Parity Objectives in Countable MDPs

Stefan Kiefer*, Richard Mayr†, Mahsa Shirmohammadi*, Dominik Wojtczak‡
*University of Oxford, UK
†University of Edinburgh, UK
‡University of Liverpool, UK

***Abstract*—We study countably infinite MDPs with parity objectives, and special cases with a bounded number of colors in the Mostowski hierarchy (including reachability, safety, Büchi and co-Büchi).**

In finite MDPs there always exist optimal memoryless deterministic (MD) strategies for parity objectives, but this does not generally hold for countably infinite MDPs. In particular, optimal strategies need not exist.

For countable infinite MDPs, we provide a complete picture of the memory requirements of optimal (resp., ϵ-optimal) strategies for all objectives in the Mostowski hierarchy.

In particular, there is a strong dichotomy between two different types of objectives. For the first type, optimal strategies, if they exist, can be chosen MD, while for the second type optimal strategies require infinite memory. (I.e., for all objectives in the Mostowski hierarchy, if finite-memory randomized strategies suffice then also MD-strategies suffice.) Similarly, some objectives admit ϵ-optimal MD-strategies, while for others ϵ-optimal strategies require infinite memory. Such a dichotomy also holds for the subclass of countably infinite MDPs that are finitely branching, though more objectives admit MD-strategies here.

***Index Terms*—countable MDPs, parity objectives, strategies, memory requirement**

I. Introduction

Markov decision processes (MDPs) are a standard model for dynamic systems that exhibit both stochastic and controlled behavior [23]. The system starts in the initial state and makes a sequence of transitions between states. Depending on the type of the current state, either the controller gets to choose an enabled transition (or a distribution over transitions), or the next transition is chosen randomly according to a defined distribution. By fixing a strategy for the controller, one obtains a probability space of plays of the MDP. The goal of the controller is to optimize the expected value of some objective function on the plays of the MDP. The fundamental questions are "what is the optimal value that the controller can achieve?", "does there exist an optimal strategy, or only ϵ-optimal approximations?", and "which types of strategies are optimal or ϵ-optimal?".

Such questions have been studied extensively for finite MDPs (see e.g. [9] for a survey) and also for certain types of countably infinite MDPs [23], [21]. However, the literature on countable MDPs is mainly focused on objective functions defined w.r.t. numeric costs (or rewards) that are assigned to transitions, e.g. (discounted) expected total reward or limit-average reward. In contrast, we study qualitative objectives that are expressed by Parity conditions and which are motived by formal verification questions.

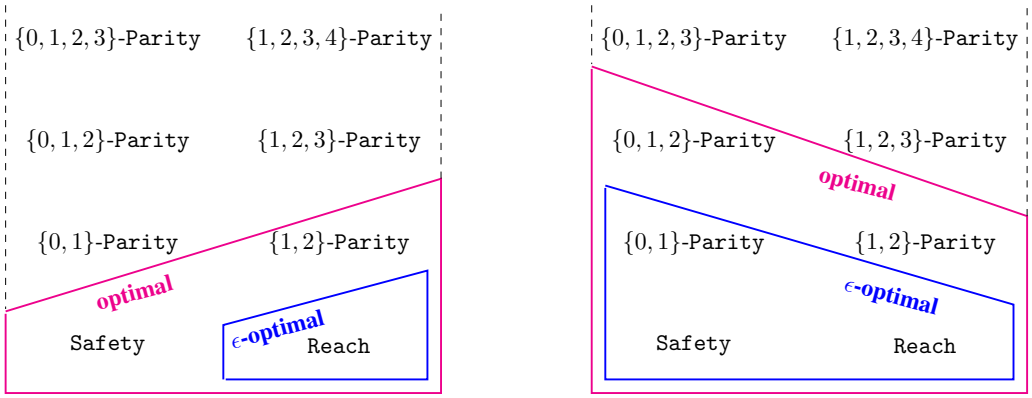
978-1-5090-3018-7/17/$31.00

There are works that studied particular classes of countably infinite, but finitely branching, MDPs that arise from models in automata theory [13], [2], [7], [5], [1]. In each of these papers, a crucial part of the analysis is establishing the existence of optimal strategies of particular structure and memory requirements, but none of them looked at proving such properties for general countable MDPs. Countable MDPs also naturally occur in the analysis of queueing systems [17], gambling [3], and branching processes [22], which have multiple applications. They also show up in the analysis of finite-state models, e.g. in two-player stochastic games [24], [12] when reasoning about an optimal strategy against a fixed (randomised and memory-full) strategy of the opponent.

Finite MDPs vs. Infinite MDPs: It should be noted that many standard properties (and proof techniques) of finite MDPs do *not* carry over to infinite MDPs.

E.g., given some objective, consider the set of all states in an MDP that have nonzero value. If the MDP is finite then this set is finite and thus there exists some minimal nonzero value. This property does *not carry over* to infinite MDPs. Here the set of states is infinite and the infimum over the nonzero values can be zero. As a consequence, even for a reachability objective, it is possible that all states have value > 0, but still the value of some states is < 1. Such phenomena appear already in infinite-state Markov chains like the classic Gambler's ruin problem with unfair coin tosses in the player's favor (0.6 win, 0.4 lose). The value, i.e., the probability of ruin, is always > 0, but still < 1 in every state except the ruin state itself; cf. [14] (Chapt. 14). Another difference is that optimal strategies need not exist, even for qualitative objectives like reachability or parity. Even if some state has value 1, there might not be any single strategy that attains the value 1, but only an infinite family of ϵ-optimal strategies for every $\epsilon > 0$.

Parity objectives: We study general 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 [15]. Every state has a *color*, out of a finite set of colors encoded as natural numbers. An infinite play is winning iff the highest color that is seen infinitely often in the play is even. The controller wants to maximize the probability of winning plays. Subclasses of parity objectives are defined by restricting the set of used colors; these are classified in the Mostowski hierarchy [20]



a) Infinitely branching MDPs **b)** Finitely branching MDPs

Fig. 1: For countable MDPs, these diagrams show the memory requirements of optimal and ϵ-optimal strategies for objectives in the Mostowski hierarchy. An objective in a level of the hierarchy subsumes all objectives in lower levels, e.g., $\{0,1,2\}$-Parity subsumes $\{1,2\}$-Parity. We have extended the Mostowski hierarchy to include reachability and safety. The magenta (resp., blue) regions enclose objectives where memoryless deterministic (MD) strategies are sufficient for optimal (resp., ϵ-optimal) strategies; for objectives outside the regions, infinite-memory strategies are necessary. The left diagram is for infinitely branching MDPs; e.g., ϵ-optimal strategies for all but reachability objectives require infinite memory, whereas MD-strategies are sufficient for reachability. The right diagram is for finitely branching MDPs; e.g., optimal strategies (if they exist) can be chosen MD for all objectives subsumed by $\{0,1,2\}$-Parity.

which includes, e.g., Büchi and co-Büchi objectives. Such prefix-independent infinitary objectives cannot generally be encoded by numeric transition rewards as in [23], though both types subsume the simpler reachability and safety objectives.

There are different types of strategies, depending on whether one can take the whole history of the play 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.

The type of strategy needed for an optimal (resp. ϵ-optimal) strategy for some objective is also called the *strategy complexity* of the objective. For finite MDPs, MD-strategies are sufficient for all types of qualitative and quantitative parity objectives [8], [10], but the picture is more complex for countably infinite MDPs.

Since optimal strategies need not exist in general, we consider both the strategy complexity of ϵ-optimal strategies, and the strategy complexity of optimal strategies under the assumption that they exist. E.g., if an optimal strategy exists, can it be chosen MD?

We provide a complete picture of the memory requirements for objectives in the Mostowski hierarchy, which is summarized in Figure 1.

In particular, our results show that there is a strong dichotomy between two different classes of objectives. For objectives of the first class, optimal strategies, where they exist, can be chosen MD. For objectives of the second class, optimal strategies require infinite memory in general, in the sense that all FR-strategies achieve the objective only with probability zero. A similar dichotomy applies to ϵ-optimal strategies. For certain objectives, ϵ-optimal MD-strategies exist, while for all others even ϵ-optimal strategies require infinite memory in general. This is a strong dichotomy because there are no objectives in the Mostowski hierarchy for which other types of strategies (MR, FD, or FR) are both necessary and sufficient. Put differently, for all objectives in the Mostowski hierarchy, if FR-strategies suffice then MD-strategies suffice as well.

We also consider the subclass of countable MDPs that are finitely branching. (Note that these generally still have an infinite number of states.) The above mentioned dichotomies apply here as well, though the classes of objectives where optimal (resp. ϵ-optimal) strategies can be chosen MD are larger than for general countable MDPs.

Outline of the results: In Section II we define countably infinite MDPs, strategies and parity objectives. In Section III we show examples that demonstrate that certain objectives require infinite memory. For some of these we refer to previous work. The main new result in this section is Theorem 1 that shows that even almost-sure $\{1,2,3\}$-Parity on finitely branching MDPs requires infinite memory. These negative results highlight the questions which other objectives still allow MD-strategies. Apart from the case of reachability objectives, these questions were open. We provide complete answers in several steps. First, in Section IV, we prove a general result (Theorem 5) that relates the strategy complexity of almost-sure winning strategies and optimal strategies. The complexity of the proof is due to the fact that we consider *infinite* MDPs (which do not satisfy basic properties of finite MDPs

in general; see above). We then use this theorem to establish MD-strategies for Büchi, co-Büchi and $\{0,1,2\}$-Parity objectives in the following sections. In Section V we show that optimal strategies for Büchi objectives, where they exist, can be chosen MD, even for infinitely branching MDPs. In Section VI we consider finitely branching MDPs. We show that optimal strategies for $\{0,1,2\}$-Parity, where they exist, can be chosen MD (Theorem 16). This is a very general result. E.g., this question had been open (and is non-trivial) even for almost-sure co-Büchi objectives. Moreover, we show that ϵ-optimal strategies for co-Büchi objectives can be chosen MD (Theorem 19). We conclude the paper with a discussion of how some results change when one considers uncountable MDPs.

Missing proofs can be found in the full version of this paper [16].

II. Preliminaries

A *probability distribution* over a countable (not necessarily finite) set S is a function $f : S \to [0,1]$ s.t. $\sum_{s\in S} f(s) = 1$. We use $\mathsf{supp}(f) = \{s \in S \mid f(s) > 0\}$ to denote the *support* of f. Let $\mathcal{D}(S)$ be the set of all probability distributions over S.

We consider countably infinite Markov decision processes (MDPs) $\mathcal{M} = (S, S_\Box, S_\bigcirc, \longrightarrow, P)$ where the countable set S of *states* is partitioned into the set $S_\Box$ of states of the player and *random states* $S_\bigcirc$. The relation $\longrightarrow \subseteq S \times S$ is the transition relation. We write $s \longrightarrow s'$ if $(s,s') \in \longrightarrow$, and we assume that each state s has a *successor* state s' with $s \longrightarrow s'$. The probability function $P : S_\bigcirc \to \mathcal{D}(S)$ assigns to each random state $s \in S_\bigcirc$ a probability distribution over its successor states. A set $T \subseteq S$ is a *sink* in $\mathcal{M}$ if for all $s \in T$ all successors of s are in T. The MDP $\mathcal{M}$ is called *finitely branching* if each state has only finitely many successors; otherwise, it is *infinitely branching*. A Markov chain is an MDP where $S_\Box = \emptyset$, i.e., all states are random states.

We describe the behavior of an MDP as a one-player stochastic game played for infinitely many rounds. The game starts in a given initial state s_0. In each round, if the game is in state $s \in S_\Box$ then the player (or controller) chooses a successor state s' with $s \longrightarrow s'$; otherwise the game is in a random state $s \in S_\bigcirc$ and proceeds randomly to s' with probability $P(s)(s')$.

Strategies. A *play* w is an infinite sequence $s_0 s_1 \cdots$ of states such that $s_i \longrightarrow s_{i+1}$ for all $i \geq 0$; let $w(i) = s_i$ denote the i-th state along w. A *partial play* is a finite prefix of a play. We say that (partial) play w *visits* s if $s = w(i)$ for some i, and that w starts in s if $s = w(0)$. A *strategy* is a function $\sigma : S^* S_\Box \to \mathcal{D}(S)$ that assigns to partial plays $ws \in S^* S_\Box$ a distribution over the successors $\{s' \in S \mid s \longrightarrow s'\}$. The set of all strategies in $\mathcal{M}$ is denoted by $\Sigma_\mathcal{M}$ (we omit the subscript and write Σ if $\mathcal{M}$ is clear). A (partial) play $s_0 s_1 \cdots$ is induced by strategy σ if $s_{i+1} \in \mathsf{supp}(\sigma(s_0 s_1 \cdots s_i))$ for all i with $s_i \in S_\Box$, and $s_{i+1} \in \mathsf{supp}(P(s_i))$ for all i with $s_i \in S_\bigcirc$.

Since this paper focuses on the memory requirements of strategies, we present an equivalent formulation of strategies, emphasizing the amount of memory required to implement a strategy. Strategies can be implemented by probabilistic transducers $\mathsf{T} = (\mathsf{M}, \mathsf{m}_0, \pi_u, \pi_s)$ where M is a countable set (the memory of the strategy), $\mathsf{m}_0 \in \mathsf{M}$ is the *initial memory mode* and S is the input and output alphabet. The probabilistic transition function $\pi_u : \mathsf{M} \times S \to \mathcal{D}(\mathsf{M})$ updates the memory mode of transducer. The probabilistic successor function $\pi_s : \mathsf{M} \times S_\Box \to \mathcal{D}(S)$ outputs the next successor, where $s' \in \mathsf{supp}(\pi_s(\mathsf{m}, s))$ implies $s \longrightarrow s'$. We extend π_u to $\mathcal{D}(\mathsf{M}) \times S \to \mathcal{D}(\mathsf{M})$ and π_s to $\mathcal{D}(\mathsf{M}) \times S_\Box \to \mathcal{D}(S)$, in the natural way. Moreover, we extend π_u to paths by $\pi_u(\mathsf{m}, \varepsilon) = \mathsf{m}$ and $\pi_u(\mathsf{m}, s_0 \cdots s_n) = \pi_u(\pi_u(s_0 \cdots s_{n-1}, \mathsf{m}), s_n)$. The strategy $\sigma_\mathsf{T} : S^* S_\Box \to \mathcal{D}(S)$ induced by the transducer T is given by $\sigma_\mathsf{T}(s_0 \cdots s_n) := \pi_s(s_n, \pi_u(s_0 \cdots s_{n-1}, \mathsf{m}_0))$. Note that such strategies allow for randomized memory updates and probabilistic successor functions.

Strategies are in general *history dependent* (H) and *randomized* (R). An H-strategy σ is *finite memory* (F) if there exists some transducer T with memory M such that $\sigma_\mathsf{T} = \sigma$ and $|\mathsf{M}| < \infty$; otherwise we say σ *requires infinite memory*. An F-strategy is *memoryless* (M) (also called *positional*) if $|\mathsf{M}| = 1$. We may view M-strategies as functions $\sigma : S_\Box \to \mathcal{D}(S)$. An R-strategy σ is *deterministic* (D) if π_u and π_s map to Dirac distributions; it implies that $\sigma(w)$ is a Dirac distribution for all partial plays w. All combinations of the properties in $\{\mathrm{M}, \mathrm{F}, \mathrm{H}\} \times \{\mathrm{D}, \mathrm{R}\}$ are possible, e.g., MD stands for memoryless deterministic. HR strategies are the most general type.

Probability Measures. An MDP $\mathcal{M} = (S, S_\Box, S_\bigcirc, \longrightarrow, P)$, an initial state s_0, and a strategy σ induce a standard probability measure on sets of infinite plays. We write $\mathcal{P}_{\mathcal{M},s_0,\sigma}(\mathfrak{R})$ for the probability of a measurable set $\mathfrak{R} \subseteq s_0 S^\omega$ of plays starting from s_0. It is defined, as usual, by first defining it 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 play induced by σ then set $\mathcal{P}_{\mathcal{M},s_0,\sigma}(s_0 s_1 \ldots s_n S^\omega) = 0$; otherwise set $\mathcal{P}_{\mathcal{M},s_0,\sigma}(s_0 s_1 \ldots s_n S^\omega) = \prod_{i=0}^{n-1} \bar{\sigma}(s_0 s_1 \ldots s_i)(s_{i+1})$, where $\bar{\sigma}$ is the map that extends σ by $\bar{\sigma}(ws) = P(s)$ for any $ws \in S^* S_\bigcirc$. Using Carathéodory's extension theorem [4], this defines a unique probability measure $\mathcal{P}_{\mathcal{M},s_0,\sigma}$ on measurable subsets of $s_0 S^\omega$.

Objectives. Let $\mathcal{M} = (S, S_\Box, S_\bigcirc, \longrightarrow, P)$ be an MDP. 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 [11]. Formulas are interpreted on the structure $(S, \longrightarrow)$. We use $[\![\varphi]\!]^s \subseteq s S^\omega$ to denote the set of plays starting from s that satisfy the LTL formula φ. This set is measurable [25], and we just write $\mathcal{P}_{\mathcal{M},s,\sigma}(\varphi)$ instead of $\mathcal{P}_{\mathcal{M},s,\sigma}([\![\varphi]\!]^s)$. We also write $[\![\varphi]\!]$ for $\bigcup_{s\in S} [\![\varphi]\!]^s$.

Given a target set $T \subseteq S$, the *reachability objective* is defined by $\mathtt{Reach}(T) = [\![\mathsf{F}\, T]\!]$, i.e., $s_0 s_1 \cdots \in \mathtt{Reach}(T) \Leftrightarrow \exists i.\, s_i \in T$. The *safety objective* is defined by $\mathtt{Safety}(T) = [\![\mathsf{G}\, \neg T]\!]$, i.e., $s_0 s_1 \cdots \in \mathtt{Safety}(T) \Leftrightarrow \forall i.\, s_i \notin T$. Given a reachability or a safety objective, we can assume without loss of generality that T is a sink in $\mathcal{M}$.

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 $Col(s)$. For $n \in \mathbb{N}$, $\triangleright \in \{<, \leq, =, \geq, >\}$ and $Q \subseteq S$, let $[Q]^{Col \triangleright n} := \{s \in Q \mid Col(s) \triangleright n\}$ be the set of states in Q with color $\triangleright n$. The *parity objective* is defined by

$$\texttt{Parity}(Col) := \Big[\!\Big[\bigvee_{i \in \mathcal{C}} \big(\mathsf{GF}[S]^{Col=2\cdot i} \wedge \mathsf{FG}[S]^{Col \leq 2\cdot i}\big) \Big]\!\Big],$$

i.e., $\texttt{Parity}(Col)$ is the set of infinite plays such that the largest color that occurs infinitely often along the play is even.

The Mostowski hierarchy [20] classifies parity objectives by restricting the range of the function Col to a set of colors $\mathcal{C} \subseteq \mathbb{N}$. We write $\mathcal{C}$-$\texttt{Parity}$ for such restricted parity objectives. In particular, Büchi objectives correspond to $\{1,2\}$-$\texttt{Parity}$, and co-Büchi objectives correspond to $\{0,1\}$-$\texttt{Parity}$. The objectives $\{0,1,2\}$-$\texttt{Parity}$ and $\{1,2,3\}$-$\texttt{Parity}$ are incomparable, but they both subsume (modulo renaming of colors) Büchi and co-Büchi objectives. Moreover, both $\{0,1\}$-$\texttt{Parity}$ and $\{1,2\}$-$\texttt{Parity}$ subsume the reachability objective $\texttt{Reach}(T)$ (for MDPs with a sink T), by defining the color function so that $Col(s) = 1 \Leftrightarrow s \notin T$. Similarly, both $\{0,1\}$-$\texttt{Parity}$ and $\{1,2\}$-$\texttt{Parity}$ subsume $\texttt{Safety}(T)$, by defining $Col(s) = 1 \Leftrightarrow s \in T$.

Optimal and ϵ-Optimal Strategies. Given an objective φ, the value of state s in an MDP $\mathcal{M}$, denoted by $\texttt{val}_{\mathcal{M}}(s)$, is the supremum probability of achieving φ, i.e., $\texttt{val}_{\mathcal{M}}(s) := \sup_{\sigma \in \Sigma} \mathcal{P}_{\mathcal{M},s,\sigma}(\varphi)$. For $\epsilon \geq 0$ and $s \in S$, we say that a strategy σ is *ϵ-optimal* iff $\mathcal{P}_{\mathcal{M},s,\sigma}(\varphi) \geq \texttt{val}_{\mathcal{M}}(s) - \epsilon$. A 0-optimal strategy is called *optimal*. An optimal strategy is *almost-surely winning* if $\texttt{val}_{\mathcal{M}}(s) = 1$. Unlike in finite-state MDPs, optimal strategies need not exist in countable MDPs, not even for reachability objectives in finitely branching MDPs. However, by the definition of the value, for all $\epsilon > 0$, an ϵ-optimal strategy exists.

For an objective φ and $\triangleright \in \{\geq, >\}$ and $c \in [0,1]$, we define $\big[\varphi\big]^{\triangleright c}$ as the set of states s for which there exists a strategy σ with $\mathcal{P}_{\mathcal{M},s,\sigma}(\varphi) \triangleright c$. We call a state s *almost-surely winning* if $s \in \big[\varphi\big]^{\geq 1}$, and we call s *limit-surely winning* if $s \in \big[\varphi\big]^{\geq c}$ for every constant $c < 1$ (which is iff $\texttt{val}_{\mathcal{M}}(s) = 1$). On infinite arenas, limit-surely winning states are not necessarily almost-surely winning.

III. Objectives that require infinite memory

In this section we consider those objectives in the Mostowski hierarchy where optimal (resp., ϵ-optimal) strategies require infinite memory. In each such case we construct an MDP that witnesses this requirement. In these MDPs, all FR-strategies achieve the objective only with probability 0, while some HD-strategy achieves the objective almost-surely (resp., with arbitrarily high probability).

Theorem 1. *Let $\varphi = \{1,2,3\}$-$\texttt{Parity}$. There exists a finitely branching MDP $\mathcal{M}$ with initial state s_0 such that*

- *for all FR-strategies σ, we have $\mathcal{P}_{\mathcal{M},s_0,\sigma}(\varphi) = 0$,*
- *there exists an HD-strategy σ such that $\mathcal{P}_{\mathcal{M},s_0,\sigma}(\varphi) = 1$.*

Hence, optimal (and even almost-surely winning) and ϵ-optimal strategies require infinite memory for $\{1,2,3\}$-$\texttt{Parity}$, even in finitely branching MDPs.

The MDP in Theorem 1 is depicted in Figure 2 (left), where $Col(s_i) = 1$ and $Col(r_i) = 2$ for all $i \in \mathbb{N}$, and $Col(t) = 3$. For every FR-strategy there is a uniform lower bound on the probability of visiting t between consecutive visits to s_0. Hence, unless the strategy with positive probability eventually always stays in states s_i (and thus also loses the almost-sure parity objective), in the long-run, the probability of visiting t (with color three) tends to 1, and the parity condition is satisfied with probability 0. Although the player cannot win by any FR-strategy, we construct an HD-strategy σ such that $\mathcal{P}_{\mathcal{M},s_0,\sigma}(\varphi) = 1$. This strategy is such that upon the i^{th} visit to s_0, the ladder $s_0 s_1 \cdots s_i$ is traversed and the transition $s_i \longrightarrow r_i$ is chosen. Moving further along the ladder $s_0 s_1 s_2 \cdots$ decreases the probability of visiting t between the previous and successive visits to s_0. Hence, the probability of visiting color three infinitely often is 0.

Remark 1. *A strict subclass of finitely branching MDPs are 1-counter MDPs, where a finite-state MDP is augmented with an integer counter [5]. The MDP in Theorem 1 (plus some auxiliary states) is implementable by a 1-counter MDP.*

Remark 2. *The classical Rabin and Streett conditions can encode $\{1,2,3\}$-$\texttt{Parity}$. Thus, optimal and ϵ-optimal strategies for Rabin/Streett require infinite memory, even in finitely branching countable MDPs.*

On finite MDPs, optimal strategies can be chosen MD for parity and Rabin objectives, but not for Streett objectives. Optimal strategies for Streett objectives can be chosen MR or FD [8].

Proof. For an infinite play π^∞, let $\mathsf{Inf}(\pi^\infty)$ be the set of states that π^∞ visits infinitely often. Let us recall the Rabin and Streett conditions.

Given a Rabin condition $\{(E_1, F_1), (E_2, F_2), \cdots, (E_n, F_n)\}$ with n pairs (or n disjunctions), an infinite play π^∞ satisfies the Rabin condition if there exists a pair (E_i, F_i) such that $\mathsf{Inf}(\pi^\infty) \cap E_i = \emptyset$ and $\mathsf{Inf}(\pi^\infty) \cap F_i \neq \emptyset$. The Rabin condition

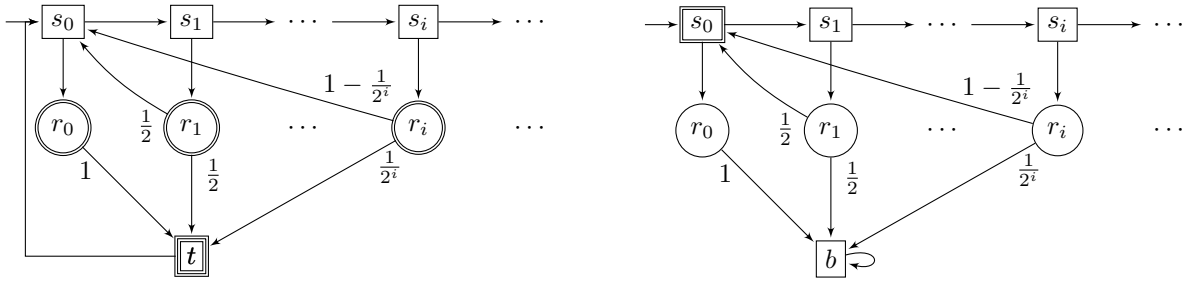
$$\{([S]^{Col=3}, [S]^{Col=2})\}$$

encodes $\{1,2,3\}$-$\texttt{Parity}$, since all satisfying runs must visit states with color 2 infinitely often and states with color 3 only finitely often. Note that $\{1,2,3\}$-$\texttt{Parity}$ is encoded in a Rabin condition with only one disjunction.

Given a Streett condition $\{(E_1, F_1), (E_2, F_2), \cdots, (E_n, F_n)\}$ with n pairs (or n conjunctions), an infinite play π^∞ satisfies the Streett condition if $\mathsf{Inf}(\pi^\infty) \cap E_i = \emptyset$ implies $\mathsf{Inf}(\pi^\infty) \cap F_i = \emptyset$ for all pairs (E_i, F_i). The Streett condition

$$\{([S]^{Col=2}, S), (\emptyset, [S]^{Col=3})\}$$

encodes $\{1,2,3\}$-$\texttt{Parity}$, since all satisfying runs must visit states with color 2 infinitely often and states with color 3 only finitely often.



a) Almost-sure $\{1,2,3\}$-Parity **b)** Limit-sure Büchi

Fig. 2: Two finitely branching MDPs where the states $s \in S_\Box$ of the player are drawn as squares and random states $s \in S_\bigcirc$ as circles. The color $Col(s)$ of s is indicated with the number of boundaries; for example, a double boundary for color 2. State s_0 in the MDP on the left is almost-surely winning for $\{1,2,3\}$-Parity, but all almost-surely winning strategies require infinite memory. The MDP on the right is such that, for all $c > 0$, strategies that achieve Büchi with probability at least c require infinite memory.

Note that a conjunction of *two* Streett pairs are needed to encode $\{1,2,3\}$-Parity. A single Streett pair $\{(X,Y)\}$ means "infinitely often X or only finitely often Y", which can be encoded as a $\{0,1,2\}$-Parity condition by assigning color 2 to X and color 1 to Y. Unlike for $\{1,2,3\}$-Parity, optimal strategies for $\{0,1,2\}$-Parity (and thus also for a single Streett pair) can be chosen MD in finitely branching MDPs (Theorem 16). □

It was known that quantitative Büchi objectives require infinite memory [18], [2]. For the sake of completeness, we present an example MDP for Proposition 2 in Figure 2 (right).

Proposition 2 ([18])**.** *Let $\varphi = \{1,2\}$-Parity be the Büchi objective. There exists a finitely branching MDP $\mathcal{M}$ with initial state s_0 such that*

- *for all FR-strategies σ, we have $\mathcal{P}_{\mathcal{M},s_0,\sigma}(\varphi) = 0$,*
- *for every $c \in [0,1)$, there exists an HD-strategy σ such that $\mathcal{P}_{\mathcal{M},s_0,\sigma}(\varphi) \geq c$.*

Hence, ϵ-optimal strategies for Büchi objectives require infinite memory.

Theorem 3. *Let $\varphi = \mathtt{Safety}(T)$. There exists an infinitely branching MDP $\mathcal{M}$ with initial state s such that*

- *for all FR-strategies σ, we have $\mathcal{P}_{\mathcal{M},s,\sigma}(\varphi) = 0$,*
- *for every $c \in [0,1)$, there exists an HD-strategy σ such that $\mathcal{P}_{\mathcal{M},s,\sigma}(\varphi) \geq c$.*

Hence, ϵ-optimal strategies for safety require infinite memory.

The MDP in Theorem 3, depicted in Figure 3 (left), was first introduced in [19]. Since our notion of finite-memory strategies allows for randomized memory updates (in contrast to [19]), our proof is somewhat more general. The target is $T = \{t\}$. For every FR-strategy there is a uniform lower bound on the probability of reaching t between consecutive visits to s_0. Since t is absorbing, it will be reached with probability 1. Thus every FR-strategy satisfies the safety objective with probability 0. However, for all $n \in \mathbb{N}$, we construct an HD-strategy σ_n such that $\mathcal{P}_{\mathcal{M},s,\sigma_n}(\mathtt{Safety}(\{t\})) \geq 1 - \frac{1}{2^n}$. This strategy is such that upon the i^{th} visit to s, the transition $s \longrightarrow r_{i+n}$ is chosen. Hence, the probability of visiting t between two successive visits to s decreases. A more detailed analysis shows that the probability of ever visiting t is bounded by $\frac{1}{2^n}$.

Theorem 4. *Let $\varphi = \{0,1\}$-Parity be the co-Büchi objective. There exists an infinitely branching MDP $\mathcal{M}$ with initial state s such that*

- *for all FR-strategies σ, we have $\mathcal{P}_{\mathcal{M},s,\sigma}(\varphi) = 0$,*
- *there exists an HD-strategy σ such that $\mathcal{P}_{\mathcal{M},s,\sigma}(\varphi) = 1$.*

Hence, optimal (and even almost-surely winning) strategies and ϵ-optimal strategies for co-Büchi require infinite memory.

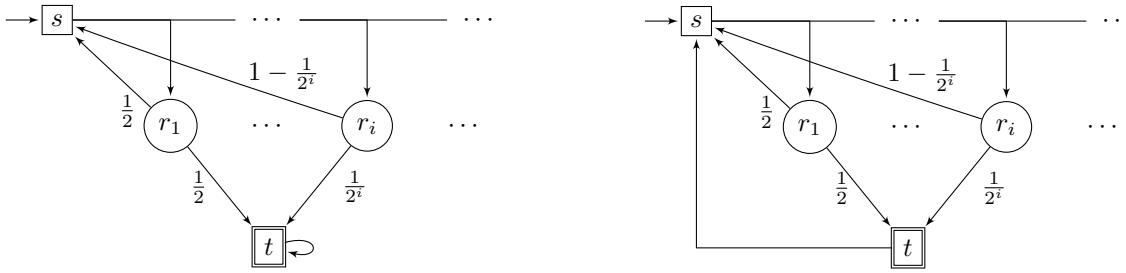
The MDP in Theorem 4 is depicted in Figure 3 (right). By a similar argument as in Theorem 3, every FR-strategy achieves co-Büchi with probability 0. However, the HD-strategy σ that chooses the transition $s \longrightarrow r_i$ upon the i^{th} visit to s is such that $\mathcal{P}_{\mathcal{M},s,\sigma}(\varphi) = 1$.

IV. From almost-sure winning to optimal strategies

In this section we prove Theorem 5. It says that, for certain objectives, if almost-surely winning strategies (where they exist) can be chosen MD, then optimal strategies (where they exist) can also be chosen MD.

We call a class $\mathcal{C}$ of MDPs *downward-closed* if every MDP whose transition relation is a subset of the transition relation of some MDP in $\mathcal{C}$ is also in $\mathcal{C}$. The class of finitely branching MDPs is downward-closed, and so is the class of MDPs with a fixed sink T.

We call an objective φ *prefix-independent in* $\mathcal{C}$ (where $\mathcal{C}$ is a class of MDPs) if for all $w_1, w_2 \in S^*$ and all $w \in S^\omega$ such that $w_1 w$ and $w_2 w$ are infinite plays in an MDP in $\mathcal{C}$ we have $w_1 w \in [\![\varphi]\!] \iff w_2 w \in [\![\varphi]\!]$. Parity objectives are prefix-independent in the class of all MDPs. Both objectives $\mathtt{Reach}(T)$ and $\mathtt{Safety}(T)$ are prefix-independent in the class of MDPs with sink T.



a) Limit-sure safety with target $\{t\}$ **b)** Almost-sure co-Büchi

Fig. 3: In the infinitely branching MDP on the left, all ϵ-optimal strategies for Safety require infinite memory. In the infinitely branching MDP on the right, all optimal (and thus almost-surely winning) strategies for co-Büchi require infinite memory.

The following theorem provides, under certain conditions, an optimal MD-strategy for all states that have an optimal strategy. In fact, a single MD-strategy is optimal for all states that have an optimal strategy:

Theorem 5. *Let φ be an objective that is prefix-independent in a downward-closed class $\mathcal{C}$ of MDPs. Suppose that for any $\mathcal{M} = (S, S_\Box, S_\bigcirc, \longrightarrow, P) \in \mathcal{C}$ and any $s \in S$ and any strategy σ with $\mathcal{P}_{\mathcal{M},s,\sigma}(\varphi) = 1$ there exists an MD-strategy σ' with $\mathcal{P}_{\mathcal{M},s,\sigma'}(\varphi) = 1$.*

Under this condition, for each $\mathcal{M} \in \mathcal{C}$ there is an MD-strategy σ' such that for all $s \in S$:

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

The remainder of the section is devoted to the proof of Theorem 5.

For prefix-independent winning conditions, whenever an optimal strategy visits some state, it achieves the value of this state; see Lemma 20 in the full version of the paper [16]. We use this to show that the MDP constructed in the following lemma is well-defined. This MDP, $\mathcal{M}_*$, will be crucial for the proof of Theorem 5. Loosely speaking, $\mathcal{M}_*$ is the MDP $\mathcal{M}$ conditioned under φ.

Lemma 6. *Let φ be an objective that is prefix-independent in a class $\mathcal{C}$ of MDPs. Let $\mathcal{M} = (S, S_\Box, S_\bigcirc, \longrightarrow, P) \in \mathcal{C}$. Construct an MDP $\mathcal{M}_* = (S_*, S_{*\Box}, S_{*\bigcirc}, \longrightarrow_*, P_*)$ by setting*

$$S_* = \{s \in S \mid \exists \sigma.\, \mathcal{P}_{\mathcal{M},s,\sigma}(\varphi) = \mathtt{val}_{\mathcal{M}}(s) > 0\}$$

and $S_{*\Box} = S_* \cap S_\Box$ and $S_{*\bigcirc} = S_* \cap S_\bigcirc$ and

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

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 $s \longrightarrow_* t$. Then:

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

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

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

The following lemma provides, under certain conditions, a *uniform* almost-surely winning MD-strategy, i.e., one that works for all initial states at the same time:

Lemma 7. *Let $\mathcal{M} = (S, S_\Box, S_\bigcirc, \longrightarrow, P)$ 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 $\mathcal{P}_{\mathcal{M},s,\sigma}(\varphi) = 1$ there exists an MD-strategy σ' with $\mathcal{P}_{\mathcal{M},s,\sigma'}(\varphi) = 1$. Then there is an MD-strategy σ' such that for all $s \in S$:*

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

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 $\mathcal{P}_{\mathcal{M},s,\sigma'}(\varphi) = 1$.

Fix an arbitrary state $s_1 \in S$. By assumption there is an MD-strategy σ_1 with $\mathcal{P}_{\mathcal{M},s_1,\sigma_1}(\varphi) = 1$. 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 $\mathcal{P}_{\mathcal{M},s_1,\sigma_1}([\![\varphi]\!]^{s_1} \cap U_1^\omega) = 1$. In fact, for any $s \in U_1$ and any strategy σ that agrees with σ_1 on U_1 we have $\mathcal{P}_{\mathcal{M},s,\sigma}([\![\varphi]\!]^{s} \cap U_1^\omega) = 1$.

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 $\mathcal{P}_{\mathcal{M}_1,s,\sigma_1'}(\varphi) = 1$. Indeed, let $s \in S$ be any state. Recall that there is an MD-strategy σ with $\mathcal{P}_{\mathcal{M},s,\sigma}(\varphi) = 1$. 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 play 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 $\mathcal{P}_{\mathcal{M}_1,s,\sigma_1'}(\varphi) = 1$.

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 $\mathcal{P}_{\mathcal{M}_1,s_2,\sigma_2}([\![\varphi]\!]^{s_2} \cap U_2^\omega) = 1$. 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 $\mathcal{P}_{\mathcal{M},s,\sigma}([\![\varphi]\!]^s \cap U_2^\omega) = 1$.

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 $s \in S_\Box$ we have $\sigma'(s) = \sigma_i(s)$ for the smallest i with $s \in U_i$. Thus, if $s \in U_i$, we have $\mathcal{P}_{\mathcal{M},s,\sigma'}(\varphi) \geq \mathcal{P}_{\mathcal{M},s,\sigma'}([\![\varphi]\!]^s \cap U_i^\omega) = 1$. □

The following measure-theoretic lemma will be used to connect probability measures induced by the MDPs $\mathcal{M}$ and $\mathcal{M}_*$ from Lemma 6.

Lemma 8. *Let S be countable and $s \in S$. Call a set of the form swS^ω for $w \in S^*$ a* cylinder. *Let $\mathcal{P}, \mathcal{P}'$ be probability measures on sS^ω defined in the standard way, i.e., first on cylinders and then extended to all measurable sets $\mathfrak{R} \subseteq sS^\omega$. Suppose there is $x \geq 0$ such that $x \cdot \mathcal{P}(\mathfrak{C}) \leq \mathcal{P}'(\mathfrak{C})$ for all cylinders $\mathfrak{C}$. Then $x \cdot \mathcal{P}(\mathfrak{R}) \leq \mathcal{P}'(\mathfrak{R})$ holds for all measurable $\mathfrak{R} \subseteq sS^\omega$.*

We are ready to prove Theorem 5.

Proof of Theorem 5. As in the statement of the theorem, suppose that φ is an objective that is prefix-independent in a downward-closed class $\mathcal{C}$ of MDPs so that for any $\mathcal{M} = (S, S_\Box, S_\bigcirc, \longrightarrow, P) \in \mathcal{C}$ and any $s \in S$ and any strategy σ with $\mathcal{P}_{\mathcal{M},s,\sigma}(\varphi) = 1$ there exists an MD-strategy σ' with $\mathcal{P}_{\mathcal{M},s,\sigma'}(\varphi) = 1$. Let $\mathcal{M} = (S, S_\Box, S_\bigcirc, \longrightarrow, P) \in \mathcal{C}$. Let $\mathcal{M}_* = (S_*, S_{*\Box}, S_{*\bigcirc}, \longrightarrow_*, P_*)$ be the MDP defined in Lemma 6. Since $\mathcal{C}$ is downward-closed, we have $\mathcal{M}_* \in \mathcal{C}$. In particular, φ is prefix-independent in $\{\mathcal{M}_*\}$.

First we show that for any $s \in S_*$ there exists an MD-strategy σ' with $\mathcal{P}_{\mathcal{M}_*,s,\sigma'}(\varphi) = 1$. Indeed, let $s \in S_*$. By the definition of S_*, there is a strategy σ with $\mathcal{P}_{\mathcal{M},s,\sigma}(\varphi) = \mathtt{val}_{\mathcal{M}}(s) > 0$. By Lemma 6.2, we have $\mathcal{P}_{\mathcal{M}_*,s,\sigma}(\varphi) = 1$. By our assumption on $\mathcal{C}$ there exists an MD-strategy σ' with $\mathcal{P}_{\mathcal{M}_*,s,\sigma'}(\varphi) = 1$.

By Lemma 7, it follows that there is an MD-strategy σ' with $\mathcal{P}_{\mathcal{M}_*,s,\sigma'}(\varphi) = 1$ 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 play in $\mathcal{M}_*$ then, by Lemma 6.1,

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

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

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

If $s_0 s_1 \cdots s_n$ is not a partial play in $\mathcal{M}_*$ then $\mathcal{P}_{\mathcal{M}_*,s_0,\sigma'}(s_0 s_1 \cdots s_n S^\omega) = 0$ and the previous inequality holds as well. Therefore, by Lemma 8, we get for all measurable sets $\mathfrak{R} \subseteq s_0 S^\omega$:

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

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

V. When MD-strategies suffice in general countable MDPs

Ornstein [21] shows that ϵ-optimal and optimal strategies for reachability can be chosen MD:

Theorem 9 (from Theorem B in [21])**.** *For every countable MDP $\mathcal{M}$ there exist uniform ϵ-optimal MD-strategies for reachability objectives $\varphi = \mathtt{Reach}(T)$, i.e., for every $\epsilon > 0$ there is an MD-strategy σ_ϵ such that for all $s \in S$ we have $\mathcal{P}_{\mathcal{M},s,\sigma_\epsilon}(\varphi) \geq \mathtt{val}_{\mathcal{M}}(s) - \epsilon$.*

Theorem 10 (follows from Proposition B in [21])**.** *Let $\mathcal{M} = (S, S_\Box, S_\bigcirc, \longrightarrow, P)$ be an MDP, and $\varphi = \mathtt{Reach}(T)$. Let $s_0 \in S$ and σ be a strategy with $\mathcal{P}_{\mathcal{M},s_0,\sigma}(\varphi) = 1$. Then there is an MD-strategy $\hat{\sigma}$ with $\mathcal{P}_{\mathcal{M},s_0,\hat{\sigma}}(\varphi) = 1$.*

Both theorems are due to [21]; we give an alternative proof of Theorem 10 in the full version of the paper [16]. We generalize Theorem 10 to Büchi objectives, using the principle that Büchi is repeated reachability:

Proposition 11. *Let $\mathcal{M} = (S, S_\Box, S_\bigcirc, \longrightarrow, P)$ be an MDP, and $s_0 \in S$, and σ a strategy, and $Col : S \to \{1, 2\}$, and $\varphi = \mathtt{Parity}(Col)$. Suppose $\mathcal{P}_{\mathcal{M},s_0,\sigma}(\varphi) = 1$. Then there is an MD-strategy σ' with $\mathcal{P}_{\mathcal{M},s_0,\sigma'}(\varphi) = 1$.*

By appealing to Theorem 5 it follows:

Theorem 12. *Let $\mathcal{M}$ be an MDP, $Col : S \to \{1, 2\}$, and $\varphi = \mathtt{Parity}(Col)$ be a Büchi-objective (subsuming reachability and safety). Then there exists an MD-strategy σ' that is optimal for all states that have an optimal strategy:*

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

VI. When MD-strategies suffice in finitely branching MDPs

In this section we prove that optimal strategies for $\{0, 1, 2\}$-Parity, where they exist, can be chosen MD (Theorem 16) and that ϵ-optimal strategies for co-Büchi objectives can be chosen MD (Theorem 19). To prepare the ground for these results, we first consider safety objectives.

A. Optimal MD-strategies for Safety

The following proposition asserts in particular that for safety in finitely branching MDPs, there is no need for merely ϵ-optimal strategies, as there always exists an optimal MD-strategy.

Proposition 13 (from Theorem 7.3.6(a) in [23])**.** *Let $\mathcal{M} = (S, S_\Box, S_\bigcirc, \longrightarrow, P)$ be a finitely branching MDP, and $T \subseteq S$, and $\varphi = \texttt{Safety}(T)$. Define an MD-strategy $\sigma_{opt\text{-}av}$ (for "optimal avoiding") that, in each state s, picks a successor state with the largest value $\texttt{val}_\mathcal{M}(s) = \sup_{\sigma \in \Sigma} \mathcal{P}_{\mathcal{M},s,\sigma}(\varphi)$. Then for all states $s \in S$ we have $\mathcal{P}_{\mathcal{M},s,\sigma_{opt\text{-}av}}(\varphi) = \texttt{val}_\mathcal{M}(s)$, i.e., $\sigma_{opt\text{-}av}$ is uniformly optimal.*

Note that, for infinitely branching MDPs, this definition of $\sigma_{opt\text{-}av}$ would be unsound, as "the largest value" might not exist.

Definition 1. *Let $\mathcal{M} = (S, S_\Box, S_\bigcirc, \longrightarrow, P)$ be a finitely branching MDP, $Col : S \to \mathbb{N}$ a color function, $\varphi = \texttt{Safety}([S]^{Col \neq 0})$, $\sigma_{opt\text{-}av}$ the strategy from Proposition 13 and $\tau \in [0,1]$. We define*

$$\mathit{Safe}_\mathcal{M}(\tau) := \{ s \in S \mid \mathcal{P}_{\mathcal{M},s,\sigma_{opt\text{-}av}}(\varphi) \geq \tau \} ,$$

i.e., $\mathit{Safe}_\mathcal{M}(\tau)$ is the set of states from which the player can remain within color-0 states forever with probability $\geq \tau$. We drop the subscript $\mathcal{M}$ when the MDP $\mathcal{M}$ is understood.

Loosely speaking, the following lemma gives a lower bound on the probability that, starting from a "safe" state, "unsafe" states are forever avoided by $\sigma_{opt\text{-}av}$:

Lemma 14. *Let $\mathcal{M} = (S, S_\Box, S_\bigcirc, \longrightarrow, P)$ be a finitely branching MDP, $Col : S \to \mathbb{N}$ a color function and $\sigma_{opt\text{-}av}$ the strategy from Proposition 13. Let $0 < \tau_1 \leq \tau_2 \leq 1$, and $s \in \mathit{Safe}(\tau_2)$. Then $\mathcal{P}_{\mathcal{M},s,\sigma_{opt\text{-}av}}(\mathsf{G}\mathit{Safe}(\tau_1)) \geq \frac{\tau_2 - \tau_1}{1 - \tau_1}$.*

Proof. We compute probabilities conditioned under the event $\mathsf{G}\mathit{Safe}(\tau_1)$. Since $\mathit{Safe}(\tau_1) \subseteq [S]^{Col=0}$, we have $\mathcal{P}_{\mathcal{M},s,\sigma_{opt\text{-}av}}(\mathsf{G}[S]^{Col=0} \mid \mathsf{G}\mathit{Safe}(\tau_1)) = 1$. From the definition of $\mathit{Safe}(\tau_1)$ and the Markov property we get $\mathcal{P}_{\mathcal{M},s,\sigma_{opt\text{-}av}}(\mathsf{G}[S]^{Col=0} \mid \neg\mathsf{G}\mathit{Safe}(\tau_1)) \leq \tau_1$. Applying the law of total probability and writing x for $\mathcal{P}_{\mathcal{M},s,\sigma_{opt\text{-}av}}(\mathsf{G}\mathit{Safe}(\tau_1))$ we obtain:

$$\begin{aligned}
\tau_2 \;&\leq\; \mathcal{P}_{\mathcal{M},s,\sigma_{opt\text{-}av}}(\mathsf{G}[S]^{Col=0}) && \text{Def. 1}\\
&=\; \mathcal{P}_{\mathcal{M},s,\sigma_{opt\text{-}av}}(\mathsf{G}[S]^{Col=0} \mid \mathsf{G}\mathit{Safe}(\tau_1)) \cdot x \\
&\quad + \mathcal{P}_{\mathcal{M},s,\sigma_{opt\text{-}av}}(\mathsf{G}[S]^{Col=0} \mid \neg\mathsf{G}\mathit{Safe}(\tau_1)) \cdot (1 - x) \\
&\leq\; x + \tau_1 \cdot (1 - x)
\end{aligned}$$

It follows $x \geq \frac{\tau_2 - \tau_1}{1 - \tau_1}$. □

The following lemma states for all $\tau < 1$ that eventually remaining in color-0 states but outside $\mathit{Safe}(\tau)$ has probability zero.

Lemma 15. *Let $\mathcal{M} = (S, S_\Box, S_\bigcirc, \longrightarrow, P)$ be a finitely branching MDP, and $Col : S \to \mathbb{N}$ a color function. Let s be a state, and σ a strategy, and $\tau < 1$. Then $\mathcal{P}_{\mathcal{M},s,\sigma}(\mathsf{FG}\neg\mathit{Safe}(\tau) \wedge \mathsf{FG}[S]^{Col=0}) = 0$.*

B. Optimal MD-strategies for $\{0,1,2\}$-Parity

Theorem 16. *Let $\mathcal{M}$ be a finitely branching MDP, $Col : S \to \{0,1,2\}$, and $\varphi = \texttt{Parity}(Col)$. Then there exists an MD-strategy σ' that is optimal for all states that have an optimal strategy:*

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

By appealing to Theorem 5 it suffices to show:

Proposition 17. *Let $\mathcal{M} = (S, S_\Box, S_\bigcirc, \longrightarrow, P)$ be a finitely branching MDP, and $s_0 \in S$, and σ a strategy, and $Col : S \to \{0,1,2\}$, and $\varphi = \texttt{Parity}(Col)$. Suppose $\mathcal{P}_{\mathcal{M},s_0,\sigma}(\varphi) = 1$. Then there is an MD-strategy σ' with $\mathcal{P}_{\mathcal{M},s_0,\sigma'}(\varphi) = 1$.*

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

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

We are ready to prove Proposition 17.

Proof of Proposition 17. 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 $\mathcal{P}_{\mathcal{M},s,\sigma}(\varphi) = 1$ for some σ.

We will define an MD-strategy σ' with $\mathcal{P}_{\mathcal{M},s,\sigma'}(\varphi) = 1$ for all $s \in S$. We first define the MD-strategy σ' partially for the states in $\mathit{Safe}_\mathcal{M}(\frac{1}{3})$ and then extend the definition of σ' to all states. For the states in $\mathit{Safe}_\mathcal{M}(\frac{1}{3})$ define $\sigma' := \sigma_{opt\text{-}av}$ as in Proposition 13, see Figure 4. Let $\mathcal{M}'$ be the MDP obtained from $\mathcal{M}$ by restricting the transition relation as prescribed by the partial MD-strategy σ'.

For any $\tau \in [0,1]$, we have $\mathit{Safe}_\mathcal{M}(\tau) = \mathit{Safe}_{\mathcal{M}'}(\tau)$. Indeed, since $\mathcal{M}'$ restricts the options of the player, we have $\mathit{Safe}_\mathcal{M}(\tau) \supseteq \mathit{Safe}_{\mathcal{M}'}(\tau)$. Conversely, let $s \in \mathit{Safe}_\mathcal{M}(\tau)$. The strategy $\sigma_{opt\text{-}av}$ from Proposition 13 achieves $\mathcal{P}_{\mathcal{M},s,\sigma_{opt\text{-}av}}(\mathsf{G}[S]^{Col=0}) \geq \tau$. Since $\sigma_{opt\text{-}av}$ can be applied in $\mathcal{M}'$, and results in the same Markov chain as applying it in $\mathcal{M}$, we conclude $s \in \mathit{Safe}_{\mathcal{M}'}(\tau)$. This justifies to write $\mathit{Safe}(\tau)$ for $\mathit{Safe}_\mathcal{M}(\tau) = \mathit{Safe}_{\mathcal{M}'}(\tau)$ 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 $\mathcal{P}_{\mathcal{M}',s,\sigma_1}(\varphi) = 1$. This strategy σ_1 is defined as follows. First play according to a strategy σ from the statement of the theorem. If and when the play visits $\mathit{Safe}(\frac{1}{3})$, switch to the MD-strategy $\sigma_{opt\text{-}av}$ from Proposition 13. If and when the play then visits $[S]^{Col \neq 0}$, switch back to a strategy σ from the statement of the theorem, and so forth.

We show that σ_1 achieves $\mathcal{P}_{\mathcal{M}',s,\sigma_1}(\varphi) = 1$. To this end we will use Lemma 18. We partition the runs of sS^ω in three events $\mathfrak{R}_0, \mathfrak{R}_1, \mathfrak{R}_2$ as follows:

- $\mathfrak{R}_0$ contains the runs where σ_1 switches between $\sigma_{opt\text{-}av}$ and σ infinitely often.
- $\mathfrak{R}_1$ contains the runs where σ_1 eventually only plays according to $\sigma_{opt\text{-}av}$.

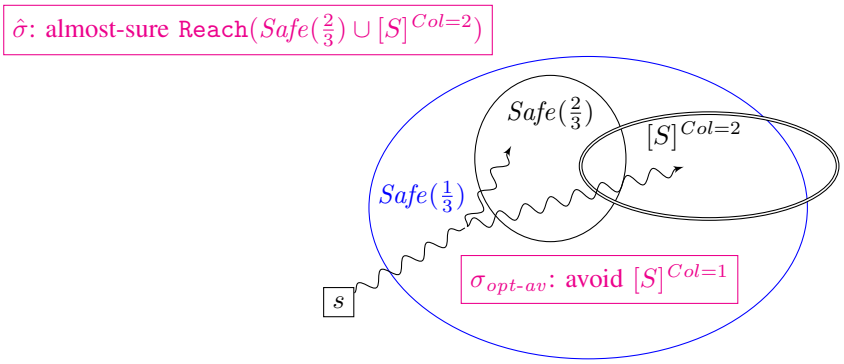


Fig. 4: The almost-surely winning MD-strategy σ' for $\{0,1,2\}$-Parity is obtained by combining the MD-strategies $\sigma_{opt\text{-}av}$ and $\hat{\sigma}$: play $\sigma_{opt\text{-}av}$ inside $Safe(\frac{1}{3})$ and $\hat{\sigma}$ outside that set. A key point is that fixing $\sigma_{opt\text{-}av}$ inside $Safe(\frac{1}{3})$ does not prevent $\hat{\sigma}$ from achieving its objective.

- $\mathfrak{R}_2$ contains the runs where σ_1 eventually only plays according to σ.

Each time σ_1 switches to $\sigma_{opt\text{-}av}$, there is, by Proposition 13, 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 $\mathcal{P}_{\mathcal{M}',s,\sigma_1}(\mathfrak{R}_0) = 0$. By the definition of $\sigma_{opt\text{-}av}$ we have $\mathfrak{R}_1 \subseteq [\![\mathsf{FG}[S]^{Col=0}]\!] \subseteq [\![\varphi]\!]$, and hence $\mathcal{P}_{\mathcal{M}',s,\sigma_1}(\mathfrak{R}_1 \cap [\![\varphi]\!]) = \mathcal{P}_{\mathcal{M}',s,\sigma_1}(\mathfrak{R}_1)$. Since $\mathcal{P}_{\mathcal{M},s,\sigma}(\varphi) = 1$ and φ is prefix-independent, we have $\mathcal{P}_{\mathcal{M}',s,\sigma_1}(\mathfrak{R}_2 \cap [\![\varphi]\!]) = \mathcal{P}_{\mathcal{M}',s,\sigma_1}(\mathfrak{R}_2)$. Using Lemma 18, we obtain $\mathcal{P}_{\mathcal{M}',s,\sigma_1}(\varphi) = 1$.

Next we show that for all $s \in S$ the strategy σ_1 defined above achieves $\mathcal{P}_{\mathcal{M}',s,\sigma_1}(\mathsf{F}Safe(\frac{2}{3}) \vee \mathsf{F}[S]^{Col=2}) = 1$. To this end we will use Lemma 18 again. We partition the runs of sS^ω into three events $\mathfrak{R}'_1, \mathfrak{R}'_2, \mathfrak{R}'_0$ as follows:

- $\mathfrak{R}'_1 = [\![\mathsf{FG}[S]^{Col=0}]\!]^s$
- $\mathfrak{R}'_2 = [\![\mathsf{GF}[S]^{Col=2}]\!]^s$
- $\mathfrak{R}'_0 = sS^\omega \setminus [\![\varphi]\!]^s$

We have previously shown that $\mathcal{P}_{\mathcal{M}',s,\sigma_1}(\varphi) = 1$, hence $\mathcal{P}_{\mathcal{M}',s,\sigma_1}(\mathfrak{R}'_0) = 0$. By Lemma 15, almost all runs in $\mathfrak{R}'_1$ satisfy $\mathsf{GF}Safe(\frac{2}{3})$. Since $[\![\mathsf{GF}Safe(\frac{2}{3})]\!] \subseteq [\![\mathsf{F}Safe(\frac{2}{3})]\!]$, we have $\mathcal{P}_{\mathcal{M}',s,\sigma_1}(\mathfrak{R}'_1 \cap [\![\mathsf{F}Safe(\frac{2}{3}) \vee \mathsf{F}[S]^{Col=2}]\!]) = \mathcal{P}_{\mathcal{M}',s,\sigma_1}(\mathfrak{R}'_1)$. Since $\mathfrak{R}'_2 \subseteq [\![\mathsf{F}[S]^{Col=2}]\!]$, we also have $\mathcal{P}_{\mathcal{M}',s,\sigma_1}(\mathfrak{R}'_2 \cap [\![\mathsf{F}Safe(\frac{2}{3}) \vee \mathsf{F}[S]^{Col=2}]\!]) = \mathcal{P}_{\mathcal{M}',s,\sigma_1}(\mathfrak{R}'_2)$. Using Lemma 18 we obtain $\mathcal{P}_{\mathcal{M}',s,\sigma_1}(\mathsf{F}Safe(\frac{2}{3}) \vee \mathsf{F}[S]^{Col=2}) = 1$.

Writing $T = Safe(\frac{2}{3}) \cup [S]^{Col=2}$ we have just shown that for all $s \in S$ there is a strategy σ_1 with $\mathcal{P}_{\mathcal{M}',s,\sigma_1}(\mathsf{F}T) = 1$. By Lemma 7 there is an MD-strategy $\hat{\sigma}$ for $\mathcal{M}'$ with $\mathcal{P}_{\mathcal{M}',s,\hat{\sigma}}(\mathsf{F}T) = 1$ 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 $\mathcal{P}_{\mathcal{M},s,\sigma'}(\mathsf{F}T) = 1$.

It remains to show that for all $s \in S$ we have $\mathcal{P}_{\mathcal{M},s,\sigma'}(\varphi) = 1$. To this end we will use Lemma 18 again. We partition the runs of sS^ω in two events $\mathfrak{R}''_1, \mathfrak{R}''_2$:

- $\mathfrak{R}''_1 = [\![\mathsf{GF}Safe(\frac{2}{3})]\!]^s$, i.e., $\mathfrak{R}''_1$ contains the runs that visit $Safe(\frac{2}{3})$ infinitely often.
- $\mathfrak{R}''_2 = [\![\mathsf{FG}\neg Safe(\frac{2}{3})]\!]^s$, i.e., $\mathfrak{R}''_2$ contains the runs that from some point on never visit $Safe(\frac{2}{3})$.

Every time a run enters $Safe(\frac{2}{3})$, by Lemma 14, the probability is at least $\frac{1}{2}$ that the run remains in $Safe(\frac{1}{3})$ forever. It follows that almost all runs in $\mathfrak{R}''_1$ eventually remain in $Safe(\frac{1}{3})$ forever, i.e., $\mathcal{P}_{\mathcal{M},s,\sigma'}(\mathfrak{R}''_1 \cap [\![\mathsf{FG}Safe(\frac{1}{3})]\!]) = \mathcal{P}_{\mathcal{M},s,\sigma'}(\mathfrak{R}''_1)$. 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 $\mathcal{P}_{\mathcal{M},s,\sigma'}(\mathfrak{R}''_1 \cap [\![\varphi]\!]) = \mathcal{P}_{\mathcal{M},s,\sigma'}(\mathfrak{R}''_1)$.

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

A final application of Lemma 18 yields $\mathcal{P}_{\mathcal{M},s,\sigma'}(\varphi) = 1$ for all $s \in S$. □

C. ϵ-Optimal MD-strategies for Co-Büchi

Theorem 19. *Let $\mathcal{M} = (S, S_\Box, S_\bigcirc, \longrightarrow, P)$ be a finitely branching MDP, $Col : S \to \{0,1\}$, and $\varphi = \mathtt{Parity}(Col)$ be the co-Büchi objective. Then there exist uniform ϵ-optimal MD-strategies. I.e., for every $\epsilon > 0$ there is an MD-strategy σ_ϵ with $\mathcal{P}_{\mathcal{M},s_0,\sigma_\epsilon}(\varphi) \geq \mathtt{val}_{\mathcal{M}}(s_0) - \epsilon$ for every $s_0 \in S$.*

Proof. Let $\epsilon_1 > 0$ be a suitably small number (to be determined later), $\tau_1 := 1 - \epsilon_1$ and $Safe_{\mathcal{M}}(\tau_1)$ defined as in Definition 1. Let $\sigma_{opt\text{-}av}$ be the MD-strategy from Proposition 13. From $\mathcal{M}$ we obtain a modified MDP $\mathcal{M}'$ by fixing all player choices from states in $Safe_{\mathcal{M}}(\tau_1)$ according to $\sigma_{opt\text{-}av}$.

We show that $\mathtt{val}_{\mathcal{M}'}(s_0) \geq \mathtt{val}_{\mathcal{M}}(s_0) - \epsilon_1$. By definition of the value $\mathtt{val}_{\mathcal{M}}(s_0)$, for every $\delta > 0$ there exists a strategy σ_δ in $\mathcal{M}$ s.t. $\mathcal{P}_{\mathcal{M},s_0,\sigma_\delta}(\varphi) \geq \mathtt{val}_{\mathcal{M}}(s_0) - \delta$. 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 $Safe_{\mathcal{M}}(\tau_1)$ is reached, play like $\sigma_{opt\text{-}av}$. This is possible, since no moves from states outside $Safe_{\mathcal{M}}(\tau_1)$ have been fixed in $\mathcal{M}'$, and all moves from states

inside $Safe_{\mathcal{M}}(\tau_1)$ have been fixed according to $\sigma_{opt\text{-}av}$. Then we have:

$$\begin{aligned}
&\mathcal{P}_{\mathcal{M}',s_0,\sigma'_\delta}(\varphi)\\
&= \mathcal{P}_{\mathcal{M},s_0,\sigma_\delta}(\varphi)\\
&\quad - \mathcal{P}_{\mathcal{M},s_0,\sigma_\delta}(\mathsf{F}Safe_{\mathcal{M}}(\tau_1)) \cdot \mathcal{P}_{\mathcal{M},s_0,\sigma_\delta}(\varphi|\mathsf{F}Safe_{\mathcal{M}}(\tau_1))\\
&\quad + \mathcal{P}_{\mathcal{M},s_0,\sigma_\delta}(\mathsf{F}Safe_{\mathcal{M}}(\tau_1)) \cdot \mathcal{P}_{\mathcal{M},s_0,\sigma'_\delta}(\varphi|\mathsf{F}Safe_{\mathcal{M}}(\tau_1))\\
&\geq \mathcal{P}_{\mathcal{M},s_0,\sigma_\delta}(\varphi)\\
&\quad - \mathcal{P}_{\mathcal{M},s_0,\sigma_\delta}(\mathsf{F}Safe_{\mathcal{M}}(\tau_1)) \cdot \mathcal{P}_{\mathcal{M},s_0,\sigma_\delta}(\varphi|\mathsf{F}Safe_{\mathcal{M}}(\tau_1))\\
&\quad + \mathcal{P}_{\mathcal{M},s_0,\sigma_\delta}(\mathsf{F}Safe_{\mathcal{M}}(\tau_1)) \cdot \tau_1\\
&\geq \mathtt{val}_{\mathcal{M}}(s_0) - \delta - \mathcal{P}_{\mathcal{M},s_0,\sigma_\delta}(\mathsf{F}Safe_{\mathcal{M}}(\tau_1))(1-\tau_1)\\
&\geq \mathtt{val}_{\mathcal{M}}(s_0) - \delta - \epsilon_1
\end{aligned}$$

Since this holds for every $\delta > 0$ we obtain $\mathtt{val}_{\mathcal{M}'}(s_0) \geq \mathtt{val}_{\mathcal{M}}(s_0) - \epsilon_1$.

Now let $\tau_2 := 1 - \epsilon_1/k$ for a suitably large $k \geq 1$ (to be determined later) and $Safe_{\mathcal{M}'}(\tau_2)$ be defined as in Definition 1. In particular, $Safe_{\mathcal{M}'}(\tau_2) = Safe_{\mathcal{M}}(\tau_2)$ (by the same argument as in the proof of Proposition 17).

By definition of the value, for every $\epsilon_2 > 0$ there exists a strategy σ_{ϵ_2} in $\mathcal{M}'$ with $\mathcal{P}_{\mathcal{M}',s_0,\sigma_{\epsilon_2}}(\varphi) \geq \mathtt{val}_{\mathcal{M}'}(s_0) - \epsilon_2$. Moreover, by Lemma 15 and $\tau_2 < 1$, $\mathcal{P}_{\mathcal{M}',s_0,\sigma}(\mathsf{F}Safe_{\mathcal{M}'}(\tau_2)) \geq \mathcal{P}_{\mathcal{M}',s_0,\sigma}(\varphi)$ for every strategy σ and thus in particular for σ_{ϵ_2}. Therefore, $\mathcal{P}_{\mathcal{M}',s_0,\sigma_{\epsilon_2}}(\mathsf{F}Safe_{\mathcal{M}'}(\tau_2)) \geq \mathtt{val}_{\mathcal{M}'}(s_0) - \epsilon_2$. By Theorem 9, for every $\epsilon_3 > 0$ there exists an MD-strategy σ' in $\mathcal{M}'$ with $\mathcal{P}_{\mathcal{M}',s_0,\sigma'}(\mathsf{F}Safe_{\mathcal{M}'}(\tau_2)) \geq \mathtt{val}_{\mathcal{M}'}(s_0) - \epsilon_2 - \epsilon_3$. In particular, σ' must coincide with $\sigma_{opt\text{-}av}$ at all states in $Safe_{\mathcal{M}}(\tau_1)$, since in $\mathcal{M}'$ these choices are already fixed.

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 σ' at all states outside $Safe_{\mathcal{M}}(\tau_1)$ and like $\sigma_{opt\text{-}av}$ at all states inside $Safe_{\mathcal{M}}(\tau_1)$.

In order to show that σ_ϵ has the required property $\mathcal{P}_{\mathcal{M},s_0,\sigma_\epsilon}(\varphi) \geq \mathtt{val}_{\mathcal{M}}(s_0) - \epsilon$, we first estimate the probability that a play according to σ_ϵ will never leave the set $Safe_{\mathcal{M}}(\tau_1)$ after having visited a state in $Safe_{\mathcal{M}'}(\tau_2)$.

Let $s \in Safe_{\mathcal{M}'}(\tau_2)$. Then, by Lemma 14,

$$\begin{aligned}
\mathcal{P}_{\mathcal{M},s,\sigma_{opt\text{-}av}}(\mathsf{G}Safe(\tau_1)) &\geq \frac{\tau_2 - \tau_1}{1 - \tau_1}\\
&= \frac{(1 - \epsilon_1/k) - (1 - \epsilon_1)}{\epsilon_1}\\
&= 1 - \frac{1}{k}\,.
\end{aligned}$$

In particular we also have $\mathcal{P}_{\mathcal{M},s,\sigma_\epsilon}(\mathsf{G}Safe(\tau_1)) \geq 1 - \frac{1}{k}$, since σ_ϵ coincides with $\sigma_{opt\text{-}av}$ inside the set $Safe_{\mathcal{M}}(\tau_1)$. Finally we obtain:

$$\begin{aligned}
\mathcal{P}_{\mathcal{M},s_0,\sigma_\epsilon}(\varphi) &\geq \mathcal{P}_{\mathcal{M},s_0,\sigma_\epsilon}(\mathsf{F}Safe_{\mathcal{M}'}(\tau_2))\\
&\quad \cdot \mathcal{P}_{\mathcal{M},s_0,\sigma_\epsilon}(\mathsf{FG}Safe_{\mathcal{M}}(\tau_1)|\mathsf{F}Safe_{\mathcal{M}'}(\tau_2))\\
&\geq \mathcal{P}_{\mathcal{M}',s_0,\sigma'}(\mathsf{F}Safe_{\mathcal{M}'}(\tau_2)) \cdot (1 - 1/k)\\
&\geq (\mathtt{val}_{\mathcal{M}'}(s_0) - \epsilon_2 - \epsilon_3) \cdot (1 - 1/k)\\
&\geq (\mathtt{val}_{\mathcal{M}}(s_0) - \epsilon_1 - \epsilon_2 - \epsilon_3) \cdot (1 - 1/k)
\end{aligned}$$

This holds for every $\epsilon_1, \epsilon_2, \epsilon_3 > 0$ and every $k \geq 1$, and moreover $\mathtt{val}_{\mathcal{M}}(s_0) \leq 1$. Thus we can set $\epsilon_1 = \epsilon_2 = \epsilon_3 := \epsilon/6$ and $k := \frac{2}{\epsilon}$ and obtain $\mathcal{P}_{\mathcal{M},s_0,\sigma_\epsilon}(\varphi) \geq \mathtt{val}_{\mathcal{M}}(s_0) - \epsilon$ for every $s_0 \in S$ as required. □

VII. Discussion

Our results on the memory requirements of (ϵ)-optimal strategies (Figure 1) directly imply how much memory is needed to win quantitative objectives of type $\left[\varphi\right]^{\rhd c}$ (considered, e.g., in [6]). For $c < 1$ the assumed winning strategy might have to be an ϵ-optimal one, since optimal strategies do not always exist. Thus MD-strategies are only sufficient for reachability objectives in countable MDPs (resp., for $\{0,1\}$-Parity, safety and reachability objectives in finitely branching MDPs). In the special case of $\left[\varphi\right]^{\geq 1}$ objectives (i.e., winning almost-surely), the winning strategy (assuming it exists) must be optimal. Thus MD-strategies are only sufficient for safety, reachability and $\{1,2\}$-Parity in countable MDPs (resp., for all objectives subsumed by $\{0,1,2\}$-Parity in finitely branching MDPs).

In this paper we have studied countable MDPs. Not all our results carry over to uncountable MDPs. The first issue is measurability. The probabilities are only well-defined if the strategies are measurable functions, which might not exist without further conditions on the MDP; cf. Section 2.3 in [23]. Another issue is that strategies cannot generally be chosen *uniform*, i.e., independent of the initial state. E.g., in countable MDPs ϵ-optimal strategies for reachability can be chosen uniform MD (Theorem 9), but this does not carry over to uncountable MDPs (Thm. A in [21]). However, optimal strategies for reachability, if they exist, can be chosen uniform MD (Proposition B in [21]).

Acknowledgements. This work was partially supported by the EPSRC through grants EP/M027287/1, EP/M027651/1, EP/P020909/1 and EP/M003795/1 and by St. John's College, Oxford.

References

[1] P.A. Abdulla, R. Ciobanu, R. Mayr, A. Sangnier, and J. Sproston. Qualitative analysis of VASS-induced MDPs. In *Proc. of FOSSACS 2016*, volume 9634 of *LNCS*, 2016.

[2] C. Baier, N. Bertrand, and Ph. Schnoebelen. Verifying nondeterministic probabilistic channel systems against omega-regular linear-time properties. *ACM Transactions on Computational Logic*, 9, 2007.

[3] N. Berger, N. Kapur, L. J. Schulman, and V. V. Vazirani. Solvency games. In Ramesh Hariharan, Madhavan Mukund, and V. Vinay, editors, *IARCS Annual Conference on Foundations of Software Technology and Theoretical Computer Science, FSTTCS 2008, December 9-11, 2008, Bangalore, India*, pages 61–72, 2008.

[4] P. Billingsley. *Probability and Measure*. Wiley, New York, NY, 1995. Third Edition.

[5] T. Brázdil, V. Brožek, K. Etessami, A. Kučera, and D. Wojtczak. One-counter Markov decision processes. In *SODA'10*, pages 863–874. SIAM, 2010.

[6] T. Brázdil, V. Brožek, A. Kučera, and J. Obdrzálek. Qualitative reachability in stochastic BPA games. *Information and Computation*, 209, 2011.

[7] T. Brázdil, V. Brožek, V. Forejt, and A. Kučera. Reachability in recursive Markov decision processes. *Information and Computation*, 206(5):520–537, 2008.

[8] K. Chatterjee, L. de Alfaro, and T. Henzinger. Trading memory for randomness. In *Proceedings of the First Annual Conference on Quantitative Evaluation of Systems (QEST)*, pages 206–217. IEEE Computer Society Press, 2004.

[9] K. Chatterjee and T. Henzinger. A survey of stochastic ω-regular games. *Journal of Computer and System Sciences*, 78(2):394–413, 2012.

[10] K. Chatterjee, M. Jurdziński, and T. Henzinger. Quantitative stochastic parity games. In *Proceedings of the Fifteenth Annual ACM-SIAM Symposium on Discrete Algorithms*, SODA '04, pages 121–130, Philadelphia, PA, USA, 2004. Society for Industrial and Applied Mathematics.

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

[12] A. Condon. The complexity of stochastic games. *Information and Computation*, 96(2):203–224, 1992.

[13] K. Etessami and M. Yannakakis. Recursive Markov decision processes and recursive stochastic games. In *ICALP'05*, volume 3580 of *LNCS*, pages 891–903. Springer, 2005.

[14] W. Feller. *An Introduction to Probability Theory and Its Applications*, volume 1. Wiley & Sons, second edition, 1966.

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

[16] 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`.

[17] M.Y. Kitaev and V.V. Rykov. *Controlled queueing system*. CRC press, 1995.

[18] J. Krčál. Determinacy and Optimal Strategies in Stochastic Games. Master's thesis, Masaryk University, School of Informatics, Brno, Czech Republic, 2009.

[19] A. Kučera. Turn-based stochastic games. In Krzysztof R. Apt and Erich Grädel, editors, *Lectures in Game Theory for Computer Scientists*. Cambridge University Press, 2011.

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

[21] D. Ornstein. On the existence of stationary optimal strategies. *Proc. Am. Math. Soc.*, 20:563–569, 1969.

[22] S.P. Pliska. Optimization of multitype branching processes. *Management Science*, 23(2):117–124, 1976.

[23] M. L. Puterman. *Markov Decision Processes*. Wiley, 1994.

[24] L.S. Shapley. Stochastic games. *Proceedings of the national academy of sciences*, 39(10):1095–1100, 1953.

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