Stochastic Shortest Paths and Weight-Bounded Properties in Markov Decision Processes

Christel Baier
Technische Universität Dresden, Germany

Nathalie Bertrand
Univ Rennes, Inria, CNRS, IRISA, France

Clemens Dubslaff
Daniel Gburek
Technische Universität Dresden, Germany

Ocan Sankur
Univ Rennes, Inria, CNRS, IRISA, France

Abstract

The paper deals with finite-state Markov decision processes (MDPs) with integer weights assigned to each state-action pair. New algorithms are presented to classify end components according to their limiting behavior with respect to the accumulated weights. These algorithms are used to provide solutions for two types of fundamental problems for integer-weighted MDPs. First, a polynomial-time algorithm for the classical stochastic shortest path problem is presented, generalizing known results for special classes of weighted MDPs. Second, qualitative probability constraints for weight-bounded (repeated) reachability conditions are addressed. Among others, it is shown that the problem to decide whether a disjunction of weight-bounded reachability conditions holds almost surely under some scheduler belongs to NP ∩ coNP, is solvable in pseudo-polynomial time and is at least as hard as solving two-player mean-payoff games, while the corresponding problem for universal quantification over schedulers is solvable in polynomial time.

ACM Reference Format:
Christel Baier, Nathalie Bertrand, Clemens Dubslaff, Daniel Gburek, and Ocan Sankur. 2018. Stochastic Shortest Paths and Weight-Bounded Properties in Markov Decision Processes. In *LICS '18: 33rd Annual ACM/IEEE Symposium on Logic in Computer Science, July 9–12, 2018, Oxford, United Kingdom.* ACM, New York, NY, USA, 9 pages. https://doi.org/10.1145/3209108.3209184

1 Introduction

Markov decision processes (MDPs) are a prominent model used, *e.g.*, in operations research, artificial intelligence, robotics and the formal analysis of probabilistic nondeterministic programs. Various types of stochastic shortest (or longest) path problems can be formalized as an optimization problem for MDPs with integer or rational weights for the transitions where the task is to determine an optimal scheduling policy for the MDP until reaching a target. Here, optimality is understood with respect to the expected accumulated weight or the probability of reaching the target under weight constraints. Such problems can be seen as a control-synthesis problem that, *e.g.*, asks to implement a decision-making routine for a robot so that the robot eventually reaches a safe state almost surely, while providing guarantees on the achieved utility.

Stochastic shortest (or longest) path problems are well understood and supported by various tools for finite-state MDPs with nonnegative weights only, for which the algorithms can rely on the monotonicity of accumulated weights along the prefixes of paths. In this case, schedulers that maximize or minimize the expected accumulated weight until reaching the target can be determined in polynomial time based on a preprocessing of end components (*i.e.*, strongly connected sub-MDPs) and linear programs [5, 12]. One can compute schedulers maximizing the probability for reaching the target within a given cost in pseudo-polynomial time using an iterative approach that successively increases the weight bound and treats zero-weight loops by linear-programming techniques [3, 19]. The corresponding decision problem is PSPACE-hard, even for acyclic MDPs [14].

For MDPs with arbitrary integer weights, the lack of monotonicity of accumulated weights makes analogous questions much harder. Even for finite-state Markov chains with integer weights, the set of relevant configurations (*i.e.*, states augmented with the weight that has been accumulated so far) can be infinite and, in MDPs with integer weights optimal or ε-optimal schedulers might require an infinite amount of memory. The latter is known from energy-MDPs [7, 9, 17] where one aims at finding a scheduler under which the system never runs out of energy (*i.e.*, the accumulated weight plus some initial credit is always positive) and satisfies an ω-regular property (*e.g.*, a parity condition) with probability 1 or maximizes the expected mean payoff. Another indication for the additional difficulties that arise when switching from nonnegative weights to integers is given by the work on one-counter MDPs [6], which can be seen as MDPs where all weights are in $\{-1, 0, +1\}$ and that terminate as soon as the counter value is 0. Among others, [6] establishes PSPACE-hardness and an EXPTIME upper bound for the almost-sure termination problem under some scheduler, while the corresponding weight-bounded (control-state) reachability problem in nonnegative MDPs is in P [19].

This paper addresses several fundamental problems for MDPs with integer weights. Our main contributions are as follows. First, we show that the classical stochastic shortest path problem, where the task is to *minimize the expected weight* until reaching a target, is solvable in polynomial time for arbitrary integer-weighted MDPs. We hereby extend previous results for restricted classes of MDPs [5, 12], while the general case was open. Second, we study

The authors are partly supported by the DFG through the collaborative research centre HAEC (SFB 912), the Excellence Initiative by the German Federal and State Governments (cluster of excellence cfAED), the Research Training Group QuantLA (GRK 1763), and the DFG-project BA-1679/11-1. The collaboration is supported by Inria associate team programme.

LICS '18, July 9–12, 2018, Oxford, United Kingdom

ACM ISBN 978-1-4503-5583-4/18/07…$15.00
https://doi.org/10.1145/3209108.3209184

disjunctions of *weight-bounded reachability conditions* with qualitative probability bounds and existential or universal scheduler quantification. The problem to check the existence of a scheduler satisfying a disjunction of weight-bounded reachability conditions almost surely (referred to as decision problem $\mathrm{DWR}^{\exists,=1}$) is shown to be in NP∩coNP, solvable in pseudo-polynomial time, and as hard as non-stochastic two-player mean-payoff games (and therefore not known to be in P). The same complexity results are achieved for checking whether a disjunction of weight-bounded reachability conditions holds with positive probability under all schedulers (problem $\mathrm{DWR}^{\forall,>0}$). In contrast, problem $\mathrm{DWR}^{\forall,=1}$ that asks whether a disjunctive weight-bounded reachability condition holds almost surely under all schedulers is shown to be in P. We also present algorithms for computing optimal weight-bounds with analogous time complexities: pseudo-polynomial for the optimization variants of $\mathrm{DWR}^{\exists,=1}$ and $\mathrm{DWR}^{\forall,>0}$ and polynomial for $\mathrm{DWR}^{\forall,=1}$. These results should be contrasted with the polynomial-time decidability of $\mathrm{DWR}^{\exists,=1}$ and $\mathrm{DWR}^{\forall,>0}$ for MDPs where all weights are nonnegative [19].

Although several other problems for integer-weighted MDPs are known to be in NP ∩ coNP and as hard as nonstochastic two-player mean-payoff games (see, *e.g.*, [8, 9, 17] and the discussion on related work in Section 5.3), our techniques crucially depart from previous work by heavily relying on new algorithms to classify end components (ECs) of MDPs. We see these results on the *classification of ECs* as a further main contribution as it provides a useful vehicle for reasoning about different problems for integer-weighted MDPs. An indication for the latter is that we use these classification algorithms not only to establish the results listed above for $\mathrm{DWR}^{\exists,=1}$ and $\mathrm{DWR}^{\forall,=1}$, but also to prove the polynomial-time solvability of the classical shortest path problem in general integer-weighted MDPs and to deal with weight-bounded Büchi conditions.

Our classification of ECs is according to the existence of schedulers that increase the weight to infinity (*pumping ECs*), or ensure that the weight eventually exceeds any threshold possibly without converging to $+\infty$ (*weight-divergent ECs*), or have oscillating behavior (*gambling ECs*), or keep the accumulated weights within a compact interval (*bounded ECs*). A sufficient and necessary criterion for the pumping property is that the maximal expected mean payoff is positive, which is decidable in polynomial time by computing the maximal expected mean payoff using linear-programming techniques [15, 18]. While this observation has been made by several other authors, we are not aware of earlier algorithms for checking the gambling or boundedness property. For checking weight-divergence, the results of [6] for one-counter MDPs without boundary yield a polynomial time bound for the special case of MDPs where all weights are in $\{+1, 0, -1\}$ and a pseudo-polynomial time bound in the general case. We improve this result by presenting a polynomial-time algorithm for deciding weight-divergence for MDPs with arbitrary integer weights. Moreover, in case that the given MDP $\mathcal{M}$ is not weight-divergent, the algorithm generates a new MDP $\mathcal{N}$ with the same state space that has no 0-ECs (*i.e.*, end components where the accumulated weight of all cycles is 0) and that is equivalent to $\mathcal{M}$ for all properties that are invariant with respect to behaviors inside 0-ECs. The generation of such an MDP $\mathcal{N}$ relies on an iterative technique to flatten 0-ECs. This new technique, called *spider construction*, can be seen as a generalization of the method proposed in [11, 12] to eliminate 0-ECs in nonnegative MDPs. There, all states that belong to some maximal end component of the sub-MDP built by state-action pairs with weight 0 are collapsed. This technique obviously fails for integer-weighted MDPs as 0-ECs can contain state-action pairs with negative and positive weights. The spider construction maintains the state space, but turns the graph structure of maximal 0-ECs into an acyclic graph with a single sink state that captures the original behavior of all other states in the same maximal 0-EC. Besides deciding weight-divergence, the spider construction will be the key to solve the classical shortest path problem for arbitrary integer-weighted MDPs.

Checking the gambling property is NP-complete in the general case, but can be decided in polynomial time using the spider construction, provided that the maximal expected mean payoff is 0. The latter is the relevant case for solving problems $\mathrm{DWR}^{\exists,=1}$ and $\mathrm{DWR}^{\forall,=1}$ as well as corresponding problems for weight-bounded Büchi conditions. We establish an analogous result for the boundedness property, shown to be equivalent to the existence of 0-ECs in cases where the given end component has maximal expected mean payoff 0.

Outline. Section 3 presents the classification of end components and corresponding algorithms. Our results on the stochastic shortest path problem and weight-bounded (repeated) reachability properties will be presented in Sections 4 and 5, respectively. For full proofs we refer to the extended version of this paper [2].

2 Preliminaries

We briefly define our notations. For more details see, *e.g.*, [4, 18].

Definition 2.1 (Markov decision processes (MDPs).)**.** An *MDP* is a tuple $\mathcal{M} = (S, Act, P, wgt)$ where S is a finite set of states, Act is a finite set of actions, $P\colon S \times Act \times S \to [0,1] \cap \mathbb{Q}$ is a probabilistic transition function satisfying $\sum_{t\in S} P(s,\alpha,t) \in \{0,1\}$ for all $(s,\alpha) \in S \times Act$, and $wgt\colon S \times Act \to \mathbb{Z}$ is a weight function.

Action α is *enabled* in s if $\sum_{t\in S} P(s,\alpha,t) = 1$, in which case (s,α) is called a *state-action pair* of $\mathcal{M}$. $Act(s)$ denotes the set of actions enabled in s. State s is called a *trap* if $Act(s) = \varnothing$.

Let $\|\mathcal{M}\|$ denote the number of state-action pairs in $\mathcal{M}$. The *size* of MDP $\mathcal{M}$ is $\|\mathcal{M}\|$ plus the sum of the logarithmic lengths of the probabilities and weights in $\mathcal{M}$.

A *path* in an MDP $\mathcal{M} = (S, Act, P, wgt)$ is an alternating sequence of states and actions, that can be finite $\pi = s_0\,\alpha_0\,s_1\,\alpha_1\,s_2 \ldots s_n$ or infinite $\varsigma = s_0\,\alpha_0\,s_1\,\alpha_1\,s_2\,\alpha_2\ldots$, such that for every index i, $\alpha_i \in Act(s_i)$ and $P(s_i,\alpha_i,s_{i+1}) > 0$. A path is called *maximal* if it is infinite or ends in a trap. *FPaths*, *IPaths* and *MPaths* denote the set of finite, infinite and maximal paths, respectively. The *weight* of a finite path $\pi = s_0\,\alpha_0\,s_1\,\alpha_1\,\ldots\,\alpha_{n-1}\,s_n$ is $wgt(\pi) = \sum_{i=0}^{n-1} wgt(s_i,\alpha_i)$. For any path $\pi = s_0\,\alpha_0\,s_1\,\alpha_1\,s_2\,\alpha_2\ldots$, we write $pref(\pi,i)$ for its prefix up to state s_i. The first (resp. last) state of a finite path π is denoted $first(\pi)$ (resp. $last(\pi)$). If ς is infinite, $\lim(\varsigma)$ is the set of state-action pairs occurring infinitely often in ς.

A *scheduler* resolves nondeterminism in MDPs. Formally, a scheduler for $\mathcal{M}$ is a partial function $\mathfrak{S}\colon FPaths \to Distr(Act)$ that maps every finite path π where $t = last(\pi)$ is not a trap to a distribution over $Act(t)$. Given a scheduler $\mathfrak{S}$ and a state s, the behavior of $\mathcal{M}$ under $\mathfrak{S}$ with starting state s can be formalized by a (possibly infinite-state) Markov chain. $\Pr^{\mathfrak{S}}_{\mathcal{M},s}$ denotes the induced probability measure. We use standard notions for deterministic, memoryless,

finite- and infinite-memory schedulers. Thus, memoryless deterministic (MD) schedulers can be viewed as functions assigning actions to non-trap states and the induced Markov chain is finite.

The analysis of the behaviors in MDPs often relies on their end components. An *end component* of $\mathcal{M}$ is a pair $\mathcal{E} = (T, \mathfrak{A})$ consisting of a set of states $T \subseteq S$ and a function $\mathfrak{A} : T \to 2^{Act}$ such that (1) $\emptyset \neq \mathfrak{A}(s) \subseteq Act(s)$ for each $s \in T$, (2) $\{t \in S : P(s, \alpha, t) > 0\} \subseteq T$ for each $s \in T$ and $\alpha \in \mathfrak{A}(s)$, and (3) the sub-MDP induced by $(T, \mathfrak{A})$ is strongly connected. We often identify end components with their sets of state-action pairs. That is, if $\mathcal{E} = (T, \mathfrak{A})$ is as above, we identify $\mathcal{E}$ with the set $\{(t, \alpha) : t \in T, \alpha \in \mathfrak{A}(t)\}$ and rely on the fact that for each scheduler the limit $\lim(\varsigma)$ of almost all infinite $\mathfrak{S}$-paths ς constitutes an end component [11]. $\mathcal{E}$ is a *maximal end component* (MEC) if there is no end component $\mathcal{F}$ such that $\mathcal{E}$ is strictly contained in $\mathcal{F}$. MECs of an MDP are computable in polynomial time [10, 11]. All notations introduced for MDPs can be used for end components, which are themselves strongly connected MDPs.

Specifying properties. We use the term *properties* to denote measurable subsets of $(S \times \mathbb{Z})^\omega \cup (S \times \mathbb{Z})^* \times S$ with respect to the standard cylindrical sigma-algebra. To reason about probabilities of properties concerning the measure $\Pr^{\mathfrak{S}}_{\mathcal{M},s}$ where $\mathfrak{S}$ is a scheduler and s is a starting state, every path (state-action sequence) in $\mathcal{M}$ is naturally mapped to a state-integer sequence. Temporal properties with weight constraints will be described by LTL-like formulas. The atoms of such formulas are (sets of) states or weight expressions of the form $\mathrm{wgt} \bowtie w$ where $\bowtie \in \{\leqslant, <, \geqslant, >, =\}$ is a comparison operator and $w \in \mathbb{Z}$ is a threshold. Such formulas are interpreted over path-position pairs. More precisely, given a path $\varsigma = s_0\, \alpha_0\, s_1\, \alpha_1\, s_2\, \alpha_2 \ldots$ in $\mathcal{M}$ and $i \in \mathbb{N}$ $(\varsigma, i) \models \mathrm{wgt} \bowtie w$ iff $wgt(pref(\varsigma, i)) \bowtie w$, and as usual, $\varsigma \models \varphi$ is a shortcut for $(\varsigma, 0) \models \varphi$. Towards an example, let *goal* be a state in $\mathcal{M}$. Then $\varsigma \models \Diamond(goal \wedge (\mathrm{wgt} \geqslant w))$ iff ς has a finite prefix π such that $last(\pi) = goal$ and $wgt(\pi) \geqslant w$.

To reason about optimal probabilities of a property φ, let

$$\Pr^{\sup}_{\mathcal{M},s}(\varphi) = \sup_{\mathfrak{S}} \Pr^{\mathfrak{S}}_{\mathcal{M},s}(\varphi) \quad \text{and} \quad \Pr^{\inf}_{\mathcal{M},s}(\varphi) = \inf_{\mathfrak{S}} \Pr^{\mathfrak{S}}_{\mathcal{M},s}(\varphi)$$

where $\mathfrak{S}$ ranges over all schedulers for $\mathcal{M}$. We write $\Pr^{\max}_{\mathcal{M},s}(\varphi)$ rather than $\Pr^{\sup}_{\mathcal{M},s}(\varphi)$ if the supremum is indeed a maximum. This is the case, *e.g.*, if φ is an ordinary LTL formula (without weight constraints). Note that the maximum/minimum might not exist for weight-bounded properties.
In any case, $\Pr^{\max}_{\mathcal{M},s}(\varphi) = 1$ (resp. $\Pr^{\max}_{\mathcal{M},s}(\varphi) > 0$) indicates the existence of a scheduler $\mathfrak{S}$ with $\Pr^{\mathfrak{S}}_{\mathcal{M},s}(\varphi) = 1$ (resp. $\Pr^{\mathfrak{S}}_{\mathcal{M},s}(\varphi) > 0$).

Given a random variable f,

$$\mathbb{E}^{\sup}_{\mathcal{M},s}(f) = \sup_{\mathfrak{S}} \mathbb{E}^{\mathfrak{S}}_{\mathcal{M},s}(f) \text{ and } \mathbb{E}^{\inf}_{\mathcal{M},s}(f) = \inf_{\mathfrak{S}} \mathbb{E}^{\mathfrak{S}}_{\mathcal{M},s}(f)$$

denote the extremal expectations of f, where sup and inf take values in $\mathbb{R} \cup \{-\infty, +\infty\}$, while, for instance, $\mathbb{E}^{\max}_{\mathcal{M},s}(f)$ will be used when the maximum exists. In particular, we will use the random variable associated with the mean payoff, defined on infinite paths by $\mathrm{MP}(\varsigma) = \limsup_{n\to\infty} \frac{wgt(pref(\varsigma,n))}{n}$. Recall that the maximal expected mean payoff in strongly connected MDPs does not depend on the starting state and that there exist MD-schedulers with a single *bottom strongly connected component* (BSCC) maximizing the expected mean payoff. When $\mathcal{M}$ is strongly connected, we omit the starting state and write $\mathbb{E}^{\max}_{\mathcal{M}}(\mathrm{MP})$.

3 Classification of End Components

As basic building blocks of our algorithms, we define four types of schedulers and end components of MDPs. The *pumping* end components have a scheduler that let the accumulated weight almost surely diverge to infinity; positively (resp. negatively) *weight-divergent* ones have a scheduler where almost surely the limsup (resp. liminf) of the accumulated sum is infinity (resp. minus infinity); the *gambling* ones have schedulers with expected mean payoff 0 and where the accumulated weight approaches both plus and minus infinity with probability 1; while the *zero end components* only have 0 cycles, so the weight stays bounded with probability 1.

Definition 3.1. An infinite path ς in an MDP $\mathcal{M}$ is called

- *pumping* if $\liminf_{n\to\infty} wgt(pref(\varsigma, n)) = +\infty$,
- *positively weight-divergent*, or briefly *weight-divergent*, if $\limsup_{n\to\infty} wgt(pref(\varsigma, n)) = +\infty$,
- *negatively weight-divergent* if $\liminf_{n\to\infty} wgt(pref(\varsigma, n)) = -\infty$,
- *gambling* if ς is positively and negatively weight-divergent,
- *bounded from below* if $\liminf_{n\to\infty} wgt(pref(\varsigma, n)) \in \mathbb{Z}$.

A scheduler $\mathfrak{S}$ for $\mathcal{M}$ is called *pumping from state* s if $\Pr^{\mathfrak{S}}_{\mathcal{M},s}\{\varsigma \in IPaths : \varsigma \text{ is pumping}\} = 1$, *i.e.*, almost all $\mathfrak{S}$-paths from s are pumping. $\mathfrak{S}$ is called *pumping* if it is pumping from all states s. The MDP $\mathcal{M}$ itself is said to be *pumping* if it has at least one pumping scheduler. $\mathcal{M}$ is called *universally pumping* if all schedulers of $\mathcal{M}$ are pumping.

The notions of weight-divergent (or negatively weight-divergent or bounded from below) schedulers and MDPs are defined analogously. *Gambling* schedulers are those where almost all paths are gambling and where the expected mean payoff is 0. A strongly connected MDP $\mathcal{M}$ is called *gambling* if $\mathbb{E}^{\max}_{\mathcal{M}}(\mathrm{MP}) = 0$ and $\mathcal{M}$ has a gambling scheduler (see Fig. 1).

Obviously, a strongly connected MDP $\mathcal{M}$ is pumping (universal pumping or weight-divergent or gambling, respectively) from some state iff $\mathcal{M}$ is pumping (universal pumping or weight-divergent or gambling, respectively).

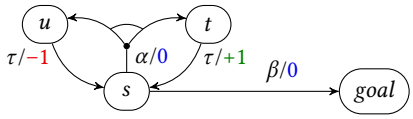

Figure 1. EC $\mathcal{E} = \{(s, \alpha), (u, \tau), (t, \tau)\}$ is gambling in case all distributions are uniform. The MD scheduler that always takes (s, α) is gambling. Moreover, *goal* can be reached almost surely for any weight threshold, using the infinite-memory scheduler that takes (s, α) if below the threshold, and (s, β) otherwise. One can show that this cannot be achieved with a finite-memory scheduler.

A *zero end component* (0-EC) is an end component $\mathcal{E}$ where $wgt(\xi) = 0$ for each cycle ξ in $\mathcal{E}$ and use the term *0-BSCC* when $\mathcal{E}$ contains at most one state-action pair (s, α) for each state s in $\mathcal{E}$. Thus, each 0-BSCC is a bottom strongly connected component of an MD-scheduler. A cycle ξ in $\mathcal{M}$ is called positive if $wgt(\xi) > 0$, and negative if $wgt(\xi) < 0$. Recall characterizations of these notions for Markov chains:

Lemma 3.2 (Folklore – see, *e.g.*, [16])**.** *Let C be a strongly connected finite Markov chain.*

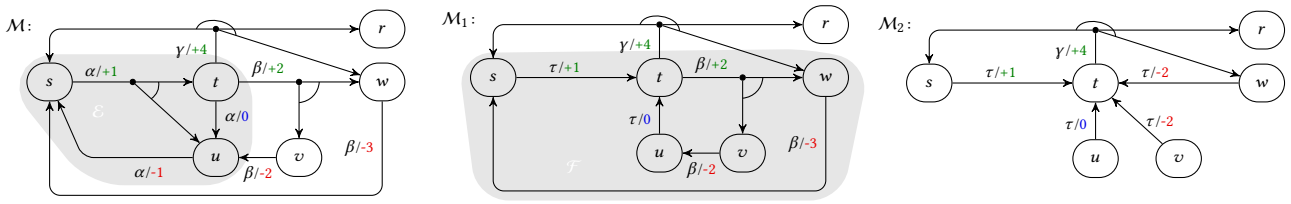

Figure 2. Illustration of the spider construction: $\mathcal{M}_1 = \mathsf{Spider}_{\mathcal{E},t}(\mathcal{M})$ and $\mathcal{M}_2 = \mathsf{Spider}_{\mathcal{F},t}(\mathcal{M}_1)$.

(a) C is pumping iff $\mathbb{E}_C(\mathrm{MP}) > 0$.
(b) $\mathbb{E}_C(\mathrm{MP}) = 0$ iff C is a 0-BSCC or C is gambling.
(c) If $\mathbb{E}_C(\mathrm{MP}) = 0$ then the following statements are equivalent: (1) C is gambling, (2) C is positively weight-divergent, (3) C is negatively weight-divergent, (4) C has a positive cycle, (5) C has a negative cycle.
(d) If $\mathbb{E}_C(\mathrm{MP}) = 0$ then the following are equivalent: (1) C is a 0-BSCC, (2) C is bounded from below, (3) the set of paths bounded from below has positive measure.

The goal of this section is to provide an analogous characterization for strongly connected MDPs and efficient algorithms to decide whether an MDP is of a given type.

This is simple for the existential and universal pumping property, checkable in polynomial time :

Lemma 3.3. *Let $\mathcal{M}$ be a strongly connected MDP. Then, $\mathcal{M}$ is pumping iff $\mathcal{M}$ has a pumping MD-scheduler iff $\mathbb{E}^{\max}_{\mathcal{M}}(\mathrm{MP}) > 0$. Likewise, $\mathcal{M}$ is universally pumping iff all MD-schedulers are pumping iff $\mathbb{E}^{\min}_{\mathcal{M}}(\mathrm{MP}) > 0$.*

The remainder of this section addresses the tasks to check weight-divergence, the gambling property and the computation of all states belonging to a 0-EC.[1] We start with an observation on weight-divergence:

Lemma 3.4. *Let $\mathcal{M}$ be a strongly connected MDP. If $\mathcal{M}$ is positively weight-divergent then $\mathbb{E}^{\max}_{\mathcal{M}}(\mathrm{MP}) \geqslant 0$. Conversely, if $\mathbb{E}^{\max}_{\mathcal{M}}(\mathrm{MP}) > 0$, then $\mathcal{M}$ is positively weight-divergent.*

3.1 Spider Construction for Flattening 0-ECs

In this section, we present a method to eliminate a given 0-EC from an MDP by "flattening" it, crucial for our algorithms. This so-called *spider construction* preserves the state space and all properties of interest, in particular, those that are invariant by adding or removing path segments of weight 0. It will be used for checking weight-divergence (Section 3.2) and for the stochastic shortest path algorithm (Section 4).

Let $\mathcal{M}$ be an MDP and $\mathcal{E}$ a 0-BSCC of $\mathcal{M}$, *i.e.*, for each state s in $\mathcal{E}$ there is a unique action $\alpha_s \in Act(s)$ such that $(s, \alpha_s) \in \mathcal{E}$. The spider construction for $\mathcal{M}$ and $\mathcal{E}$ works as follows. As $\mathcal{E}$ is a 0-EC, all paths in $\mathcal{E}$ from s to some state t in $\mathcal{E}$ have the same weight, say $w(s,t)$. Note that then each path from t to s has weight $w(t,s) = -w(s,t)$.

Definition 3.5. Let $\mathcal{M}$ be an MDP, $\mathcal{E}$ a 0-BSCC of $\mathcal{M}$, and s_0 a reference state in $\mathcal{E}$. The *spider MDP* $\mathcal{N} = \mathsf{Spider}_{\mathcal{E},s_0}(\mathcal{M})$ (or shortly $\mathsf{Spider}_{\mathcal{E}}(\mathcal{M})$) results from $\mathcal{M}$ by

(i) removing the state-action pairs (s, α_s) for all states s in $\mathcal{E}$;

(ii) adding state-action pairs (s, τ) for each state s in $\mathcal{E}$ with $s \neq s_0$ where $P_{\mathcal{N}}(s, \tau, s_0) = 1$ and $wgt_{\mathcal{N}}(s, \tau) = w(s, s_0)$; and

(iii) for each state $s \neq s_0$ in $\mathcal{E}$ and action $\beta \in Act_{\mathcal{M}}(s) \backslash \{\alpha_s\}$, replacing (s, β) with (s_0, β) s.t. $P_{\mathcal{N}}(s_0, \beta, u) = P_{\mathcal{M}}(s, \beta, u)$ for all states u in $\mathcal{M}$ and $wgt_{\mathcal{N}}(s_0, \beta) = w(s_0, s) + wgt_{\mathcal{M}}(s, \beta)$.

Example 3.6. We exemplify the spider construction in Figure 2: Starting with an MDP $\mathcal{M}$, we apply the spider construction twice, each with reference state $s_0 = t$.
First, we chose the 0-BSCC $\mathcal{E} = \{(s,\alpha),(t,\alpha),(u,\alpha)\}$ of $\mathcal{M}$ and obtain $\mathcal{M}_1 = \mathsf{Spider}_{\mathcal{E},t}(\mathcal{M})$. Second, choosing the 0-BSCC $\mathcal{F} = \{(s,\tau),(t,\beta),(u,\tau),(v,\beta),(w,\beta)\}$ of $\mathcal{M}_1$ yields to an MDP $\mathcal{M}_2 = \mathsf{Spider}_{\mathcal{F},t}(\mathcal{M}_1)$ that does not contain any non-trivial 0-EC anymore. In each step, the chosen 0-EC turns into a sub-MDP where the reference state is the only sink. ∎

To formally state the equivalence of $\mathcal{M}$ and $\mathsf{Spider}_{\mathcal{E}}(\mathcal{M})$, we define the notion of $\mathcal{E}$-*invariant properties*. Given a path $\varsigma = t_0\,\alpha_0\,t_1\ldots$, let $purge_{\mathcal{E}}(\varsigma) \in (S \times \mathbb{Z})^\omega \cup (S \times \mathbb{Z})^* \times S$ be obtained from ς by (1) replacing each fragment $t_i\,\alpha_i\,\ldots\,\alpha_j\,t_{j+1}$ of ς such that (a) either $i = 0$ or $(t_{i-1}, \alpha_{i-1}) \notin \mathcal{E}$, (b) $(t_j, \alpha_j) \notin \mathcal{E}$, and (c) $(t_\ell, \alpha_\ell) \in \mathcal{E}$ for $\ell = i, i{+}1, \ldots, j{-}1$ with $t_i\,w\,t_{j+1}$ where $w = w(t_i, t_j) + wgt(t_j, \alpha_j)$ and (2) replacing each action α_i in the resulting sequence with $wgt(t_i, \alpha_i)$. A property φ is called $\mathcal{E}$-invariant if for all maximal paths ς we have: (I1) if ς has an infinite suffix of state-action pairs in $\mathcal{E}$, then $\varsigma \not\models \varphi$ and (I2) if $\varsigma \models \varphi$ and ς' is a maximal path with $purge_{\mathcal{E}}(\varsigma) = purge_{\mathcal{E}}(\varsigma')$ then $\varsigma' \models \varphi$. Weight-divergence and the pumping property are $\mathcal{E}$-invariant properties, and so are properties of the form $\Diamond(t \wedge (\mathrm{wgt} \bowtie K))$ where t is a trap, $\bowtie$ a comparison operator (*e.g.*, $=$ or $\geqslant$) and $K \in \mathbb{Z}$.

Lemma 3.7. *The spider construction generates an MDP $\mathsf{Spider}_{\mathcal{E}}(\mathcal{M})$ that satisfies the following properties:*

(S1) $\mathcal{M}$ and $\mathsf{Spider}_{\mathcal{E}}(\mathcal{M})$ have the same state space and we have $\|\mathsf{Spider}_{\mathcal{E}}(\mathcal{M})\| = \|\mathcal{M}\|{-}1$.
(S2) If $\mathcal{E} \neq \mathcal{M}$ and $\mathcal{M}$ is strongly connected then $\mathsf{Spider}_{\mathcal{E}}(\mathcal{M})$ has a single MEC that is reachable from all states.
(S3) $\mathcal{M}$ and $\mathsf{Spider}_{\mathcal{E}}(\mathcal{M})$ are equivalent for $\mathcal{E}$-invariant properties in the following sense:
  (S3.1) For each scheduler $\mathfrak{T}$ for $\mathsf{Spider}_{\mathcal{E}}(\mathcal{M})$ there is a scheduler $\mathfrak{S}$ for $\mathcal{M}$ with $\Pr^{\mathfrak{S}}_{\mathcal{M},s}(\varphi) = \Pr^{\mathfrak{T}}_{\mathsf{Spider}_{\mathcal{E}}(\mathcal{M}),s}(\varphi)$ for all states s and all $\mathcal{E}$-invariant properties φ. If $\mathfrak{T}$ is MD, then $\mathfrak{S}$ can be chosen MD.
  (S3.2) For each scheduler $\mathfrak{S}$ for $\mathcal{M}$ there exists a scheduler $\mathfrak{T}$ for $\mathsf{Spider}_{\mathcal{E}}(\mathcal{M})$ such that
  $$\Pr^{\mathfrak{S}}_{\mathcal{M},s}(\varphi) \leqslant \Pr^{\mathfrak{T}}_{\mathsf{Spider}_{\mathcal{E}}(\mathcal{M}),s}(\varphi) \leqslant \Pr^{\mathfrak{S}}_{\mathcal{M},s}(\varphi) + p^{\mathfrak{S}}_s$$
  for all states s and all $\mathcal{E}$-invariant properties φ. Here, $p^{\mathfrak{S}}_s = \Pr^{\mathfrak{S}}_{\mathcal{M},s}\{\varsigma \in IPaths : \lim(\varsigma) = \mathcal{E}\}$.

[1] We focus here on results for (positive) weight-divergence. The negative case can be obtained analogously by multiplying all weights with -1.

(S4) Suppose that $\mathcal{E}$ is contained in an MEC $\mathcal{G}$ of $\mathcal{M}$ with $\mathbb{E}^{\max}_{\mathcal{G}}(\mathrm{MP}) = 0$. Then for each state s with $s \notin \mathcal{E}$: s belongs to a 0-EC of $\mathcal{M}$ iff s belongs to a 0-EC of $\mathrm{Spider}_{\mathcal{E}}(\mathcal{M})$. Likewise, for each state-action pair (s,α) of $\mathcal{M}$: (s,α) belongs to a 0-EC of $\mathcal{M}$ iff $(s,\alpha) \in \mathcal{E}$ or (s_0,α) belongs to a 0-EC of $\mathrm{Spider}_{\mathcal{E}}(\mathcal{M})$.

The main property of the spider construction is that it eliminates the given 0-BSCC while maintaining all other 0-EC, as stated in (S4). (S3) states an equivalence between $\mathcal{M}$ and $\mathrm{Spider}_{\mathcal{E}}(\mathcal{M})$ with respect to $\mathcal{E}$-invariant properties. While any scheduler for $\mathrm{Spider}_{\mathcal{E}}(\mathcal{M})$ can be transformed to an equivalent scheduler for $\mathcal{M}$ (case (S3.1)), the converse direction (case (S3.2)) is more involved and requires restrictions, which are, however, sufficient for our applications.

As a consequence of the equivalence stated in (S3) we obtain that weight-divergent and pumping end components are preserved by the spider construction:

Corollary 3.8. *If $\mathcal{M}$ is strongly connected and $\mathcal{E}$ is a 0-BSCC of $\mathcal{M}$ then $\mathcal{M}$ is weight-divergent (resp. pumping) iff $\mathrm{Spider}_{\mathcal{E}}(\mathcal{M})$ is weight-divergent (resp. pumping).*

3.2 Checking Weight-Divergence

We present an algorithm to check the weight-divergence of an end component (see Algorithm 1). Such end components will be useful, *e.g.*, when solving weight-bounded reachability problems that require the accumulated weight to be above a threshold. Given a strongly connected MDP $\mathcal{M}$ we first compute $\mathbb{E}^{\max}_{\mathcal{M}}(\mathrm{MP})$ and an MD-scheduler $\mathfrak{S}$ maximizing the expected mean payoff. If $\mathbb{E}^{\max}_{\mathcal{M}}(\mathrm{MP}) > 0$ then $\mathcal{M}$ is pumping (Lemma 3.3) and therefore positively weight-divergent. If $\mathbb{E}^{\max}_{\mathcal{M}}(\mathrm{MP}) < 0$ then all schedulers for $\mathcal{M}$ are negatively weight-divergent (Lemma 3.3 with weights multiplied by -1), and hence, $\mathcal{M}$ is not positively weight-divergent. If $\mathbb{E}^{\max}_{\mathcal{M}}(\mathrm{MP}) = 0$ and $\mathfrak{S}$ has a gambling BSCC then $\mathcal{M}$ is gambling and therefore positively weight-divergent. Otherwise, each BSCC of the Markov chain induced by $\mathfrak{S}$ is a 0-BSCC (Lemma 3.2) and we pick such a 0-BSCC $\mathcal{E}$ of $\mathfrak{S}$. In case $\mathcal{M} = \mathcal{E}$ then $\mathcal{M}$ is a 0-EC, hence not weight-divergent, and the algorithm terminates. If $\mathcal{M} \neq \mathcal{E}$, we apply the spider construction to generate the MDP $\mathrm{Spider}_{\mathcal{E}}(\mathcal{M})$ that contains a unique maximal end component $\mathcal{F}$ ((S2) in Lemma 3.7). Repeating the procedure recursively on $\mathcal{F}$ etc. thus generates a sequence of MDPs $\mathcal{M}_0 = \mathcal{M}, \mathcal{M}_1, \ldots, \mathcal{M}_\ell$ with $\mathcal{M}_{i+1} = \mathrm{Spider}_{\mathcal{E}_i}(\mathcal{M}_i)$ for some 0-BSCC $\mathcal{E}_i$ of $\mathcal{M}_i$. All $\mathcal{M}_i$'s have the same state space and the number of state-action pairs is strictly decreasing, *i.e.*, we have $\|\mathcal{M}_0\| > \|\mathcal{M}_1\| > \ldots > \|\mathcal{M}_\ell\|$ by property (S1) in Lemma 3.7. Moreover, $\mathcal{M}_i$ is weight-divergent iff $\mathcal{M}$ is weight-divergent (see Corollary 3.8).

Algorithm 1: Wgtdiv($\cdot$)

input : strongly connected MDP $\mathcal{M}$
output : "yes" if $\mathcal{M}$ is weight divergent and "no" otherwise

1 Compute $e := \mathbb{E}^{\max}_{\mathcal{M}}(\mathrm{MP})$ and $\mathfrak{S}$ with $\mathbb{E}^{\mathfrak{S}}_{\mathcal{M}}(\mathrm{MP}) = e$
2 **if** $e < 0$ **then return** "no"
3 **if** $e > 0$ *or* $\mathfrak{S}$ *has a gambling BSCC* **then return** "yes"
4 Pick a 0-BSCC $\mathcal{E}$ of $\mathfrak{S}$
5 **if** $\mathcal{M} = \mathcal{E}$ **then return** "no"
6 Compute the MEC $\mathcal{F}$ of $\mathrm{Spider}_{\mathcal{E}}(\mathcal{M})$ that is reachable from all states and **return** Wgtdiv($\mathcal{F}$)

As each iteration takes polynomial time and the size of each $\mathcal{M}_i$ is polynomially bounded by the size of $\mathcal{M}$, the algorithm runs in polynomial time. Using an inductive argument and Lemma 3.7, we obtain:

Theorem 3.9. *The algorithm for checking weight-divergence of a strongly connected MDP $\mathcal{M}$ runs in polynomial time. If $\mathcal{M}$ is weight-divergent then it either finds a pumping or a gambling MD-scheduler. If $\mathcal{M}$ is not weight-divergent, then it generates an MDP $\mathcal{N}$ without 0-ECs on the same state space as $\mathcal{M}$, and is equivalent to $\mathcal{M}$ w.r.t. all properties that are $\mathcal{E}$-invariant for all 0-ECs $\mathcal{E}$ of $\mathcal{M}$ in the sense of (S3) in Lemma 3.7.*

Observe the following consequence of this theorem:

Corollary 3.10. *Let $\mathcal{M}$ be a strongly connected MDP with $\mathbb{E}^{\max}_{\mathcal{M}}(\mathrm{MP}) = 0$. Then, $\mathcal{M}$ is weight-divergent iff $\mathcal{M}$ is gambling iff $\mathcal{M}$ has a gambling MD-scheduler.*

However, an MDP can have gambling schedulers even when it has no gambling MD-scheduler: Consider the MDP over a single state s with state-action pairs (s,α), (s,β), where $P(s,\alpha,s) = 1, P(s,\beta,s) = 1$, $wgt(s,\alpha) = -wgt(s,\beta) = 1$. Then, $\mathbb{E}^{\max}_{\mathcal{M}}(\mathrm{MP}) = +\infty$ and there is no gambling MD-scheduler, while the randomized memoryless scheduler $\mathfrak{S}$ with $\mathfrak{S}(s)(\alpha) = \mathfrak{S}(s)(\beta) = \frac{1}{2}$ is gambling.

Given a strongly connected MDP $\mathcal{M}$ with $\mathbb{E}^{\max}_{\mathcal{M}}(\mathrm{MP}) = 0$, $\mathcal{M}$ is gambling iff $\mathcal{M}$ is weight-divergent. Thus, the gambling property for strongly connected MDPs with maximal expected mean payoff 0 can be checked in polynomial time using Theorem 3.9, which yields part (a) of the next theorem.

Theorem 3.11. *Given a strongly connected MDP $\mathcal{M}$, the existence of a gambling MD-scheduler is (a) decidable in polynomial time if $\mathbb{E}^{\max}_{\mathcal{M}}(\mathrm{MP}) = 0$, and (b) NP-complete in general.*

One can compute an MD-scheduler in polynomial time that maximizes the probability of weight-divergence. In fact, one can compute weight-divergent MECs (and corresponding weight-divergent MD-schedulers) and maximize the probability of reaching one of these components. Likewise, the minimal probability of weight-divergence equals the maximal probability to reach the set V of states of all trap states and all states belonging to an MEC $\mathcal{E}$ where either $\mathbb{E}^{\min}_{\mathcal{E}}(\mathrm{MP}) < 0$ or $\mathbb{E}^{\min}_{\mathcal{E}}(\mathrm{MP}) = 0$ and $\mathcal{E}$ has a 0-EC. Theorem 3.12 below shows that set V is computable in polynomial time. This yields a polynomial-time algorithm for finding an MD-scheduler minimizing the weight-divergence probability.

Previous work established the polynomial-time computability of maximal weight-divergence probabilities in special cases. In fact, [6, Theorem 3.1] presents an algorithm to compute an MD-scheduler maximizing the probability for weight-divergent paths in a given MDP where the weights belong to $\{-1, 0, 1\}$. Thus, [6] yields a *pseudo-polynomial* time bound for deciding weight-divergence or computing the maximal weight-divergence probabilities in MDPs with integer weights. Theorem 3.9 and the previous paragraph improve this result by establishing a polynomial time bound. Moreover, our algorithm is different; while [6] uses transformations to incorporate accumulated weights in the state space (up to some threshold), our algorithm uses the spider construction and maintains the state space.

3.3 Reasoning about 0-ECs

We are now interested in checking the existence of 0-ECs and computing all state-action pairs inside some 0-EC, useful, *e.g.*, to deal with weight-bounded constraints (see Section 5).

In MDPs without weight-divergent end components, the weight-divergence algorithm can be used to determine all state-action pairs belonging to a 0-EC in polynomial time. However, this does not work in general as the algorithm stops as soon as a weight-divergent end component is found.

To check whether a given strongly connected MDP $\mathcal{M}$ with $\mathbb{E}^{\max}_{\mathcal{M}}(\mathrm{MP}) = 0$ contains a 0-EC, we use an iterative approach: we apply standard algorithms to compute an MD-scheduler $\mathfrak{S}$ with a single BSCC $\mathcal{B}$ maximizing the expected mean payoff (in particular, $\mathbb{E}_{\mathcal{B}}(\mathrm{MP}) = 0$) and checks whether $\mathcal{B}$ is a 0-BSCC. If yes, $\mathcal{B}$ is a 0-EC of $\mathcal{M}$. Otherwise, $\mathcal{B}$ is gambling (see Lemma 3.2). In this case, we give a transformation that modifies the transition probabilities in $\mathcal{B}$ to obtain an MDP $\mathcal{M}'$ with the same structure as $\mathcal{M}$ (in particular, with the same 0-ECs) such that $\mathcal{M}'$ has fewer gambling MD-schedulers than $\mathcal{M}$. Thus, if $\mathbb{E}^{\max}_{\mathcal{M}'}(\mathrm{MP}) < 0$ then $\mathcal{M}$ has no 0-EC. Otherwise, we repeat the procedure on $\mathcal{M}'$.

This transformation is crucial in several results that follow.

Theorem 3.12. *Given a strongly connected MDP $\mathcal{M}$, the existence of 0-ECs is (a) decidable in polynomial time if $\mathbb{E}^{\max}_{\mathcal{M}}(\mathrm{MP}) = 0$, and (b) NP-complete in the general case.*

Combining the above decision algorithm and the iterative elimination of 0-ECs, we can also compute the set of all 0-ECs in polynomial time. An important notion in our algorithms is the *recurrence value* defined as follows. For a state s of a 0-EC in a strongly connected MDP $\mathcal{M}$ with $\mathbb{E}^{\max}_{\mathcal{M}}(\mathrm{MP}) = 0$, $rec(s)$ is the maximal integer K s.t. $\Pr^{\mathfrak{S}}_{\mathcal{M},s}(\Box(\mathrm{wgt} \geqslant K) \wedge \Box\Diamond s) = 1$ for some $\mathfrak{S}$ that only uses actions belonging to some 0-EC. In fact, to ensure that the accumulated weight stays above 0, it does not suffice to enter a 0-EC with nonnegative weight, as 0-ECs can contain state-action pairs with negative weight.

Lemma 3.13. *If $\mathcal{M}$ is strongly connected and $\mathbb{E}^{\max}_{\mathcal{M}}(\mathrm{MP}) = 0$ then the set ZeroEC consisting of all states s that belong to some 0-EC, as well as the recurrence values $rec(s)$ for the states $s \in$ ZeroEC are computable in polynomial time.*

3.4 Universal Negative Weight-Divergence and Boundedness

We now show how to determine end components that are bounded from below and those that are universally negatively weight-divergent. Part (a) of the following theorem is the MDP-analogue of part (d) of Lemma 3.2.

Theorem 3.14. *Let $\mathcal{M}$ be a strongly connected MDP with $\mathbb{E}^{\max}_{\mathcal{M}}(\mathrm{MP}) = 0$. Then, (a) $\mathcal{M}$ contains a 0-EC iff $\mathcal{M}$ has a scheduler where the measure of infinite paths that are bounded from below is positive iff $\mathcal{M}$ has a scheduler that is bounded from below; (b) $\mathcal{M}$ has no 0-EC iff each scheduler for $\mathcal{M}$ is negatively weight-divergent.*

Given a strongly connected MDP $\mathcal{M}$, universal (positive) weight-divergence of $\mathcal{M}$ can be checked in polynomial time. In fact, if $\mathbb{E}^{\min}_{\mathcal{E}}(\mathrm{MP}) > 0$, then $\mathcal{M}$ is universally weight-divergent, and if $\mathbb{E}^{\min}_{\mathcal{E}}(\mathrm{MP}) < 0$, it is not. If $\mathbb{E}^{\max}_{\mathcal{M}}(\mathrm{MP}) = 0$, we use Theorem 3.14 (by multiplying the weights by -1) and check the nonexistence of 0-ECs by Theorem 3.12. We get:

Corollary 3.15. *Universal (positive) weight-divergence of an MDP can be checked in polynomial time.*

Remark 3.16. The set of states s of an arbitrary MDP $\mathcal{M}$ that belongs to an end component bounded from below can be computed in polynomial time as follows. We first determine the MECs of $\mathcal{M}$ and their maximal expected mean payoff. MECs $\mathcal{E}$ with $\mathbb{E}^{\max}_{\mathcal{E}}(\mathrm{MP}) > 0$ are pumping and therefore bounded from below. MECs $\mathcal{E}$ with either $\mathbb{E}^{\max}_{\mathcal{E}}(\mathrm{MP}) < 0$ or $\mathbb{E}^{\max}_{\mathcal{E}}(\mathrm{MP}) = 0$ and $\mathcal{E}$ has no 0-EC are universally negatively weight-divergent (Theorem 3.14). Hence, none of their states belongs to an end component that is bounded from below. Otherwise, *i.e.*, if $\mathbb{E}^{\max}_{\mathcal{E}}(\mathrm{MP}) = 0$ and $\mathcal{E}$ has 0-ECs, we compute the maximal 0-ECs using the techniques presented in Section 3.3 (see Lemma 3.13).

4 Stochastic Shortest Paths

We present an algorithm to solve the stochastic shortest path problem that relies on the classification of end components presented above. The classical shortest path problem for MDPs is to compute the *minimal expected accumulated weight* until reaching a goal state *goal*. Here, the infimum is taken over all *proper* schedulers. These are schedulers $\mathfrak{S}$ that reach *goal* almost surely, *i.e.*, $\Pr^{\mathfrak{S}}_{\mathcal{M},s}(\Diamond goal) = 1$ for all states $s \in S$.

We assume, w.l.o.g., that *goal* is a trap, and that all states s are reachable from an initial state s_{init} and can reach *goal*. We write $\diamondplus\, goal$ for the random variable that represents the accumulated weight until reaching *goal*: it assigns to each path reaching *goal* its accumulated weight, and is undefined otherwise. Formally, $(\diamondplus\, goal)(\varsigma) = wgt(\varsigma)$ if $\varsigma \models \Diamond goal$ and undefined if $\varsigma \not\models \Diamond goal$. The *stochastic shortest path problem* aims at computing the minimal expected accumulated weight until reaching *goal*:

$$\mathbb{E}^{\inf}_{\mathcal{M},s_{init}}(\diamondplus\, goal) \;=\; \inf_{\mathfrak{S}\text{ proper}} \mathbb{E}^{\mathfrak{S}}_{\mathcal{M},s_{init}}(\diamondplus\, goal)\ .$$

Although for each proper scheduler this quantity is finite, the infimum may be $-\infty$. We describe a polynomial-time algorithm to check whether $\mathbb{E}^{\inf}_{\mathcal{M},s_{init}}(\diamondplus\, goal)$ is finite and to compute it, both using our classification of end components.

It is well known (see, *e.g.*, [15]) that if $\mathcal{M}$ is *contracting*, *i.e.*, if all schedulers are proper, then $\mathbb{E}^{\inf}_{\mathcal{M},s_{init}}(\diamondplus\, goal) > -\infty$ and one can compute $\mathbb{E}^{\inf}_{\mathcal{M},s_{init}}(\diamondplus\, goal)$ using linear-programming techniques. To relax the assumption of $\mathcal{M}$ being contracting, Bertsekas and Tsitsiklis [5] identified conditions that guarantee the finiteness of the values $\mathbb{E}^{\inf}_{\mathcal{M},s_{init}}(\diamondplus\, goal)$, the existence of a minimizing MD-scheduler, and the computability of the vector $(\mathbb{E}^{\inf}_{\mathcal{M},s}(\diamondplus\, goal))_{s\in S}$ as the unique solution of a linear program (or using value and policy iteration). The assumptions of [5], written (BT) in the sequel, are: (i) existence of a proper scheduler, and (ii) under each non-proper scheduler the expected accumulated weight is $+\infty$ from at least one state. While these assumptions are sound, they are incomplete in the sense that there are MDPs where $\mathbb{E}^{\inf}_{\mathcal{M},s}(\diamondplus\, goal)$ is finite for all states s, but (BT) does not hold.

Orthogonally, De Alfaro [12] showed that in MDPs where the weights are either all nonnegative or all nonpositive, one can decide in polynomial time whether $\mathbb{E}^{\inf}_{\mathcal{M},s_{init}}(\diamondplus\, goal)$ is finite. Moreover, when this is the case, $\mathcal{M}$ can be transformed into another MDP that has proper schedulers, satisfies (BT) and preserves the minimal expected accumulated weight. Using the classification of

end components, we generalize De Alfaro's result and provide a characterization of finiteness of the minimal expected accumulated weight.

Lemma 4.1. *Let $\mathcal{M}$ be an MDP with a distinguished initial state s_{init} and a trap state goal such that all states are reachable from s_{init} and can reach goal. Then, $\mathbb{E}^{\inf}_{\mathcal{M},s_{init}}(\oplus\, goal)$ is finite iff $\mathcal{M}$ has no negatively weight-divergent end component. If so, then $\mathcal{M}$ satisfies* (BT) *iff $\mathcal{M}$ has no 0-EC.*

The above lemma allows us to derive our algorithm by first determining if $\mathbb{E}^{\inf}_{\mathcal{M},s_{init}}(\oplus\, goal)$ is finite, and then using the iterative spider construction to transform $\mathcal{M}$ into an equivalent new MDP satisfying BT. More precisely, one can check in polynomial time whether $\mathbb{E}^{\inf}_{\mathcal{M},s_{init}}(\oplus\, goal) > -\infty$ by applying Theorem 3.9 to the maximal end components of $\mathcal{M}$ (in fact, checking negative weight-divergence reduces to checking positive weight-divergence after multiplication of all weights by -1). If so, by the iterative spider construction to flatten 0-ECs (see Section 3.1), we obtain in polynomial time an MDP $\mathcal{N}$ such that $\mathcal{N}$ satisfies condition (BT) and $\mathbb{E}^{\inf}_{\mathcal{N},s}(\oplus\, goal) = \mathbb{E}^{\inf}_{\mathcal{M},s}(\oplus\, goal)$ for each state s. To establish this result, we rely on the equivalence of $\mathcal{M}$ and $\mathcal{N}$ w.r.t. properties that are $\mathcal{E}$-invariant for each 0-EC $\mathcal{E}$ ((S3) in Lemma 3.7). This yields:

Theorem 4.2. *Given an arbitrary MDP $\mathcal{M}$, one can compute in polynomial time $\mathbb{E}^{\inf}_{\mathcal{M},s_{init}}(\oplus\, goal)$ as well as an MD scheduler achieving the minimum when this value is finite.*

Analogous results are obtained for $\mathbb{E}^{\sup}_{\mathcal{M},s_{init}}(\oplus\, goal)$ by multiplying all weights in $\mathcal{M}$ with -1.

5 Qualitative Weight-Bounded Properties

5.1 Disjunctive Weight-Bounded Reachability

We consider properties that combine reachability objectives with quantitative constraints on the accumulated weight when reaching the targets.

Definition 5.1. A *disjunctive weight-bounded reachability* property, DWR-property for short, is defined by a set $T \subseteq S$ of target states, and for each $t \in T$ a weight threshold $K_t \in \mathbb{Z} \cup \{-\infty\}$ as $\varphi = \bigvee_{t\in T} \Diamond\big(t \wedge (\mathrm{wgt} \geqslant K_t)\big)$.

Our objective is to study the following decision problems: Given an MDP $\mathcal{M}$, a state s in $\mathcal{M}$ and a DWR-property φ

$$\mathrm{DWR}^{\exists,=1}\colon\ \exists \mathfrak{S} \text{ s.t. } \Pr^{\mathfrak{S}}_{\mathcal{M},s}(\varphi) = 1?$$
$$\mathrm{DWR}^{\exists,>0}\colon\ \exists \mathfrak{S} \text{ s.t. } \Pr^{\mathfrak{S}}_{\mathcal{M},s}(\varphi) > 0?$$

as well as their variants $\mathrm{DWR}^{\forall,=1}$ and $\mathrm{DWR}^{\forall,>0}$ with universal quantification over schedulers. Let $T^* = \{t \in T : K_t = -\infty\}$ denote the set of states for which no accumulated weight constraint is specified. For corresponding optimization problems, we assume $T\backslash T^* = \{goal\}$ to be a singleton, write φ_K for φ with $K = K_{goal}$, and ask to compute

$$K^{\exists,=1}_{\mathcal{M},s} = \sup\{\, K \in \mathbb{Z} \mid \exists \mathfrak{S} \text{ s.t. } \Pr^{\mathfrak{S}}_{\mathcal{M},s}(\varphi_K) = 1 \,\},$$
$$K^{\exists,>0}_{\mathcal{M},s} = \sup\{\, K \in \mathbb{Z} \mid \exists \mathfrak{S} \text{ s.t. } \Pr^{\mathfrak{S}}_{\mathcal{M},s}(\varphi_K) > 0 \,\},$$

and the analogous values $K^{\forall,=1}_{\mathcal{M},s}$ and $K^{\forall,>0}_{\mathcal{M},s}$ where the supremum belongs to $\mathbb{Z} \cup \{\pm\infty\}$.

Deciding $\mathrm{DWR}^{\exists,>0}$ and computing $K^{\exists,>0}_{\mathcal{M},s}$ can be done using standard shortest-path algorithms in weighted graphs. Thus, $\mathrm{DWR}^{\exists,>0}$ belongs to P and the value $K^{\exists,>0}_{\mathcal{M},s}$ is computable in polynomial time.

In contrast, we do not know if $\mathrm{DWR}^{\forall,>0}$ is in P, but show that it is as hard as mean-payoff games, and is polynomially reducible to mean-payoff Büchi games.

Theorem 5.2. *The decision problem $DWR^{\forall,>0}$ is in NP∩ coNP, and at least as hard as (non-stochastic) mean-payoff games. The value $K^{\forall,>0}_{\mathcal{M},s}$ is computable in pseudo-polynomial time.*

We now give a polynomial-time algorithm for $\mathrm{DWR}^{\forall,=1}$. In the case where all states of T are traps, we show that $\Pr^{\mathfrak{S}}_{\mathcal{M},s}(\varphi) = 1$ for all schedulers $\mathfrak{S}$ iff (i) $\Pr^{\min}_{\mathcal{M},s}(\Diamond T) = 1$ and (ii) $wgt(\pi) \geqslant K_t$ for each path π from s to some state $t \in T\backslash T^*$. (In particular, (ii) implies that the paths from s to some state in $T\backslash T^*$ do not contain negative cycles.) Thus, this case can be solved with standard MDP and shortest-path algorithms in graphs. The general case requires an analysis of end components. If each end component containing $t \in T\backslash T^*$ is weight-divergent, then the weight-constraint is useless and we may set $K_t = +\infty$. Otherwise we show that t can be treated as a trap. To check whether all end components containing t are weight-divergent we consider the MECs $\mathcal{E}$ containing t and distinguish cases where $\mathbb{E}^{\min}_{\mathcal{M},\mathcal{E}}(\mathrm{MP}) > 0$ or $\mathbb{E}^{\min}_{\mathcal{M},\mathcal{E}}(\mathrm{MP}) = 0$ and $\mathcal{E}$ does not have a 0-EC containing t.

Theorem 5.3. *The decision problem $DWR^{\forall,=1}$ belongs to P and the value $K^{\forall,=1}_{\mathcal{M},s}$ is computable in polynomial time.*

The remaining case $\mathrm{DWR}^{\exists,=1}$ is perhaps the most interesting case; it is also our main and most technical result. First, we observe that infinite memory can be necessary.

Example 5.4. Let $\mathcal{M}$ be the MDP depicted left in Figure 3.

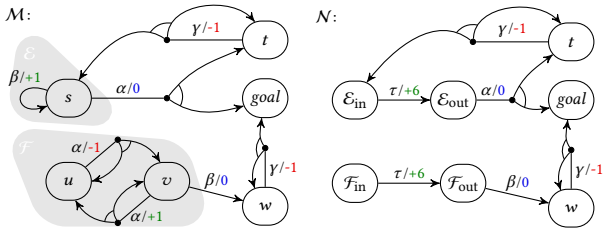

Figure 3. Resolution of $\mathrm{DWR}^{\exists,=1}$ on an example.

Consider the weight-bounded reachability property $\varphi_K = \Diamond(goal \wedge (\mathrm{wgt} \geqslant K))$. Given $K \in \mathbb{Z}$, a scheduler $\mathfrak{S}_K$ ensuring $\Pr^{\mathfrak{S}_K}_{\mathcal{M},s}(\varphi_K) = 1$ acts as follows: for a finite path π ending in state s with accumulated weight k, $\mathfrak{S}_K$ schedules $K-k$ times action β, followed by α. Thus, all $\mathfrak{S}_K$-paths from s ending in state t or $goal$ have weight at least K and $K^{\exists,=1}_{\mathcal{M},s} = +\infty$. However, for every finite-memory scheduler $\mathfrak{S}$, there is no $K \in \mathbb{Z}$ with $\Pr^{\mathfrak{S}}_{\mathcal{M},s}(\varphi_K) = 1$. ∎

Theorem 5.5. *The decision problem $DWR^{\exists,=1}$ is in NP∩ coNP, and at least as hard as (non-stochastic) mean-payoff games. The value $K^{\exists,=1}_{\mathcal{M},s}$ is computable in pseudo-polynomial time.*

Proof sketch. We sketch the proof for the upper bound. The general case easily reduces to the same problem for $T \backslash T^* = \{goal\}$ is a singleton; so we make this assumption.

First, in the case where $\mathcal{M}$ has no positively weight-divergent end components, we give a polynomial-time reduction to mean payoff games which can be solved in NP ∩ coNP.

For the general case, let us write $\mathcal{E}_1, \ldots, \mathcal{E}_k$ for the maximal positively weight-divergent end components of $\mathcal{M}$. They can be computed by first determining the MECs and checking weight-divergence for each of them by Theorem 3.9. We then show that there exists $K_i \in \{+\infty, -\infty\}$ such that for all states s in $\mathcal{E}_i$ we have $K^{\exists,=1}_{\mathcal{M},s} = K_i$. This observation follows from the fact that any scheduler can be modified to have a first phase where the weight is increased by a desired constant inside a weight-divergent end component.

We compute the set $GoodEC = \{\mathcal{E}_i : K_i = +\infty\}$ using the greatest fixed point of a monotonic operator $\Omega\colon 2^{\mathfrak{E}} \to 2^{\mathfrak{E}}$ where $\mathfrak{E} = \{\mathcal{E}_1, \ldots, \mathcal{E}_k\}$ using the techniques for MDPs without positively weight-divergent end components. To define this operator Ω, we switch from $\mathcal{M}$ to a new MDP $\mathcal{N}$ obtained from $\mathcal{M}$ by replacing each $\mathcal{E} \in \mathfrak{E}$ with two fresh states $\mathcal{E}_{in}$ and $\mathcal{E}_{out}$. The actions enabled in $\mathcal{E}_{out}$ serve to mimic $\mathcal{M}$'s state-action pairs (s, α) where s is a state of $\mathcal{E}$ and $P_{\mathcal{M}}(s, \alpha, s') > 0$ for at least one state s' outside $\mathcal{E}$. A single action τ is enabled in $\mathcal{E}_{in}$ with $P_{\mathcal{N}}(\mathcal{E}_{in}, \tau, \mathcal{E}_{out}) = 1$ whose weight is chosen large enough to ensure that $\mathcal{E}_{in}$ and $\mathcal{E}_{out}$ do not belong to a negative simple cycle. The construction is illustrated in Fig. 3. $\mathcal{N}$ has no positively weight-divergent end components by construction. However, the values in $\mathcal{N}$ can be used as lower bounds of those in $\mathcal{M}$. In particular, we may have $K^{\exists,=1}_{\mathcal{N},r} = -\infty$ and $K^{\exists,=1}_{\mathcal{M},r'} = +\infty$ where r and r' are corresponding states in $\mathcal{M}$ and $\mathcal{N}$ (*e.g.*, state s in Fig. 3 has value $+\infty$ in $\mathcal{M}$ but $\mathcal{E}_{out}$ has value $-\infty$ in $\mathcal{N}$). Despite this, we can identify end components in *GoodEC*, *i.e.*, with value $+\infty$, using $\mathcal{N}$ via a fixed-point computation. Namely, we define the operator Ω that assigns to each $X \subseteq \mathfrak{E}$ the set of end components $\mathcal{E} \in \mathfrak{E}$ for which there is $K \in \mathbb{Z}$ with $\Pr^{\max}_{\mathcal{N},\mathcal{E}_{out}}\big(\varphi_K[X]\big) = 1$ where

$$\varphi_K[X] \;=\; \Diamond\big(T^* \cup \{\mathcal{E}_{in} : \mathcal{E} \in X\}\big) \vee \Diamond\big(goal \wedge (\mathrm{wgt} \geqslant K)\big).$$

Intuitively, these are states from which almost surely we either satisfy φ, or reach another weight-divergent end component that allows to increase the weight and start again. This fixed-point computation applied to $\mathcal{N}$ in Fig. 3 yields, *e.g.*, $X_0 = \{\mathcal{E}, \mathcal{F}\}$, $\Omega(X_0) = \{\mathcal{E}\}$, $\Omega(\Omega(X_0)) = \{\mathcal{E}\}$. In fact, from $\mathcal{E}$ one can either immediately reach *goal* or go back to $\mathcal{E}$; while from $\mathcal{F}$ there is no bound on the accumulated weight towards reaching *goal*.

The above computation yields the values of the states of weight-divergent end components; in fact, we show that $K^{\exists,=1}_{\mathcal{M},s} = +\infty$ iff $\Pr^{\max}_{\mathcal{M},s}(\Diamond(T^* \cup GoodEC)) = 1$. For other states, we show that the maximal K such that $\Pr^{\max}_{\mathcal{N},s}(\phi_K[GoodEC]) = 1$ corresponds to $K^{\exists,=1}_{\mathcal{M},s'}$ where s and s' are corresponding states. Here, $\phi_K[GoodEC]$ is an instance of $\mathrm{DWR}^{\exists,=1}$and $\mathcal{N}$ has no weight-divergent end components, so we can use the NP ∩ coNP algorithm described at the beginning. □

5.2 Weight-Bounded Repeated Reachability

Beyond weight-bounded reachability, we address a Büchi weight condition in conjunction with a standard Büchi condition. Given an MDP $\mathcal{M}$ without traps, a set $F \cup \{s\}$ of states in $\mathcal{M}$ and $K \in \mathbb{Z}$, we consider the problems

$$\mathrm{WB}^{\exists,=1}\colon \quad \exists \mathfrak{S} \text{ s.t. } \Pr^{\mathfrak{S}}_{\mathcal{M},s}(\Box\Diamond(\mathrm{wgt} \geqslant K) \wedge \Box\Diamond F) = 1?$$
$$\mathrm{WB}^{\exists,>0}\colon \quad \exists \mathfrak{S} \text{ s.t. } \Pr^{\mathfrak{S}}_{\mathcal{M},s}(\Box\Diamond(\mathrm{wgt} \geqslant K) \wedge \Box\Diamond F) > 0?$$

and the corresponding problems $\mathrm{WB}^{\forall,=1}$ and $\mathrm{WB}^{\forall,>0}$ with universal quantification over schedulers. The two existential problems are polynomially reducible to the respective existential DWR problems, maintaining the same complexity classes. The universal problems can be solved using techniques to treat existential problems for coBüchi weight constraints, which again are polynomially reducible to $\mathrm{DWR}^{\exists,>0}$ and $\mathrm{DWR}^{\exists,=1}$, respectively.

Theorem 5.6. *$\mathrm{WB}^{\exists,>0}$ and $\mathrm{WB}^{\forall,=1}$ are decidable in polynomial time. $\mathrm{WB}^{\exists,=1}$ and $\mathrm{WB}^{\forall,>0}$ are in NP ∩ coNP, decidable in pseudo-polynomial time, and at least as hard as mean-payoff games.*

The proof of Theorem 5.6 heavily uses the concepts of Section 3. Let us briefly describe the reduction of $\mathrm{WB}^{\exists,=1}$ and $\mathrm{WB}^{\exists,>0}$ to $\mathrm{DWR}^{\exists,=1}$ and $\mathrm{DWR}^{\exists,>0}$ for some DWR formula $\varphi = \bigvee_{t\in T} \Diamond\big(t \wedge (\mathrm{wgt} \geqslant K_t)\big)$. We define T^* as the set of all states in maximal weight-divergent end components containing at least one state in F and $T \backslash T^*$ as the set of states belonging to a maximal 0-EC $\mathcal{Z}$ of a maximal end component $\mathcal{E}$ with $\mathbb{E}^{\max}_{\mathcal{E}}(\mathrm{MP}) = 0$ and $\mathcal{Z} \cap F \neq \varnothing$. Note that both T^* and $T \backslash T^*$ are computable in polynomial time (due to Theorem 3.9 and Lemma 3.13). For the states in $T \backslash T^*$, we let $K_t = K$, where K is taken from the input of $\mathrm{WB}^{\exists,=1}$ or $\mathrm{WB}^{\exists,>0}$.

To solve problem $\mathrm{WB}^{\forall,=1}$ we rely on the observation that $\mathrm{WB}^{\forall,=1}$ holds iff (i) $\Pr^{\min}_{\mathcal{M},s}(\Box\Diamond F) = 1$ and (ii) there is no scheduler $\mathfrak{S}$ with $\Pr^{\mathfrak{S}}_{\mathcal{M}^-,s}\big(\Diamond\Box(\mathrm{wgt} \geqslant L)\big) > 0$ where $\mathcal{M}^-$ results from $\mathcal{M}$ by multiplying all weights with -1 and $L = -(K-1)$. While (i) can be checked in polynomial time, (ii) is equivalent to the complement of $\mathrm{DWR}^{\exists,>0}$ for $\mathcal{M}^-$ and $\bigvee_{t\in T} \Diamond\big(t \wedge (\mathrm{wgt} \geqslant K_t)\big)$ where T^* denotes the set of states belonging to a pumping end component of $\mathcal{M}^-$ and $T \backslash T^*$ is the set of states belonging to the set *ZeroEC* and $K_t = L - rec(t)$. Here *ZeroEC* is the set of states that belong to a maximal 0-EC $\mathcal{Z}$ of a maximal end component $\mathcal{E}$ of $\mathcal{M}$ or $\mathcal{M}^-$ with $\mathbb{E}^{\max}_{\mathcal{E}}(\mathrm{MP}) = 0$ and moreover, $rec(t)$ refers to this maximal end component $\mathcal{E}$.

For problem $\mathrm{WB}^{\forall,=1}$ we transform $\mathcal{M}^-$ into a new MDP $\mathcal{N}$ such that $\mathrm{WB}^{\forall,=1}$ holds for $\mathcal{M}$ iff there is no scheduler for $\mathcal{N}$ where the coBüchi weight constraint $\Diamond\Box(\mathrm{wgt} \geqslant L)$ holds almost surely, which can be checked applying the algorithm for $\mathrm{DWR}^{\exists,=1}$ for $\mathcal{N}$ and the same DWR property as for $\mathrm{DWR}^{\forall,>0}$. Here L is as above and $\mathcal{N}$ arises from $\mathcal{M}^-$ by identifying all states that belong to an end component not containing an F-state and replacing their enabled actions with a self-loop of weight 0.

The optimization problems of $\mathrm{WB}^{\exists,=1}$ and $\mathrm{WB}^{\forall,>0}$ are computable in pseudo-polynomial time, and optimal weight bounds for $\mathrm{WB}^{\exists,>0}$ and $\mathrm{WB}^{\forall,=1}$in polynomial time.

5.3 Discussion on Related Work

To the best of our knowledge, problems $\mathrm{DWR}^{\exists,=1}$, $\mathrm{DWR}^{\forall,>0}$ and $\mathrm{DWR}^{\forall,=1}$ or the variants for Büchi weight constraints have not been studied before for general integer-weighted MDPs. Qualitative weight-bounded reachability properties in MDPs with only nonnegative weights are decidable in polynomial time [19]. This result relies on the monotonicity of accumulated weights along all paths. The lack of monotonicity in the general case rules out analogous algorithms.

For Markov chains, qualitative weight-bounded reachability properties can be treated in polynomial time [16]. This result uses expected mean payoff in BSCCs, variants of shortest-path algorithms and the continued-fraction method. In MDPs, however, optimal schedulers might need infinite memory (see Example 5.4) so these algorithms cannot be adapted. In fact, our algorithms crucially rely on the classification of end components.

Let us point out the similarities and differences between the problems we considered and the ones for energy MDPs [9, 17]. Rephrased for our notations, the energy-MDP problem is to check whether $\Pr^{\max}_{\mathcal{M},s}\big(\Box(\text{wgt} \geqslant K) \wedge \phi\big) = 1$ where φ is a parity condition and $K \in \mathbb{Z}$. This problem is in NP $\cap$ coNP and at least as hard as two-player mean-payoff games, even if $\phi = \mathit{true}$. The complement of the energy-MDP problem asks whether $\Pr^{\min}_{\mathcal{M},s}\big(\Diamond(\text{wgt} < K) \vee \neg\phi\big) > 0$, which corresponds to $\Pr^{\min}_{\mathcal{M},s}\big(\Diamond(\text{wgt} \geqslant K) \vee \neg\phi\big) > 0$ when switching from wgt to $-wgt$ and from K to $-(K-1)$. However, although in the spirit of this problem, $\text{DWR}^{\forall,>0}$ asks whether $\Pr^{\min}_{\mathcal{M},s}\big(\Diamond(\mathit{goal} \wedge (\text{wgt} \geqslant K))\big) > 0$, in the case $T^* = \varnothing$ and $T \setminus T^* = \{\mathit{goal}\}$. Given the similarities of these questions, and our decision procedure that reduces $\text{DWR}^{\forall,>0}$ to mean-payoff Büchi games, it is no surprise that the problem $\text{DWR}^{\forall,>0}$ is at least as hard as mean-payoff games.

Nevertheless, the instances $\text{DWR}^{\exists,=1}$ and $\text{DWR}^{\forall,=1}$ are of different nature than energy-MDPs. These can rather be seen as variants of the *termination problem* for *one-counter MDPs* [6, 13]. One-counter MDPs have their weights in $\{-1, 0, +1\}$, while we allow arbitrary weights. Moreover, a one-counter MDPs halts whenever the counter reaches 0, but there is no lower bound on the accumulated weight in our setting. Following [6], we refer to these one-counter MDPs as *one-counter MDP with boundary* and to MDPs in our setting with weights in $\{-1, 0, +1\}$ as *boundaryless one-counter MDPs*.

We commented on [6] in the paragraph following Theorem 3.11. For one-counter MDPs $\mathfrak{M}$ with boundary, [6] also provides an exponential-time algorithm for checking $\Pr^{\max}_{\mathfrak{M},s}\big(\bigvee_{t\in T} \Diamond(t \wedge (\text{wgt} = 0))\big) = 1$ and shows PSPACE-hardness. This contrasts with our NP $\cap$ coNP upper bound for $\text{DWR}^{\exists,=1}$ with arbitrary integer weights (Theorem 5.5). Besides the differences "boundary vs boundaryless" and "integer vs unit weights", we consider objectives imposing lower bounds on the accumulated weights. Considering $\Diamond(t \wedge (\text{wgt} = K_t))$ would raise the complexity in our setting at least to EXPTIME-hardness, by [14] which shows that for MDPs $\mathcal{M}$ with non-negative integer weights and $\Pr^{\min}_{\mathcal{M},s}(\Diamond \mathit{goal}) = 1$, checking whether $\Pr^{\max}_{\mathcal{M},s}(\Diamond(\mathit{goal} \wedge (\text{wgt} = K))) = 1$ for some given $K \in \mathbb{N}$ is EXPTIME-complete.

Nondeterministic and probabilistic models for vector addition systems (VASS-MDPs) can be seen as boundary MDPs with multiple weight functions. Decidable results on VASS-MDPs include the existence of a scheduler that almost surely ensures some property expressible in μ-calculus (with no constraint on the accumulated weights) [1]. The decision algorithms rely on the termination of fixed-point computations thanks to well-quasi orderings, thus yielding much higher complexity than our techniques.

6 Conclusion

We provided a classification of end components according to their behaviors with respect to the accumulated weight. This allowed us to solve the general stochastic shortest path problem and to derive algorithms for weight-bounded properties. We believe our classification helps better understanding the accumulated weights in MDPs, and can be helpful for other problems and perhaps simplify existing results.

An interesting future work is to address analogous questions for quantitative probability thresholds. This appears to be challenging as the probabilities for weight-bounded properties can be irrational, even in Markov chains [6, 13].

References

[1] Parosh Aziz Abdulla, Radu Ciobanu, Richard Mayr, Arnaud Sangnier, and Jeremy Sproston. Qualitative analysis of VASS-induced MDPs. In FoSSaCS'16, LNCS 9634, p. 319–334. Springer, 2016.

[2] Christel Baier, Nathalie Bertrand, Clemens Dubslaff, Daniel Gburek, and Ocan Sankur. Stochastic shortest paths and weight-bounded properties in markov decision processes (extended version). https://arxiv.org/abs/1804.11301, 2018.

[3] Christel Baier, Marcus Daum, Clemens Dubslaff, Joachim Klein, and Sascha Klüppelholz. Energy-utility quantiles. In NFM'14, LNCS 8430, p. 285–299. Springer, 2014.

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

[5] Dimitri P. Bertsekas and John N. Tsitsiklis. An analysis of stochastic shortest path problems. *Mathematics of Operations Research*, 16(3):580–595, 1991.

[6] Tomás Brázdil, Václav Brozek, Kousha Etessami, Antonín Kucera, and Dominik Wojtczak. One-counter Markov decision processes. In SODA'10, p. 863–874. SIAM, 2010.

[7] Tomás Brázdil, Antonín Kucera, and Petr Novotný. Optimizing the expected mean payoff in energy Markov decision processes. In ATVA'16, LNCS 9938, p. 32–49, 2016.

[8] Véronique Bruyère, Emmanuel Filiot, Mickael Randour, and Jean-François Raskin. Meet your expectations with guarantees: Beyond worst-case synthesis in quantitative games. *Information and Computation*, 254:259–295, 2017.

[9] Krishnendu Chatterjee and Laurent Doyen. Energy and mean-payoff parity Markov decision processes. In MFCS'11, LNCS 6907, p. 206–218. Springer, 2011.

[10] Krishnendu Chatterjee and Monika Henzinger. Faster and dynamic algorithms for maximal end-component decomposition and related graph problems in probabilistic verification. In SODA'11, p. 1318–1336. SIAM, 2011.

[11] Luca de Alfaro. *Formal Verification of Probabilistic Systems*. PhD thesis, Stanford University, Department of Computer Science, 1997.

[12] Luca de Alfaro. Computing minimum and maximum reachability times in probabilistic systems. In CONCUR'99, LNCS 1664, p. 66–81, 1999.

[13] Kousha Etessami, Dominik Wojtczak, and Mihalis Yannakakis. Quasi-birth-death processes, tree-like qbds, probabilistic 1-counter automata, and pushdown systems. In QEST'08, p. 243–253. IEEE Computer Society, 2008.

[14] Christoph Haase and Stefan Kiefer. The odds of staying on budget. In ICALP'15, LNCS 9135, p. 234–246. Springer, 2015.

[15] Lodewijk Kallenberg. *Markov Decision Processes*. Lecture Notes. University of Leiden, 2011.

[16] Daniel Krähmann, Jana Schubert, Christel Baier, and Clemens Dubslaff. Ratio and weight quantiles. In MFCS'15, LNCS 9234, p. 344–356. Springer, 2015.

[17] Richard Mayr, Sven Schewe, Patrick Totzke, and Dominik Wojtczak. MDPs with energy-parity objectives. In LICS'17, IEEE Computer Society, IEEE Computer Society, p. 1–12, 2017.

[18] Martin L. Puterman. *Markov Decision Processes: Discrete Stochastic Dynamic Programming*. John Wiley & Sons, Inc., New York, NY, 1994.

[19] Michael Ummels and Christel Baier. Computing quantiles in Markov reward models. In FoSSaCS'13, LNCS 7794, p. 353–368. Springer, 2013.