{e-2}. We apply Lévy's zero-one law to deduce that the winning runs of $\mathsf{F}\theta_e$ likely reach a finite set R_e of states that have a high value for θ_e. In other words, most runs that eventually win through color e, while eventually avoiding Fix_{e-2}, will reach R_e within a bounded number of steps.

▶ **Lemma 10.** *Let $s_0 \in S$ and $\mathcal{E}$ be a suffix-closed objective. For all $\varepsilon, \varepsilon' > 0$, there exist n and a finite set $F \subseteq Safe_{\mathcal{E}}(1-\varepsilon)$ such that $\mathbb{P}_{s_0}(\mathsf{F}\mathcal{E} \wedge \mathsf{F}^{\leq n} F) \geq \mathbb{P}_{s_0}(\mathsf{F}\mathcal{E}) - \varepsilon'$.*

By Lemma 10, there exist n_{ℓ_0} and a finite set $R_{\ell_0} \subseteq Safe_{\mathcal{L}_{e-2},\theta_e}(\alpha)$ such that

$$\mathbb{P}_{\mathcal{L}_{e-2},\ell_0,\sigma_{\ell_0}}(\mathsf{F}\theta_e \wedge \mathsf{F}^{\leq n_{\ell_0}} R_{\ell_0}) \geq \mathbb{P}_{\mathcal{L}_{e-2},\ell_0,\sigma_{\ell_0}}(\mathsf{F}\theta_e) - \frac{\gamma}{2}. \tag{8}$$

Define $n_e \stackrel{\text{def}}{=} \max_{\ell_0 \in L_0}(n_{\ell_0})$ and $R \stackrel{\text{def}}{=} \bigcup_{\ell_0 \in L_0} R_{\ell_0}$. Write $R_e \stackrel{\text{def}}{=} \{(s,0) \mid \exists b \cdot (s,b) \in R\}$ for the projection of R_e on the layer 0.

▶ **Remark 11.** Suppose $\mathcal{E}' \subseteq \mathcal{E}$ and $\varepsilon > 0$ are such that $\mathbb{P}(\mathcal{E}') \geq \mathbb{P}(\mathcal{E}) - \varepsilon$. Then, for any $\mathcal{R}$, we have $\mathbb{P}(\mathcal{E}' \cap \mathcal{R}) \geq \mathbb{P}(\mathcal{E} \cap \mathcal{R}) - \varepsilon$.

Proof. We have:

$$\mathbb{P}(\mathcal{E}' \cap \mathcal{R}) = \mathbb{P}(\mathcal{E}') - \mathbb{P}(\mathcal{E}' \setminus \mathcal{R}) \geq \mathbb{P}(\mathcal{E}) - \varepsilon - \mathbb{P}(\mathcal{E}' \setminus \mathcal{R}) \geq \mathbb{P}(\mathcal{E}) - \varepsilon - \mathbb{P}(\mathcal{E} \setminus \mathcal{R}) = \mathbb{P}(\mathcal{E} \cap \mathcal{R}) - \varepsilon.$$

◀

We apply Remark 11 to Equation (8) to get

$$\mathbb{P}_{\mathcal{L}_{e-2},\ell_0,\sigma_{\ell_0}}(\mathsf{F}\theta_e \wedge \mathsf{G}\neg\mathsf{Fix}_{e-2} \wedge \mathsf{F}\,Cl(R_e)) \geq \mathbb{P}_{\mathcal{L}_{e-2},\ell_0,\sigma_{\ell_0}}(\mathsf{F}\theta_e \wedge \mathsf{G}\neg\mathsf{Fix}_{e-2}) - \frac{\gamma}{2}.$$

Since $\mathsf{FG}\neg\mathsf{Fix}_{e-2} \wedge \mathsf{G}\neg\mathsf{Fix}_{e-2} = \mathsf{G}\neg\mathsf{Fix}_{e-2}$ and $\chi_e = \mathsf{F}\varphi_e \wedge \mathsf{G}\neg\mathsf{Fix}_{e-2}$,

$$\mathbb{P}_{\mathcal{L}_{e-2},\ell_0,\sigma_{\ell_0}}(\chi_e \wedge \mathsf{F}\,Cl(R_e)) \geq \mathbb{P}_{\mathcal{L}_{e-2},\ell_0,\sigma_{\ell_0}}(\chi_e) - \frac{\gamma}{2}. \tag{9}$$

We think of $\mathsf{GF}[S]^{Col=e}$ as a Büchi condition on a slightly modified MDP. This allows us to apply the following theorem from [12] about the strategy complexity of Büchi objectives.

▶ **Theorem 12** (Theorem 5 in [12])**.** *For every acyclic countable MDP $\mathcal{M}$, a Büchi objective φ, finite set I of initial states and $\varepsilon > 0$, there exists a deterministic 1-bit strategy that is ε-optimal from every $s \in I$.*

Using Theorem 12, we prove the following.

▷ **Claim 13.**   In MDP $\mathcal{L}_{e-2}$, there is an MD strategy τ_e, that is $(\alpha - \beta)$-optimal for θ_e from R_e.

Notice that τ_e is used to define regions $core_e \subseteq fix_e$; see Equation (6) and Figure 3. Since $\mathtt{val}_{\mathcal{L}_{e-2}, \theta_e}(\ell) = \mathtt{val}_{\mathcal{L}_{e-2}, \theta_e}(\ell')$ holds for all siblings ℓ and ℓ', all states in R_e have value $\geq \alpha$ w.r.t. θ_e. We have chosen τ_e to be $(\alpha - \beta)$-optimal, which implies $\mathbb{P}_{\mathcal{L}_{e-2}, \ell, \tau_e}(\theta_e) \geq \beta$ for all $\ell \in R_e$. This shows that $R_e \subseteq fix_e$. Strategy τ_e is also used to obtain $\mathcal{L}_e$ from $\mathcal{L}_{e-2}$: for all controlled states $\ell \in fix_e$, the successor is fixed to be $\tau_e(\ell)$ in $\mathcal{L}_e$, see Equation (7).

Invariant (3):   Given a state $\ell_0 \in L_0$, this invariant states that, for all colors e, $\mathtt{val}_{\mathcal{L}_e, \psi_e}(\ell_0) \geq \mathtt{val}_{\mathcal{L}, \varphi}(\ell_0) - e \cdot \gamma$ holds. Recall that $\psi_0 = \varphi$ and $\mathcal{L}_0 = \mathcal{L}$. To prove the invariant, by an induction on even colors e, it suffices to prove the following:

$$\mathtt{val}_{\mathcal{L}_e, \psi_e}(\ell_0) \geq \mathtt{val}_{\mathcal{L}_{e-2}, \psi_{e-2}}(\ell_0) - 2\gamma.$$

We construct a strategy π for ψ_e in $\mathcal{L}_e$ such that $\mathbb{P}_{\mathcal{L}_e, \ell_0, \pi}(\psi_e) \geq \mathtt{val}_{\mathcal{L}_{e-2}, \psi_{e-2}}(\ell_0) - 2\gamma$. Intuitively speaking, π enforces that most runs that win through colors e', with $e' \leq e$, eventually reach the $core_{e'}$-region and most remaining winning runs always avoid the Fix_e-region.

The strategy π is defined by combining σ_{ℓ_0} and τ_e; recall that the strategy σ_{ℓ_0} is $\frac{\gamma}{2}$-optimal w.r.t. ψ_{e-2} starting from ℓ_0 in $\mathcal{L}_{e-2}$. We define π such that it starts by following σ_{ℓ_0}. If it ever enters $Cl(fix_e)$ then we ensure that it enters fix_e as well (in at most one more step). Then π continues by playing as τ_e does forever.

The following claim concludes the proof of **Invariant 3**.

▷ **Claim 14.** $\mathbb{P}_{\mathcal{L}_e, \ell_0, \pi}(\psi_e) \geq \mathtt{val}_{\mathcal{L}_{e-2}, \psi_{e-2}}(\ell_0) - 2\gamma.$

We summarize the main steps in the proof of Claim 14 here. We first prove the claim that if π ever enters $Cl(fix_e)$ then it is possible to define it in such a way that it actually enters fix_e.

Comparing ψ_e with ψ_{e-2}, one notices that two significant terms in the symmetric difference of these two objectives are χ_e and $\mathsf{F}core_e$. Roughly speaking, we use Equation (9) to move from χ_e to $\mathsf{F}\,Cl(fix_e)$. Then we move from $\mathsf{F}\,Cl(fix_e)$ to $\mathsf{F}core_e$ by proving that $\mathbb{P}_{\mathcal{L}_e, \ell_0, \pi}(\mathsf{F}core_e)$ is almost as high as $\mathbb{P}_{\mathcal{L}_{e-2}, \ell_0, \pi}(\mathsf{F}\,Cl(fix_e))$, modulo small errors. To derive the latter, we rely on two facts: another application of Lévy's zero-one law that guarantees $\mathbb{P}_{\mathcal{L}_e, \ell_0, \pi}(\theta_e \wedge \mathsf{F}core_e)$ is equal to $\mathbb{P}_{\mathcal{L}_e, \ell_0, \pi}(\theta_e)$; and the fact that, as soon as π visits the first state $\ell \in fix_e$, it switches to τ_e forever, and thus attains θ_e with probability at least β.

Reach iteration: After all $\frac{e_{\max}}{2}$-iterations for even colors and the fixing, by **Invariant** (3), for all $\ell_0 \in L_0$, we have:

$$\mathtt{val}_{\mathcal{L}_{e_{\max}}, \psi_{e_{\max}}}(\ell_0) \geq \mathtt{val}_{\mathcal{L}, \varphi}(\ell_0) - e_{\max}\gamma. \tag{10}$$

Recall that $\psi_{e_{\max}} = \bigvee_{e \in even(\mathcal{C})} \mathsf{F}core_e$. At this last iteration, we fix the choice of all remaining states in $\mathcal{L}_{e_{\max}}$ such that the probability of $\psi_{e_{\max}}$ is maximized. Recall that there are uniformly ε-optimal MD strategies for reachability objectives [16]. Hence, there is a single MD strategy τ_{reach} in $\mathcal{L}_{e_{\max}}$ that is uniformly γ-optimal w.r.t. $\psi_{e_{\max}}$; in particular, τ_{reach} is γ-optimal from every state $\ell_0 \in L_0$.

Let $\mathcal{L}' \stackrel{\text{def}}{=} \mathcal{L}_{e_{\max}}[\tau_{\mathrm{reach}}, L]$. Let $\hat{\sigma}$ be the MD strategy in $\mathcal{L}$ that plays from L_0 as prescribed by all the fixings in $\mathcal{L}'$. Since all choices in all the fix_e-region are resolved according to τ_e, $e \in \{2, \cdots, e_{\max}\}$, we can apply Lévy's zero-one law another time.
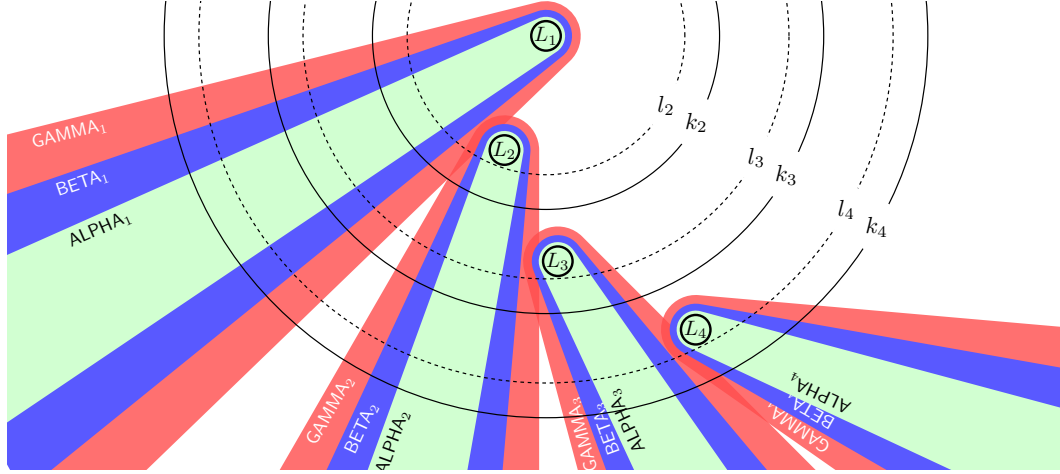
Figure 4 Initial segment of the sea urchin construction. $\mathcal{L}_i$ is the result of fixing τ_i inside BETA_i and then ρ_i inside the k_i-bubble (the set of states reachable from the initial state(s) in $\leq k_i$ steps). Drawn here for $i = 1, 2, 3, 4$.

▶ **Lemma 15.** *Let $0 < \beta_1 < \beta_2 \leq 1$ and $\mathcal{E}$ a tail objective. For $s \in Safe_{\mathcal{E}}(\beta_2)$, the following holds:* $\mathbb{P}_s(\mathsf{G}\, Safe_{\mathcal{E}}(\beta_1)) \geq \frac{\beta_2 - \beta_1}{1 - \beta_1}$.

By Lemma 15, for all states $\ell \in core_e$,

$$\mathbb{P}_{\mathcal{L}_{e_{\max}}, \ell, \tau_e}(\mathsf{G}fix_e) \geq \frac{\alpha - \beta}{1 - \beta} \geq 1 - \gamma. \tag{11}$$

States in fix_e have a high value for θ_e and thus also for $\mathsf{F}\varphi_e$.

▶ **Lemma 16.** *Let $0 < \beta < 1$ and $\mathcal{E}$ a tail objective. For all states $s \in Safe_{\mathcal{E}}(\beta)$:*
1. $\mathbb{P}_s(\mathsf{FG}Safe_{\mathcal{E}}(\beta) \setminus \mathcal{E}) = 0$; *and*
2. $\mathbb{P}_s(\mathcal{E} \setminus \mathsf{FG}Safe_{\mathcal{E}}(\beta)) = 0$.

By Lemma 16.2, we satisfy $\mathsf{F}\varphi_e$ almost surely:

$$\mathbb{P}_{\mathcal{L}_{e_{\max}}, \ell, \tau_e}(\mathsf{F}\varphi_e \mid \mathsf{G}fix_e) = 1. \tag{12}$$

Using Equations (10) and (11), we prove that following.

▷ **Claim 17.** The MD strategy $\hat{\sigma}$ is ε-optimal for parity objective φ, from every state $\ell_0 \in L_0$. This concludes the proof of Lemma 9.

4 Optimal Strategies for Parity

In this section we show Theorem 2, i.e., that optimal strategies for parity, where they exist, can be chosen deterministic 1-bit Markov.

First we show the main technical result of this section.

▶ **Lemma 18.** *Let $\mathcal{L}(\mathcal{M})$ be the layered MDP obtained from an acyclic and finitely branching MDP $\mathcal{M}$ and a coloring Col such that all states are almost surely winning for $\varphi = \mathtt{Parity}(Col)$ (i.e., every state s has a strategy σ_s such that $\mathbb{P}_{\mathcal{L}(\mathcal{M}), s, \sigma_s}(\varphi) = 1$).*

For every initial state s_0 there exists an MD strategy σ that almost surely wins, i.e., $\mathbb{P}_{\mathcal{L}(\mathcal{M}), s_0, \sigma}(\varphi) = 1$.

Proof sketch. The full version of this rather complex proof can be found in Appendix E.

For some intuition consider Figure 4. The sea urchin construction is a plastering construction with infinitely many iterations where MD strategies are fixed in larger and larger subspaces. Its name comes from the shape of the subspace in which player choices are fixed up-to iteration i: A growing finite body of states that are reachable from the initial state s_0 within $\leq k_i$ steps, plus i different spikes of infinite size. Each spike is composed of nested subsets $\mathsf{ALPHA}_i \subseteq \mathsf{BETA}_i$ (and $\subseteq \mathsf{GAMMA}_i$, which is used only in the correctness argument) that correspond to different levels of attainment of certain ε-optimal MD strategies τ_i, obtained from Lemma 9. Strategy τ_i is then fixed in BETA_i (and thus in ALPHA_i). Other MD strategies ρ_i are fixed elsewhere in the finite body, up-to horizon k_i. Using Lévy's zero-one law, we prove that, once inside ALPHA_i, there is a high chance of never leaving the i-th spike BETA_i. Moreover, almost all runs that stay in the i-th spike satisfy parity. Finally, the strategies ρ_i ensure that at least $1/2$ (by probability mass) of the runs from s_0 that don't stay in one of the first i spikes will eventually stay in the $(i+1)$-th spike and satisfy parity there. Thus, at the stage with i spikes, the fixed MD strategy attains parity with some probability $\geq 1 - 2^{-i}$ already *inside* this fixed subspace. In the limit of $i \to \infty$, the resulting MD strategy attains parity almost surely. ◀

▶ **Definition 19.** *For a tail objective φ and an MDP $\mathcal{M} = (S, S_\square, S_\bigcirc, \longrightarrow, P)$, we define the* conditioned version *of $\mathcal{M}$ w.r.t. φ to be the MDP $\mathcal{M}_* = (S_*, S_{*\square}, S_{*\bigcirc}, \longrightarrow_*, P_*)$ with $S_* = \{s \in S \mid \exists \sigma. \, \mathbb{P}_{\mathcal{M},s,\sigma}(\varphi) = \mathtt{val}_\mathcal{M}(s) > 0\}$ and $S_{*\square} = S_* \cap S_\square$ and $S_{*\bigcirc} = S_* \cap S_\bigcirc$ and*

$$\longrightarrow_* = \{(s,t) \in S_* \times S_* \mid s \longrightarrow t \text{ and if } s \in S_{*\square} \text{ then } \mathtt{val}_\mathcal{M}(s) = \mathtt{val}_\mathcal{M}(t)\}$$

and $P_* : S_{*\bigcirc} \to \mathcal{D}(S_*)$ so that $P_*(s)(t) = P(s)(t) \cdot \frac{\mathtt{val}_\mathcal{M}(t)}{\mathtt{val}_\mathcal{M}(s)}$ for all $s \in S_{*\bigcirc}$ and $t \in S_*$ with $s \longrightarrow_* t$.

See Appendix C for a proof that $P_*(s)$ is a probability distribution for all $s \in S_{*\bigcirc}$ and therefore that the conditioned MDP $\mathcal{M}_*$ is well-defined. The name stems from a useful property (cf. Lemma 33.2) that for all strategies that are optimal for φ in $\mathcal{M}$, the probability in $\mathcal{M}_*$ of any event is the same as that of its probability in $\mathcal{M}$ conditioned under φ.

The following theorem is a very slight generalization of [13, Theorem 5] (cf. Appendix E). It gives a sufficient condition under which we can conclude the existence of MD optimal strategies from the existence of MD almost-sure winning strategies.

▶ **Theorem 20.** *Let φ be a tail objective. Let $\mathcal{M} = (S, S_\square, S_\bigcirc, \longrightarrow, P)$ be an MDP and $\mathcal{M}_* = (S_*, S_{*\square}, S_{*\bigcirc}, \longrightarrow_*, P_*)$ its conditioned version wrt. φ. Then:*
1. *For all $s \in S_*$ there exists a strategy σ with $\mathbb{P}_{\mathcal{M}_*,s,\sigma}(\varphi) = 1$.*
2. *Suppose that for every $s \in S_*$ there exists an MD strategy σ'' with $\mathbb{P}_{\mathcal{M}_*,s,\sigma''}(\varphi) = 1$. Then there is an MD strategy σ' such that for all $s \in S$:*

$$\left(\exists \sigma \in \Sigma. \, \mathbb{P}_{\mathcal{M},s,\sigma}(\varphi) = \mathtt{val}_\mathcal{M}(s) \right) \implies \mathbb{P}_{\mathcal{M},s,\sigma'}(\varphi) = \mathtt{val}_\mathcal{M}(s)$$

▶ **Theorem 21.** *Consider an acyclic MDP $\mathcal{M}$ and a parity objective.*

There exists a deterministic 1-bit strategy that is optimal from all states that have an optimal strategy.

Proof. Consider the corresponding layered system $\mathcal{L}(\mathcal{M})$ (cf. Definition 6), which is also acyclic. Let S_{opt} be the subset of states that have an optimal strategy in $\mathcal{M}$. Thus all states in $S_{opt} \times \{0, 1\}$ have an optimal strategy in $\mathcal{L}(\mathcal{M})$ by Lemma 7.

We now use Theorem 20 to obtain an MD strategy σ' in $\mathcal{L}(\mathcal{M})$ that is optimal for all states in $\mathcal{L}(\mathcal{M})$ that have an optimal strategy. First, the parity objective is tail. Second, in

$\mathcal{L}(\mathcal{M})$, any two siblings have the same value w.r.t. parity by Remark 8. Therefore the changes from $\mathcal{L}(\mathcal{M})$ to its conditioned version $\mathcal{L}(\mathcal{M})_*$ (wrt. the parity objective) are symmetric in the two layers. Thus $\mathcal{L}(\mathcal{M})_*$ is also a layered acyclic MDP (i.e., there exists some acyclic MDP $\mathcal{M}'$ s.t. $\mathcal{L}(\mathcal{M})_* = \mathcal{L}(\mathcal{M}')$), and by Theorem 20.1 all states in $\mathcal{L}(\mathcal{M})_*$ are almost surely winning. Now we can apply Lemma 18 (generalized to infinitely branching acyclic layered MDPs by Lemma 3) to $\mathcal{L}(\mathcal{M})_*$ and obtain that for every state in $\mathcal{L}(\mathcal{M})_*$ there is an MD strategy that almost surely wins. By Theorem 20.2 there is an MD strategy σ' in $\mathcal{L}(\mathcal{M})$ that is optimal for all states that have an optimal strategy. In particular, σ' is optimal for the states in $S_{opt} \times \{0, 1\}$ in $\mathcal{L}(\mathcal{M})$. By Lemma 7, this yields a deterministic 1-bit strategy in $\mathcal{M}$ that is optimal for all states in S_{opt}. ◀

In Theorem 21 the initial memory mode of the 1-bit strategy is irrelevant (recall Remark 8). Theorem 2 now follows directly from Theorem 21 and Lemma 4(3).

5 Optimal Strategies for $\{0, 1, 2\}$-Parity

▶ **Theorem 22.** *Let $\mathcal{M} = (S, S_\square, S_\bigcirc, \longrightarrow, P)$ be an MDP, φ a $\{0, 1, 2\}$-Parity objective and $\mathcal{M}_* = (S_*, S_{*\square}, S_{*\bigcirc}, \longrightarrow_*, P_*)$ its conditioned version wrt. φ. Assume that in $\mathcal{M}_*$ for every safety objective (given by some target $T \subseteq S_*$) and $\varepsilon > 0$ there exists a uniformly ε-optimal MD strategy. Let S_{opt} be the subset of states that have an optimal strategy for φ in $\mathcal{M}$.*

Then there exists an MD strategy in $\mathcal{M}$ that is optimal for φ from every state in S_{opt}.

The above result generalizes [13, Theorem 16], which considers only finitely-branching MDPs and uses the fact that for every safety objective, an MD strategy exists that is uniformly *optimal*. This is not generally true for infinitely-branching acyclic MDPs [13]. To prove Theorem 22, we adjust the construction so that it only requires uniformly ε-*optimal* MD strategies for safety objectives (in the conditioned MDP $\mathcal{M}_*$).

In order to apply Theorem 22 to infinitely-branching acyclic MDPs, we now show that acyclicity guarantees the existence of uniformly ε-optimal MD strategies for safety objectives.

▶ **Lemma 23.** *For every acyclic MDP with a safety objective and every $\varepsilon > 0$ there exists an MD strategy that is uniformly ε-optimal.*

While we defined ε-optimality wrt. additive errors (cf. Section 2), our proof of Lemma 23 shows that the claim holds even wrt. multiplicative errors (in the style of [16]).

▶ **Theorem 24.** *Consider an MDP $\mathcal{M}$ with a $\{0, 1, 2\}$-Parity objective and let S_{opt} be the subset of states that have an optimal strategy.*
1. *If $\mathcal{M}$ is acyclic then there exists an MD strategy that is optimal from every state in S_{opt}.*
2. *There exists a deterministic Markov strategy that is optimal from every state in S_{opt}.*

Proof. Towards item 1, if $\mathcal{M}$ is acyclic then also its conditioned version $\mathcal{M}_*$ (wrt. $\{0, 1, 2\}$-Parity) is acyclic. Thus, by Lemma 23, in $\mathcal{M}_*$ for every $\varepsilon > 0$ and every safety objective there is a uniformly ε-optimal MD strategy. The result now follows from Theorem 22.

Item 2 follows from Item 1 and Lemma 4 (item 3 with $k = 0$). ◀

6 ε-Optimal Strategies for $\{0, 1\}$-Parity (co-Büchi)

▶ **Theorem 25.** *Suppose that $\mathcal{M} = (S, S_\square, S_\bigcirc, \longrightarrow, P)$ is an MDP such that for every safety objective (given by some target $T \subseteq S$) and $\varepsilon > 0$ there exists a uniformly ε-optimal MD strategy.*

Then for every co-Büchi objective (given by some coloring $Col : S \to \{0, 1\}$) and $\varepsilon > 0$ there exists a uniformly ε-optimal MD strategy.

The precondition of Theorem 25 is satisfied by many classes of MDPs. Indeed, we obtain the following.

▶ **Corollary 26.** *Consider an MDP $\mathcal{M}$ and a co-Büchi objective.*
1. *If $\mathcal{M}$ is acyclic then, for every $\varepsilon > 0$, there exists a uniformly ε-optimal MD strategy.*
2. *If $\mathcal{M}$ is finitely branching then, for every $\varepsilon > 0$, there exists a uniformly ε-optimal MD strategy.*
3. *For every $\varepsilon > 0$ there exists a deterministic Markov strategy that, from every initial state s, attains at least $\mathtt{val}_{\mathcal{M}}(s) - \varepsilon$.*

Proof. Towards (1), for acyclic MDPs, uniformly ε-optimal strategies for safety can be chosen MD by Lemma 23. Towards (2), for finitely branching MDPs there always exists even a uniformly optimal MD strategy for every safety objective. In both cases the claim then follows from Theorem 25. Claim (3) follows directly from (1) and Lemma 4 (item 2 with $k = 0$). ◀

References

1 P. Abbeel and A. Y. Ng. Learning first-order Markov models for control. In *Advances in Neural Information Processing Systems 17*, pages 1–8. MIT Press, 2004. URL: `http://papers.nips.cc/paper/2569-learning-first-order-markov-models-for-control`.

2 C. Baier and J.-P. Katoen. *Principles of Model Checking*. MIT Press, 2008.

3 P. Billingsley. *Probability and Measure*. Wiley, New York, NY, 1995. Third Edition.

4 V. D. Blondel and J. N. Tsitsiklis. A survey of computational complexity results in systems and control. *Automatica*, 36(9):1249–1274, 2000.

5 N. Bäuerle and U. Rieder. *Markov Decision Processes with Applications to Finance*. Springer-Verlag Berlin Heidelberg, 2011.

6 K. Chatterjee, M. Jurdziński, and T. Henzinger. Quantitative stochastic parity games. In *Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 121–130, Philadelphia, PA, USA, 2004. Society for Industrial and Applied Mathematics. URL: `http://dl.acm.org/citation.cfm?id=982792.982808`.

7 E. M. Clarke, T. A. Henzinger, H. Veith, and R. Bloem, editors. *Handbook of Model Checking*. Springer, 2018. URL: `https://doi.org/10.1007/978-3-319-10575-8`, `doi:10.1007/978-3-319-10575-8`.

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

9 E. Grädel, W. Thomas, and T. Wilke, editors. *Automata, Logics, and Infinite Games*, volume 2500 of *LNCS*, 2002.

10 T. P. Hill. On the existence of good Markov strategies. *Transactions of the American Mathematical Society*, 247:157–176, 1979. `doi:https://doi.org/10.1090/S0002-9947-1979-0517690-9`.

11 T. P. Hill. Goal problems in gambling theory. *Revista de Matemática: Teoría y Aplicaciones*, 6(2):125–132, 1999.

12 S. Kiefer, R. Mayr, M. Shirmohammadi, and P. Totzke. Büchi objectives in countable MDPs. In *International Colloquium on Automata, Languages and Programming*, volume 132. LIPIcs, 2019. A technical report is available at `https://arxiv.org/abs/1904.11573`.

13 S. Kiefer, R. Mayr, M. Shirmohammadi, and D. Wojtczak. Parity objectives in countable MDPs. In *Annual IEEE Symposium on Logic in Computer Science*, 2017.

14 S. Kiefer, R. Mayr, M. Shirmohammadi, and D. Wojtczak. Parity objectives in countable MDPs. Technical report, arxiv.org, 2017. Available at `https://arxiv.org/pdf/1704.04490.pdf`.

15 A. Mostowski. Regular expressions for infinite trees and a standard form of automata. In *Computation Theory*, volume 208 of *LNCS*, pages 157–168, 1984.

16 D. Ornstein. On the existence of stationary optimal strategies. *Proceedings of the American Mathematical Society*, 20:563–569, 1969.

17 M. L. Puterman. *Markov Decision Processes: Discrete Stochastic Dynamic Programming*. John Wiley & Sons, Inc., New York, NY, USA, 1st edition, 1994.

18 M. Schäl. Markov decision processes in finance and dynamic options. In *Handbook of Markov Decision Processes*, pages 461–487. Springer, 2002.

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

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

21 M.Y. Vardi. Automatic verification of probabilistic concurrent finite-state programs. In *Proc. of FOCS'85*, pages 327–338, 1985.

22 W. Zielonka. Infinite games on finitely coloured graphs with applications to automata on infinite trees. *Theoretical Computer Science*, 200(1-2):135–183, 1998.

23 W. Zielonka. Perfect-information stochastic parity games. In *Foundations of Software Science and Computation Structures*, volume 2987 of *LNCS*, pages 499–513. Springer, 2004.

A   Reductions in Section 3 and related Lemmas

By the following lemma, the strategy complexity of general parity objectives does not depend on the branching degree of the MDPs. However, this does not hold for particular parity objectives with a restricted set of colors, since the construction introduces an extra color.

▶ **Lemma 3.**
1. *Suppose that for every finitely branching acyclic MDP with a finite set S_0 of initial states, and a parity objective, there exist ε-optimal deterministic 1-bit strategies from S_0.*
   Then even for every infinitely branching acyclic MDP with a finite set S_0 of initial states and a parity objective, there exist ε-optimal deterministic 1-bit strategies from S_0.
2. *Suppose that for every finitely branching acyclic MDP with a parity objective, there exists a deterministic 1-bit strategy that is optimal from all states that have an optimal strategy. Then even for every infinitely branching acyclic MDP with a parity objective, there exists a deterministic 1-bit strategy that is optimal from all states that have an optimal strategy.*

Proof. Towards item (1), we encode an infinitely branching acyclic MDP $\mathcal{M}$ into a finitely branching acyclic MDP $\mathcal{M}'$. 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., $p_i' = p_i/(\prod_{j=1}^{i-1}(1-p_j'))$. The fresh states are labeled with an unfavorable color that is smaller than all other colors, e.g., -1.

We take an ε-optimal deterministic 1-bit strategy σ' for parity from all states $s \in S_0$ in $\mathcal{M}'$. 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. 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 all states in the gadgets are labeled with color -1, σ performs at least as well in $\mathcal{M}$ as σ' in $\mathcal{M}'$ and is thus ε-optimal from every $s \in S_0$.

Towards item (2), the proof is almost identical, expect that we consider optimal strategies from initial states s that have an optimal strategy.   ◀

In order to show the existence of Markov (resp. 1-bit Markov) strategies, it suffices to show the existence of memoryless (resp. 1-bit) strategies in an MDP that is made acyclic

by encoding a step counter in the state space. (Note that deterministic 0-bit strategies are MD strategies and 0-bit Markov strategies are Markov strategies.) This idea appears already in [12] and can be formally stated as follows.

▶ **Lemma 4.** *Consider MDPs with a parity objective and $k \in \mathbb{N}$.*

1. *Suppose that for every acyclic MDP $\mathcal{M}'$ and every finite set of initial states S_0' and $\varepsilon > 0$, there exists a deterministic k-bit strategy that is ε-optimal from all states $s \in S_0'$. Then for every MDP $\mathcal{M}$ and every finite set of initial states S_0 and $\varepsilon > 0$, there exists a deterministic k-bit Markov strategy that is ε-optimal from all states $s \in S_0$.*

2. *Suppose that for every acyclic MDP $\mathcal{M}'$ and $\varepsilon > 0$, there exists a deterministic k-bit strategy that is ε-optimal from all states. Then for every MDP $\mathcal{M}$ and $\varepsilon > 0$, there exists a deterministic k-bit Markov strategy that is ε-optimal from all states.*

3. *Suppose that for every acyclic MDP $\mathcal{M}'$, where S_{opt}' is the subset of states that have an optimal strategy, there exists a deterministic k-bit strategy that is optimal from all states $s \in S_{opt}'$. Then for every MDP $\mathcal{M}$, where S_{opt} is the subset of states that have an optimal strategy, there exists a deterministic k-bit Markov strategy that is optimal from all states $s \in S_{opt}$.*

Proof. The construction is similar for all three items.

Consider an MDP $\mathcal{M} = (S, S_\square, S_\bigcirc, \longrightarrow, P)$ with sets of initial states S_0 (finite), S and S_{opt}, respectively.

We transform it into an acyclic MDP $\mathcal{M}'$ by encoding a step-counter into the states, i.e., $\mathcal{M}' = (S', S_\square', S_\bigcirc', \longrightarrow', P')$ where $S' \stackrel{\text{def}}{=} S \times \mathbb{N}$, $S_\square' \stackrel{\text{def}}{=} S_\square \times \mathbb{N}$, $S_\bigcirc' \stackrel{\text{def}}{=} S_\bigcirc \times \mathbb{N}$, $Col((s, n)) \stackrel{\text{def}}{=} Col(s)$, $(s, n) \longrightarrow'(s', n+1)$ iff $s \longrightarrow s'$ and $P'((s, n))((s', n+1)) \stackrel{\text{def}}{=} P(s)(s')$.

For every deterministic k-bit strategy σ' in $\mathcal{M}'$ there is a corresponding deterministic k-bit Markov strategy σ in $\mathcal{M}$, and vice-versa. At any state s, σ in memory mode m and step-counter n plays exactly like σ' in memory mode m at state (s, n).

It follows from the definition of the colorings that σ (with memory mode m) attains the same from any initial state s as σ' (with memory mode m) attains from $(s, 0)$. Moreover, every state s has the same value as its corresponding state $(s, 0)$.

1. In $\mathcal{M}'$ we consider the set of initial states $S_0' \stackrel{\text{def}}{=} S_0 \times \{0\}$, which is finite since S_0 is finite. By our assumption, for every $\varepsilon > 0$, there exists a deterministic k-bit strategy σ' in $\mathcal{M}'$ that is ε-optimal from all states $s \in S_0'$. Thus σ is ε-optimal from all states $s \in S_0$.

2. Like above, except that the set of initial states $S_0 \stackrel{\text{def}}{=} S$ is not finite. Since σ' is assumed to be ε-optimal from all states in $\mathcal{M}'$, in particular it is ε-optimal from all states in $S_0' \stackrel{\text{def}}{=} S \times \{0\}$. Thus σ is ε-optimal from all states $s \in S$.

3. Here the set of initial states is S_{opt}. Every state $s \in S_{opt}$ has the same value as its corresponding state $(s, 0) \in S_{opt} \times \{0\}$ and the corresponding strategies σ and σ' attain the same from s and $(s, 0)$, respectively. Therefore $S_{opt} \times \{0\} \subseteq S_{opt}'$. Since the strategy σ' is assumed to be optimal from all states $s \in S_{opt}'$, it is optimal from all states in $S_{opt} \times \{0\}$, and thus σ is optimal from all states in S_{opt}.

◀

For ease of presentation, we will, instead of showing the existence of 1-bit strategies in an acyclic MDP $\mathcal{M}$, show the existence of MD strategies in the corresponding *layered* MDP $\mathcal{L}(\mathcal{M})$, which encodes the two memory modes into the states by having two copies of $\mathcal{M}$ (called *layers* 0 and 1). The transitions and probability functions, as well as whether a state is randomized, and its (parity) color, are lifted naturally.

The next lemma shows the correspondence between deterministic 1-bit strategies in $\mathcal{M}$ and MD strategies in $\mathcal{L}(\mathcal{M})$.

▶ **Lemma 7.** *Consider an acyclic MDP $\mathcal{M} = (S, S_\square, S_\bigcirc, \longrightarrow, P)$ with a parity objective $\varphi = \texttt{Parity}(Col)$ and let $\mathcal{L}(\mathcal{M})$ be the corresponding layered MDP.*

For every deterministic 1-bit strategy $u[\mathsf{m_0}]$ in $\mathcal{M}$ there is a corresponding MD strategy τ in $\mathcal{L}(\mathcal{M})$, and vice-versa, such that for every $s_0 \in S$, $\mathbb{P}_{\mathcal{L}(\mathcal{M}),(s_0,\mathsf{m_0}),\tau}(\varphi) = \mathbb{P}_{\mathcal{M},s_0,u[\mathsf{m_0}]}(\varphi)$.

Proof. For the "$\mathcal{M} \implies \mathcal{L}(\mathcal{M})$" direction, given $u[m_0]$, we define the MD strategy τ to play in $\mathcal{L}(\mathcal{M})$ as follows. For $b, b' \in \{0, 1\}$,

- for a controlled state $s \in S_\square$, if $u[m_0](b, s) = (b', s')$ meaning that $u[m_0]$ chooses s' at s, by taking a transition $t = (s, s')$, and updates the bit to b', we define $\tau((s, b)) \stackrel{\text{def}}{=} (t, b)$ and $\tau((t, b)) \stackrel{\text{def}}{=} (s', b')$;
- for a random state $s \in S_\bigcirc$, if $u[m_0]$ updates the memory bit to b' in case the random successor resolves to s', by taking a transition $t = (s, s')$, we define $\tau((t, b)) \stackrel{\text{def}}{=} (s', b')$.

Similarly, for the "$\mathcal{M} \impliedby \mathcal{L}(\mathcal{M})$" direction, given τ in $\mathcal{L}(\mathcal{M})$, we define an update function u, such that for all initial bit $m_0 \in \{0, 1\}$ the deterministic 1-bit strategy $u[m_0]$ in $\mathcal{M}$ plays from any state $s \in S$ as τ plays in $\mathcal{L}(\mathcal{M})$ from (s, m_0). The construction is as follows. For all $b, b' \in \{0, 1\}$ and all transitions $t = (s, s')$,

- if $s \in S_\square$, and if $\tau((s, b)) = (t, b)$ and $\tau((t, b)) = (s', b')$, we define $u(b, s) \stackrel{\text{def}}{=} (b', s')$;
- if $s \in S_\bigcirc$, and if $\tau((t, b)) = (s', b')$, we define $u(b, s)(b', s') \stackrel{\text{def}}{=} P(s)(s')$.

Denote by $\mathcal{C}^\tau$ the Markov chain obtained from $\mathcal{L}(\mathcal{M})$ after fixing τ, and by $\mathcal{C}^{u[m_0]}$ the Markov chain obtained from $\mathcal{M}$ after fixing $u[m_0]$. Observe there is a clear bijection between the runs in the Markov chains $\mathcal{C}^\tau$ and $\mathcal{C}^{u[m_0]}$. Since the parity colors are lifted accordingly, we conclude that $\mathbb{P}_{\mathcal{L}(\mathcal{M}),(s_0,m_0),\tau}(\varphi) = \mathbb{P}_{\mathcal{M},s_0,u[m_0]}(\varphi)$, as required. ◀

B   Lévy's zero-one law

We fix a finitely branching Markov chain $\mathcal{C}$ with state space S. We use the probability measure $\mathbb{P}_s$ when starting in a state s.

For an event $\mathcal{E} \in \mathcal{F}$, the indicator function $\mathbf{1}_\mathcal{E} : S^\omega \to \{0, 1\}$ is defined by

$$\mathbf{1}_\mathcal{E}(\rho) = \begin{cases} 1 & \text{if } \rho \in \mathcal{E}, \\ 0 & \text{otherwise.} \end{cases}$$

Below we recall Lévy's zero-one law; we state this result for a specific family of sub σ-algebras that is used throughout our proofs. Consider the simplest sequence of sub σ-algebras $(\mathcal{F}_i)_{i\in\mathbb{N}}$ of $\mathcal{F}$ where each $\mathcal{F}_i$ is the σ-algebra generated by all events that depend only on the length-i prefixes. Formally, for all $i \in \mathbb{N}$, define the sub σ-algebra

$$\mathcal{F}_i = \{A \cdot S^\omega \subseteq S^\omega \mid A \subseteq S^i\}.$$

Observe that $\mathcal{F}_1 \subset \mathcal{F}_2 \subset \cdots \subset \mathcal{F}_\infty$ where $\mathcal{F}_\infty = \mathcal{F}$ is the smallest σ-algebra containing all the $\mathcal{F}_i$. The sub σ-algebra $\mathcal{F}_i$, $i \in \mathbb{N}$, introduces an equivalence class $\sim_i$ on S^ω where $\rho \sim_i \rho'$ if and only if for all $\mathcal{E} \in \mathcal{F}_i$, the condition $\rho \in \mathcal{E} \Leftrightarrow \rho' \in \mathcal{E}$ is met. Given a run ρ, denote by $[\rho]_{\sim_i}$ the equivalence class of ρ. By definition of the $\mathcal{F}_i$, if $\rho \in s_1 \cdots s_i S^\omega$ then $[\rho]_{\sim_i} = s_1 \cdots s_i S^\omega$.

Given a state s, define the random variable $\mathbb{P}_s(\mathcal{E} \mid \mathcal{F}_i) : S^\omega \to [0,1] \cup \{\bot\}$ such that, for all runs $\rho \in s_1 \cdots s_i S^\omega$,

$$\mathbb{P}_s(\mathcal{E} \mid \mathcal{F}_i)(\rho) = \begin{cases} \mathbb{P}_s(\mathcal{E} \mid [\rho]_{\sim_i}) & \text{if } \mathbb{P}_s(s_1 \cdots s_i) \neq 0; \\ \bot \text{ (read as undefined)} & \text{otherwise.} \end{cases} \tag{13}$$

By Lévy's zero-one law for all events $\mathcal{E} \subseteq \mathcal{F}_\infty$ we have that

$$\lim_{i \to \infty} \mathbb{P}_s(\mathcal{E} \mid \mathcal{F}_i) = \mathbf{1}_{\mathcal{E}}$$

holds $\mathbb{P}$-almost-surely.

▶ **Remark 27.** Given a suffix-closed objective $\mathcal{E}$ and a run $\rho \in s_1 \cdots s_i S^\omega$, if $\mathbb{P}_s(\mathcal{E} \mid \mathcal{F}_i)(\rho)$ is defined, then

$$\begin{aligned} \mathbb{P}_{s_1}(\mathcal{E} \mid \mathcal{F}_i)(\rho) &= \mathbb{P}_{s_1}(\mathcal{E} \mid [\rho]_{\sim_i}) \\ &= \mathbb{P}_{s_1}(\mathcal{E} \mid s_1 \cdots s_i S^\omega) \\ &\leq \mathbb{P}_{s_i}(\mathcal{E} \mid s_i S^\omega) \\ &= \mathbb{P}_{s_i}(\mathcal{E}). \end{aligned}$$

If $\mathcal{E}$ is tail then $\mathbb{P}_{s_1}(\mathcal{E} \mid \mathcal{F}_i)(\rho) = \mathbb{P}_{s_i}(\mathcal{E})$.

For the fixed Markov chain $\mathcal{C}$ and $\varepsilon > 0$, we define $Safe_{\mathcal{E}}(1 - \varepsilon) \stackrel{\text{def}}{=} \{s \mid \mathbb{P}_s(\mathcal{E}) \geq 1 - \varepsilon\}$.

▶ **Lemma 28.** *Let $s_0 \in S$ and $\mathcal{E}$ be a suffix-closed objective and $\varepsilon > 0$. Then $\mathbb{P}_{s_0}(\mathsf{F}\mathcal{E} \wedge \neg \mathsf{F}Safe_{\mathcal{E}}(1 - \varepsilon)) = 0$.*

Proof. Let $s_0 \in S$. We have:

$$\begin{aligned} \llbracket \mathsf{G}\neg Safe_{\mathcal{E}}(1 - \varepsilon) \rrbracket^{s_0} &= \{s_0 s_1 \cdots \mid \forall i . \mathbb{P}_{s_i}(\mathcal{E}) < 1 - \varepsilon\} \\ &\subseteq \{\rho \in s_0 S^\omega \mid \forall i . \mathbb{P}_{s_0}(\mathcal{E} \mid \mathcal{F}_i)(\rho) < 1 - \varepsilon\} \qquad \text{by Remark 27} \\ &\subseteq \{\rho \in s_0 S^\omega \mid \lim_{i \to \infty} \mathbb{P}_{s_0}(\mathcal{E} \mid \mathcal{F}_i)(\rho) \neq 1\} \end{aligned}$$

It follows

$$\mathbb{P}_{s_0}(\mathcal{E} \wedge \mathsf{G}\neg Safe_{\mathcal{E}}(1 - \varepsilon)) \leq \mathbb{P}_{s_0}(\mathcal{E} \cap \{\rho \in s_0 S^\omega \mid \lim_{i \to \infty} \mathbb{P}_{s_0}(\mathcal{E} \mid \mathcal{F}_i)(\rho) \neq 1\}) = 0 \tag{14}$$

by Lévy's zero-one law.

Let $s_0 \in S$. We have:

$$\begin{aligned} \mathbb{P}_{s_0}&(\mathsf{F}\mathcal{E} \wedge \neg \mathsf{F}Safe_{\mathcal{E}}(1 - \varepsilon)) \\ &= \mathbb{P}_{s_0}(\mathsf{F}\mathcal{E} \wedge \mathsf{G}\neg Safe_{\mathcal{E}}(1 - \varepsilon)) \\ &= \mathbb{P}_{s_0}\left(\bigcup_{s_1 \cdots s_i \in S^*} s_0 s_1 \cdots s_{i-1}(\mathcal{E} \cap s_i S^\omega) \wedge \mathsf{G}\neg Safe_{\mathcal{E}}(1 - \varepsilon) \right) \\ &\leq \sum_{s_1 \cdots s_i \in S^*} \mathbb{P}_{s_0}(s_0 s_1 \cdots s_{i-1}(\mathcal{E} \cap s_i S^\omega) \wedge \mathsf{G}\neg Safe_{\mathcal{E}}(1 - \varepsilon)) \qquad \text{union bound} \\ &\leq \sum_{s_1 \cdots s_i \in S^*} \mathbb{P}_{s_i}((\mathcal{E} \cap s_i S^\omega) \wedge \mathsf{G}\neg Safe_{\mathcal{E}}(1 - \varepsilon)) \\ &= \sum_{s_1 \cdots s_i \in S^*} \mathbb{P}_{s_i}(\mathcal{E} \wedge \mathsf{G}\neg Safe_{\mathcal{E}}(1 - \varepsilon)) \\ &= 0 \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \text{by Equation (14)} \end{aligned}$$

◀

▶ **Lemma 10.** *Let $s_0 \in S$ and $\mathcal{E}$ be a suffix-closed objective. For all $\varepsilon, \varepsilon' > 0$, there exist n and a finite set $F \subseteq Safe_{\mathcal{E}}(1 - \varepsilon)$ such that $\mathbb{P}_{s_0}(\mathsf{F}\mathcal{E} \wedge \mathsf{F}^{\leq n} F) \geq \mathbb{P}_{s_0}(\mathsf{F}\mathcal{E}) - \varepsilon'$.*

Proof. By Lemma 28 we have

$$\mathbb{P}_{s_0}(\mathsf{F}\mathcal{E} \wedge \mathsf{F} Safe_{\mathcal{E}}(1 - \varepsilon)) = \mathbb{P}_{s_0}(\mathsf{F}\mathcal{E}).$$

By continuity of measures it follows that there is n such that

$$\mathbb{P}_{s_0}(\mathsf{F}\mathcal{E} \wedge \mathsf{F}^{\leq n} Safe_{\mathcal{E}}(1 - \varepsilon)) \geq \mathbb{P}_{s_0}(\mathsf{F}\mathcal{E}) - \varepsilon'.$$

Let $\mathsf{bubble}_n(s_0)$ be the set of states that can be reached from s_0 within at most n steps. Since the Markov chain $\mathcal{C}$ is finitely branching, $F \overset{\text{def}}{=} Safe_{\mathcal{E}}(1 - \varepsilon) \cap \mathsf{bubble}_n(s_0)$ is a finite set. Then we have $[\![\mathsf{F}^{\leq n} F]\!] = [\![\mathsf{F}^{\leq n} Safe_{\mathcal{E}}(1 - \varepsilon)]\!]$ and the statement of the lemma follows.   ◀

▶ **Lemma 16.** *Let $0 < \beta < 1$ and $\mathcal{E}$ a tail objective. For all states $s \in Safe_{\mathcal{E}}(\beta)$:*
1. $\mathbb{P}_s(\mathsf{FG} Safe_{\mathcal{E}}(\beta) \setminus \mathcal{E}) = 0$; *and*
2. $\mathbb{P}_s(\mathcal{E} \setminus \mathsf{FG} Safe_{\mathcal{E}}(\beta)) = 0$.

Proof. By Lévy's zero-one law,

$$\mathbb{P}_s(\{\rho \mid \lim_{i \to \infty} \mathbb{P}_s(\mathcal{E} \mid \mathcal{F}_i)(\rho) = \mathbf{1}_{\mathcal{E}}(\rho)\}) = 1, \text{ and}$$
$$\mathbb{P}_s(\{\rho \mid \lim_{i \to \infty} \mathbb{P}_s(\mathcal{E} \mid \mathcal{F}_i)(\rho) \neq \mathbf{1}_{\mathcal{E}}(\rho)\}) = 0. \tag{15}$$

On one hand Equation (15) implies that

$$\mathbb{P}_s(\{\rho \mid \lim_{i \to \infty} \mathbb{P}_s(\mathcal{E} \mid \mathcal{F}_i)(\rho) = 0 \wedge \mathbf{1}_{\mathcal{E}}(\rho) = 0\}$$
$$\cup \{\rho \mid \lim_{i \to \infty} \mathbb{P}_s(\mathcal{E} \mid \mathcal{F}_i)(\rho) = 1 \wedge \mathbf{1}_{\mathcal{E}}(\rho) = 1\}) = 1$$
$$\Rightarrow \mathbb{P}_s(\{\rho \mid \lim_{i \to \infty} \mathbb{P}_s(\mathcal{E} \mid \mathcal{F}_i)(\rho) = 0\} \cup \{\rho \mid \mathbf{1}_{\mathcal{E}}(\rho) = 1\}) = 1$$
$$\Leftrightarrow \mathbb{P}_s(\{\rho \mid \forall \varepsilon > 0 \, \exists n \, \forall i \geq n \, \mathbb{P}_s(\mathcal{E} \mid \mathcal{F}_i)(\rho) < \varepsilon\} \cup \mathcal{E}) = 1$$
$$\Rightarrow \mathbb{P}_s(\{\rho \mid \exists n \, \forall i \geq n \, \mathbb{P}_s(\mathcal{E} \mid \mathcal{F}_i)(\rho) < \beta\} \cup \mathcal{E}) = 1$$
$$\Rightarrow \mathbb{P}_s(\mathsf{FG} \neg Safe_{\mathcal{E}}(\beta) \cup \mathcal{E}) = 1 \qquad\qquad\qquad \text{by Remark 27}$$

since $\left[\!\left[\mathsf{FG} \neg Safe_{\mathcal{E}}(\beta)\right]\!\right] \subseteq \left[\!\left[\neg \mathsf{FG} Safe_{\mathcal{E}}(\beta)\right]\!\right]$

$$\Rightarrow \mathbb{P}_s(\neg \mathsf{FG} Safe_{\mathcal{E}}(\beta) \cup \mathcal{E}) = 1$$
$$\Leftrightarrow \mathbb{P}_s(\mathsf{FG} Safe_{\mathcal{E}}(\beta) \cap \neg \mathcal{E}) = 0$$
$$\Leftrightarrow \mathbb{P}_s(\mathsf{FG} Safe_{\mathcal{E}}(\beta) \setminus \mathcal{E}) = 0$$

On the other hand Equation (15) implies that

$$\mathbb{P}_s(\{\rho \mid \lim_{i \to \infty} \mathbb{P}_s(\mathcal{E} \mid \mathcal{F}_i)(\rho) \neq 1 \wedge \mathbf{1}_{\mathcal{E}}(\rho) = 1\}) = 0$$
$$\Rightarrow \mathbb{P}_s(\{\rho \mid \lim_{i \to \infty} \mathbb{P}_s(\mathcal{E} \mid \mathcal{F}_i)(\rho) \neq 1\} \cap \{\rho \mid \mathbf{1}_{\mathcal{E}}(\rho) = 1\}) = 0$$
$$\Rightarrow \mathbb{P}_s(\{\rho \mid \forall n \, \exists i \geq n \, \mathbb{P}_s(\mathcal{E} \mid \mathcal{F}_i)(\rho) < \beta\} \cap \{\rho \mid \mathbf{1}_{\mathcal{E}}(\rho) = 1\}) = 0$$
$$\Leftrightarrow \mathbb{P}_s(\neg \mathsf{FG} Safe_{\mathcal{E}}(\beta) \cap \mathcal{E}) = 0 \qquad\qquad\qquad \text{by Remark 27}$$
$$\Leftrightarrow \mathbb{P}_s(\mathcal{E} \setminus \mathsf{FG} Safe_{\mathcal{E}}(\beta)) = 0$$

◀

▶ **Corollary 29.** *Let $0 < \beta < 1$ and $\mathcal{E}$ a tail objective. For all states $s \in Safe_{\mathcal{E}}(\beta)$, we have $\mathbb{P}_s(\mathcal{E} \mid \mathsf{G}Safe_{\mathcal{E}}(\beta)) = 1$.*

Proof. Since $\mathsf{G}Safe_{\mathcal{E}}(\beta)$ is contained in $\mathsf{FG}Safe_{\mathcal{E}}(\beta)$, Lemma 16 leads to $\mathbb{P}_s(\mathsf{G}Safe_{\mathcal{E}}(\beta) \setminus \mathcal{E}) = 0$. Then,

$$\mathbb{P}_s(\mathcal{E} \cap \mathsf{G}Safe_{\mathcal{E}}(\beta)) = \mathbb{P}_s(\mathsf{G}Safe_{\mathcal{E}}(\beta)) - \mathbb{P}_s(\mathsf{G}Safe_{\mathcal{E}}(\beta) \setminus \mathcal{E})$$
$$= \mathbb{P}_s(\mathsf{G}Safe_{\mathcal{E}}(\beta)).$$

By the above equality, we get that $\mathbb{P}_s(\mathcal{E} \mid \mathsf{G}Safe_{\mathcal{E}}(\beta)) = \frac{\mathbb{P}_s(\mathcal{E} \cap \mathsf{G}Safe_{\mathcal{E}}(\beta))}{\mathbb{P}_s(\mathsf{G}Safe_{\mathcal{E}}(\beta))} = 1$. ◀

▶ **Corollary 30.** *Let $0 < \beta_1, \beta_2 < 1$ and $\mathcal{E}$ a tail objective. For all states s we have*

$$\mathbb{P}_s(\mathsf{FG}\,Safe_{\mathcal{E}}(\beta_1) \setminus \mathsf{FG}\,Safe_{\mathcal{E}}(\beta_2)) = 0$$

Proof. We have that

$$\mathsf{FG}\,Safe_{\mathcal{E}}(\beta_1) \setminus \mathsf{FG}\,Safe_{\mathcal{E}}(\beta_2)$$
$$= [\mathcal{E} \cap (\mathsf{FG}\,Safe_{\mathcal{E}}(\beta_1) \setminus \mathsf{FG}\,Safe_{\mathcal{E}}(\beta_2))] \tag{16}$$
$$\cup [(\mathsf{FG}\,Safe_{\mathcal{E}}(\beta_1) \setminus \mathsf{FG}\,Safe_{\mathcal{E}}(\beta_2)) \setminus \mathcal{E}]$$
$$\subseteq (\mathcal{E} \setminus \mathsf{FG}\,Safe_{\mathcal{E}}(\beta_2)) \cup (\mathsf{FG}\,Safe_{\mathcal{E}}(\beta_1) \setminus \mathcal{E}).$$

Thus $\mathbb{P}_s(\mathsf{FG}\,Safe_{\mathcal{E}}(\beta_1) \setminus \mathsf{FG}\,Safe_{\mathcal{E}}(\beta_2)) \leq \mathbb{P}_s(\mathcal{E} \setminus \mathsf{FG}\,Safe_{\mathcal{E}}(\beta_2)) + \mathbb{P}_s(\mathsf{FG}\,Safe_{\mathcal{E}}(\beta_1) \setminus \mathcal{E}) = 0$, by Lemma 16. ◀

▶ **Lemma 15.** *Let $0 < \beta_1 < \beta_2 \leq 1$ and $\mathcal{E}$ a tail objective. For $s \in Safe_{\mathcal{E}}(\beta_2)$, the following holds: $\mathbb{P}_s(\mathsf{G}\,Safe_{\mathcal{E}}(\beta_1)) \geq \frac{\beta_2 - \beta_1}{1 - \beta_1}$.*

Proof. Write x for $\mathbb{P}_s(\mathsf{G}\,Safe_{\mathcal{E}}(\beta_1))$. We condition the probability of $\mathcal{E}$ under $\mathsf{G}\,Safe_{\mathcal{E}}(\beta_1)$. By the law of total probability, we have

$$\beta_2 \leq \mathbb{P}_s(\mathcal{E}) = \mathbb{P}_s(\mathcal{E} \mid \mathsf{G}\,Safe_{\mathcal{E}}(\beta_1)) \cdot x + \mathbb{P}_s(\mathcal{E} \mid \neg\mathsf{G}\,Safe_{\mathcal{E}}(\beta_1)) \cdot (1 - x).$$

By Corollary 29, we have $\mathbb{P}_s(\mathcal{E} \mid \mathsf{G}Safe_{\mathcal{E}}(\beta_1)) = 1$. Hence we have $\beta_2 \leq x + \beta_1 \cdot (1 - x)$; and $x \geq \frac{\beta_2 - \beta_1}{1 - \beta_1}$ follows. ◀

C The Conditioned MDP

In this section we adapt some results from [13].

We will need the following lemma, which is a variant of [14, Lemma 20]:

▶ **Lemma 31.** *Let φ be a tail objective. Let $\mathcal{M} = (S, S_\square, S_\bigcirc, \longrightarrow, P)$ be an MDP, and $s_0 \in S$, and σ be a strategy with $\mathbb{P}_{\mathcal{M}, s_0, \sigma}(\varphi) = \mathtt{val}_{\mathcal{M}}(s_0)$. Suppose that $s_0 s_1 \cdots s_n$ for some $n \geq 0$ is a partial run starting in s_0 and induced by σ. Then:*
1. *$\mathtt{val}_{\mathcal{M}}(s_n) = \mathbb{P}_{\mathcal{M}, s_0, \sigma}(\llbracket \varphi \rrbracket^{s_0} \mid s_0 s_1 \cdots s_n S^\omega)$.*
2. *If $s_n \in S_\bigcirc$ then $\mathtt{val}_{\mathcal{M}}(s_n) = \sum_{s_{n+1} \in S} P(s_n)(s_{n+1}) \cdot \mathtt{val}_{\mathcal{M}}(s_{n+1})$.*
3. *If $s_n \in S_\square$ then $\mathtt{val}_{\mathcal{M}}(s_n) = \mathtt{val}_{\mathcal{M}}(s_{n+1})$ for all $s_{n+1} \in \mathtt{supp}(\sigma(s_0 s_1 \cdots s_n))$.*

Proof. First we show $\mathbb{P}_{\mathcal{M}, s_0, \sigma}(\llbracket \varphi \rrbracket^{s_0} \mid s_0 s_1 \cdots s_n S^\omega) \leq \mathtt{val}_{\mathcal{M}}(s_n)$. Define a strategy $\sigma' : S^* S_\square \to \mathcal{D}(S)$ by $\sigma'(w) = \sigma(s_0 s_1 \cdots s_{n-1} w)$ for all $w \in S^* S_\square$. Then we have $\mathbb{P}_{\mathcal{M}, s_0, \sigma}(\llbracket \varphi \rrbracket^{s_0} \mid s_0 s_1 \cdots s_n S^\omega) = \mathbb{P}_{\mathcal{M}, s_n, \sigma'}(\llbracket \varphi \rrbracket^{s_n}) \leq \mathtt{val}_{\mathcal{M}}(s_n)$.

Next we show $\mathtt{val}_{\mathcal{M}}(s_n) \leq \mathbb{P}_{\mathcal{M}, s_0, \sigma}(\llbracket \varphi \rrbracket^{s_0} \mid s_0 s_1 \cdots s_n S^\omega)$. Towards a contradiction, suppose that $\mathtt{val}_{\mathcal{M}}(s_n) > \mathbb{P}_{\mathcal{M}, s_0, \sigma}(\llbracket \varphi \rrbracket^{s_0} \mid s_0 s_1 \cdots s_n S^\omega)$. Then, by the definition of $\mathtt{val}_{\mathcal{M}}(s_n)$,

there is a strategy σ' with $\mathbb{P}_{\mathcal{M},s_n,\sigma'}(\llbracket\varphi\rrbracket^{s_n}) > \mathbb{P}_{\mathcal{M},s_0,\sigma}(\llbracket\varphi\rrbracket^{s_0} \mid s_0s_1\cdots s_n S^\omega)$. Define a strategy σ'' that plays according to σ; if and when partial run $s_0s_1\cdots s_n$ is played, then σ'' acts like σ' henceforth; otherwise σ'' continues with σ forever. Using the tail property we get:

$$
\begin{aligned}
&\mathbb{P}_{\mathcal{M},s_0,\sigma''}(\llbracket\varphi\rrbracket^{s_0})\\
&= \mathbb{P}_{\mathcal{M},s_0,\sigma''}(\llbracket\varphi\rrbracket^{s_0} \mid s_0s_1\cdots s_n S^\omega) \cdot \mathbb{P}_{\mathcal{M},s_0,\sigma''}(s_0s_1\cdots s_n S^\omega)\\
&\quad + \mathbb{P}_{\mathcal{M},s_0,\sigma''}(\llbracket\varphi\rrbracket^{s_0} \setminus s_0s_1\cdots s_n S^\omega)\\
&= \mathbb{P}_{\mathcal{M},s_n,\sigma'}(\llbracket\varphi\rrbracket^{s_n}) \cdot \mathbb{P}_{\mathcal{M},s_0,\sigma}(s_0s_1\cdots s_n S^\omega)\\
&\quad + \mathbb{P}_{\mathcal{M},s_0,\sigma}(\llbracket\varphi\rrbracket^{s_0} \setminus s_0s_1\cdots s_n S^\omega) && \text{def. of } \sigma''\\
&> \mathbb{P}_{\mathcal{M},s_0,\sigma}(\llbracket\varphi\rrbracket^{s_0} \mid s_0s_1\cdots s_n S^\omega) \cdot \mathbb{P}_{\mathcal{M},s_0,\sigma}(s_0s_1\cdots s_n S^\omega)\\
&\quad + \mathbb{P}_{\mathcal{M},s_0,\sigma}(\llbracket\varphi\rrbracket^{s_0} \setminus s_0s_1\cdots s_n S^\omega) && \text{def. of } \sigma'\\
&= \mathbb{P}_{\mathcal{M},s_0,\sigma}(\llbracket\varphi\rrbracket^{s_0})\\
&= \mathtt{val}_{\mathcal{M}}(s_0) && \text{def. of } \sigma
\end{aligned}
$$

This contradicts the definition of $\mathtt{val}_{\mathcal{M}}(s_0)$. Hence we have shown item 1.

Towards items 2 and 3, we extend $\sigma : S^*S_\square \to \mathcal{D}(S)$ to $\sigma : S^*S \to \mathcal{D}(S)$ by defining $\sigma(ws) = P(s)$ for $w \in S^*$ and $s \in S_\bigcirc$. Then we have for all $s_{n+1} \in S$:

$$
\mathbb{P}_{\mathcal{M},s_0,\sigma}(s_0s_1\cdots s_n s_{n+1} S^\omega) = \mathbb{P}_{\mathcal{M},s_0,\sigma}(s_0s_1\cdots s_n S^\omega) \cdot \sigma(s_0s_1\cdots s_n)(s_{n+1}) \tag{17}
$$

Further we have:

$$
\begin{aligned}
&\mathtt{val}_{\mathcal{M}}(s_n)\\
&= \mathbb{P}_{\mathcal{M},s_0,\sigma}(\llbracket\varphi\rrbracket^{s_0} \mid s_0s_1\cdots s_n S^\omega) && \text{by item 1}\\
&= \frac{\mathbb{P}_{\mathcal{M},s_0,\sigma}(\llbracket\varphi\rrbracket^{s_0} \cap s_0s_1\cdots s_n S^\omega)}{\mathbb{P}_{\mathcal{M},s_0,\sigma}(s_0s_1\cdots s_n S^\omega)}\\
&= \frac{\sum_{s_{n+1}\in S} \mathbb{P}_{\mathcal{M},s_0,\sigma}(\llbracket\varphi\rrbracket^{s_0} \cap s_0s_1\cdots s_n s_{n+1} S^\omega)}{\mathbb{P}_{\mathcal{M},s_0,\sigma}(s_0s_1\cdots s_n S^\omega)}\\
&= \frac{1}{\mathbb{P}_{\mathcal{M},s_0,\sigma}(s_0s_1\cdots s_n S^\omega)} \cdot \sum_{s_{n+1}\in S} \mathbb{P}_{\mathcal{M},s_0,\sigma}(s_0s_1\cdots s_n s_{n+1} S^\omega) \cdot\\
&\qquad\qquad\qquad\qquad\qquad\qquad \cdot \mathbb{P}_{\mathcal{M},s_0,\sigma}(\llbracket\varphi\rrbracket^{s_0} \mid s_0s_1\cdots s_n s_{n+1} S^\omega)\\
&= \sum_{s_{n+1}\in S} \sigma(s_0s_1\cdots s_n)(s_{n+1}) \cdot \mathbb{P}_{\mathcal{M},s_0,\sigma}(\llbracket\varphi\rrbracket^{s_0} \mid s_0s_1\cdots s_n s_{n+1} S^\omega) && \text{by (17)}\\
&= \sum_{s_{n+1}\in S} \sigma(s_0s_1\cdots s_n)(s_{n+1}) \cdot \mathtt{val}_{\mathcal{M}}(s_{n+1}) && \text{by item 1}
\end{aligned}
$$

Thus we have shown item 2. Towards item 3, suppose $s_n \in S_\square$. Then, by the tail property, $\mathtt{val}_{\mathcal{M}}(s_n) \geq \mathtt{val}_{\mathcal{M}}(s_{n+1})$ for all s_{n+1} with $s_n \longrightarrow s_{n+1}$. Since $\sigma(s_0s_1\cdots s_n)$ is a probability distribution, the equality chain above shows that $\mathtt{val}_{\mathcal{M}}(s_n) = \mathtt{val}_{\mathcal{M}}(s_{n+1})$ for all $s_{n+1} \in \mathtt{supp}(\sigma(s_0s_1\cdots s_n))$. Thus we have shown item 3. ◀

▶ **Lemma 32.** *The conditioned version $\mathcal{M}_*$ of $\mathcal{M}$ w.r.t. tail objective φ (cf. Definition 19 is well defined.*

Proof. By Lemma 31.2 we have that $P_*(s)$ is a probability distribution for all $s \in S_{*\bigcirc}$; hence the conditioned MDP $\mathcal{M}_*$ is well-defined. ◀

The following lemma is a reformulation of [13, Lemma 6]:

▶ **Lemma 33.** *Let φ be a tail objective. Let $\mathcal{M} = (S, S_\square, S_\bigcirc, \longrightarrow, P)$ be an MDP, and let $\mathcal{M}_* = (S_*, S_{*\square}, S_{*\bigcirc}, \longrightarrow_*, P_*)$ be its conditioned version. Then:*

1. *For all $\sigma \in \Sigma_{\mathcal{M}_*}$ and all $n \geq 0$ and all $s_0, \ldots, s_n \in S_*$ with $s_0 \longrightarrow_* s_1 \longrightarrow_* \cdots \longrightarrow_* s_n$:*

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

2. *For all $s_0 \in S_*$ and all $\sigma \in \Sigma_{\mathcal{M}}$ with $\mathbb{P}_{\mathcal{M}, s_0, \sigma}(\varphi) = \mathtt{val}_{\mathcal{M}}(s_0) > 0$ and all measurable $\mathcal{R} \subseteq s_0 S^\omega$ we have $\mathbb{P}_{\mathcal{M}_*, s_0, \sigma}(\mathcal{R}) = \mathbb{P}_{\mathcal{M}, s_0, \sigma}(\mathcal{R} \mid [\![\varphi]\!]^{s_0})$.*

Proof. We prove item 1 by induction on n. For $n = 0$ it is trivial. For the step, suppose that the equality in item 1 holds for some n. If $s_n \in S_{*\bigcirc}$ then we have:

$$\mathbb{P}_{\mathcal{M}_*, s_0, \sigma}(s_0 s_1 \cdots s_n s_{n+1} S^\omega)$$
$$= \mathbb{P}_{\mathcal{M}_*, s_0, \sigma}(s_0 s_1 \cdots s_n S^\omega) \cdot P_*(s_n)(s_{n+1})$$
$$= \mathbb{P}_{\mathcal{M}, s_0, \sigma}(s_0 s_1 \cdots s_n S^\omega) \cdot \frac{\mathtt{val}_{\mathcal{M}}(s_n)}{\mathtt{val}_{\mathcal{M}}(s_0)} \cdot P_*(s_n)(s_{n+1}) \qquad \text{ind. hyp.}$$
$$= \mathbb{P}_{\mathcal{M}, s_0, \sigma}(s_0 s_1 \cdots s_n S^\omega) \cdot \frac{\mathtt{val}_{\mathcal{M}}(s_n)}{\mathtt{val}_{\mathcal{M}}(s_0)} \cdot P(s_n)(s_{n+1}) \cdot \frac{\mathtt{val}_{\mathcal{M}}(s_{n+1})}{\mathtt{val}_{\mathcal{M}}(s_n)} \qquad \text{def. of } P_*$$
$$= \mathbb{P}_{\mathcal{M}, s_0, \sigma}(s_0 s_1 \cdots s_n s_{n+1} S^\omega) \cdot \frac{\mathtt{val}_{\mathcal{M}}(s_{n+1})}{\mathtt{val}_{\mathcal{M}}(s_0)}$$

Let now $s_n \in S_{*\square}$. If $\sigma(s_0 s_1 \ldots s_n)(s_{n+1}) = 0$ then the inductive step is trivial. Otherwise we have:

$$\mathbb{P}_{\mathcal{M}_*, s_0, \sigma}(s_0 s_1 \cdots s_n s_{n+1} S^\omega)$$
$$= \mathbb{P}_{\mathcal{M}_*, s_0, \sigma}(s_0 s_1 \cdots s_n S^\omega) \cdot \sigma(s_0 s_1 \ldots s_n)(s_{n+1})$$
$$= \mathbb{P}_{\mathcal{M}, s_0, \sigma}(s_0 s_1 \cdots s_n S^\omega) \cdot \frac{\mathtt{val}_{\mathcal{M}}(s_n)}{\mathtt{val}_{\mathcal{M}}(s_0)} \cdot \sigma(s_0 s_1 \ldots s_n)(s_{n+1}) \qquad \text{ind. hyp.}$$
$$= \mathbb{P}_{\mathcal{M}, s_0, \sigma}(s_0 s_1 \cdots s_n S^\omega) \cdot \frac{\mathtt{val}_{\mathcal{M}}(s_{n+1})}{\mathtt{val}_{\mathcal{M}}(s_0)} \cdot \sigma(s_0 s_1 \ldots s_n)(s_{n+1}) \qquad \text{def. of } \longrightarrow_*$$
$$= \mathbb{P}_{\mathcal{M}, s_0, \sigma}(s_0 s_1 \cdots s_n s_{n+1} S^\omega) \cdot \frac{\mathtt{val}_{\mathcal{M}}(s_{n+1})}{\mathtt{val}_{\mathcal{M}}(s_0)}$$

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

Towards item 2, let $s_0 \in S_*$ and $\sigma \in \Sigma_{\mathcal{M}}$ such that $\mathbb{P}_{\mathcal{M}, s_0, \sigma}(\varphi) = \mathtt{val}_{\mathcal{M}}(s_0) > 0$. Observe that σ can be applied also in the MDP $\mathcal{M}_*$. Indeed, for any $s \in S_{*\square}$, if t is a possible successor state of s under σ, then $\mathtt{val}_{\mathcal{M}}(s) = \mathtt{val}_{\mathcal{M}}(t)$ by Lemma 31.3 and thus $t \in S_*$.

Let again $n \geq 0$ and $s_0, s_1, \ldots, s_n \in S$.

▪ Suppose $s_0 s_1 \cdots s_n$ is a partial run in $\mathcal{M}_*$ induced by σ. Then we have:

$$\mathbb{P}_{\mathcal{M}_*, s_0, \sigma}(s_0 s_1 \cdots s_n S^\omega) \cdot \mathbb{P}_{\mathcal{M}, s_0, \sigma}(\varphi)$$
$$= \mathbb{P}_{\mathcal{M}, s_0, \sigma}(s_0 s_1 \cdots s_n S^\omega) \cdot \frac{\mathtt{val}_{\mathcal{M}}(s_n)}{\mathtt{val}_{\mathcal{M}}(s_0)} \cdot \mathbb{P}_{\mathcal{M}, s_0, \sigma}(\varphi) \qquad \text{item 1}$$
$$= \mathbb{P}_{\mathcal{M}, s_0, \sigma}(s_0 s_1 \cdots s_n S^\omega) \cdot \mathtt{val}_{\mathcal{M}}(s_n) \qquad \text{assumption on } \sigma$$
$$= \mathbb{P}_{\mathcal{M}, s_0, \sigma}(s_0 s_1 \cdots s_n S^\omega) \cdot \mathbb{P}_{\mathcal{M}, s_0, \sigma}([\![\varphi]\!]^{s_0} \mid s_0 s_1 \cdots s_n S^\omega) \qquad \text{Lemma 31.1}$$
$$= \mathbb{P}_{\mathcal{M}, s_0, \sigma}([\![\varphi]\!]^{s_0} \cap s_0 s_1 \cdots s_n S^\omega)$$

▪ Suppose $s_0 s_1 \cdots s_n$ is not a partial run in $\mathcal{M}_*$ induced by σ. Hence $\mathbb{P}_{\mathcal{M}_*, s_0, \sigma}(s_0 s_1 \cdots s_n S^\omega) = 0$. If $s_0 s_1 \cdots s_n$ is not a partial run in $\mathcal{M}$ induced by σ then $\mathbb{P}_{\mathcal{M}, s_0, \sigma}(s_0 s_1 \cdots s_n S^\omega) = 0$.

Otherwise, since σ is optimal, there is $i \leq n$ with $\mathtt{val}_{\mathcal{M}}(s_i) = 0$, hence $\mathbb{P}_{\mathcal{M},s_0,\sigma}(\llbracket \varphi \rrbracket^{s_0} \cap s_0 s_1 \cdots s_n S^{\omega})$. In either case we have $\mathbb{P}_{\mathcal{M}_*,s_0,\sigma}(s_0 s_1 \cdots s_n S^{\omega}) \cdot \mathbb{P}_{\mathcal{M},s_0,\sigma}(\varphi) = 0 = \mathbb{P}_{\mathcal{M},s_0,\sigma}(\llbracket \varphi \rrbracket^{s_0} \cap s_0 s_1 \cdots s_n S^{\omega})$.

In either case we have the equality $\mathbb{P}_{\mathcal{M}_*,s_0,\sigma}(\mathcal{R}) = \mathbb{P}_{\mathcal{M},s_0,\sigma}(\mathcal{R} \mid \llbracket \varphi \rrbracket^{s_0})$ for cylinders $\mathcal{R} = s_0 s_1 \cdots s_n S^{\omega}$. Since probability measures extend uniquely from cylinders [3], the equality holds for all measurable $\mathcal{R} \subseteq s_0 S^{\omega}$. Thus we have shown item 2. ◄

The following lemma is [13, Lemma 7].

▶ **Lemma 34.** *Let $\mathcal{M} = (S, S_{\square}, S_{\bigcirc}, \longrightarrow, P)$ be an MDP. Let φ be an objective that is prefix-independent in $\{\mathcal{M}\}$. Suppose that for any $s \in S$ and any strategy σ with $\mathbb{P}_{\mathcal{M},s,\sigma}(\varphi) = 1$ there exists an MD-strategy σ' with $\mathbb{P}_{\mathcal{M},s,\sigma'}(\varphi) = 1$. Then there is an MD-strategy σ' such that for all $s \in S$:*

$$\left(\exists \sigma \in \Sigma. \, \mathbb{P}_{\mathcal{M},s,\sigma}(\varphi) = 1 \right) \implies \mathbb{P}_{\mathcal{M},s,\sigma'}(\varphi) = 1$$

Proof. We can assume that all states are almost-surely winning, since in order to achieve an almost-sure winning objective, the player must forever remain in almost-surely winning states. So we need to define an MD-strategy σ' so that for all $s \in S$ we have $\mathbb{P}_{\mathcal{M},s,\sigma'}(\varphi) = 1$.

Fix an arbitrary state $s_1 \in S$. By assumption there is an MD-strategy σ_1 with $\mathbb{P}_{\mathcal{M},s_1,\sigma_1}(\varphi) = 1$. Let $U_1 \subseteq S$ be the set of states that occur in plays that both start from s_1 and are induced by σ_1. We have $\mathbb{P}_{\mathcal{M},s_1,\sigma_1}(\llbracket \varphi \rrbracket^{s_1} \cap U_1^{\omega}) = 1$. In fact, for any $s \in U_1$ and any strategy σ that agrees with σ_1 on U_1 we have $\mathbb{P}_{\mathcal{M},s,\sigma}(\llbracket \varphi \rrbracket^s \cap U_1^{\omega}) = 1$.

If $U_1 = S$ we are done. Otherwise, consider the MDP $\mathcal{M}_1$ obtained from $\mathcal{M}$ by fixing σ_1 on U_1 (i.e., in $\mathcal{M}_1$ we can view the states in U_1 as random states). We argue that, in $\mathcal{M}_1$, for any state s there is an MD-strategy σ_1' with $\mathbb{P}_{\mathcal{M}_1,s,\sigma_1'}(\varphi) = 1$. Indeed, let $s \in S$ be any state. Recall that there is an MD-strategy σ with $\mathbb{P}_{\mathcal{M},s,\sigma}(\varphi) = 1$. Let σ_1' be the MD-strategy obtained by restricting σ to the non-U_1 states (recall that the U_1 states are random states in $\mathcal{M}_1$). This strategy σ_1' almost surely generates a run that *either* satisfies φ without ever entering U_1 *or* at some point enters U_1. In the latter case, φ is satisfied almost surely: this follows from prefix-independence and the fact that σ_1' agrees with σ_1 on U_1. We conclude that $\mathbb{P}_{\mathcal{M}_1,s,\sigma_1'}(\varphi) = 1$.

Let $s_2 \in S \setminus U_1$. We repeat the argument from above, with s_2 instead of s_1, and with $\mathcal{M}_1$ instead of $\mathcal{M}$. This yields an MD-strategy σ_2 and a set $U_2 \ni s_2$ with $\mathbb{P}_{\mathcal{M}_1,s_2,\sigma_2}(\llbracket \varphi \rrbracket^{s_2} \cap U_2^{\omega}) = 1$. In fact, for any $s \in U_2$ and any strategy σ that agrees with σ_2 on U_2 and with σ_1 on U_1 we have $\mathbb{P}_{\mathcal{M},s,\sigma}(\llbracket \varphi \rrbracket^s \cap U_2^{\omega}) = 1$.

If $U_1 \cup U_2 = S$ we are done. Otherwise we continue in the same manner, and so forth. Since S is countable, we can pick $s_1, s_2, \ldots$ to have $\bigcup_{i \geq 1} U_i = S$. Define an MD-strategy σ' such that for any $s \in S_{\square}$ we have $\sigma'(s) = \sigma_i(s)$ for the smallest i with $s \in U_i$. Thus, if $s \in U_i$, we have $\mathbb{P}_{\mathcal{M},s,\sigma'}(\varphi) \geq \mathbb{P}_{\mathcal{M},s,\sigma'}(\llbracket \varphi \rrbracket^s \cap U_i^{\omega}) = 1$. ◄

The following lemma is [14, Lemma 8].

▶ **Lemma 35.** *Let S be countable and $s \in S$. Call a set of the form swS^{ω} for $w \in S^*$ a cylinder. Let $\mathbb{P}, \mathbb{P}'$ be probability measures on sS^{ω} defined in the standard way, i.e., first on cylinders and then extended to all measurable sets $\mathcal{R} \subseteq sS^{\omega}$. Suppose there is $x \geq 0$ such that $x \cdot \mathbb{P}(\mathfrak{C}) \leq \mathbb{P}'(\mathfrak{C})$ for all cylinders $\mathfrak{C}$. Then $x \cdot \mathbb{P}(\mathcal{R}) \leq \mathbb{P}'(\mathcal{R})$ holds for all measurable $\mathcal{R} \subseteq sS^{\omega}$.*

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 the set of finite disjoint unions of cylinders [3, Section 2]. Hence $x \cdot \mathbb{P}(\mathcal{R}) \leq \mathbb{P}'(\mathcal{R})$ for all $\mathcal{R} \in \mathcal{C}_*$.

Define

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

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 $\mathcal{R}_1, \mathcal{R}_2, \ldots \in \mathcal{Q}$, then $\mathcal{R}_1 \subseteq \mathcal{R}_2 \subseteq \cdots$ implies $\bigcup_i \mathcal{R}_i \in \mathcal{Q}$, and $\mathcal{R}_1 \supseteq \mathcal{R}_2 \supseteq \cdots$ implies $\bigcap_i \mathcal{R}_i \in \mathcal{Q}$. Suppose $\mathcal{R}_1, \mathcal{R}_2, \ldots \in \mathcal{Q}$ and $\mathcal{R}_1 \subseteq \mathcal{R}_2 \subseteq \cdots$. Then:

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

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

Now the *monotone class theorem* (see, e.g., [3, Theorem 3.4]) implies that $\sigma(\mathcal{C}) \subseteq \mathcal{Q}$, thus $\mathcal{Q} = \sigma(\mathcal{C})$. Hence $x \cdot \mathbb{P}(\mathcal{R}) \leq \mathbb{P}'(\mathcal{R})$ for all $\mathcal{R} \in \sigma(\mathcal{C})$. ◄

The following theorem is a variant of [13, Theorem 5].

▶ **Theorem 20.** *Let φ be a tail objective. Let $\mathcal{M} = (S, S_\square, S_\bigcirc, \longrightarrow, P)$ be an MDP and $\mathcal{M}_* = (S_*, S_{*\square}, S_{*\bigcirc}, \longrightarrow_*, P_*)$ its conditioned version wrt. φ. Then:*
1. *For all $s \in S_*$ there exists a strategy σ with $\mathbb{P}_{\mathcal{M}_*, s, \sigma}(\varphi) = 1$.*
2. *Suppose that for every $s \in S_*$ there exists an MD strategy σ'' with $\mathbb{P}_{\mathcal{M}_*, s, \sigma''}(\varphi) = 1$. Then there is an MD strategy σ' such that for all $s \in S$:*

$$\big(\exists \sigma \in \Sigma. \, \mathbb{P}_{\mathcal{M}, s, \sigma}(\varphi) = \mathtt{val}_\mathcal{M}(s)\big) \implies \mathbb{P}_{\mathcal{M}, s, \sigma'}(\varphi) = \mathtt{val}_\mathcal{M}(s)$$

Proof. Towards item 1, let $s \in S_*$. By the definition of S_*, there is a strategy σ with $\mathbb{P}_{\mathcal{M}, s, \sigma}(\varphi) = \mathtt{val}_\mathcal{M}(s) > 0$. By Lemma 33.2, we have $\mathbb{P}_{\mathcal{M}_*, s, \sigma}(\varphi) = 1$, as desired.

It remains to prove item 2. Suppose that for any $s \in S_*$ there exists an MD-strategy σ'' with $\mathbb{P}_{\mathcal{M}_*, s, \sigma''}(\varphi) = 1$. By Lemma 34, it follows that there is an MD-strategy σ' with $\mathbb{P}_{\mathcal{M}_*, s, \sigma'}(\varphi) = 1$ for all $s \in S_*$. We show that this strategy σ' satisfies the property claimed in the statement of the theorem.

To this end, let $n \geq 0$ and $s_0, s_1, \ldots, s_n \in S$. If $s_0 s_1 \cdots s_n$ is a partial run in $\mathcal{M}_*$ then, by Lemma 33.1,

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

and thus, as $\mathtt{val}_\mathcal{M}(s_n) \leq 1$,

$$\mathtt{val}_\mathcal{M}(s_0) \cdot \mathbb{P}_{\mathcal{M}_*, s_0, \sigma'}(s_0 s_1 \cdots s_n S^\omega) \leq \mathbb{P}_{\mathcal{M}, s_0, \sigma'}(s_0 s_1 \cdots s_n S^\omega) \, .$$

If $s_0 s_1 \cdots s_n$ is not a partial run in $\mathcal{M}_*$ then $\mathbb{P}_{\mathcal{M}_*, s_0, \sigma'}(s_0 s_1 \cdots s_n S^\omega) = 0$ and the previous inequality holds as well. Therefore, by Lemma 35, we get for all measurable sets $\mathcal{R} \subseteq s_0 S^\omega$:

$$\mathtt{val}_\mathcal{M}(s_0) \cdot \mathbb{P}_{\mathcal{M}_*, s_0, \sigma'}(\mathcal{R}) \leq \mathbb{P}_{\mathcal{M}, s_0, \sigma'}(\mathcal{R})$$

In particular, since $\mathbb{P}_{\mathcal{M}_*, s_0, \sigma'}(\varphi) = 1$, we obtain $\mathtt{val}_\mathcal{M}(s_0) \leq \mathbb{P}_{\mathcal{M}, s_0, \sigma'}(\varphi)$. The converse inequality $\mathbb{P}_{\mathcal{M}, s_0, \sigma'}(\varphi) \leq \mathtt{val}_\mathcal{M}(s_0)$ holds by the definition of $\mathtt{val}_\mathcal{M}(s_0)$, hence we conclude $\mathbb{P}_{\mathcal{M}, s_0, \sigma'}(\varphi) = \mathtt{val}_\mathcal{M}(s_0)$. ◄

D   Missing proofs in Section 3

We first recall our results [12] on the strategy complexity of Büchi objectives:

▶ **Theorem 12** (Theorem 5 in [12]). *For every acyclic countable MDP $\mathcal{M}$, a Büchi objective φ, finite set I of initial states and $\varepsilon > 0$, there exists a deterministic 1-bit strategy that is ε-optimal from every $s \in I$.*

We will prove that

▷ **Claim 13.**   In MDP $\mathcal{L}_{e-2}$, there is an MD strategy τ_e, that is $(\alpha - \beta)$-optimal for θ_e from R_e.

Proof. Consider the original MDP $\mathcal{M}$. Given a set $B \subseteq L$ in $\mathcal{L}$, we use $\mathrm{project}(B) \stackrel{\text{def}}{=} \{s \mid (s, b) \in B, b \in \{0, 1\}\}$ to project the set into $\mathcal{M}$.

We first slightly modify $\mathcal{M}$ to obtain $\mathcal{M}'$. The modification guarantees that, for all states s and runs ρ of $\mathcal{M}'$,

$$s\rho \in \left[\!\left[\mathsf{GF}[S]^{Col=e}\right]\!\right] \quad \text{if and only if} \quad s\rho \in \left[\!\left[\mathsf{GF}[S]^{Col=e} \wedge \mathsf{G}[S]^{Col\leq e} \wedge \mathsf{G}\neg\mathrm{project}(\mathsf{Fix}_e)\right]\!\right].$$

We redirect all out-going transitions of states $s' \in \mathrm{project}(\mathsf{Fix}_e)$ or s' with $Col(s') > e$ to an infinite chain $q_0 q_1 q_2 \cdots$ of controlled states where $Col(q_i) = 1$ and $q_i \longrightarrow q_{i+1}$. We also update the color of all states s with $Col(s) < e$ to 1.

The objective $\mathsf{GF}[S]^{Col=e}$ is a Büchi Objective in $\mathcal{M}'$. By Theorem 12, given the *finite* set $\mathrm{project}(R_e)$ of initial states, there exists a deterministic 1-bit strategy σ in $\mathcal{M}'$ that is $(\alpha - \beta)$-optimal w.r.t. $\mathsf{GF}[S]^{Col=e}$ for every state $s \in \mathrm{project}(R_e)$ (with the memory bit initially set to 0).

Since the fixed choices in $\mathcal{L}_{e-2}$ are only in the Fix_{e-2}-region, strategy σ can be translated in a natural way to a deterministic memoryless strategy σ' in $\mathcal{L}_{e-2}$: For a state $s \in S_\square$ and $b \in \{0, 1\}$, if σ chooses the successor state s', by taking a transition $t = (s, s')$, and updates the bit to b', we define $\sigma'((s, b)) \stackrel{\text{def}}{=} (t, b)$ and $\sigma'((t, b)) = (s', b')$. For a random state $s \in S_\bigcirc$ and $b \in \{0, 1\}$, if the strategy σ updates the memory bit to b' in case the random successor resolves to s', by taking a transition $t = (s, s')$, we define $\sigma'((t, b)) \stackrel{\text{def}}{=} (s', b')$. Recall that the bit is initially set to 0 in σ. Consequently, the strategy σ' is $(\alpha - \beta)$-optimal for θ_e from every state $\ell \in R_e$ in the layered MDP $\mathcal{L}_{e-2}$.   ◁

We next prove the main technical claim in Section 3:

▷ **Claim 14.** $\mathbb{P}_{\mathcal{L}_e, \ell_0, \pi}(\psi_e) \geq \mathtt{val}_{\mathcal{L}_{e-2}, \psi_{e-2}}(\ell_0) - 2\gamma.$

Proof. Recall the definition of π: it starts by following σ_{ℓ_0}. If it ever enters $Cl(\mathit{fix}_e)$ then we ensure that it enters fix_e as well (in at most one more step). Then π continues by playing as τ_e does forever.

Below we argue that if π ever enters $Cl(\mathit{fix}_e)$ then it is in fact possible to choose the layer in such a way that π enters fix_e instead. Assume π enters $Cl(\mathit{fix}_e)$ at q after taking a transition from p to q. Let $\bar{q} \in \mathit{fix}_e$ be the sibling of q. By construction,

1. either $p \in \longrightarrow_1 \times \{0, 1\}$ is controlled: the controller switches the layer in p, by choosing $\bar{q}$ rather than q and enters fix_e;
2. or $q \in \longrightarrow_1 \times \{0, 1\}$ is controlled. By definition (6), the MD strategy τ_e attains a high value from state $\bar{q}$ for θ_e. Hence, $\tau_e(\bar{q}) \in \mathit{fix}_e$. Hence, the controller can switch the layer in q by playing $\tau_e(\bar{q})$ and enters fix_e.

For all $e' \in \{2, 4, \cdots, e_{\max}\}$ define

$$\chi_{e'} \stackrel{\text{def}}{=} \mathsf{F}\varphi_{e'} \wedge \mathsf{G}\neg\mathsf{Fix}_{e-2} \qquad\qquad \tilde{\chi}_{e'} \stackrel{\text{def}}{=} \mathsf{F}\varphi_{e'} \wedge \mathsf{G}\neg\mathsf{Fix}_e.$$

We define

$$
\begin{aligned}
\psi &\stackrel{\text{def}}{=} \bigvee_{e'<e} \mathsf{F}\,core_{e'} \vee \bigvee_{e'>e} \chi_{e'} \\
\psi' &\stackrel{\text{def}}{=} \bigvee_{e'<e} \mathsf{F}\,core_{e'} \vee \bigvee_{e'>e} \tilde{\chi}_{e'}
\end{aligned}
\tag{18}
$$

By definition of ψ_{e-2} and ψ_e, see definition (4), we have $\psi_{e-2} = \psi \vee \chi_e$ and $\psi_e = \psi' \vee \mathsf{F}\,core_e$. For brevity, further define $\rho \stackrel{\text{def}}{=} \mathsf{F}\,Cl(fix_e)$. Observe that $[\![\psi \wedge \neg\rho]\!] \subseteq [\![\psi']\!]$.

We first have that

$$
\begin{aligned}
&\mathbb{P}_{\mathcal{L}_e,\ell_0,\pi}(\mathsf{F}\,core_e) \\
&\quad \geq \mathbb{P}_{\mathcal{L}_e,\ell_0,\pi}(\mathsf{F}\,core_e \wedge \rho) \\
&\quad \geq \mathbb{P}_{\mathcal{L}_e,\ell_0,\pi}(\neg Cl(fix_e) \text{ until } (Cl(fix_e) \wedge \mathsf{F}\varphi_e \wedge \mathsf{F}\,core_e \wedge \mathsf{G}\neg\mathsf{Fix}_{e-2})) \\
&\quad = \sum_{\ell \in Cl(fix_e)} \mathbb{P}_{\mathcal{L}_e,\ell_0,\pi}(\neg Cl(fix_e) \text{ until } \ell) \cdot \mathbb{P}_{\mathcal{L}_e,\ell,\tau_e}(\theta_e \wedge \mathsf{F}\,core_e) \\
&\quad = \sum_{\ell \in Cl(fix_e)} \mathbb{P}_{\mathcal{L}_e,\ell_0,\pi}(\neg Cl(fix_e) \text{ until } \ell) \cdot \mathbb{P}_{\mathcal{L}_e,\ell,\tau_e}(\theta_e) \qquad \text{by Lemma 16.2} \\
&\quad \geq \sum_{\ell \in Cl(fix_e)} \mathbb{P}_{\mathcal{L}_e,\ell_0,\pi}(\neg Cl(fix_e) \text{ until } \ell) \cdot \beta \\
&\quad = \mathbb{P}_{\mathcal{L}_e,\ell_0,\pi}(\rho) \cdot \beta \\
&\quad = \mathbb{P}_{\mathcal{L}_{e-2},\ell_0,\sigma_{\ell_0}}(\rho) \cdot \beta \\
&\quad = \mathbb{P}_{\mathcal{L}_{e-2},\ell_0,\sigma_{\ell_0}}(\rho) \cdot (1-\gamma) \\
&\quad \geq \mathbb{P}_{\mathcal{L}_{e-2},\ell_0,\sigma_{\ell_0}}(\rho) - \gamma
\end{aligned}
\tag{19}
$$

We use the law of total probability:

$$\mathbb{P}_{\mathcal{L}_{e-2},\ell_0,\sigma_{\ell_0}}(\psi_{e-2}) = \mathbb{P}_{\mathcal{L}_{e-2},\ell_0,\sigma_{\ell_0}}(\psi \wedge \neg\rho) + \mathbb{P}_{\mathcal{L}_{e-2},\ell_0,\sigma_{\ell_0}}(\chi_e) + \mathbb{P}_{\mathcal{L}_{e-2},\ell_0,\sigma_{\ell_0}}(\psi \wedge \rho) \tag{20}$$

In one hand, since $\mathcal{L}_e$ and $\mathcal{L}_{e-2}$ only differ in the fix_e-region, and since π plays as σ_{ℓ_0} on all runs contained in $\neg\rho$:

$$\mathbb{P}_{\mathcal{L}_{e-2},\ell_0,\sigma_{\ell_0}}(\psi \wedge \neg\rho) = \mathbb{P}_{\mathcal{L}_e,\ell_0,\pi}(\psi \wedge \neg\rho) \leq \mathbb{P}_{\mathcal{L}_e,\ell_0,\pi}(\psi')$$

In the other hand, by Equation (9):

$$\mathbb{P}_{\mathcal{L}_{e-2},\ell_0,\sigma_{\ell_0}}(\chi_e) \leq \mathbb{P}_{\mathcal{L}_{e-2},\ell_0,\sigma_{\ell_0}}(\chi_e \wedge \rho) + \frac{\gamma}{2}$$

Applying the above to Equation (20) yields:

$$
\mathbb{P}_{\mathcal{L}_{e-2},\ell_0,\sigma_{\ell_0}}(\psi_{e-2})
$$

$$
\begin{aligned}
&\leq \mathbb{P}_{\mathcal{L}_e,\ell_0,\pi}(\psi') + \mathbb{P}_{\mathcal{L}_{e-2},\ell_0,\sigma_{\ell_0}}(\chi_e \wedge \rho) + \mathbb{P}_{\mathcal{L}_{e-2},\ell_0,\sigma_{\ell_0}}(\psi \wedge \rho) + \frac{\gamma}{2} \\
&= \mathbb{P}_{\mathcal{L}_e,\ell_0,\pi}(\psi') + \mathbb{P}_{\mathcal{L}_{e-2},\ell_0,\sigma_{\ell_0}}((\chi_e \vee \psi) \wedge \rho) + \frac{\gamma}{2} && \text{since } \chi_e \text{ and } \psi \text{ are disjoint} \\
&\leq \mathbb{P}_{\mathcal{L}_e,\ell_0,\pi}(\psi') + \mathbb{P}_{\mathcal{L}_{e-2},\ell_0,\sigma_{\ell_0}}(\rho) + \frac{\gamma}{2} \\
&\leq \mathbb{P}_{\mathcal{L}_e,\ell_0,\pi}(\psi') + \mathbb{P}_{\mathcal{L}_{e-2},\ell_0,\pi}(\mathsf{F}\,core_e) + \gamma + \frac{\gamma}{2} && \text{by Equation (19)} \\
&= \mathbb{P}_{\mathcal{L}_e,\ell_0,\pi}(\psi_e) + \frac{3\gamma}{2}
\end{aligned}
$$

To conclude the proof we recall that σ_{ℓ_0} is $\frac{\gamma}{2}$-optimal w.r.t ψ_{e-2}. $\triangleleft$

$\triangleright$ **Claim 17.** The MD strategy $\hat{\sigma}$ is ε-optimal for parity objective φ, from every state $\ell_0 \in L_0$.

Proof. For the MD strategy $\hat{\sigma}$, by the law of total probability, we have

$$
\mathbb{P}_{\mathcal{L},\ell_0,\hat{\sigma}}(\varphi) \geq \sum_{e \in even(\mathcal{C})} \mathbb{P}_{\mathcal{L},\ell_0,\hat{\sigma}}(\mathsf{F}\varphi_e \wedge \mathsf{F}\,core_e)\,.
$$

Let $\mathcal{C}'$ be the set of even colors e where $\mathbb{P}_{\mathcal{L},\ell_0,\hat{\sigma}}(\mathsf{F}\,core_e) > 0$. Then:

$$
\begin{aligned}
&= \sum_{e \in \mathcal{C}'} \mathbb{P}_{\mathcal{L},\ell_0,\hat{\sigma}}(\mathsf{F}\varphi_e \mid \mathsf{F}\,core_e) \cdot \mathbb{P}_{\mathcal{L},\ell_0,\hat{\sigma}}(\mathsf{F}\,core_e) \\
&\geq \sum_{e \in \mathcal{C}'} \mathbb{P}_{\mathcal{L},\ell_0,\hat{\sigma}}(\mathsf{G}\,fix_e \mid \mathsf{F}\,core_e) \cdot \mathbb{P}_{\mathcal{L},\ell_0,\hat{\sigma}}(\mathsf{F}\,core_e) && \text{by Equation (12)} \\
&\geq \sum_{e \in \mathcal{C}'} (1-\gamma) \cdot \mathbb{P}_{\mathcal{L},\ell_0,\hat{\sigma}}(\mathsf{F}\,core_e) && \text{by Equation (11)} \\
&= (1-\gamma) \cdot \sum_{e \in even(\mathcal{C})} \mathbb{P}_{\mathcal{L},\ell_0,\hat{\sigma}}(\mathsf{F}\,core_e) \\
&\geq (1-\gamma) \cdot \mathbb{P}_{\mathcal{L},\ell_0,\hat{\sigma}}(\psi_{e_{\max}})
\end{aligned}
$$

since τ_{reach} is γ-optimal and by Equation (10),

$$
\begin{aligned}
&\geq (1-\gamma) \cdot (\mathtt{val}_{\mathcal{L},\varphi}(\ell_0) - e_{\max}\gamma - \gamma) \\
&\geq \mathtt{val}_{\mathcal{L},\varphi}(\ell_0) - (e_{\max} + 2)\gamma
\end{aligned}
$$

Recall that $\varepsilon = (e_{\max} + 2)\gamma$. Thus we have shown that the MD strategy $\hat{\sigma}$ is ε-optimal w.r.t. φ from every state $\ell_0 \in L_0$. $\triangleleft$

E   Missing proofs in Section 4

▶ **Definition 36** (Bubbles). *Let $\mathcal{M}$ be an MDP with states S, $R \subseteq S$, $l \in \mathbb{N}$. The l-bubble around R is the set*

$$
\mathsf{bubble}(\mathcal{M}, R, l) \stackrel{def}{=} \{s \mid \exists s_0 \in R.\exists\tau.\mathbb{P}_{\mathcal{M},s_0,\tau}(\mathsf{F}^{\leq l}s) > 0\}
$$

of states that can be reached from R in at most l steps. Any bubble around a closed set $R \subseteq L$ is closed.

Recall that for an MD strategy τ, we write $\mathcal{M}[\tau, R]$ for the MDP obtained from $\mathcal{M}$ by fixing the strategy τ for all states in R. We will simply write $\mathcal{M}[\tau]$ for $\mathcal{M}[\tau, S]$, where τ is fixed everywhere, and $\mathcal{M}[\tau, R, l] \stackrel{\text{def}}{=} \mathcal{M}[\tau, \mathsf{bubble}(\mathcal{M}, R, l)]$ that fixes τ in the l-bubble around R.

▶ **Lemma 18.** *Let $\mathcal{L}(\mathcal{M})$ be the layered MDP obtained from an acyclic and finitely branching MDP $\mathcal{M}$ and a coloring Col such that all states are almost surely winning for $\varphi = \mathtt{Parity}(Col)$ (i.e., every state s has a strategy σ_s such that $\mathbb{P}_{\mathcal{L}(\mathcal{M}),s,\sigma_s}(\varphi) = 1$).*

For every initial state s_0 there exists an MD strategy σ that almost surely wins, i.e., $\mathbb{P}_{\mathcal{L}(\mathcal{M}),s_0,\sigma}(\varphi) = 1$.

Proof. Directly from Lemma 37 (let $L_0 \stackrel{\text{def}}{=} Cl(\{s_0\})$). ◀

▶ **Lemma 37.** *Let $\mathcal{L}(\mathcal{M})$ be the layered MDP obtained from an acyclic and finitely branching MDP $\mathcal{M}$ and a coloring Col such that all states are almost surely winning for $\varphi = \mathtt{Parity}(Col)$ (i.e., every state s has a strategy σ_s such that $\mathbb{P}_{\mathcal{L}(\mathcal{M}),s,\sigma_s}(\varphi) = 1$).*

For every finite closed set L_0 of initial states there exists an MD strategy $\hat{\sigma}$ that almost surely wins from every state $s_0 \in L_0$. That is, $\forall s_0 \in L_0. \ \mathbb{P}_{\mathcal{L}(\mathcal{M}),s_0,\hat{\sigma}}(\varphi) = 1$.

Proof. We iteratively produce an infinite sequence $\mathcal{L}_0, \mathcal{L}_1, \mathcal{L}_2, \ldots$ of layered MDPs. They have the same structure as $\mathcal{L}(\mathcal{M})$, but in each step from $\mathcal{L}_i$ to $\mathcal{L}_{i+1}$ the choices in some subset of states (reachable from L_0) are fixed. In the limit all choices from all controlled states reachable from L_0 are fixed. Hence this prescribes an MD strategy $\hat{\sigma}$ from L_0 in $\mathcal{L}(\mathcal{M})$. It is not sufficient that these fixings of MD strategies in subspaces are compatible with some strategy almost sure winning for φ, since progress (e.g., towards visiting a particular color) might only be made outside of the fixed subspace, and thus be delayed forever. Instead we prove the stronger property that $\hat{\sigma}$ ensures φ with some probability $p_i(s_0)$ from $s_0 \in L_0$ already in the fixed subspace of $\mathcal{L}_i$ alone, and that $\lim_{i\to\infty} p_i(s_0) = 1$. This then implies that $\hat{\sigma}$ is almost surely winning for φ in $\mathcal{L}(\mathcal{M})$.

The sea urchin construction.

Its name comes from the shape of the subspace where strategies are fixed: a finite body H_i out of which come finitely many spikes (BETA_i, where each spike is infinite). As the body grows, more spikes are added. Eventually the sea urchin covers the entire space; see Figure 4.

The construction uses some global thresholds $1 > \alpha > \beta > \gamma > 0$, to be determined later. Moreover, in each step from $\mathcal{L}_i$ to $\mathcal{L}_{i+1}$ we will define the following notions.

- Small error thresholds $\varepsilon_j > 0$ for $j \in \{0, 1, 2, 3\}$).
- Thresholds $l_i, k_i \in \mathbb{N}$ of a number of steps from L_0.
- Finite closed subsets of states H_i where $H_0 \stackrel{\text{def}}{=} \emptyset$ and $H_i \stackrel{\text{def}}{=} \mathsf{bubble}(\mathcal{L}, L_0, k_i)$ for $i > 0$. (H_i is finite, because $\mathcal{L}$ is finitely branching.)
- Finite subsets $L_i \subseteq L$ as starting sets for certain modified objectives φ_i (see below).
- MD strategies τ_i (for $i > 0$) and subsets of states $\mathsf{ALPHA}_i \subseteq \mathsf{BETA}_i \subseteq \mathsf{GAMMA}_i \subseteq L$, where ALPHA_i (resp. BETA_i, GAMMA_i) are the sets of states from which τ_i attains $\geq \alpha$ (resp. $\geq \beta$, $\geq \gamma$) for objective φ_i (see below) in $\mathcal{L}_{i-1}$ (and $\mathcal{L}_i$). Let $\mathsf{ALPHA}_0 = \mathsf{BETA}_0 = \mathsf{GAMMA}_0 \stackrel{\text{def}}{=} \emptyset$, $\mathsf{ALPHA}_i \stackrel{\text{def}}{=} Safe_{\mathcal{L}_{i-1}, \tau_i, \varphi_i}(\alpha)$ (and similar for BETA_i, GAMMA_i).
  We write $\mathsf{ALPHA}_{\leq i} \stackrel{\text{def}}{=} \bigcup_{j \leq i} \mathsf{ALPHA}_j$ and similar for $\mathsf{BETA}_{\leq i}$, $\mathsf{GAMMA}_{\leq i}$.
- Let $\mathsf{FIX}_i \stackrel{\text{def}}{=} \mathsf{BETA}_{\leq i} \cup H_i$. This is the subspace where choices are fixed in rounds up-to i.
- Modified objectives φ_i with $\varphi_0 \stackrel{\text{def}}{=} \varphi$ and $\varphi_{i+1} \stackrel{\text{def}}{=} \varphi \wedge \mathsf{G}(L \setminus Cl(\mathsf{FIX}_i))$. For $i > 0$ the φ_i are not strictly tail objectives, but they still enjoy the same properties as tail objectives wrt. the Levy zero-one law; cf. Remark 11.

- In $\mathcal{L}_i$ the choices inside FIX_i are already fixed. Inside BETA_i the strategy τ_i is fixed, and inside $H_i \setminus \mathsf{BETA}_{\leq i}$ the choices are fixed according to another MD strategy ρ_i.
- It follows from the properties above that we have the invariant

$$\mathsf{GAMMA}_{i+1} \cap Cl(\mathsf{FIX}_i) = \emptyset \tag{21}$$

In particular, the sets BETA_j are disjoint for different j. However, for $j' > j$, it is possible that GAMMA_j overlaps with $\mathsf{BETA}_{j'}$ (and $\mathsf{GAMMA}_{j'}$).

Base case.

We start with the MDP $\mathcal{L}_0 \overset{\text{def}}{=} \mathcal{L} \overset{\text{def}}{=} \mathcal{L}(\mathcal{M})$. By assumption, in $\mathcal{L}_0$ all states are almost surely winning for $\varphi_0 \overset{\text{def}}{=} \varphi$ (the unrestricted parity objective). The invariant (21) is trivially satisfied for $i = 0$, since $Cl(\mathsf{FIX}_0) = \emptyset$.

Step.

Now we define the step from $\mathcal{L}_i$ to $\mathcal{L}_{i+1}$ for $i \geq 0$. We assume that for all $j \leq i$ the MD strategies τ_j and the sets $\mathsf{ALPHA}_j \subseteq \mathsf{BETA}_j \subseteq \mathsf{GAMMA}_j$ and H_j are already defined. Moreover, τ_j is fixed inside BETA_j, and in $\mathcal{L}_j$ the strategy τ_j attains at least β for objective φ_j from each state $s \in \mathsf{BETA}_j$. Moreover, some other MD strategy is fixed in $H_j \setminus \mathsf{BETA}_{\leq j}$. (All this trivially holds for the base case $i = 0$. For $i > 0$ our construction will ensure these properties.)

    We now consider $\mathcal{L}_i$. By s_0 we denote initial states in L_0. (General states are denoted by s.) We show that in $\mathcal{L}_i$, all initial states $s_0 \in L_0$ are still almost surely winning for φ, as witnessed by a *resetting strategy* σ defined below (where σ is generally not MD, except inside the subspace FIX_i). First we need a basic property of $\mathsf{ALPHA}_j, \mathsf{BETA}_j$.

$\triangleright$ **Claim 38.**   Let $0 \leq j \leq i$ and σ be an arbitrary strategy in $\mathcal{L}_i$. If $s \in \mathsf{ALPHA}_j$ then $\mathbb{P}_{\mathcal{L}_i, s, \sigma}(\mathsf{G}\ \mathsf{BETA}_j) \geq \frac{\alpha - \beta}{1 - \beta}$.

Proof. By Lemma 15, since σ behaves just like τ_j in the relevant subspaces already fixed to τ_j in $\mathcal{L}_i$. $\blacktriangleleft$

    Recall that for every state $s \in L$ there exists an almost surely winning strategy $\sigma(s)$ for φ in $\mathcal{L}$. The *resetting strategy* σ in $\mathcal{L}_i$ starts in L_0 and behaves as specified in the three different modes $\mathsf{m}_1, \mathsf{m}_2, \mathsf{m}_3$ as follows. For all $j \leq i$:

1. In H_j it plays as prescribed by the fixing there, (starting in memory mode m_1).
2. Whenever σ enters a set $Cl(\mathsf{BETA}_j) \setminus H_i$ then it switches to mode m_2 and chooses the layer in such a way that it enters even BETA_j (in at most one more step) and continues playing τ_j, as required by the fixing inside BETA_j. [1] [2] Inside BETA_j, it plays τ_j that is fixed in BETA_j. It continues to play τ_j even in $\mathsf{GAMMA}_j \setminus (\mathsf{BETA}_{\leq i} \cup H_i)$.
3. While playing in mode m_2 (or m_1), upon reaching an unfixed state s outside of GAMMA_j (and outside of H_i), it goes to mode m_3 and resets to an almost surely winning strategy $\sigma(s)$ for φ in $\mathcal{L}$. It keeps playing $\sigma(s)$ until (and if) it reaches the fixed part $\mathsf{BETA}_{\leq i} \cup H_i$, whereupon it continues as before with mode m_2.

---

[1] By Definition 6, either the current state or the next state allows to switch between layers; cf. the proof of Claim 14.

[2] Remember that (21) implies that the sets BETA_j are disjoint.

We will see that, not only is the resetting strategy σ almost surely winning for φ, but every time it re-enters $\mathsf{BETA}_{\leq i}$ it has a lower-bounded chance of *eventually* staying in $\mathsf{ALPHA}_{\leq i}$ forever.

We now classify the runs induced by the resetting strategy σ (from some initial state $s_0 \in L_0$) according to how often which modes $\mathsf{m}_1, \mathsf{m}_2, \mathsf{m}_3$ are used.

First we note that, since $\mathcal{L}_i$ is acyclic, under any strategy (and in particular σ), any run can visit any finite set (in particular H_i) only finitely often and therefore has an infinite suffix that is always outside H_i. Thus σ is eventually always not in mode m_1.

By our invariant (21), playing in BETA_i and GAMMA_i is not restricted by our previous fixings in $\mathsf{BETA}_{\leq i-1} \cup H_{i-1}$. Thus, when playing from $s \in \mathsf{BETA}_i$ in mode m_2, we keep playing τ_i even in GAMMA_i. Analogously to Claim 38, the chance of staying in the set GAMMA_i can be lower bounded.

$$\forall s \in \mathsf{BETA}_i \, \mathbb{P}_{\mathcal{L}_i, s, \sigma[\mathsf{m}_2]}(\mathsf{G}\,\mathsf{GAMMA}_i) \geq \frac{\beta - \gamma}{1 - \gamma} > 0 \tag{22}$$

This again follows from Lemma 15, observing that σ behaves just like τ_i even inside GAMMA_i while staying in mode m_2.

When playing from $s \in \mathsf{BETA}_j$ for some $j < i$, a similar property holds. If a run visits some state $s' \in Cl(\mathsf{BETA}_{j'})$, for some $j' > j$, then we can assume that we have even $s' \in \mathsf{BETA}_{j'}$ by our assumption on σ above, because outside of the fixed region the layer can be chosen freely. Then the strategy switches from τ_j to $\tau_{j'}$ from $s' \in \mathsf{BETA}_{j'}$. Otherwise we keep playing τ_j while in GAMMA_j, i.e., by Lemma 15, we get, for every $s \in \mathsf{BETA}_j$, that

$$\mathbb{P}_{\mathcal{L}_i, s, \sigma[\mathsf{m}_2]}(\mathsf{G}\,\mathsf{GAMMA}_j \vee \mathsf{F}\,\mathsf{BETA}_{>j}) \geq \frac{\beta - \gamma}{1 - \gamma} > 0. \tag{23}$$

From (22) and (23), we obtain that the set of runs that infinitely often switch from mode m_2 to m_3 are a null-set. Moreover, as shown above, every run has an infinite suffix where the mode is not m_1. It follows that, except for a null-set, all runs either have an infinite suffix in mode m_2 or an infinite suffix in mode m_3. Let $\mathcal{R}_2$ and $\mathcal{R}_3$ denote these subsets of runs, respectively. I.e., we have $\forall s_0 \in L_0$

$$\mathbb{P}_{\mathcal{L}_i, s_0, \sigma}(\mathcal{R}_2 \cup \mathcal{R}_3) = 1. \tag{24}$$

In mode m_3 the resetting strategy σ plays an almost surely winning strategy for φ outside of $Cl(\mathsf{FIX}_i)$ that is not impeded by the fixings in $\mathcal{L}_i$, and φ is a tail objective. Thus, for all $s_0 \in L_0$,

$$\mathbb{P}_{\mathcal{L}_i, s_0, \sigma}(\mathcal{R}_3) = \mathbb{P}_{\mathcal{L}_i, s_0, \sigma}(\mathcal{R}_3 \wedge \varphi \wedge \mathsf{FG}(L \setminus Cl(\mathsf{FIX}_i))). \tag{25}$$

From the property that φ is tail and the definition of φ_{i+1} as $\varphi \wedge \mathsf{G}(L \setminus Cl(\mathsf{FIX}_i))$ we obtain that, for all $s_0 \in L_0$,

$$\mathbb{P}_{\mathcal{L}_i, s_0, \sigma}(\mathcal{R}_3) = \mathbb{P}_{\mathcal{L}_i, s_0, \sigma}(\mathcal{R}_3 \wedge \mathsf{F}\varphi_{i+1}) \tag{26}$$

In mode m_2 the resetting strategy σ plays some MD strategy τ_j in GAMMA_j (for some $j \leq i$). Thus, for all $s_0 \in L_0$,

$$\mathbb{P}_{\mathcal{L}_i, s_0, \sigma}(\mathcal{R}_2) = \mathbb{P}_{\mathcal{L}_i, s_0, \sigma}(\mathcal{R}_2 \wedge \mathsf{FG}\,\mathsf{GAMMA}_{\leq i}). \tag{27}$$

Since in mode m_2 the resetting strategy σ plays some MD strategy τ_j with attainment $\geq \gamma$ (resp. $\geq \alpha$) in GAMMA_j (resp. ALPHA_j), we can apply Levy's zero-one law (Corollary 30) and obtain even

$$\mathbb{P}_{\mathcal{L}_i, s_0, \sigma}(\mathcal{R}_2) = \mathbb{P}_{\mathcal{L}_i, s_0, \sigma}(\mathcal{R}_2 \wedge \mathsf{FG}\,\mathsf{ALPHA}_{\leq i}) \tag{28}$$

By (24), (26) and (28) we obtain

$$\mathbb{P}_{\mathcal{L}_i,s_0,\sigma}(\mathsf{FG\,ALPHA}_{\leq i} \vee \mathsf{F}\varphi_{i+1}) = 1 \tag{29}$$

σ plays like τ_j inside ALPHA_j which attains $\geq \alpha$ for φ. By using Levy's zero-one law (Lemma 16(1)) for safety sets at level α, we obtain that $\mathbb{P}_{\mathcal{L}_i,s_0,\sigma}(\varphi \wedge \mathsf{FG\,ALPHA}_{\leq i}) = \mathbb{P}_{\mathcal{L}_i,s_0,\sigma}(\mathsf{FG\,ALPHA}_{\leq i})$. Since $\mathsf{F}\varphi_{i+1} \subseteq \varphi$ it follows from Equation (29) that $\mathbb{P}_{\mathcal{L}_i,s_0,\sigma}(\varphi) = 1$, i.e., the resetting strategy σ wins φ almost surely.

For $s_0 \in L_0$ let

$$p_i(s_0) \stackrel{\text{def}}{=} \mathbb{P}_{\mathcal{L}_i,s_0,\sigma}(\varphi \wedge \mathsf{G\,FIX}_i)$$

be the attainment for φ inside the fixed region FIX_i of $\mathcal{L}_i$.

Since H_i is finite and $\mathcal{L}_i$ is acyclic, almost surely H_i is eventually left forever. Moreover, the sets ALPHA_j are safety sets (at level α) for φ. It follows from Levy's zero-one law (cf. Corollary 29) that

$$p_i(s_0) = \mathbb{P}_{\mathcal{L}_i,s_0,\sigma}(\varphi \wedge \mathsf{G\,FIX}_i) = \mathbb{P}_{\mathcal{L}_i,s_0,\sigma}(\mathsf{G\,FIX}_i) \tag{30}$$

Let's now consider only those runs from states $s_0 \in L_0$ that do *not* satisfy $\mathsf{G\,FIX}_i$ (the rest satisfy φ already inside the fixed part of $\mathcal{L}_i$ by (30)). From (29) we obtain

$$\mathbb{P}_{\mathcal{L}_i,s_0,\sigma}((\mathsf{FG\,ALPHA}_{\leq i} \vee \mathsf{F}\varphi_{i+1}) \wedge \neg\mathsf{G\,FIX}_i)$$
$$= \mathbb{P}_{\mathcal{L}_i,s_0,\sigma}(\neg\mathsf{G\,FIX}_i) \tag{31}$$

Using Lemma 10, we show the following claim.

▷ **Claim 39.** For every $\varepsilon_1, \varepsilon_2 > 0$, there must exist a threshold l_{i+1} and a finite set

$$L'_{i+1} \subseteq Safe_{\mathcal{L}_i,(\mathsf{G\,ALPHA}_{\leq i} \vee \varphi_{i+1})}(1 - \varepsilon_1)$$

such that, following σ from any state $s_0 \in L_0$, the chance of satisfying $\neg\mathsf{G\,FIX}_i$ and within at most l_{i+1} steps reaching a state s in L'_{i+1} is at least $\mathbb{P}_{\mathcal{L}_i,s_0,\sigma}(\neg\mathsf{G\,FIX}_i)(1 - \varepsilon_2)$.

$$\mathbb{P}_{\mathcal{L}_i,s_0,\sigma}\left(\neg\mathsf{G\,FIX}_i \wedge \mathsf{F}^{\leq l_{i+1}} L'_{i+1}\right) \geq \mathbb{P}_{\mathcal{L}_i,s_0,\sigma}\left(\neg\mathsf{G\,FIX}_i\right)(1 - \varepsilon_2) \tag{32}$$

Proof. For those $s_0 \in L_0$ where $\mathbb{P}_{\mathcal{L}_i,s_0,\sigma}(\neg\mathsf{G\,FIX}_i) = 0$ the claim holds trivially.

We now consider the remaining cases of those states $s_0 \in L_0$ where $\mathbb{P}_{\mathcal{L}_i,s_0,\sigma}(\neg\mathsf{G\,FIX}_i) > 0$. Let

$$\delta \stackrel{\text{def}}{=} \varepsilon_2 \cdot \min_{s_0 \in L_0}\left\{\mathbb{P}_{\mathcal{L}_i,s_0,\sigma}(\neg\mathsf{G\,FIX}_i) > 0\right\} \tag{33}$$

where $\delta > 0$ since L_0 is finite. Let $\mathcal{E} \stackrel{\text{def}}{=} \mathsf{G\,ALPHA}_{\leq i} \vee \varphi_{i+1}$. By (29) we have for every $s_0 \in L_0$

$$\mathbb{P}_{\mathcal{L}_i,s_0,\sigma}(\mathsf{F}\mathcal{E}) = 1$$

We now consider the finitely many Markov chains $\mathcal{C}_{s_0}$ induced by playing σ in $\mathcal{L}_i$ from the finitely many initial states $s_0 \in L_0$. Thus we obtain for every $s_0 \in L_0$

$$\mathbb{P}_{\mathcal{C}_{s_0}}(\mathsf{F}\mathcal{E}) = 1 \tag{34}$$

Since $\mathcal{E}$ is suffix-closed, we can apply Lemma 10 to each Markov chain $\mathcal{C}_{s_0}$. Thus there exist thresholds l^{s_0} and finite sets

$$L^{s_0} \subseteq Safe_{\mathcal{C}_{s_0},(\mathcal{E})}(1 - \varepsilon_1)$$

such that

$$\mathbb{P}_{\mathcal{C}_{s_0}}\left(\mathsf{F}^{\leq l^{s_0}} L^{s_0}\right) \geq \mathbb{P}_{\mathcal{C}_{s_0}}\left(\mathsf{F}\mathcal{E}\right) - \delta = 1 - \delta \tag{35}$$

where the last equality is due to (34). Let now $L'_{i+1} \stackrel{\text{def}}{=} \bigcup_{s_0 \in L_0} L^{s_0}$ (which is finite, since it is a finite union of finite sets) and $l_{i+1} \stackrel{\text{def}}{=} \max_{s_0 \in L_0} l^{s_0}$ (which is finite as the maximum of a finite set of numbers).

For every $s_0 \in L_0$ we have

$$Safe_{\mathcal{C}_{s_0},(\mathcal{E})}(1 - \varepsilon_1) \subseteq Safe_{\mathcal{L}_i,(\mathsf{G}\,\mathsf{ALPHA}_{\leq i} \vee \varphi_{i+1})}(1 - \varepsilon_1)$$

since the required value for $\mathcal{E} = \mathsf{G}\,\mathsf{ALPHA}_{\leq i} \vee \varphi_{i+1}$ is witnessed by the strategy σ, and thus

$$L'_{i+1} \subseteq Safe_{\mathcal{L}_i,(\mathsf{G}\,\mathsf{ALPHA}_{\leq i} \vee \varphi_{i+1})}(1 - \varepsilon_1)$$

as required. From (35) we obtain that for all $s_0 \in L_0$

$$\mathbb{P}_{\mathcal{L}_i,s_0,\sigma}\left(\mathsf{F}^{\leq l_{i+1}} L'_{i+1}\right) \geq 1 - \delta \tag{36}$$

Now we are ready to show (32). We have for all $s_0 \in L_0$

$$\begin{aligned}
\mathbb{P}_{\mathcal{L}_i,s_0,\sigma}&\left(\neg\mathsf{G}\,\mathsf{FIX}_i \wedge \mathsf{F}^{\leq l_{i+1}} L'_{i+1}\right) \\
&= \mathbb{P}_{\mathcal{L}_i,s_0,\sigma}\left(\mathsf{F}^{\leq l_{i+1}} L'_{i+1}\right) - \mathbb{P}_{\mathcal{L}_i,s_0,\sigma}\left(\mathsf{G}\,\mathsf{FIX}_i \wedge \mathsf{F}^{\leq l_{i+1}} L'_{i+1}\right) &&\text{law of total prob.} \\
&\geq \mathbb{P}_{\mathcal{L}_i,s_0,\sigma}\left(\mathsf{F}^{\leq l_{i+1}} L'_{i+1}\right) - \mathbb{P}_{\mathcal{L}_i,s_0,\sigma}\left(\mathsf{G}\,\mathsf{FIX}_i\right) \\
&\geq 1 - \delta - (1 - \mathbb{P}_{\mathcal{L}_i,s_0,\sigma}\left(\neg\mathsf{G}\,\mathsf{FIX}_i\right)) &&\text{by (36)} \\
&\geq 1 - \varepsilon_2 \cdot \mathbb{P}_{\mathcal{L}_i,s_0,\sigma}\left(\neg\mathsf{G}\,\mathsf{FIX}_i\right) - (1 - \mathbb{P}_{\mathcal{L}_i,s_0,\sigma}\left(\neg\mathsf{G}\,\mathsf{FIX}_i\right)) &&\text{by (33)} \\
&= \mathbb{P}_{\mathcal{L}_i,s_0,\sigma}\left(\neg\mathsf{G}\,\mathsf{FIX}_i\right)(1 - \varepsilon_2)
\end{aligned}$$

◀

Notice that $L_{i+1} \cap \mathsf{FIX}_i = \emptyset$, because every state in L_{i+1} must have a value $\geq 1 - \varepsilon_1$ for φ_{i+1}. (In the special case of $i = 0$ we have $\mathsf{FIX}_0 = \emptyset$ and $\varphi_1 = \varphi$ and thus $l_1 = 0$ and $L_1 = L'_1 = L_0$.) Also recall that

$$\begin{aligned}
L'_{i+1} &\subseteq Safe_{\mathcal{L}_i,(\mathsf{G}\,\mathsf{ALPHA}_{\leq i} \vee \varphi_{i+1})}(1 - \varepsilon_1) \\
&\subseteq \mathsf{ALPHA}_{\leq i} \cup Safe_{\mathcal{L}_i,\varphi_{i+1}}(1 - \varepsilon_1).
\end{aligned}$$

We define L_{i+1} as $L_{i+1} \stackrel{\text{def}}{=} L'_{i+1} \setminus \mathsf{ALPHA}_{\leq i}$.

Since $\varphi_{i+1} \stackrel{\text{def}}{=} \varphi \wedge \mathsf{G}(L \setminus Cl(\mathsf{FIX}_i))$ and φ is a parity objective, we can, by Lemma 9 and Remark 11, pick an MD strategy τ_{i+1} that is ε_0-optimal for φ_{i+1} from all states in L_{i+1}.

Based on this strategy τ_{i+1} and parameters $\alpha > \beta > \gamma > 0$, we define $\mathsf{ALPHA}_{i+1} \subseteq \mathsf{BETA}_{i+1} \subseteq \mathsf{GAMMA}_{i+1} \subseteq L$ to be the sets of states from which τ_{i+1} attains at least values α, β and γ, for φ_{i+1}, respectively. E.g.,

$$\mathsf{BETA}_{i+1} \stackrel{\text{def}}{=} Safe_{\mathcal{L}_i,\tau_{i+1},\varphi_{i+1}}(\beta)$$

In particular, this definition satisfies our invariant (21), i.e., $\mathsf{GAMMA}_{i+1} \cap Cl(\mathsf{FIX}_i) = \emptyset$, because a high attainment γ for $\varphi_{i+1} = \varphi \wedge \mathsf{G}(L \setminus Cl(\mathsf{FIX}_i))$ requires that $Cl(\mathsf{FIX}_i)$ is not visited.

W.l.o.g., by choosing $\varepsilon_1, \varepsilon_0$ sufficiently small, we can assume that $\alpha < (1 - \varepsilon_1 - \varepsilon_0)$, and therefore that $L_{i+1} \subseteq \mathsf{ALPHA}_{i+1} \subseteq \mathsf{BETA}_{i+1}$ (we only need $\subseteq \mathsf{BETA}_{i+1}$).

Let $\mathcal{L}'_i \stackrel{\text{def}}{=} \mathcal{L}_i[\tau_{i+1}, \mathsf{BETA}_{i+1}]$. Note that in $\mathcal{L}'_i$ the strategy σ might not be able to reach L_{i+1} with the same probability as in $\mathcal{L}_i$, because the choices in BETA_{i+1} are now fixed. However, a similar strategy σ' can reach BETA_{i+1} in $\mathcal{L}'_i$ with at least the probability by which σ reaches L_{i+1} in $\mathcal{L}_i$. We now define a new resetting strategy σ' in $\mathcal{L}'_i$. It behaves like the previous strategy σ until (and if) it reaches $Cl(\mathsf{BETA}_{i+1})$. Without restriction we can assume that it reaches even BETA_{i+1} in this case (similar to the argument for σ above). Then it plays like τ_{i+1} while in GAMMA_{i+1}. This is possible, since $\mathsf{GAMMA}_{i+1} \cap Cl(\mathsf{FIX}_i) = \emptyset$ by our invariant (21). If and when it exits GAMMA_{i+1} at some state s then it resets to some almost surely winning strategy $\sigma(s)$ for φ in $\mathcal{L}$ until it reaches BETA_{i+1} (or another previously fixed part) again, etc.

From Claim 39 (Equation (32)) and the fact that σ' behaves like σ until it reaches BETA_{i+1} we obtain that

$$
\begin{aligned}
&\mathbb{P}_{\mathcal{L}'_i, s_0, \sigma'} \left(\neg\mathsf{G}\,\mathsf{FIX}_i \wedge \mathsf{F}^{\leq l_{i+1}} \left(\mathsf{ALPHA}_{\leq i} \vee \mathsf{BETA}_{i+1}\right)\right) \\
&\geq \mathbb{P}_{\mathcal{L}_i, s_0, \sigma} \left(\neg\mathsf{G}\,\mathsf{FIX}_i\right) \left(1 - \varepsilon_2\right) \\
&= \mathbb{P}_{\mathcal{L}'_i, s_0, \sigma'} \left(\neg\mathsf{G}\,\mathsf{FIX}_i\right) \left(1 - \varepsilon_2\right),
\end{aligned} \tag{37}
$$

where the last equality holds because $\mathcal{L}_i$ and $\mathcal{L}'_i$ (resp. σ and σ') coincide inside FIX_i. Analogously to Claim 38, from any state in BETA_{i+1}, the chance of staying in the set GAMMA_{i+1} can be lower-bounded.

$$
\forall s \in \mathsf{BETA}_{i+1}\; \mathbb{P}_{\mathcal{L}'_i, s, \sigma'}(\mathsf{G}\,\mathsf{GAMMA}_{i+1}) \geq \frac{\beta - \gamma}{1 - \gamma} > 0 \tag{38}
$$

$\mathsf{GAMMA}_{i+1} \cap Cl(\mathsf{FIX}_i) = \emptyset$ by (21) and σ' continues to play τ_{i+1} in GAMMA_{i+1}. Since $\mathsf{G}\,\mathsf{GAMMA}_{i+1} \subseteq \mathsf{FG}\,\mathsf{GAMMA}_{i+1}$, we can apply Levy's zero-one law (Corollary 30) to (38) and obtain

$$
\forall s \in \mathsf{BETA}_{i+1}\; \mathbb{P}_{\mathcal{L}'_i, s, \sigma'}(\mathsf{FG}\,\mathsf{ALPHA}_{i+1}) \geq \frac{\beta - \gamma}{1 - \gamma} > 0. \tag{39}
$$

By combining (37) with (39), we get

$$
\begin{aligned}
&\mathbb{P}_{\mathcal{L}'_i, s_0, \sigma'} \left(\neg\mathsf{G}\,\mathsf{FIX}_i \wedge \mathsf{F}^{\leq l_{i+1}} \left(\mathsf{ALPHA}_{\leq i} \vee \mathsf{FG}\,\mathsf{ALPHA}_{i+1}\right)\right) \\
&\geq \left(\mathbb{P}_{\mathcal{L}'_i, s_0, \sigma'} \left(\neg\mathsf{G}\,\mathsf{FIX}_i\right) \left(1 - \varepsilon_2\right)\right) \frac{\beta - \gamma}{1 - \gamma}
\end{aligned}
$$

By continuity of measures (recall that $\mathsf{F}X = \bigcup_{k \in \mathbb{N}} \mathsf{F}^k X$), for every $\varepsilon_3 > 0$ there must exist a threshold $k_{i+1} \geq l_{i+1}$ of steps such that, for all $s_0 \in L_0$,

$$
\begin{aligned}
&\mathbb{P}_{\mathcal{L}'_i, s_0, \sigma'} \left(\neg\mathsf{G}\,\mathsf{FIX}_i \wedge \mathsf{F}^{\leq k_{i+1}}\mathsf{ALPHA}_{\leq i+1}\right) \\
&\geq \left(\mathbb{P}_{\mathcal{L}'_i, s_0, \sigma'} \left(\neg\mathsf{G}\,\mathsf{FIX}_i\right) \left(1 - \varepsilon_2\right)\right) \frac{\beta - \gamma}{1 - \gamma}(1 - \varepsilon_3).
\end{aligned} \tag{40}
$$

(Since L_0 is finite, we can have the same multiplicative error $(1 - \varepsilon_3)$ for all $s_0 \in L_0$.) (In the special case of $i = 0$, we have $k_1 = 0$, since $L_0 = L_1 \subseteq \mathsf{ALPHA}_1$.) Once inside $\mathsf{ALPHA}_{\leq i+1}$, there is a bounded chance $\geq \frac{\alpha - \beta}{1 - \beta}$ of staying inside $\mathsf{BETA}_{\leq i+1}$ forever, by Claim 38. Thus from (40) we get

$$
\begin{aligned}
&\mathbb{P}_{\mathcal{L}'_i, s_0, \sigma'}(\neg\mathsf{G}\,\mathsf{FIX}_i \wedge \mathsf{F}^{\leq k_{i+1}}\,\mathsf{G}\,\mathsf{BETA}_{\leq i+1}) \\
&\geq \left(\left(\mathbb{P}_{\mathcal{L}'_i, s_0, \sigma'} \left(\neg\mathsf{G}\,\mathsf{FIX}_i\right) \left(1 - \varepsilon_2\right)\right) \frac{\beta - \gamma}{1 - \gamma}(1 - \varepsilon_3)\right) \frac{\alpha - \beta}{1 - \beta}
\end{aligned} \tag{41}
$$

Consider the finite k_{i+1}-bubble $H_{i+1} \stackrel{\text{def}}{=} \mathsf{bubble}(\mathcal{L}, L_0, k_{i+1})$ around L_0. Remember that in finite MDPs, there are uniformly optimal MD strategies for reachability objectives [16]. Consequently, since H_{i+1} is finite, there exists an MD strategy ρ_{i+1} in $\mathcal{L}'_i$ that is optimal from H_{i+1} for the objective of reaching $\mathsf{ALPHA}_{\leq i+1}$ (from L_0) *inside* H_{i+1} without leaving H_{i+1}. We fix ρ_{i+1} inside H_{i+1}, and obtain our new MDP

$$\mathcal{L}_{i+1} \stackrel{\text{def}}{=} \mathcal{L}'_i[\rho_{i+1}, L_0, k_{i+1}]$$

See Figure 4 for an illustration after round $i = 3$. We define $\mathsf{FIX}_{i+1} \stackrel{\text{def}}{=} \mathsf{BETA}_{\leq i+1} \cup H_{i+1}$ as the region where the strategy is already fixed in $\mathcal{L}_{i+1}$. We now define an almost surely winning resetting strategy σ'' in $\mathcal{L}_{i+1}$, analogously as σ previously in $\mathcal{L}_i$. Similarly as in Equation (30) for $\mathcal{L}_i$, we can derive the corresponding property for $\mathcal{L}_{i+1}$.

$$\mathbb{P}_{\mathcal{L}_{i+1},s_0,\sigma''}(\varphi \wedge \mathsf{G}\,\mathsf{FIX}_{i+1}) = \mathbb{P}_{\mathcal{L}_{i+1},s_0,\sigma''}(\mathsf{G}\,\mathsf{FIX}_{i+1}) \tag{42}$$

$\mathcal{L}_{i+1}$, $\mathcal{L}'_i$ and $\mathcal{L}_i$ (resp. the strategies σ'', σ' and σ) coincide inside FIX_i. Thus by (30) we have

$$
\begin{aligned}
\mathbb{P}_{\mathcal{L}_{i+1},s_0,\sigma''}(\mathsf{G}\,\mathsf{FIX}_i) = \quad & \mathbb{P}_{\mathcal{L}'_i,s_0,\sigma'}(\mathsf{G}\,\mathsf{FIX}_i) \\
= \quad & \mathbb{P}_{\mathcal{L}_i,s_0,\sigma}(\mathsf{G}\,\mathsf{FIX}_i) \\
= \quad & p_i(s_0)
\end{aligned} \tag{43}
$$

By the optimality of the reachability strategy ρ_{i+1} that is fixed in H_{i+1} and $\mathsf{FIX}_{i+1} = \mathsf{BETA}_{\leq i+1} \cup H_{i+1}$, we obtain from this and Equation (41) that

$$
\begin{aligned}
&\mathbb{P}_{\mathcal{L}_{i+1},s_0,\sigma''}(\neg\mathsf{G}\,\mathsf{FIX}_i \wedge \mathsf{G}\,\mathsf{FIX}_{i+1}) \\
&\geq \left((1-p_i(s_0))(1-\varepsilon_2)\frac{\beta-\gamma}{1-\gamma}(1-\varepsilon_3)\right)\frac{\alpha-\beta}{1-\beta}
\end{aligned} \tag{44}
$$

The crucial question is how much σ'' attains for φ in the fixed part alone, i.e., how large is $\mathbb{P}_{\mathcal{L}_{i+1},s_0,\sigma''}(\varphi \wedge \mathsf{G}\,\mathsf{FIX}_{i+1}) = p_{i+1}(s_0)$? For all $s_0 \in L_0$ we have

$$
\begin{aligned}
&p_{i+1}(s_0) \\
&= \mathbb{P}_{\mathcal{L}_{i+1},s_0,\sigma''}(\mathsf{G}\,\mathsf{FIX}_{i+1}) \\
&= \mathbb{P}_{\mathcal{L}_{i+1},s_0,\sigma''}(\mathsf{G}\,\mathsf{FIX}_i) + \mathbb{P}_{\mathcal{L}_{i+1},s_0,\sigma''}(\neg\mathsf{G}\,\mathsf{FIX}_i \wedge \mathsf{G}\,\mathsf{FIX}_{i+1}) \\
&\geq p_i(s_0) + (1-p_i(s_0))\left((1-\varepsilon_2)\frac{\beta-\gamma}{1-\gamma}(1-\varepsilon_3)\frac{\alpha-\beta}{1-\beta}\right),
\end{aligned}
$$

where the first equality is due to (42) and the last inequation is due to Equations (43) and (44).

We can suitably choose the parameters $\alpha, \beta, \gamma, \varepsilon_2, \varepsilon_3$ such that $\left((1-\varepsilon_2)\frac{\beta-\gamma}{1-\gamma}(1-\varepsilon_3)\frac{\alpha-\beta}{1-\beta}\right)$ is arbitrarily close to 1, and thus in particular $\geq 1/2$, and obtain that $p_{i+1}(s_0) \geq p_i(s_0) + (1-p_i(s_0))/2$. Since $p_0(s_0) = 0$, we get $1 - p_i(s_0) \leq 2^{-i}$ and thus $\lim_{i\to\infty} p_i(s_0) = 1$, as required.

Finally, let $\hat{\sigma}$ be the MD strategy in $\mathcal{L}$ that plays from L_0 as prescribed by all the fixings in $\bigcup_i \mathsf{FIX}_i$ in the systems $\mathcal{L}_i$. Then, for all $s_0 \in S_0$ and every $i \in \mathbb{N}$, it holds that

$$\mathbb{P}_{\mathcal{L},s_0,\hat{\sigma}}(\varphi) \geq \mathbb{P}_{\mathcal{L}_i,s_0,\hat{\sigma}}(\varphi \wedge \mathsf{G}\,\mathsf{FIX}_i) = p_i(s_0) \geq 1 - 2^{-i}$$

Since this holds for every $i \in \mathbb{N}$ we get that $\mathbb{P}_{\mathcal{L},s_0,\hat{\sigma}}(\varphi) = 1$, i.e., the MD strategy $\hat{\sigma}$ wins φ almost surely from every $s_0 \in L_0$. ◀

F   Optimal Strategies for $\{0, 1, 2\}$-Parity

▶ **Theorem 22.** *Let $\mathcal{M} = (S, S_\square, S_\bigcirc, \longrightarrow, P)$ be an MDP, φ a $\{0, 1, 2\}$-Parity objective and $\mathcal{M}_* = (S_*, S_{*\square}, S_{*\bigcirc}, \longrightarrow_*, P_*)$ its conditioned version wrt. φ. Assume that in $\mathcal{M}_*$ for every safety objective (given by some target $T \subseteq S_*$) and $\varepsilon > 0$ there exists a uniformly ε-optimal MD strategy. Let S_{opt} be the subset of states that have an optimal strategy for φ in $\mathcal{M}$.*

Then there exists an MD strategy in $\mathcal{M}$ that is optimal for φ from every state in S_{opt}.

In the rest of this section we prove Theorem 22. It generalizes [13, Theorem 16], which considers only finitely-branching MDPs and uses the fact that for every safety objective, an MD strategy exists that is uniformly *optimal*. This is not generally true for infinitely-branching acyclic MDPs [13]. To prove Theorem 22, we adjust the construction so that it only requires uniformly ε-*optimal* MD strategies for safety objectives (in the conditioned MDP $\mathcal{M}_*$).

▶ **Theorem 40** (from Theorem B in [16]). *For every MDP $\mathcal{M}$ there exist uniform ε-optimal MD-strategies for reachability objectives.*

The following simple lemma provides a scheme for proving almost-sure properties.

▶ **Lemma 41** (Lem. 18 in [13]). *Let $\mathbb{P}$ be a probability measure over the sample space Ω. Let $(\mathcal{R}_i)_{i \in I}$ be a countable partition of Ω in measurable events. Let $\mathcal{E} \subseteq \Omega$ be a measurable event. Suppose $\mathbb{P}(\mathcal{R}_i \cap \mathcal{E}) = \mathbb{P}(\mathcal{R}_i)$ holds for all $i \in I$. Then $\mathbb{P}(\mathcal{E}) = 1$.*

We need a few lemmas about safety objectives first. Recall the definition of safe sets (Equation (1)).

▶ **Lemma 42.** *Let $\mathcal{M} = (S, S_\square, S_\bigcirc, \longrightarrow, P)$ be an MDP, $T \subseteq S$, σ a strategy from state $s \in S$ and $\tau < 1$. It holds that $\mathbb{P}_{\mathcal{M},s,\sigma}(\mathsf{FG}\neg Safe_{\mathcal{M},\mathsf{Safety}(T)}(\tau) \wedge \mathsf{FG}(S \setminus T)) = 0$.*

Proof. For any $n \in \mathbb{N}$ define $Z_n \stackrel{\text{def}}{=} (S \setminus T)^n$. That is, $Z_n S^\omega$ is the event that the first n visited states are outside T. For every state $s \notin Safe_{\mathcal{M},\mathsf{Safety}(T)}(\tau)$ and every strategy σ from s we have that $\lim_{n \to \infty} \mathbb{P}_{\mathcal{M},s,\sigma}(Z_n S^\omega) < \tau < (1 + \tau)/2$ by Equation (1). Let $n(s) \in \mathbb{N}$ be the smallest number such that $\mathbb{P}_{\mathcal{M},s,\sigma}(Z_{n(s)} S^\omega) \leq (1 + \tau)/2$. Let $L \subseteq S^*$ be the set of finite sequences $s_0 s_1 \cdots s_{n-1}$ such that $s_0 \notin Safe_{\mathcal{M},\mathsf{Safety}(T)}(\tau)$ and $n = n(s_0)$ and $\forall i < n. \, s_i \in (S \setminus T) \setminus Safe_{\mathcal{M},\mathsf{Safety}(T)}(\tau)$.

We show for all $s \in S \setminus Safe_{\mathcal{M},\mathsf{Safety}(T)}(\tau)$ and all $k \in \mathbb{N}$ that $\mathbb{P}_{\mathcal{M},s,\sigma}(L^k S^\omega) \leq \left(\frac{1+\tau}{2}\right)^k$. We proceed by induction on k. The case $k = 0$ is trivial. For the induction step let $k \geq 0$.

$$
\begin{aligned}
\mathbb{P}_{\mathcal{M},s,\sigma}(L^{k+1} S^\omega) &\leq \mathbb{P}_{\mathcal{M},s,\sigma}(Z_{n(s)} L^k S^\omega) \\
&\leq \mathbb{P}_{\mathcal{M},s,\sigma}(Z_{n(s)} S^\omega) \cdot \sup_{s' \in S \setminus Safe_{\mathcal{M},\mathsf{Safety}(T)}(\tau)} \mathbb{P}_{\mathcal{M},s',\sigma}(L^k S^\omega) \\
&\leq \mathbb{P}_{\mathcal{M},s,\sigma}(Z_{n(s)} S^\omega) \cdot \left(\frac{1+\tau}{2}\right)^k \leq \left(\frac{1+\tau}{2}\right)^{k+1}
\end{aligned}
$$

where the first inequality uses that $L \cap \{s\} S^* \subseteq Z_{n(s)}$, the third uses the induction hypothesis, and the last the definition of $n(s)$. This completes the induction proof.

Write $\varphi \stackrel{\text{def}}{=} \mathsf{G}\neg Safe_{\mathcal{M},\mathtt{Safety}(T)}(\tau) \wedge \mathsf{G}\,(S \setminus T)$. For all $s \in S$,

$$
\begin{aligned}
\mathbb{P}_{\mathcal{M},s,\sigma}(\varphi) &= \mathbb{P}_{\mathcal{M},s,\sigma}(L^\omega) & &\text{because } \llbracket\varphi\rrbracket = L^\omega \\
&= \lim_{k\to\infty} \mathbb{P}_{\mathcal{M},s,\sigma}(L^k S^\omega) & &\text{by continuity of measures} \\
&\leq \lim_{k\to\infty}\left(\frac{1+\tau}{2}\right)^k & &\text{as shown above} \\
&= 0 & &\text{because } \tau < 1
\end{aligned}
$$

It follows that $\mathbb{P}_{\mathcal{M},s,\sigma}(\mathsf{X}^j\varphi) = 0$, for all $s \in S$ and all $j \in \mathbb{N}$ and therefore that

$$
\begin{aligned}
&\mathbb{P}_{\mathcal{M},s,\sigma}(\mathsf{FG}\,Safe_{\mathcal{M},\mathtt{Safety}(T)}(\tau) \wedge \mathsf{FG}\,(S \setminus T)) \\
&= \mathbb{P}_{\mathcal{M},s,\sigma}(\mathsf{F}\varphi) \\
&= \mathbb{P}_{\mathcal{M},s,\sigma}\Big(\bigcup_{j\in\mathbb{N}}\llbracket\mathsf{X}^j\varphi\rrbracket^s\Big) \\
&\leq \sum_{j\in\mathbb{N}}\mathbb{P}_{\mathcal{M},s,\sigma}(\mathsf{X}^j\varphi) \;=\; 0 && \blacktriangleleft
\end{aligned}
$$

Now we show that if an MDP admits uniformly ε-optimal strategies for all safety objectives, then optimal strategies for $\{0,1,2\}$-$\mathtt{Parity}$ (where they exist) can be chosen MD.

▶ **Lemma 43.** *Let $\mathcal{M} = (S, S_\square, S_\bigcirc, \longrightarrow, P)$ be an MDP such that for every safety objective (given by some target set $T \subseteq S$) and $\varepsilon > 0$ there exists a uniformly ε-optimal MD strategy. Let $s_0 \in S$, $Col : S \to \{0,1,2\}$, $\varphi = \mathtt{Parity}(Col)$, and σ a strategy with $\mathbb{P}_{\mathcal{M},s_0,\sigma}(\varphi) = 1$. Then there is an MD-strategy σ' with $\mathbb{P}_{\mathcal{M},s_0,\sigma'}(\varphi) = 1$.*

Proof. To achieve an almost-sure winning objective, the player must forever remain in states from which the objective can be achieved almost surely. So we can assume without loss of generality that all states are almost-sure winning, i.e., for all $s \in S$ we have $\mathbb{P}_{\mathcal{M},s,\sigma}(\varphi) = 1$ for some strategy σ. We will define an MD-strategy σ' with $\mathbb{P}_{\mathcal{M},s,\sigma'}(\varphi) = 1$ for all $s \in S$.

Recall that $[S]^{Col\neq 0} \subseteq S$ denotes the subset of states of color 1 or 2. Let $T \stackrel{\text{def}}{=} [S]^{Col\neq 0}$ and let σ_ε be a uniformly ε-optimal MD strategy for $\mathtt{Safety}(T)$, whose existence is guaranteed by our assumption on $\mathcal{M}$. The precise $\varepsilon > 0$ is immaterial, we only need that $\varepsilon < \frac{1}{3}$. The MD-strategy σ' will be based on special subsets (Equation (1)):

$$
Safe_{\mathcal{M}}(\tau) \stackrel{\text{def}}{=} Safe_{\mathcal{M},\sigma_\varepsilon,\mathtt{Safety}(T)}(\tau) \subseteq S. \tag{45}
$$

We first define the MD-strategy σ' partially for the states in $Safe_{\mathcal{M}}(\frac{1}{3})$ and then extend the definition of σ' to all states. For the states in $Safe_{\mathcal{M}}(\frac{1}{3})$ define $\sigma' \stackrel{\text{def}}{=} \sigma_\varepsilon$ (which is MD). Let $\mathcal{M}'$ be the MDP obtained from $\mathcal{M}$ by restricting the transition relation as prescribed by the partial MD-strategy σ' in $Safe_{\mathcal{M}}(\frac{1}{3})$ (elsewhere the choices remain free). We define $Safe_{\mathcal{M}'}(\tau)$ for $\mathcal{M}'$ as in Equation (45) for $\mathcal{M}$. Thus, for any $\tau \in [0,1]$, we have $Safe_{\mathcal{M}}(\tau) = Safe_{\mathcal{M}'}(\tau)$. Indeed, since $\mathcal{M}'$ restricts the options of the player, we have $Safe_{\mathcal{M}}(\tau) \supseteq Safe_{\mathcal{M}'}(\tau)$. Conversely, let $s \in Safe_{\mathcal{M}}(\tau)$. The strategy σ_ε attains $\mathbb{P}_{\mathcal{M},s,\sigma_\varepsilon}(\mathsf{G}[S]^{Col=0}) \geq \tau$. Since σ_ε can be applied in $\mathcal{M}'$, and results in the same Markov chain as applying it in $\mathcal{M}$, we conclude $s \in Safe_{\mathcal{M}'}(\tau)$. This justifies to write $Safe(\tau)$ for $Safe_{\mathcal{M}}(\tau) = Safe_{\mathcal{M}'}(\tau)$ in the remainder of the proof.

Next we show that, also in $\mathcal{M}'$, for all states $s \in S$ there exists a strategy σ_1 with $\mathbb{P}_{\mathcal{M}',s,\sigma_1}(\varphi) = 1$. This strategy σ_1 is defined as follows. First play according to an almost-surely winning strategy σ from the statement of the theorem. If and when the play visits $Safe(\frac{1}{3})$, switch to the MD-strategy σ_ε. If and when the play then visits $[S]^{Col\neq 0}$, switch back to an almost-surely winning strategy σ from the statement of the theorem, and so forth.

We show that σ_1 attains $\mathbb{P}_{\mathcal{M}',s,\sigma_1}(\varphi) = 1$. To this end we will use Lemma 41. We partition the runs of sS^ω into three events $\mathcal{R}_0, \mathcal{R}_1, \mathcal{R}_2$ as follows:

- $\mathcal{R}_0$ contains the runs where σ_1 switches between σ_ε and σ infinitely often.
- $\mathcal{R}_1$ contains the runs where σ_1 eventually only plays according to σ_ε.
- $\mathcal{R}_2$ contains the runs where σ_1 eventually only plays according to σ.

Each time σ_1 switches to σ_ε, there is, by definition of $Safe_\mathcal{M}(\frac{1}{3})$, a probability of at least $\frac{1}{3}$ of never visiting a color-$\{1,2\}$ state again and thus of never again switching to σ. It follows that $\mathbb{P}_{\mathcal{M}',s,\sigma_1}(\mathcal{R}_0) = 0$. By the definition of the switching behavior of σ_1, we have $\mathcal{R}_1 \subseteq [\![\mathsf{FG}[S]^{Col=0}]\!] \subseteq [\![\varphi]\!]$, and hence $\mathbb{P}_{\mathcal{M}',s,\sigma_1}(\mathcal{R}_1 \cap [\![\varphi]\!]) = \mathbb{P}_{\mathcal{M}',s,\sigma_1}(\mathcal{R}_1)$. Since $\mathbb{P}_{\mathcal{M},s,\sigma}(\varphi) = 1$ and φ is tail, we have that $\mathbb{P}_{\mathcal{M}',s,\sigma_1}(\mathcal{R}_2 \cap [\![\varphi]\!]) = \mathbb{P}_{\mathcal{M}',s,\sigma_1}(\mathcal{R}_2)$. Using Lemma 41, we obtain $\mathbb{P}_{\mathcal{M}',s,\sigma_1}(\varphi) = 1$.

Next we show that for all $s \in S$ the strategy σ_1 defined above achieves $\mathbb{P}_{\mathcal{M}',s,\sigma_1}(\mathsf{F}\,Safe(\frac{2}{3}) \vee \mathsf{F}[S]^{Col=2}) = 1$. To this end we will use Lemma 41 again. We partition the runs of sS^ω into three events $\mathcal{R}'_1, \mathcal{R}'_2, \mathcal{R}'_0$ as follows:

- $\mathcal{R}'_1 = [\![\mathsf{FG}[S]^{Col=0}]\!]^s$

- $\mathcal{R}'_2 = [\![\mathsf{GF}[S]^{Col=2}]\!]^s$

- $\mathcal{R}'_0 = sS^\omega \setminus [\![\varphi]\!]^s$

We have previously shown that $\mathbb{P}_{\mathcal{M}',s,\sigma_1}(\varphi) = 1$ and hence that $\mathbb{P}_{\mathcal{M}',s,\sigma_1}(\mathcal{R}'_0) = 0$. We now invoke Lemma 42 with $\tau \overset{\text{def}}{=} \frac{2}{3} + \varepsilon < 1$ and obtain that almost all runs in $\mathcal{R}'_1$ satisfy $\mathsf{GF}\,Safe_{\mathcal{M},\mathtt{Safety}(T)}(\tau)$. Since σ_ε is uniformly ε-optimal for $\mathtt{Safety}(T)$ we have $Safe_{\mathcal{M},\mathtt{Safety}(T)}(\tau) \subseteq Safe(\tau - \varepsilon) = Safe(\frac{2}{3})$ and thus almost all runs in $\mathcal{R}'_1$ satisfy $\mathsf{GF}\,Safe(\frac{2}{3})$. Since $[\![\mathsf{GF}\,Safe(\frac{2}{3})]\!] \subseteq [\![\mathsf{F}\,Safe(\frac{2}{3})]\!]$, we observe that

$$\mathbb{P}_{\mathcal{M}',s,\sigma_1}(\mathcal{R}'_1 \cap [\![\mathsf{F}\,Safe(\frac{2}{3}) \vee \mathsf{F}[S]^{Col=2}]\!]) = \mathbb{P}_{\mathcal{M}',s,\sigma_1}(\mathcal{R}'_1).$$

Since $\mathcal{R}'_2 \subseteq [\![\mathsf{F}[S]^{Col=2}]\!]$, we also have that

$$\mathbb{P}_{\mathcal{M}',s,\sigma_1}(\mathcal{R}'_2 \cap [\![\mathsf{F}\,Safe(\frac{2}{3}) \vee \mathsf{F}[S]^{Col=2}]\!]) = \mathbb{P}_{\mathcal{M}',s,\sigma_1}(\mathcal{R}'_2).$$

By Lemma 41 we obtain $\mathbb{P}_{\mathcal{M}',s,\sigma_1}(\mathsf{F}\,Safe(\frac{2}{3}) \vee \mathsf{F}[S]^{Col=2}) = 1$.

Writing $T' \overset{\text{def}}{=} Safe(\frac{2}{3}) \cup [S]^{Col=2}$ we have just shown that for all $s \in S$ there is a strategy σ_1 with $\mathbb{P}_{\mathcal{M}',s,\sigma_1}(\mathsf{F}T') = 1$. Since this holds for all $s \in S$, it follows from Theorem 40 that there is an MD-strategy $\hat{\sigma}$ for $\mathcal{M}'$ with $\mathbb{P}_{\mathcal{M}',s,\hat{\sigma}}(\mathsf{F}T') = 1$ for all $s \in S$. We extend the (so far partially defined) strategy σ' by $\hat{\sigma}$. Thus we obtain a (fully defined) strategy σ' for $\mathcal{M}$ such that for all $s \in S$ we have $\mathbb{P}_{\mathcal{M},s,\sigma'}(\mathsf{F}T') = 1$.

It remains to show that $\mathbb{P}_{\mathcal{M},s,\sigma'}(\varphi) = 1$ holds for all $s \in S$. To this end we will use Lemma 41 again. We partition the runs of sS^ω into two events $\mathcal{R}''_1, \mathcal{R}''_2$:

- $\mathcal{R}''_1 = [\![\mathsf{GF}\,Safe(\frac{2}{3})]\!]^s$, i.e., $\mathcal{R}''_1$ contains the runs that visit $Safe(\frac{2}{3})$ infinitely often.

- $\mathcal{R}''_2 = [\![\mathsf{FG}\neg Safe(\frac{2}{3})]\!]^s$, i.e., $\mathcal{R}''_2$ contains the runs that from some point on never visit $Safe(\frac{2}{3})$.

Recall that σ' plays like σ_ε inside of $Safe(\frac{1}{3})$, that $Safe(\frac{2}{3}) \subseteq Safe(\frac{1}{3})$, and that σ_ε is an MD-strategy. Thus we can invoke Lemma 15 with $\beta_2 \overset{\text{def}}{=} \frac{2}{3}$ and $\beta_1 \overset{\text{def}}{=} \frac{1}{3}$ and conclude that every time a run (according to σ') enters $Safe(\frac{2}{3})$, the probability that the run remains in $Safe(\frac{1}{3})$ forever is at least $\frac{1}{2}$. It follows that almost all runs in $\mathcal{R}''_1$ eventually remain in $Safe(\frac{1}{3})$ forever. That is, $\mathbb{P}_{\mathcal{M},s,\sigma'}(\mathcal{R}''_1 \cap [\![\mathsf{FG}\,Safe(\frac{1}{3})]\!]) = \mathbb{P}_{\mathcal{M},s,\sigma'}(\mathcal{R}''_1)$. Since $Safe(\frac{1}{3}) \subseteq [S]^{Col=0}$, we have $[\![\mathsf{FG}\,Safe(\frac{1}{3})]\!] \subseteq [\![\mathsf{FG}[S]^{Col=0}]\!] \subseteq [\![\varphi]\!]$. Hence also $\mathbb{P}_{\mathcal{M},s,\sigma'}(\mathcal{R}''_1 \cap [\![\varphi]\!]) = \mathbb{P}_{\mathcal{M},s,\sigma'}(\mathcal{R}''_1)$.

We have previously shown that $\mathbb{P}_{\mathcal{M},s,\sigma'}(\mathsf{F}T') = 1$ holds for all $s \in S$. Hence also $\mathbb{P}_{\mathcal{M},s,\sigma'}(\mathsf{GF}T') = 1$ holds for all $s \in S$. In particular, almost all runs in $\mathcal{R}_2''$ satisfy $\mathsf{GF}T'$. By comparing the definitions of $\mathcal{R}_2''$ and T' we see that almost all runs in $\mathcal{R}_2''$ even satisfy $\mathsf{GF}[S]^{Col=2}$. Since $[\![\mathsf{GF}[S]^{Col=2}]\!] \subseteq [\![\varphi]\!]$, we obtain $\mathbb{P}_{\mathcal{M},s,\sigma'}(\mathcal{R}_2'' \cap [\![\varphi]\!]) = \mathbb{P}_{\mathcal{M},s,\sigma'}(\mathcal{R}_2'')$. A final application of Lemma 41 yields $\mathbb{P}_{\mathcal{M},s,\sigma'}(\varphi) = 1$ for all $s \in S$. ◀

We are ready to prove Theorem 22.

Proof of Theorem 22. Let $\mathcal{M} = (S, S_\square, S_\bigcirc, \longrightarrow, P)$ be an MDP and φ a $\{0, 1, 2\}$-`Parity` objective. Since φ is tail, it is possible to define (see Definition 19) the conditioned version $\mathcal{M}_* = (S_*, S_{*\square}, S_{*\bigcirc}, \longrightarrow_*, P_*)$ of $\mathcal{M}$ wrt. φ. Assume that in $\mathcal{M}_*$ for every safety objective (given by some target $T \subseteq S_*$) and $\varepsilon > 0$ there exists a uniformly ε-optimal MD strategy. Let S_{opt} be the subset of states that have an optimal strategy for φ in $\mathcal{M}$.

By Theorem 20.1, all states in $\mathcal{M}_*$ are almost sure winning for φ. By our above condition about safety objectives in $\mathcal{M}_*$, we can apply Lemma 43 to $\mathcal{M}_*$ and obtain that for every state in $\mathcal{M}_*$ there is an MD strategy that is almost surely winning for φ. By Theorem 20.2, there is an MD strategy in $\mathcal{M}$ that is optimal for φ from every state in S_{opt}, as required. ◀

In order to apply Theorem 22 to infinitely-branching acyclic MDPs, we now show that acyclicity guarantees the existence of uniformly ε-optimal MD strategies for safety objectives.

▶ **Lemma 23.** *For every acyclic MDP with a safety objective and every $\varepsilon > 0$ there exists an MD strategy that is uniformly ε-optimal.*

Proof. Let $\mathtt{Safety}(T)$ be the safety objective and shortly write $\mathtt{val}(s_0) = \mathtt{val}_{\mathcal{M},\mathtt{Safety}(T)}(s_0)$ for the value of a state s_0 w.r.t. this objective. Assume w.l.o.g. that the target $T \subseteq S$ is a sink and let $\iota : S \to \mathbb{N}$ be an enumeration of the state space.

Let σ be an MD-strategy that, at any state $s \in S_\square$, picks a successor s' such that

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

We show that $\mathbb{P}_{\mathcal{M},s_0,\sigma}(\mathtt{Safety}(T)) \geq \mathtt{val}(s_0)(1 - \varepsilon)$ holds for every initial state s_0.

Let's write $Post^*(s) \subseteq S$ for the set of states reachable from state $s \in S$ and define $L(s) \overset{\text{def}}{=} \prod_{s' \in Post^*(s)} (1 - \varepsilon 2^{-\iota(s')})$. Further, let $\mathtt{val}_s(n)$ be the random variable denoting the value of the nth state of a random run that starts in s. In particular, $\mathtt{val}_s(0) = \mathtt{val}(s)$. An induction on n using our choice of strategy gives, for every $s_0 \in S$, that

$$\mathbb{E}(\mathtt{val}_{s_0}(n)) \geq \mathtt{val}(s_0)L(s_0). \tag{46}$$

Indeed, this trivially holds for $n = 0$. For the induction step there are two cases.
Case 1: $s_0 \in S_\square$ and $\sigma(s_0) = s$. Then

$$\mathbb{E}(\mathtt{val}_{s_0}(n+1))$$

$$= \mathbb{E}(\mathtt{val}_s(n))$$

$$\geq \mathtt{val}(s)L(s) \qquad\qquad\qquad\qquad \text{ind. hyp.}$$

$$\geq \mathtt{val}(s_0)\left(1 - \varepsilon 2^{-\iota(s_0)}\right)L(s) \qquad\qquad \text{def. of } \sigma$$

$$\geq \mathtt{val}(s_0)L(s_0) \qquad\qquad\qquad\qquad \text{acyclicity; def. of } L(s_0).$$

Case 2: $s_0 \in S_\bigcirc$. Then

$$\mathbb{E}(\mathtt{val}_{s_0}(n+1))$$

$$= \sum_{s_0 \longrightarrow s} P(s_0)(s) \cdot \mathbb{E}(\mathtt{val}_s(n))$$

$$\geq \sum_{s_0 \longrightarrow s} P(s_0)(s) \cdot \mathtt{val}(s)L(s) \qquad \text{by ind. hyp.}$$

$$\geq \sum_{s_0 \longrightarrow s} P(s_0)(s) \cdot \mathtt{val}(s) \left(1 - \varepsilon 2^{-\iota(s_0)}\right) L(s)$$

$$\geq \sum_{s_0 \longrightarrow s} P(s_0)(s) \cdot \mathtt{val}(s)L(s_0) \quad \text{acyclicity; def. of } L(s_0)$$

$$= \mathtt{val}(s_0)L(s_0).$$

Together with the observation that $L(s_0) > (1 - \varepsilon)$ for every s_0, we derive that

$$\liminf_{n \to \infty} \mathbb{E}(\mathtt{val}_{s_0}(n)) \geq \mathtt{val}(s_0)(1 - \varepsilon). \tag{47}$$

To show the claim, fix $s_0 \in S$ and shortly write $\mathbb{P}$ for $\mathbb{P}_{\mathcal{M}, s_0, \sigma}$ here. Let $[\mathsf{X}^n \neg T] : S^\omega \to \{0, 1\}$ be the random variable that indicates that the nth state is not in the target set T. Note that $[\mathsf{X}^n \neg T] \geq \mathtt{val}_{s_0}(n)$ because target states have value 0. We conclude that

$$\mathbb{P}(\mathtt{Safety}(T))$$

$$= \quad \mathbb{P}\left(\bigcap_{i=0}^{\infty} [\![X^i \neg T]\!]\right) \qquad\qquad \text{semantics of } \mathtt{Safety}(T) = \mathsf{G} \neg T$$

$$= \quad \lim_{n \to \infty} \mathbb{P}\left(\bigcap_{i=0}^{n} [\![X^i \neg T]\!]\right) \qquad\qquad \text{cont. of measures from above}$$

$$= \quad \lim_{n \to \infty} \mathbb{P}\left([\![X^i \neg T]\!]\right) \qquad\qquad T \text{ is a sink}$$

$$= \quad \lim_{n \to \infty} \mathbb{E}([\mathsf{X}^n \neg T]) \qquad\qquad \text{definition of } [\mathsf{X}^n \neg T]$$

$$\geq \quad \liminf_{n \to \infty} \mathbb{E}(\mathtt{val}(n)) \qquad\qquad \text{as } [\mathsf{X}^n \neg T] \geq \mathtt{val}_{s_0}(n)$$

$$\geq \quad \mathtt{val}(s_0)(1 - \varepsilon) \qquad\qquad \text{by Equation (47).} \qquad\qquad \blacktriangleleft$$

G ε-Optimal Strategies for $\{0, 1\}$-Parity

▶ **Theorem 25.** *Suppose that $\mathcal{M} = (S, S_\square, S_\bigcirc, \longrightarrow, P)$ is an MDP such that for every safety objective (given by some target $T \subseteq S$) and $\varepsilon > 0$ there exists a uniformly ε-optimal MD strategy.*

Then for every co-Büchi objective (given by some coloring $Col : S \to \{0, 1\}$) and $\varepsilon > 0$ there exists a uniformly ε-optimal MD strategy.

Proof. Let $\mathcal{M} = (S, S_\square, S_\bigcirc, \longrightarrow, P)$ be an MDP such that for all safety objectives, uniformly ε-optimal strategies can be chosen MD. Let $Col : S \to \{0, 1\}$ be a coloring and $\varphi = \mathtt{Parity}(Col)$ be the resulting co-Büchi objective.

We show that there exist uniformly ε-optimal MD-strategies for φ. I.e., for every $\varepsilon > 0$ there is an MD-strategy σ_ε with $\forall_{s_0 \in S} \mathbb{P}_{\mathcal{M}, s_0, \sigma_\varepsilon}(\varphi) \geq \mathtt{val}_{\mathcal{M}}(s_0) - \varepsilon$.

To construct this MD-strategy strategy σ_ε, we first need several auxiliary notions.

Let $\varepsilon_1 > 0$ be a suitably small number (to be determined later) and $\tau_1 \stackrel{\text{def}}{=} 1 - \varepsilon_1 > 0$. Let $\tau_2 \stackrel{\text{def}}{=} 1 - \varepsilon_1/k + \lambda$ for a suitably large $k \geq 1$ (to be determined later) and let $\lambda < \varepsilon_1/k$ (e.g., $\lambda \stackrel{\text{def}}{=} \varepsilon_1/(2k)$). Thus $\tau_2 < 1$.

Let $T \stackrel{\text{def}}{=} Col^{-1}(\{1\})$ be the set of states with color 1 and $\mathtt{Safety}(T)$ the safety objective. We have that $\varphi = \mathsf{FG}(S \setminus T)$.

By our assumption on $\mathcal{M}$, there exists a uniformly λ-optimal MD-strategy $\sigma_{opt\text{-}av}$ for $\mathtt{Safety}(T)$. Let $S' \stackrel{\text{def}}{=} Safe_{\mathcal{M},\sigma_{opt\text{-}av},\mathtt{Safety}(T)}(\tau_1)$ be the set of states where $\sigma_{opt\text{-}av}$ achieves at least value τ_1 for $\mathtt{Safety}(T)$ (refer to Equation (1) for the definition of safe sets). In particular $S' \subseteq S \setminus T$, since $\tau_1 > 0$. From $\mathcal{M}$ we obtain a modified MDP $\mathcal{M}'$ by fixing all player choices from states in S' according to $\sigma_{opt\text{-}av}$.

We show that the value w.r.t. objective φ is only slightly smaller in $\mathcal{M}'$, i.e.,

$$\mathtt{val}_{\mathcal{M}'}(s_0) \geq \mathtt{val}_{\mathcal{M}}(s_0) - \varepsilon_1 \quad \text{for every } s_0 \in S. \tag{48}$$

Let $s_0 \in S$. By definition of the value $\mathtt{val}_{\mathcal{M}}(s_0)$, for every $\delta > 0$ there exists a strategy σ_δ in $\mathcal{M}$ from s_0 s.t. $\mathbb{P}_{\mathcal{M},s_0,\sigma_\delta}(\varphi) \geq \mathtt{val}_{\mathcal{M}}(s_0) - \delta$. We define a strategy σ'_δ in $\mathcal{M}'$ from state s_0 as follows. First play like σ_δ. If and when a state in S' is reached, then henceforth play like $\sigma_{opt\text{-}av}$. This is possible, since no moves from states outside S' have been fixed in $\mathcal{M}'$, and all moves from states inside S' have been fixed according to $\sigma_{opt\text{-}av}$. Then we have:

$$\mathbb{P}_{\mathcal{M}',s_0,\sigma'_\delta}(\varphi)$$

$$= \mathbb{P}_{\mathcal{M},s_0,\sigma_\delta}(\varphi)$$
$$\quad - \mathbb{P}_{\mathcal{M},s_0,\sigma_\delta}(\mathsf{F}S') \cdot \mathbb{P}_{\mathcal{M},s_0,\sigma_\delta}(\varphi \mid \mathsf{F}S')$$
$$\quad + \mathbb{P}_{\mathcal{M},s_0,\sigma_\delta}(\mathsf{F}S') \cdot \mathbb{P}_{\mathcal{M}',s_0,\sigma'_\delta}(\varphi \mid \mathsf{F}S')$$

$$\geq \mathbb{P}_{\mathcal{M},s_0,\sigma_\delta}(\varphi)$$
$$\quad - \mathbb{P}_{\mathcal{M},s_0,\sigma_\delta}(\mathsf{F}S') \cdot \mathbb{P}_{\mathcal{M},s_0,\sigma_\delta}(\varphi \mid \mathsf{F}S')$$
$$\quad + \mathbb{P}_{\mathcal{M},s_0,\sigma_\delta}(\mathsf{F}S') \cdot \tau_1$$

$$\geq \mathtt{val}_{\mathcal{M}}(s_0) - \delta - \mathbb{P}_{\mathcal{M},s_0,\sigma_\delta}(\mathsf{F}S')(1 - \tau_1)$$

$$\geq \mathtt{val}_{\mathcal{M}}(s_0) - \delta - \varepsilon_1$$

Since this holds for every $\delta > 0$ we have $\mathtt{val}_{\mathcal{M}'}(s_0) \geq \mathtt{val}_{\mathcal{M}}(s_0) - \varepsilon_1$, thus (48).

Let $S'' \stackrel{\text{def}}{=} Safe_{\mathcal{M},\sigma_{opt\text{-}av},\mathtt{Safety}(T)}(\tau_2 - \lambda)$ as by definition of safe sets in Equation (1). In particular, $S'' = Safe_{\mathcal{M}',\sigma_{opt\text{-}av},\mathtt{Safety}(T)}(\tau_2 - \lambda)$, since $\sigma_{opt\text{-}av}$ is a uniform MD-strategy that has been fixed on the subset S' in the step from $\mathcal{M}$ to $\mathcal{M}'$.

Let $s_0 \in S$ be an arbitrary state. By definition of the value w.r.t. φ of s_0, for every $\varepsilon' > 0$ there exists a strategy $\sigma_{\varepsilon'}$ from s_0 in $\mathcal{M}'$ with $\mathbb{P}_{\mathcal{M}',s_0,\sigma_{\varepsilon'}}(\varphi) \geq \mathtt{val}_{\mathcal{M}'}(s_0) - \varepsilon'$.

Since $\tau_2 < 1$, can we apply Lemma 42 and obtain $\mathbb{P}_{\mathcal{M}',s_0,\hat\sigma}(\mathsf{F}Safe_{\mathcal{M}',\mathtt{Safety}(T)}(\tau_2)) \geq \mathbb{P}_{\mathcal{M}',s_0,\hat\sigma}(\varphi)$ for every strategy $\hat\sigma$ from s_0 and thus in particular for $\sigma_{\varepsilon'}$.

Therefore, $\mathbb{P}_{\mathcal{M}',s_0,\sigma_{\varepsilon'}}(\mathsf{F}Safe_{\mathcal{M}',\mathtt{Safety}(T)}(\tau_2)) \geq \mathtt{val}_{\mathcal{M}'}(s_0) - \varepsilon'$. Since this holds for every $\varepsilon' > 0$, in $\mathcal{M}'$ the value of s_0 w.r.t. the reachability objective $\mathsf{F}Safe_{\mathcal{M}',\mathtt{Safety}(T)}(\tau_2)$ is $\geq \mathtt{val}_{\mathcal{M}'}(s_0)$ for every state s_0.

By Theorem 40, for every $\varepsilon_2 > 0$ there exists a uniformly ε_2-optimal MD-strategy σ' in $\mathcal{M}'$ for this reachability objective. So we obtain

$$\mathbb{P}_{\mathcal{M}',s_0,\sigma'}(\mathsf{F}Safe_{\mathcal{M}',\mathtt{Safety}(T)}(\tau_2)) \geq \mathtt{val}_{\mathcal{M}'}(s_0) - \varepsilon_2 \quad \text{for every state } s_0. \tag{49}$$

In particular, σ' must coincide with $\sigma_{opt\text{-}av}$ at all states in S', since in $\mathcal{M}'$ these choices are already fixed.

Since σ is a uniformly λ-optimal MD-strategy for $\mathtt{Safety}(T)$ in $\mathcal{M}$ and $\mathcal{M}'$, we have $Safe_{\mathcal{M}',\mathtt{Safety}(T)}(\tau_2) \subseteq Safe_{\mathcal{M}',\sigma,\mathtt{Safety}(T)}(\tau_2 - \lambda) = S''$ and thus by (49) we get

$$\mathbb{P}_{\mathcal{M}',s_0,\sigma'}(\mathsf{F}S'') \geq \mathbb{P}_{\mathcal{M}',s_0,\sigma'}(\mathsf{F}Safe_{\mathcal{M}',\mathtt{Safety}(T)}(\tau_2)) \geq \mathtt{val}_{\mathcal{M}'}(s_0) - \varepsilon_2. \tag{50}$$

We obtain the MD-strategy σ_ε in $\mathcal{M}$ by combining the MD-strategies σ' and $\sigma_{opt\text{-}av}$. The strategy σ_ε plays like $\sigma_{opt\text{-}av}$ at all states inside S' and like σ' at all states outside S' (i.e., at $S \setminus S'$).

In order to show that σ_ε has the required property $\mathbb{P}_{\mathcal{M},s_0,\sigma_\varepsilon}(\varphi) \geq \mathtt{val}_{\mathcal{M}}(s_0) - \varepsilon$, we first estimate the probability that a play according to σ_ε will never leave the set S' after having visited a state in S''.

Let $s \in S''$. Then, by applying Lemma 15 to the Markov chain obtained from applying the MD-strategy σ_ε to $\mathcal{M}$, we obtain

$$
\begin{aligned}
\mathbb{P}_{\mathcal{M},s,\sigma_{opt\text{-}av}}(\mathsf{G}S') \quad &\geq \quad \frac{(\tau_2 - \lambda) - \tau_1}{1 - \tau_1} \\
&= \quad \frac{(1 - \varepsilon_1/k) - (1 - \varepsilon_1)}{\varepsilon_1} \\
&= \quad 1 - \frac{1}{k}.
\end{aligned}
\tag{51}
$$

In particular we also have $\mathbb{P}_{\mathcal{M},s,\sigma_\varepsilon}(\mathsf{G}S') \geq 1 - \frac{1}{k}$, since σ_ε coincides with $\sigma_{opt\text{-}av}$ inside the set S'. Finally we obtain for every $s_0 \in S$

$$
\begin{aligned}
\mathbb{P}_{\mathcal{M},s_0,\sigma_\varepsilon}(\varphi) \quad &= \quad \mathbb{P}_{\mathcal{M},s_0,\sigma_\varepsilon}(\mathsf{F}\mathsf{G}(S \setminus T)) && \text{by def. of } \varphi \\
&\geq \quad \mathbb{P}_{\mathcal{M},s_0,\sigma_\varepsilon}(\mathsf{F}S'') \cdot \mathbb{P}_{\mathcal{M},s_0,\sigma_\varepsilon}(\mathsf{F}\mathsf{G}S' \mid \mathsf{F}S'') && \text{since } S' \subseteq S \setminus T \\
&\geq \quad \mathbb{P}_{\mathcal{M}',s_0,\sigma'}(\mathsf{F}S'') \cdot (1 - 1/k) && \text{by (51)} \\
&\geq \quad (\mathtt{val}_{\mathcal{M}'}(s_0) - \varepsilon_2) \cdot (1 - 1/k) && \text{by (50)} \\
&\geq \quad (\mathtt{val}_{\mathcal{M}}(s_0) - \varepsilon_1 - \varepsilon_2) \cdot (1 - 1/k) && \text{by (48)}
\end{aligned}
$$

This holds for every $1 > \varepsilon_1, \varepsilon_2 > 0$ and every $k \geq 1$, and moreover $\mathtt{val}_{\mathcal{M}}(s_0) \leq 1$. Thus we can set $\varepsilon_1 = \varepsilon_2 \overset{\text{def}}{=} \varepsilon/4$ and $k \overset{\text{def}}{=} \frac{2}{\varepsilon}$ and obtain $\mathbb{P}_{\mathcal{M},s_0,\sigma_\varepsilon}(\varphi) \geq \mathtt{val}_{\mathcal{M}}(s_0) - \varepsilon$ for every $s_0 \in S$ as required.   ◀