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.

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.

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.

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 [4, 11]. 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 [2, 18]. The corresponding decision problem is PSPACE-hard, even for acyclic MDPs [13].

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 [6, 8, 16] 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 [5], 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, [5] 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 [18].

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 [4, 11], while the general case was open. Second, we study 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 $\mathsf{DWR}^{\exists,=1}$) 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 $\mathsf{DWR}^{\forall,>0}$). In contrast, problem $\mathsf{DWR}^{\forall,=1}$ 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 $\mathsf{DWR}^{\exists,=1}$ and $\mathsf{DWR}^{\forall,>0}$ and polynomial for $\mathsf{DWR}^{\forall,=1}$. These results should be contrasted with the polynomial-time decidability of $\mathsf{DWR}^{\exists,=1}$ and $\mathsf{DWR}^{\forall,>0}$ for MDPs where all weights are nonnegative [18].

Although several other problems for integer-weighted MDPs are known to be in NP ∩ coNP and as hard as non-stochastic two-player mean-payoff games (see, *e.g.*, [7, 8, 16] 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 $\mathsf{DWR}^{\exists,=1}$ and $\mathsf{DWR}^{\forall,=1}$, 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 [14, 17]. 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 [5] 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 [10, 11] to eliminate 0-ECs in nonnegative MDPs simply by collapsing all states that belong to some maximal end component of the sub-MDP built by state-action pairs with weight 0. 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 $\mathsf{DWR}^{\exists,=1}$ and $\mathsf{DWR}^{\forall,=1}$ 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 Appendix.

2 Preliminaries

We briefly define our notations; for details see, *e.g.*, [3, 17].

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 $\pi = s_0\,\alpha_0\,s_1\,\alpha_1\,s_2\,\alpha_2 \ldots s_n$ or infinite $\varsigma = s_0\,\alpha_0\,s_1\,\alpha_1\,s_2\,\alpha_2\ldots$, such that for every index i, $\alpha_i \in Act(s_i)$. 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 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 finite or infinite 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 written $first(\pi)$ (resp. $last(\pi)$). If ς is infinite, $\lim(\varsigma)$ denotes 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 $\mathfrak{A}\colon T \to 2^{Act}$ 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 [10]. $\mathcal{E}$ 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 [9, 10]. 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 $\Pr^{\sup}_{\mathcal{M},s}(\varphi) = \sup_{\mathfrak{S}} \Pr^{\mathfrak{S}}_{\mathcal{M},s}(\varphi)$ and $\Pr^{\inf}_{\mathcal{M},s}(\varphi) = \inf_{\mathfrak{S}} \Pr^{\mathfrak{S}}_{\mathcal{M},s}(\varphi)$ 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, which 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, $\mathbb{E}^{\sup}_{\mathcal{M},s}(f) = \sup_{\mathfrak{S}} \mathbb{E}^{\mathfrak{S}}_{\mathcal{M},s}(f)$ 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).

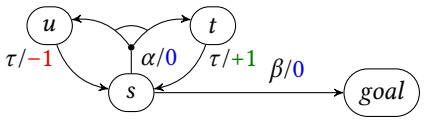


Figure 1. Gambling EC $\mathcal{E} = \{(s,\alpha),(u,\tau),(t,\tau)\}$. The MD scheduler that always takes (s,α) is gambling. Moreover, *goal* can be reached almost surely for any threshold using the infinite-memory scheduler taking (s,α) until the weight exceeds the threshold, and then (s,β). One can 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 $\mathcal{E}$. The term *0-BSCC* is used for a 0-EC containing at most one state-action pair (s,α) for each state s in $\mathcal{M}$. 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.*, [15])**.** *Let C be a strongly connected finite Markov chain.*

(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. (Lemmas B.8 and B.9 in the appendix):

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 (see App. B.3 for a proof):

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 a $\mathcal{E}$ 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* $\mathcal{N} = \mathrm{Spider}_{\mathcal{E},s_0}(\mathcal{M})$ (or shortly $\mathrm{Spider}_{\mathcal{E}}(\mathcal{M})$) 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 in $\mathcal{E}$ with $s \neq s_0$ and each 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, the 0-BSCC $\mathcal{E} = \{(s,\alpha),(t,\alpha),(u,\alpha)\}$ of $\mathcal{M}$ is chosen, obtaining $\mathcal{M}_1 = \mathrm{Spider}_{\mathcal{E},t}(\mathcal{M})$. Then, taking the 0-BSCC $\mathcal{F} = \{(s,\tau),(t,\beta),(u,\tau),(v,\beta),(w,\beta)\}$ of $\mathcal{M}_1$, we obtain $\mathcal{M}_2 = \mathrm{Spider}_{\mathcal{F},t}(\mathcal{M}_1)$. 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 $\mathrm{Spider}_{\mathcal{E}}(\mathcal{M})$, we define the notion of $\mathcal{E}$*-invariant properties*, which are properties that are invariant by adding or removing subpaths traversing $\mathcal{E}$. Given a path $\varsigma = t_0\,\alpha_0\,t_1\,\alpha_1\ldots$, 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 $t_i\,\alpha_i\ \ldots\ \alpha_{j-1}\,t_j\,\alpha_j\,t_{j+1}$ 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

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

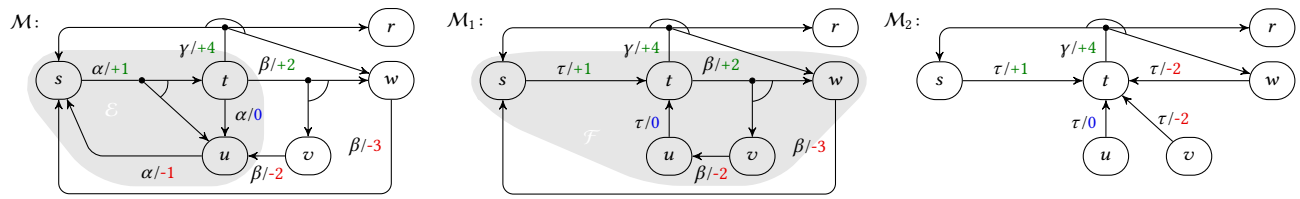


Figure 2. Illustration of the spider construction: $\mathcal{M}_1 = \text{Spider}_{\mathcal{E},t}(\mathcal{M})$ and $\mathcal{M}_2 = \text{Spider}_{\mathcal{F},t}(\mathcal{M}_1)$.

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 $\varsigma' \models \varphi$. Clearly, weight-divergence and the pumping property are $\mathcal{E}$-invariant properties, and so are properties of the form $\Diamond(t \wedge (\text{wgt} \bowtie K))$ 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* $\text{Spider}_{\mathcal{E}}(\mathcal{M})$ *that satisfies the following properties:*

(S1) $\mathcal{M}$ *and* $\text{Spider}_{\mathcal{E}}(\mathcal{M})$ *have the same state space and* $\|\text{Spider}_{\mathcal{E}}(\mathcal{M})\| = \|\mathcal{M}\|-1$.

(S2) If $\mathcal{E} \neq \mathcal{M}$ *and* $\mathcal{M}$ *is strongly connected then* $\text{Spider}_{\mathcal{E}}(\mathcal{M})$ *has a single MEC that is reachable from all states.*

(S3) $\mathcal{M}$ *and* $\text{Spider}_{\mathcal{E}}(\mathcal{M})$ *are equivalent for* $\mathcal{E}$*-invariant properties in the following sense:*

(S3.1) For each scheduler $\mathfrak{T}$ *for* $\text{Spider}_{\mathcal{E}}(\mathcal{M})$ *there is a scheduler* $\mathfrak{S}$ *for* $\mathcal{M}$ *with* $\Pr^{\mathfrak{S}}_{\mathcal{M},s}(\varphi) = \Pr^{\mathfrak{T}}_{\text{Spider}_{\mathcal{E}}(\mathcal{M}),s}(\varphi)$ *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* $\text{Spider}_{\mathcal{E}}(\mathcal{M})$ *such that*
$\Pr^{\mathfrak{S}}_{\mathcal{M},s}(\varphi) \leqslant \Pr^{\mathfrak{T}}_{\text{Spider}_{\mathcal{E}}(\mathcal{M}),s}(\varphi) \leqslant \Pr^{\mathfrak{S}}_{\mathcal{M},s}(\varphi) + p^{\mathfrak{S}}_s$
for all states s *and all* $\mathcal{E}$*-invariant properties* φ*. Here,* $p^{\mathfrak{S}}_s = \Pr^{\mathfrak{S}}_{\mathcal{M},s}\{\varsigma \in \mathit{IPaths} : \lim(\varsigma) = \mathcal{E}\}$.

(S4) Suppose that $\mathcal{E}$ *is contained in an MEC* $\mathcal{G}$ *of* $\mathcal{M}$ *with* $\mathbb{E}^{\max}_{\mathcal{G}}(\text{MP}) = 0$*. 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* $\text{Spider}_{\mathcal{E}}(\mathcal{M})$*. 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* $\text{Spider}_{\mathcal{E}}(\mathcal{M})$.

The proof is given in Appendix B.5.1. 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 $\text{Spider}_{\mathcal{E}}(\mathcal{M})$ with respect to $\mathcal{E}$-invariant properties. While any scheduler for $\text{Spider}_{\mathcal{E}}(\mathcal{M})$ 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* $\mathcal{E}$ *a 0-BSCC of* $\mathcal{M}$ *then* $\mathcal{M}$ *is weight-divergent (resp. pumping) iff* $\text{Spider}_{\mathcal{E}}(\mathcal{M})$ *is weight-divergent (resp. pumping).*

3.2 Checking Weight-Divergence

We present an algorithm to check the weight-divergence of an end component. 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 use a recursive approach that first computes $\mathbb{E}^{\max}_{\mathcal{M}}(\text{MP})$ and an MD-scheduler $\mathfrak{S}$ maximizing the expected mean payoff. If $\mathbb{E}^{\max}_{\mathcal{M}}(\text{MP}) > 0$ then $\mathcal{M}$ is pumping (Lemma 3.3) and therefore positively weight-divergent. If $\mathbb{E}^{\max}_{\mathcal{M}}(\text{MP}) < 0$ 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 $\mathbb{E}^{\max}_{\mathcal{M}}(\text{MP}) = 0$ 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). Pick such a 0-BSCC $\mathcal{E}$ of $\mathfrak{S}$ and apply the spider construction to generate the MDP $\mathcal{M}_1 = \text{Spider}_{\mathcal{E}}(\mathcal{M})$. If $\mathcal{M} = \mathcal{E}$ then $\mathcal{M}$ is a 0-EC, hence not weight-divergent, and the algorithm terminates. Otherwise $\mathcal{M} \neq \mathcal{E}$, and the MDP $\mathcal{M}_1$ contains a unique maximal end component $\mathcal{F}_1$ ((S2) in Lemma 3.7). We repeat the procedure for $\mathcal{F}_1$. The algorithm thus generates a sequence of MDPs $\mathcal{M}_0 = \mathcal{M}, \mathcal{M}_1, \ldots, \mathcal{M}_\ell$ with $\mathcal{M}_{i+1} = \text{Spider}_{\mathcal{E}_i}(\mathcal{M}_i)$ for some 0-BSCC $\mathcal{E}_i$ of $\mathcal{M}_i$. All $\mathcal{M}_i$'s have the same state space and 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 (Corollary 3.8).

As each iteration takes polynomial time, and the size of each $\mathcal{M}_i$ is polynomially bounded by the size of $\mathcal{M}$ (Lemma B.27) the algorithm runs in polynomial time. Using an inductive argument and Lemma 3.7 (see Appendix B.5.3):

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. to 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.*

As a consequence of this theorem, we obtain:

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.*

Note, however, that an MDP can have gambling schedulers, but no gambling MD-schedulers. Consider the MDP consisting of the 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 following theorem. The remaining part is provided in Appendix B.7.

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, [5, 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, [5] 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 [5] 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 to $\mathcal{M}'$.

This transformation is crucial in several results that follow. Detailed construction and the proof of the following theorem are given in Appendix Sections B.4.3 and B.4.4.

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. $\Pr^{\mathfrak{S}}_{\mathcal{M},s}\big(\Box(\mathrm{wgt} \geqslant K) \wedge \Box\Diamond s\big) = 1$ 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 (see Appendix B.6) 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 $\mathbb{E}^{\max}_{\mathcal{E}}(\text{MP}) > 0$ are pumping and therefore bounded from below. MECs $\mathcal{E}$ with either $\mathbb{E}^{\max}_{\mathcal{E}}(\text{MP}) < 0$ or $\mathbb{E}^{\max}_{\mathcal{E}}(\text{MP}) = 0$ 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 $\mathbb{E}^{\max}_{\mathcal{E}}(\text{MP}) = 0$ 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.*, [14]) 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 [4] 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 [4], 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 [11] 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, $\mathbb{E}^{\inf}_{\mathcal{M},s_{init}}(\diamondplus\, goal)$ 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 $\mathbb{E}^{\inf}_{\mathcal{M},s_{init}}(\diamondplus\, goal)$ 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 $\mathbb{E}^{\inf}_{\mathcal{M},s_{init}}(\diamondplus\, goal) > -\infty$ 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 Appendix B.5.2), we obtain in polynomial time an MDP $\mathcal{N}$ such that $\mathcal{N}$ satisfies condition (BT) and $\mathbb{E}^{\inf}_{\mathcal{N},s}(\diamondplus\, goal) = \mathbb{E}^{\inf}_{\mathcal{M},s}(\diamondplus\, goal)$ 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 $\mathbb{E}^{\inf}_{\mathcal{M},s_{init}}(\diamondplus\, goal)$ in polynomial time.*

Analogous results are obtained for maximal expected accumulated weights $\mathbb{E}^{\sup}_{\mathcal{M},s}(\diamondplus\, goal)$ by multiplying all weights with -1. Details of this section are given in Appendix C.

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 $\varphi = \bigvee_{t\in T} \Diamond\big(t \wedge (\text{wgt} \geqslant K_t)\big)$.

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

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

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. See Appendix D.1.

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 (Appendix D.2).

Theorem 5.2. *The decision problem $DWR^{\forall,>0}$ is in NP ∩ coNP, and hard for (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. 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

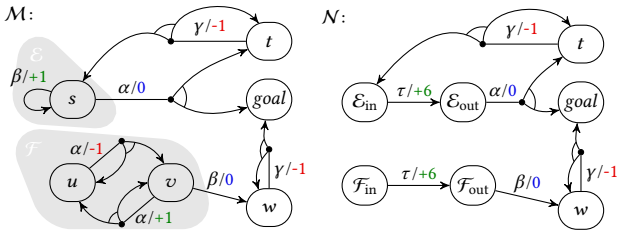


Figure 3. Resolution of $\mathrm{DWR}^{\exists,=1}$ on an example.

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 hard for (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 $\Pr^{\max}_{\mathcal{N},\mathcal{E}_{out}}(\varphi_K[X]) = 1$ 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 $\Pr^{\max}_{\mathcal{M},s}(\Diamond(T^* \cup \mathit{GoodEC})) = 1$. For other states, we show that the maximal K such that $\Pr^{\max}_{\mathcal{N},s}(\phi_K[\mathit{GoodEC}]) = 1$ corresponds to $K^{\exists,=1}_{\mathcal{M},s'}$ where s and s' are corresponding states. Here, $\phi_K[\mathit{GoodEC}]$ is an instance of $\mathsf{DWR}^{\exists,=1}$and $\mathcal{N}$ has no weight-divergent end components, so we can use the NP ∩ coNP algorithm described at the beginning. Details are given in Appendix D.4. □

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

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

and the corresponding problems $\mathsf{WB}^{\forall,=1}$ and $\mathsf{WB}^{\forall,>0}$ 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 $\mathsf{DWR}^{\exists,>0}$ and $\mathsf{DWR}^{\exists,=1}$, respectively. For details see Appendix D.5.

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

The proof of Theorem 5.6 heavily uses the concepts of Section 3. Let us briefly describe the reduction of $\mathsf{WB}^{\exists,=1}$ and $\mathsf{WB}^{\exists,>0}$ to $\mathsf{DWR}^{\exists,=1}$ and $\mathsf{DWR}^{\exists,>0}$ 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 $\mathsf{WB}^{\exists,=1}$ or $\mathsf{WB}^{\exists,>0}$.

To solve problem $\mathsf{WB}^{\forall,=1}$ we rely on the observation that $\mathsf{WB}^{\forall,=1}$ 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 $\mathsf{DWR}^{\exists,>0}$ 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 $K_t = L{-}\mathit{rec}(t)$. 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, $\mathit{rec}(t)$ refers to this maximal end component $\mathcal{E}$.

For problem $\mathsf{WB}^{\forall,=1}$ we transform $\mathcal{M}^-$ into a new MDP $\mathcal{N}$ such that $\mathsf{WB}^{\forall,=1}$ 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 $\mathsf{DWR}^{\exists,=1}$ for $\mathcal{N}$ and the same DWR property as for $\mathsf{DWR}^{\forall,>0}$. 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 $\mathsf{WB}^{\exists,=1}$ and $\mathsf{WB}^{\forall,>0}$ are computable in pseudo-polynomial time, and optimal weight bounds for $\mathsf{WB}^{\exists,>0}$ and $\mathsf{WB}^{\forall,=1}$in polynomial time.

5.3 Discussion on Related Work

To the best of our knowledge, problems $\mathsf{DWR}^{\exists,=1}$, $\mathsf{DWR}^{\forall,>0}$ and $\mathsf{DWR}^{\forall,=1}$ 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 [18]. 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 [15]. 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 [8, 16]. Rephrased for our notations, the energy-MDP problem is to check whether $\Pr^{\max}_{\mathcal{M},s}\big(\Box(\mathrm{wgt} \geqslant K) \wedge \phi\big) = 1$ where φ is a parity condition and $K \in \mathbb{Z}$. This problem is in NP ∩ coNP and hard for 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, $\mathsf{DWR}^{\forall,>0}$ asks whether $\Pr^{\min}_{\mathcal{M},s}\big(\Diamond(goal \wedge (\text{wgt} \geqslant K))\big) > 0$, in the case $T^* = \varnothing$ and $T \setminus T^* = \{goal\}$. Given the similarities of these questions, and our decision procedure that reduces $\mathsf{DWR}^{\forall,>0}$ to mean-payoff Büchi games, it is no surprise that the problem $\mathsf{DWR}^{\forall,>0}$ is hard for mean-payoff games.

Nevertheless, the instances $\mathsf{DWR}^{\exists,=1}$ and $\mathsf{DWR}^{\forall,=1}$ are of different nature than energy-MDPs. These can rather be seen as variants of the *termination problem* for *one-counter MDPs* [5, 12]. 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 [5], 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 [5] in the paragraph following Theorem 3.11. For one-counter MDPs $\mathfrak{M}$ with boundary, [5] 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 ∩ coNP upper bound for $\mathsf{DWR}^{\exists,=1}$ 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 [13] which shows that for MDPs $\mathcal{M}$ with non-negative integer weights and $\Pr^{\min}_{\mathcal{M},s}(\Diamond goal) = 1$, checking whether $\Pr^{\max}_{\mathcal{M},s}(\Diamond(goal \wedge (\text{wgt} = K))) = 1$ 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 [5, 12].

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

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

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

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

[6] 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.

[7] 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.

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

[9] 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.

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

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

[12] 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.

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

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

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

[16] 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.

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

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

Appendix

In the appendix, we provide the proofs of the results of the main paper that had to be omitted due to space constraints.

Contents - Overview of the Paper

Literature Referred to in the Appendix

[19] Christel Baier, Joachim Klein, Sascha Klüppelholz, and Sascha Wunderlich. Maximizing the conditional expected reward for reaching the goal. In TACAS'17, LNCS 10206, p. 269–285. Springer, 2017.

[20] Gilles Brassard. A note on the complexity of cryptography (corresp.). *Information Theory, IEEE Transactions on*, 25:232 – 233, 04 1979.

[21] Krishnendu Chatterjee and Laurent Doyen. Energy parity games. *Theoretical Computer Science*, 458:49–60, 2012.

[22] D.A. Martin. The determinacy of blackwell games. *The Journal of Symbolic Logic*, 63(4):1565–1581, 1998.

[23] 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.

[24] Mickael Randour, Jean-François Raskin, and Ocan Sankur. Variations on the stochastic shortest path problem. In VMCAI'15, LNCS 8931, p. 1–18. Springer, 2015.

A Additional Notations

Notations for paths (concatenation, length, fragments). Given a finite path $\pi' = s_0\alpha_0 \dots \alpha_n s_n$ and a (finite or infinite) path $\pi = t_0\beta_0 t_1\beta_1 \dots$ with $s_n = t_m$ then $\pi'; \pi$ denotes the path $s_0\alpha_0 \dots \alpha_n s_n\beta_0 t_1\beta_1 \dots$. The length of a path π, denoted $|\pi|$, is defined as the number of state-action pairs in π, *i.e.*, if π' is a finite path as above then $|\pi'| = n$, while $|\pi| = \infty$ for each infinite path. If π is a path as above and $i, n \in \mathbb{N}$ with $i \leqslant n \leqslant |\pi|$ then $\pi[n]$ denotes t_n (the $(n+1)$-st state of π) and $\pi[i \dots n]$ the finite path $t_i\ \beta_i\ t_{i+1}\ \beta_{i+1}\ \dots\ \beta_{n-1}\ t_n$. Thus, $\pi[0 \dots n] = \mathit{pref}(\pi, n)$, $\pi[n \dots n] = \pi[n]$ and $\mathit{first}(\pi) = \pi[0]$. Similarly, $\mathit{last}(\pi) = \pi[n]$ if $n = |\pi|$ is finite.

Residual schedulers. Let $\mathfrak{S}$ be a scheduler and π a finite path. The residual scheduler $\mathfrak{S}{\uparrow}\pi$ is defined by $(\mathfrak{S}{\uparrow}\pi)(\pi') = \mathfrak{S}(\pi; \pi')$ if $\mathit{first}(\pi') = \mathit{last}(\pi)$, and $(\mathfrak{S}{\uparrow}\pi)(\pi') = \mathfrak{S}(\pi')$ otherwise.

Finite-memory scheduler. A finite memory scheduler can defined as follows. Given a finite set M of memory elements, $\mathfrak{S}_u(s, m) = m'$ is an *update function* that determines the new memory element given current state s and current memory element m; and $\mathfrak{S}_n(s, m) = \alpha$ determines the action to be played at state s and if the memory contains m.

Markov chain. A *Markov chain* is an MDP $\mathcal{M} = (S, \mathit{Act}, P, \mathit{wgt})$ where Act is a singleton. We occasionally use the notation $C = (S, P', \mathit{wgt}')$ for $\mathcal{M}$, where $P' \colon S \times S \to [0, 1]$ and $\mathit{wgt}' \colon S \to \mathbb{Z}$ are defined as P and wgt but omitting the uniquely defined action.

Limit of infinite paths. The limit of an infinite path ς, denoted $\lim(\varsigma)$, is the set of state-action pairs that occur infintely often in ς. If $\mathcal{E}$ is an end component then we often write $\mathit{Limit}_{\mathcal{E}}$ for $\{\varsigma \in \mathit{IPaths} : \lim(\varsigma) = \mathcal{E}\}$. At various places, we rely on De Alfaro's result [10] stating that for each scheduler $\mathfrak{S}$, the limit of almost all infinite $\mathfrak{S}$-paths is an end component. Formally, for each scheduler $\mathfrak{S}$ and each state s, we have $\Pr^{\mathfrak{S}}_{\mathcal{M},s}\big(\bigcup_{\mathcal{E}} \mathit{Limit}_{\mathcal{E}}\big) = 1$ when $\mathcal{E}$ ranges over all (possibly nonmaximal) end components.

Probabilities. Recall that we use the notation $\Pr^{\max}_{\mathcal{M},s}(\varphi)$ when there exists a scheduler $\mathfrak{S}$ for $\mathcal{M}$ such that $\Pr^{\mathfrak{S}}_{\mathcal{M},s}(\varphi) = \Pr^{\sup}_{\mathcal{M},s}(\varphi)$.

If π is a finite path starting in s and $\mathfrak{S}$ a scheduler then $\Pr^{\mathfrak{S}}_{\mathcal{M},s}(\pi)$ is used as a shortform notation for $\Pr^{\mathfrak{S}}_{\mathcal{M},s}\big(\mathit{Cyl}(\pi)\big)$ where $\mathit{Cyl}(\pi)$ denotes the cylinder set of π, *i.e.*, the set of all maximal paths ς where π is a prefix of ς.

Properties. Let $\mathcal{M}$ be an MDP with state space S. Define $\Lambda_{\mathcal{M}} = (S \times \mathbb{Z})^{\omega} \cup (S \times \mathbb{Z})^{*} \times S$. The set $\Lambda_{\mathcal{M}}$ is equipped with the (standard) sigma-algebra generated by the cylinder sets of the finite sequences in $(S \times \mathbb{Z})^{*}$. A property is a measurable subset of $\Lambda_{\mathcal{M}}$. Of course, each path $t_0\,\alpha_0\,t_1\,\alpha_1 \dots$ in $\mathcal{M}$ naturally induces such a sequence in $\Lambda_{\mathcal{M}}$ by replacing α_i with $\mathit{wgt}(t_i, \alpha_i)$. Denote the function mapping every $t_0\,\alpha_0\,t_1\,\alpha_1 \dots$ to $t_0\ \mathit{wgt}(t_0, \alpha_0)\ t_1\ \mathit{wgt}(t_1, \alpha_1) \dots$ by f. Hence, for each scheduler $\mathfrak{S}$ and each state s the probability measure $\Pr^{\mathfrak{S}}_{\mathcal{M},s}$ on induces a probability measure $\Pr^{\mathfrak{S}}_{\mathcal{M},s,\sharp}$ on $\Lambda_{\mathcal{M}}$. Formally, for every property φ we have $\Pr^{\mathfrak{S}}_{\mathcal{M},s,\sharp}(\varphi) = \Pr^{\mathfrak{S}}_{\mathcal{M},s}\big(\{\pi \ :\ f(\pi) \in \varphi\}\big)$. To simplify notions, we identify the two probability measures $\Pr^{\mathfrak{S}}_{\mathcal{M},s}$ and $\Pr^{\mathfrak{S}}_{\mathcal{M},s,\sharp}$, *i.e.*, for every property φ we write $\Pr^{\mathfrak{S}}_{\mathcal{M},s}(\varphi)$ rather than $\Pr^{\mathfrak{S}}_{\mathcal{M},s,\sharp}(\varphi)$.

B Proofs and Complements for Section 3

B.1 Illustration of the Notions on End Components

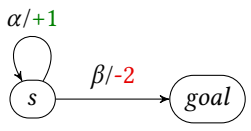


Figure 4. MDP with pumping EC $\mathcal{E} = \{(s, \alpha)\}$.

Example B.1 (Pumping EC)**.** Let $\mathcal{M}$ be the simple MDP depicted in Figure 4 consisting of the two states s and *goal*, probabilistic transitions $\Pr(s,\alpha,s) = 1$ and $\Pr(s,\beta,goal) = 1$ with weights $wgt(s,\alpha) = +1$ and $wgt(s,\beta) = -2$. The pair (s,α) constitute a maximal end component of $\mathcal{M}$ that is trivially pumping.

This example also illustrates that no MD-scheduler can ensure $\Diamond(goal \wedge (\mathrm{wgt} \geqslant 0))$ almost surely. Indeed, if $\mathfrak{S}$ is the scheduler that takes action α twice in s and then action β to move to *goal*, then $\Pr^{\mathfrak{S}}_{\mathcal{M},s}\big(\Diamond(goal \wedge (\mathrm{wgt} \geqslant 0))\big) = 1$. However, there is no MD-scheduler $\mathfrak{T}$ satisfying $\Pr^{\mathfrak{T}}_{\mathcal{M},s}\big(\Diamond(goal \wedge (\mathrm{wgt} \geqslant 0))\big) = 1$. ∎

B.2 Mean Payoff in Strongly Connected Markov Chains

Let us start by recalling some simple observations on Markov chains.

Lemma B.2 (Folkore)**.** *For each finite strongly connected Markov chain C:*

(a) If $\mathbb{E}_C(\mathrm{MP}) > 0$ then C is positively pumping.
(b) If $\mathbb{E}_C(\mathrm{MP}) < 0$ then C is negatively pumping.

In what follows, if $s \in S$ then we write "*wgt* until s" to denote the random variable that assigns to each infinite path ς the accumulated weight $wgt(\pi)$ of the shortest prefix π of ς with $last(\pi) = s$, provided that $\varsigma \models \Diamond s$. Note that "*wgt* until s" agrees with the random variable $\oplus s$ defined in the core of the paper; we use here "*wgt* until s" instead, and also "steps until s" defined below. For the infinite paths ς with $\varsigma \not\models \Diamond s$, "*wgt* until s" is undefined. Thus, if $s,t \in S$ and $\mathfrak{S}$ is a scheduler with $\Pr^{\mathfrak{S}}_{\mathcal{M},t}(\Diamond s) = 1$ then $\mathbb{E}^{\mathfrak{S}}_t(\text{"}wgt \text{ until } s\text{"})$ stands for the expected accumulated weight from t until reaching s. Similarly, "steps until s" denotes the random variable counting the number of steps until reaching state s. Thus, if $\Pr^{\mathfrak{S}}_{\mathcal{M},t}(\Diamond s) = 1$ then $\mathbb{E}^{\mathfrak{S}}_t(\text{"steps until } s\text{"})$ stands for the expected number of steps from t until reaching s with respect to scheduler $\mathfrak{S}$.

Lemma B.3 (Quotient representation of expected mean payoff in MCs)**.** *Let $C = (S, P, wgt)$ be a strongly connected Markov chain. Then, for each state s in C we have:*

$$\mathbb{E}_C(\mathrm{MP}) \;=\; \frac{wgt(s) \;+\; \sum_{t\in S} P(s,t)\cdot \mathbb{E}_{C,t}(\text{"}wgt \text{ until } s\text{"})}{1 \;+\; \sum_{t\in S} P(s,t)\cdot \mathbb{E}_{C,t}(\text{"steps until } s\text{"})}$$

Proof. The statement is a consequence of the well-known fact that for (finite-state) strongly connected Markov chains, the long-run frequencies of almost all paths converge to the steady-state probabilities. (We suppose here the definition of the steady-state probability of state s as the Cesàro limit $\lim_{n\to\infty} \frac{1}{n+1}\cdot\sum_{i=0}^{n} \Pr_{C,\iota}(\bigcirc^i s)$ where ι is an arbitrary initial distribution.) □

B.3 The Pumping Property in MDPs

We now provide the proofs for Lemma 3.3 and Lemma 3.4.

Lemma B.4. *Each strongly connected MDP $\mathcal{M}$ with $\mathbb{E}^{\max}_{\mathcal{M}}(\mathrm{MP}) > 0$ has a pumping MD-scheduler.*

Proof. Let $\mathfrak{S}$ be an MD-scheduler for $\mathcal{M}$ that maximizes the expected mean payoff and where the induced Markov chain has a single BSCC. Then:

$$\lim_{n\to\infty} \frac{1}{n} wgt(pref(\varsigma,n)) \;=\; \mathbb{E}^{\mathfrak{S}}_{\mathcal{M}}(\mathrm{MP}) \;>\; 0$$

for almost all infinite $\mathfrak{S}$-paths ς. Hence, $\mathfrak{S}$ is pumping. □

Recall that that no scheduler can exceed the maximum expected mean payoff in MDPs:

Lemma B.5 (Folklore – see, *e.g.*, [17])**.** *Let $\mathcal{M}$ be a strongly connected MDP. Then for each scheduler $\mathfrak{S}$ and each state s*

$$\Pr^{\mathfrak{S}}_{\mathcal{M},s}\Big\{\, \varsigma \in IPaths \;:\; \limsup_{n\to\infty} \frac{1}{n} wgt(pref(\varsigma,n)) \;\leqslant\; \mathbb{E}^{\max}_{\mathcal{M}}(\mathrm{MP}) \,\Big\} \;=\; 1 \;.$$

By the results of de Alfaro [10], for each scheduler $\mathfrak{S}$, the limit of almost all infinite $\mathfrak{S}$-paths is an end component. Recall that $\lim(\varsigma)$, the limit of an infinite path ς, is the set of state-action pairs that occur infinitely often in ς. Hence, we get (see, *e.g.*, [24]):

Lemma B.6. *Let $\mathcal{M}$ be a strongly connected MDP such that for some state s_0 in $\mathcal{M}$*

$$\Pr^{\max}_{\mathcal{M},s_0}\Big\{\, \varsigma \in IPaths \;:\; \limsup_{n\to\infty}\; wgt(pref(\varsigma,n)) = +\infty \,\Big\} \;>\; 0 \;.$$

Then $\mathbb{E}^{\max}_{\mathcal{M}}(\mathrm{MP}) \geqslant 0$.

Corollary B.7. *Let $\mathcal{M}$ be a strongly connected MDP $\mathcal{M}$. Then:*

(a) If $\mathbb{E}^{\max}_{\mathcal{M}}(\mathrm{MP}) < 0$ then $\mathcal{M}$ is universally negatively pumping, i.e., for each scheduler $\mathfrak{S}$ and each state s:

$$\Pr^{\mathfrak{S}}_{\mathcal{M},s}\big\{\, \varsigma \in \mathit{IPaths} \;:\; \limsup_{n\to\infty}\; \mathit{wgt}(\mathit{pref}(\varsigma,n)) \;=\; -\infty \big\} \;=\; 1$$

(b) If $\mathbb{E}^{\min}_{\mathcal{M}}(\mathrm{MP}) > 0$ then $\mathcal{M}$ is universally pumping, i.e., for each scheduler $\mathfrak{S}$ and each state s:

$$\Pr^{\mathfrak{S}}_{\mathcal{M},s}\big\{\, \varsigma \in \mathit{IPaths} \;:\; \liminf_{n\to\infty}\; \mathit{wgt}(\mathit{pref}(\varsigma,n)) \;=\; +\infty \big\} \;=\; 1$$

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.*

Proof. Part (a) follows from Lemma B.4 as each pumping scheduler is weight-divergent. Part (b) is an immediate consequence of Corollary B.7. □

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 $\mathbb{E}^{\min}_{\mathcal{M}}(\mathrm{MP}) > 0$.*

For the proof, the statement of Lemma 3.3 is split into the following two Lemmas B.8 and B.9:

Lemma B.8. *Let $\mathcal{M}$ be a strongly connected MDP. Then, the following three statements are equivalent:*

(i) $\mathcal{M}$ is pumping.
(ii) $\mathcal{M}$ has a pumping MD-scheduler.
(iii) $\mathbb{E}^{\max}_{\mathcal{M}}(\mathrm{MP}) > 0$.

Proof. The implication "(ii) $\Longrightarrow$ (i)" is trivial, while "(iii) $\Longrightarrow$ (ii)" has been shown in Lemma B.4. It remains to prove the implication "(i) $\Longrightarrow$ (iii)". Suppose $\mathcal{M}$ is pumping and let $\mathfrak{S}$ be a pumping scheduler. Then Corollary B.7 (a) implies $\mathbb{E}^{\max}_{\mathcal{M}}(\mathrm{MP}) \geqslant 0$. Let Γ denote the set of pumping paths in $\mathcal{M}$, *i.e.*,

$$\Gamma \;=\; \big\{\, \varsigma \in \mathit{IPaths} \;:\; \liminf_{n\to\infty}\; \mathit{wgt}(\mathit{pref}(\varsigma,n)) = +\infty \;\big\}\;.$$

If $\mathcal{E}$ is an end component then let $\Gamma_{\mathcal{E}} = \{\varsigma \in \Gamma : \lim(\varsigma) = \mathcal{E}\}$. For each end component $\mathcal{E}$ where $\Pr^{\mathfrak{S}}_{\mathcal{M},s_0}(\Gamma_{\mathcal{E}}) > 0$ for some state s_0 we have: $\sup_{\mathfrak{S}'} \Pr^{\mathfrak{S}'}_{\mathcal{M},s}(\Gamma_{\mathcal{E}}) = 1$, where s is an arbitrary state in $\mathcal{E}$ and $\mathfrak{S}'$ ranges over all residual schedulers $\mathfrak{S}{\restriction}\pi$ of $\mathfrak{S}$ with $\mathit{first}(\pi) = s_0$ and $\mathit{last}(\pi) = s$.

We pick an end component $\mathcal{E}$ such that $\Pr^{\mathfrak{S}}_{\mathcal{M},s_0}(\Gamma_{\mathcal{E}}) > 0$ for some state s_0 and an MD-scheduler $\mathfrak{U}$ for $\mathcal{E}$ with a single BSCC $\mathcal{B}$ such that $\mathbb{E}^{\mathfrak{U}}_{\mathcal{E}}(\mathrm{MP}) \geqslant 0$. (Note that $\mathbb{E}^{\max}_{\mathcal{E}}(\mathrm{MP}) \geqslant 0$ by Lemma B.5.) Let s be a state of $\mathcal{B}$ and $E = \min_{t\in\mathcal{E}} \mathbb{E}^{\mathfrak{U}}_{\mathcal{E},t}(\text{“}\mathit{wgt}\text{ until }s\text{”})$. We choose a positive integer $\Delta \in \mathbb{N}$, $p \in]0,1[$ and a residual scheduler $\mathfrak{S}'$ of $\mathfrak{S}$ such that:

$$p\Delta + (1{-}p)E \;>\; 0 \qquad \text{and} \qquad \Pr^{\mathfrak{S}'}_{\mathcal{M},s}\Big(\; \bigcirc\big(\Diamond(s \wedge (\mathrm{wgt} \geqslant \Delta))\big)\Big) \;>\; p\;.$$

Let Π denote the set of $\mathfrak{S}'$-paths π from s to s such that $\mathit{wgt}(\pi) \geqslant \Delta$. For $n \in \mathbb{N}$, let Π_n be the set of paths $\pi \in \Pi$ such that $|\pi| \leqslant n$. We pick some $n \in \mathbb{N}$ such that $\sum_{\pi\in\Pi_n} \Pr^{\mathfrak{S}'}_{\mathcal{M},s}(\pi) \;>\; p$ (possible due to the choice of Δ, p, and $\mathfrak{S}'$).

Let $\mathfrak{T}$ be the following scheduler operating in two modes: normal mode and recovery mode. In its normal mode, $\mathfrak{T}$ attempts to generate a path in Π_n by mimicking $\mathfrak{S}'$. If it fails, *i.e.*, if the path π that has been generated since the last switch from recovery to normal mode is not a prefix of some path $\pi \in \Pi_n$, then $\mathfrak{T}$ switches to recovery mode where it behaves as $\mathfrak{U}$ until state s has been reached. As soon as s has been reached in recovery mode, $\mathfrak{T}$ switches back to normal mode and attempts to generate a path $\pi \in \Pi_n$. If $\mathfrak{T}$ in normal mode has generated a path $\pi \in \Pi_n$ then it keeps in normal mode and restarts to attempt to generate a path in Π_n.

This scheduler $\mathfrak{T}$ is pumping and the memory requirements are finite (as Π_n is finite). According to Lemma B.3 applied to the Markov chain induced by $\mathfrak{T}$, we obtain:

$$\mathbb{E}^{\mathfrak{T}}_{\mathcal{M},s}(\mathrm{MP}) \;\geqslant\; \frac{p\Delta + (1{-}p)E}{c} \;>\; 0,$$

where c is the expected number of steps under $\mathfrak{T}$ to return to s from s in normal mode. Hence, $\mathbb{E}^{\max}_{\mathcal{M}}(\mathrm{MP}) > 0$. □

Lemma B.9. *Let $\mathcal{M}$ be a strongly connected MDP. Then, $\mathcal{M}$ is universally pumping iff $\mathbb{E}^{\min}_{\mathcal{M}}(\mathrm{MP}) > 0$.*

Proof. The implication "$\Longleftarrow$" has been stated in Corollary B.7 (b).

To prove "$\Longrightarrow$", we assume that the minimal expected mean payoff in $\mathcal{M}$ is nonpositive, (*i.e.*, $\mathbb{E}^{\min}_{\mathcal{M}}(\mathrm{MP}) \leqslant 0$) and show that $\mathcal{M}$ is not universally pumping. Pick an MD-scheduler $\mathfrak{S}$ minimizing the expected mean payoff in $\mathcal{M}$. Then, $\mathbb{E}^{\mathfrak{S}}_{\mathcal{M}}(\mathrm{MP}) \leqslant 0$. But then, the limit $\lim_{n\to\infty}\ wgt(pref(\varsigma, n))$ exists for almost all $\mathfrak{S}$-paths and equals $\mathbb{E}^{\mathfrak{S}}_{\mathcal{M}}(\mathrm{MP})$. Hence, $\mathcal{M}$ is not universally pumping. □

B.4 Properties of 0-ECs

We establish some properties of 0-ECs necessary to prove Theorems 3.12 and to establish an algorithm for computing all states belonging to some 0-EC (Lemma 3.13).

B.4.1 Maximal 0-ECs

Let $\mathcal{E}_1$ and $\mathcal{E}_2$ be two 0-ECs. If $\mathcal{E}_1$ and $\mathcal{E}_2$ are weight-divergent, then so is $\mathcal{E}_1 \cup \mathcal{E}_2$. Thus, any weight-divergent end component is contained in a maximal end-component that is weight-divergent. The same holds for pumping end components or for gambling end components.

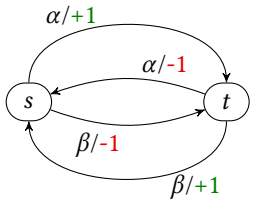


Figure 5. An example showing that the union of 0-ECs is not a 0-EC.

However, in general, the union of 0-ECs is not a 0-EC, and there are MDPs with maximal end components that are not a 0-EC but contain 0-ECs. This is the case of the MDP $\mathcal{M}$ depicted in Figure 5. The union of the 0-ECs $\mathcal{E}_\alpha = \{(s,\alpha),(t,\alpha)\}$ and $\mathcal{E}_\beta = \{(s,\beta),(t,\beta)\}$ is not a 0-EC. However, we have:

Lemma B.10 (Union of 0-ECs)**.** *Let $\mathcal{M}$ be a strongly connected MDP with $\mathbb{E}^{\max}_{\mathcal{M}}(\mathrm{MP}) = 0$, and let $\mathcal{E}_1$ and $\mathcal{E}_2$ be 0-ECs that have at least one common state. Then, $\mathcal{E}_1 \cup \mathcal{E}_2$ is a 0-EC too.*

Proof. Let $i \in \{1,2\}$. As $\mathcal{E}_i$ is a 0-EC, for each pair (s,t) of states in $\mathcal{E}_i$ there exists an integer $w_i(s,t)$ such that all paths from s to t in $\mathcal{E}_i$ have weight $w_i(s,t)$. Clearly, $w_i(t,s) = -w_i(s,t)$.

The union $\mathcal{E} = \mathcal{E}_1 \cup \mathcal{E}_2$ is an end component of $\mathcal{M}$. Hence, $\mathbb{E}^{\max}_{\mathcal{E}}(\mathrm{MP}) = 0$. Suppose by contradiction that $\mathcal{E}$ is not a 0-EC. Then, there exist two states s and t in $\mathcal{E}$ such that $w_1(s,t) \neq w_2(s,t)$, say $w_1(s,t) > w_2(s,t)$. For $i = 1,2$, let $\mathfrak{S}_i$ be an MD-scheduler for $\mathcal{E}_i$ such that s and t belong to a BSCC $\mathcal{B}_i$ of $\mathfrak{S}_i$. We now combine $\mathfrak{S}_1$ and $\mathfrak{S}_2$ to a scheduler $\mathfrak{S}$ for $\mathcal{E}$. $\mathfrak{S}$ alternates between two modes. Starting in mode 1, $\mathfrak{S}$ behaves as $\mathfrak{S}_1$ until state t is reached. It then switches to mode 2 where $\mathfrak{S}$ behaves as $\mathfrak{S}_2$ until state s is reached, in which case it switches back to mode 1.

$\mathfrak{S}$ is a finite-memory scheduler. Hence, it induces a finite Markov chain $C_{\mathfrak{S}}$. We now apply Lemma B.3 to $C_{\mathfrak{S}}$ and obtain:

$$\mathbb{E}^{\mathfrak{S}}_{\mathcal{E}}(\mathrm{MP}) \;=\; \frac{wgt(s) \;+\; \sum_{u\in S} P(s,\alpha,u)\cdot \mathbb{E}^{\mathfrak{S}}_{\mathcal{E},u}(\text{“}wgt \text{ until } s\text{”})}{1 \;+\; \sum_{u\in S} P(s,\alpha,u)\cdot \mathbb{E}^{\mathfrak{S}}_{\mathcal{E},u}(\text{“steps until } s\text{”})}$$

where $\alpha = \mathfrak{S}(s)$. By definition of $\mathfrak{S}$, for each state u with $P(s,\alpha,u) > 0$:

$$\mathbb{E}^{\mathfrak{S}}_{\mathcal{E},u}(\text{“}wgt \text{ until } s\text{”}) \;=\; \mathbb{E}^{\mathfrak{S}_1}_{\mathcal{E},u}(\text{“}wgt \text{ until } t\text{”}) \;+\; \mathbb{E}^{\mathfrak{S}_2}_{\mathcal{E},t}(\text{“}wgt \text{ until } s\text{”}) = w_1(s,t) \;+\; w_2(t,s) \;=\; w_1(s,t) \;-\; w_2(s,t) \;>\; 0\ .$$

Hence, $\mathbb{E}^{\mathfrak{S}}_{\mathcal{E}}(\mathrm{MP}) > 0$, which contradicts $\mathbb{E}^{\max}_{\mathcal{M}}(\mathrm{MP}) = 0$. □

Thus, each MDP as in Lemma B.10 has finitely many *maximal 0-ECs*, *i.e.*, 0-ECs $\mathcal{E}$ such that there is no 0-EC $\mathcal{E}'$ with $\mathcal{E} \subseteq \mathcal{E}'$ and $\mathcal{E} \neq \mathcal{E}'$. Maximal 0-ECs are non-overlapping in the sense that they do not share any state. A further consequence of the existence of maximal 0-ECs is that whenever $\mathcal{E}_1, \mathcal{E}_2$ are 0-ECs that share two states s and t then the weights of all paths from s to t in $\mathcal{E}_1$ and $\mathcal{E}_2$ are the same.

B.4.2 Criterion for 0-BSCCs

We start by the following observation for strongly connected Markov chains.

Lemma B.11 (Criterion for 0-BSCCs). *Let C be a strongly connected Markov chain with $\mathbb{E}_C(\mathrm{MP}) = 0$ such that*

$$\mathbb{E}_{C,t}(\text{“}wgt \text{ until } s\text{”}) \;=\; \mathbb{E}_{C,u}(\text{“}wgt \text{ until } s\text{”}) \tag{$\dagger$}$$

for all states s, t, u in C with $P(s,t) > 0$, $P(s,u) > 0$. Then, $wgt(\xi) = 0$ for all cycles ξ in C.

Proof. Using Lemma B.3, condition (†) together with $\mathbb{E}_C(\mathrm{MP}) = 0$ implies that

$$\mathbb{E}_{C,t}(\text{“}wgt \text{ until } s\text{”}) \;=\; -wgt(s) \tag{C1}$$

for all states s, $t \in C$ with $P(s,t) > 0$.

Let $Post(v) = \{v' \in S : P(v,v') > 0\}$ denote the set of direct successors of a state v. States v_1, v_2 are called *siblings* if there is some state $v \in S$ such that $v_1, v_2 \in Post(v)$. We now show that (†) propagates to arbitrary siblings, *i.e.*, whenever s, v_1, v_2 are states in C then:

$$\mathbb{E}_{C,v_1}(\text{“}wgt \text{ until } s\text{”}) \;=\; \mathbb{E}_{C,v_2}(\text{“}wgt \text{ until } s\text{”}) \quad \text{if } v_1, v_2 \text{ are siblings} \;. \tag{C2}$$

Suppose by contradiction that $s, v, v_1, v_2 \in S$ are states such that $v_1, v_2 \in Post(v)$ and

$$\mathbb{E}_{C,v_1}(\text{“}wgt \text{ until } s\text{”}) \;>\; \mathbb{E}_{C,v_2}(\text{“}wgt \text{ until } s\text{”}) \;.$$

By assumption (†), we have $s \neq v$.

Let C' be the Markov chain that has the same graph and weight structure as C, but $P'(v,v_1) = P(v,v_1) + \delta$, $P'(v,v_2) = P(v,v_2) - \delta$ for some δ with $0 < \delta < \min\{1 - P(v,v_1), P(v,v_2)\}$ and $P'(\cdot) = P(\cdot)$ in all other cases. As only the probabilities of the transitions from v are modified, we get for all states t

$$\mathbb{E}_{C',t}(\text{“}wgt \text{ until } v\text{”}) \;=\; \mathbb{E}_{C,t}(\text{“}wgt \text{ until } v\text{”}) \;.$$

In particular, if $t \in Post(v)$ then $\mathbb{E}_{C',t}(\text{“}wgt \text{ until } v\text{”}) = -wgt(v)$ where we use (C1). Hence:

$$\sum_{t \in S} P'(v,t) \cdot \mathbb{E}_{C',t}(\text{“}wgt \text{ until } v\text{”}) \;=\; \sum_{t \in Post(v)} P'(v,t) \cdot \underbrace{\mathbb{E}_{C,t}(\text{“}wgt \text{ until } v\text{”})}_{=-wgt(v)} \;=\; -wgt(v) \;.$$

But then the quotient representation of the mean payoff in C' applied to state v (Lemma B.3) yields $\mathbb{E}_{C'}(\mathrm{MP}) = 0$.

We now show that for all states $t \in S$:

$$\mathbb{E}_{C',t}(\text{“}wgt \text{ until } s\text{”}) \;\geqslant\; \mathbb{E}_{C,t}(\text{“}wgt \text{ until } s\text{”}) \;. \tag{C3}$$

To prove (C3), we consider the function $\Upsilon\colon \mathbb{R}^S \to \mathbb{R}^S$ defined as follows. If $f = (f_t)_{t \in S}$ is a vector then $\Upsilon(f) = \big(\Upsilon_t(f)\big)_{t \in S}$ where $\Upsilon_s(f) = 0$ and for $t \in S \setminus \{s\}$: $\Upsilon_t(f) = wgt(t) + \sum_{u \in S} P'(t,u) \cdot f_u$. Let $e = (e_t)_{t \in S}$ denote the vector with $e_t = \mathbb{E}_{C,t}(\text{“}wgt \text{ until } s\text{”})$, and $e' = (e'_t)_{t \in S}$ the corresponding vector for C', *i.e.*, $e'_t = \mathbb{E}_{C',t}(\text{“}wgt \text{ until } s\text{”})$ for all states t. It is well known that:

(i) e' is the unique fixed point of Υ

(ii) $f \leqslant \Upsilon(f)$ implies $f \leqslant \Upsilon(f) \leqslant e'$

where we use the element-wise natural order on vectors, *i.e.*, $(f_t)_{t \in S} \leqslant (g_t)_{t \in S}$ iff $f_t \leqslant g_t$ for all $t \in S$. Using the assumption $e_{v_1} > e_{v_2}$ we obtain:

$$\begin{aligned}
\Upsilon_v(e) \;&=\; wgt(v) \;+\; \sum_{t \in Post(v)} P'(v,t) \cdot e_t \\
&=\; wgt(v) \;+\; \sum_{\substack{t \in Post(v) \\ t \notin \{v_1, v_2\}}} P(v,t) \cdot e_t \;+\; (P(v_1,t) + \delta) \cdot e_{v_1} \;+\; (P(v_2,t) - \delta) \cdot e_{v_2} \\
&=\; \underbrace{wgt(v) \;+\; \sum_{t \in Post(v)} P(v,t) \cdot e_t}_{=e_v} \;+\; \delta \cdot \underbrace{(e_{v_1} - e_{v_2})}_{>0} \;>\; e_v
\end{aligned}$$

and $\Upsilon_t(e) = e_t$ for $t \in S \setminus \{v\}$. Thus, $e \leqslant \Upsilon(e)$ and therefore $e \leqslant e'$. This yields statement (C3).

Moreover, the above calculation shows $e_v < \Upsilon_v(e)$. This yields $e_v < e'_v$ by statement (ii). Hence, for all states $u \in S$:

$$\mathbb{E}_{C',u}(\text{“}wgt \text{ until } s\text{”}) \;>\; \mathbb{E}_{C,u}(\text{“}wgt \text{ until } s\text{”}) \quad \text{if } u \models \exists((\neg s)\,\mathbf{U}\,v) \tag{C4}$$

where we use the CTL-notation $\exists((\neg s)\,\mathbf{U}\,v)$ to denote the existence of a finite path to v that does not traverse s. In particular, (C4) holds for $u = v$. As C is strongly connected, state v is accessible from s. Let $\pi = s_0\, s_1 \ldots s_k$ be a shortest path from $s = s_0$ to $s_k = v$, where "shortest" refers to the standard length (number of transitions) rather than the accumulated weight. As $s \neq v$ we have $k \geqslant 1$. Moreover, $k = 1$ iff $v = s_1 \in \mathit{Post}(s)$. Otherwise, *i.e.*, if $k \geqslant 2$, then $s_1 \ldots s_k$ is a path from state $s_1 = u \in \mathit{Post}(s)$ to v that does not traverse s. In both cases, $u \models \exists((\neg s)\,\mathbf{U}\,v)$ for some state $u \in \mathit{Post}(s)$. As $s \neq v$ we have $P'(s,t) = P(s,t)$ for all states t. By (C3) and (C4) we obtain:

$$\begin{aligned}
& wgt(s) \;+\; \sum_{t \in S} P'(s,t) \cdot \mathbb{E}_{C',t}(\text{“}wgt \text{ until } s\text{”}) \\
= \quad & wgt(s) \;+\; \sum_{\substack{t \in \mathit{Post}(s) \\ t \neq u}} P(s,t) \cdot \underbrace{\mathbb{E}_{C',t}(\text{“}wgt \text{ until } s\text{”})}_{\geqslant \mathbb{E}_{C,t}(\text{“}wgt \text{ until } s\text{”})} \;+\; P(s,u) \cdot \underbrace{\mathbb{E}_{C',u}(\text{“}wgt \text{ until } s\text{”})}_{> \mathbb{E}_{C,u}(\text{“}wgt \text{ until } s\text{”})} \\
> \quad & wgt(s) \;+\; \sum_{t \in S} P(s,t) \cdot \underbrace{\mathbb{E}_{C,t}(\text{“}wgt \text{ until } s\text{”})}_{= -wgt(s)} \;\overset{\text{(C1)}}{=}\; 0
\end{aligned}$$

The quotient representation of the mean payoff (Lemma B.3) yields $\mathbb{E}_{C'}(\mathrm{MP}) > 0$. Contradiction. This completes the proof of statement (C2).

We now fix a state s and show by induction on the length (number of transitions) of paths ξ starting from s that

$$\mathbb{E}_{C,\mathit{last}(\xi)}(\text{“}wgt \text{ until } s\text{”}) \;=\; -wgt(\xi) \tag{C5}$$

In the basis of induction we consider a path of length 1, *i.e.*, ξ consists of a single transition from s to some state $t \in \mathit{Post}(s)$. But then $wgt(\xi) = wgt(s)$ and the claim follows directly from (C1). In the step of induction $k \implies k{+}1$, we regard a path $\xi = s_0\, s_1 \ldots s_k\, s_{k+1}$ of length $k{+}1$ starting in $s_0 = s$. By induction hypothesis we have:

$$\mathbb{E}_{C,s_k}(\text{“}wgt \text{ until } s\text{”}) \;=\; -wgt(s_0\, s_1 \ldots s_k)$$

Recall that $wgt(s_0\, s_1 \ldots s_k) = \sum_{i=0}^{k-1} wgt(s_i)$. Suppose first $s_k = s$. Then, $\mathbb{E}_{C,s_k}(\text{“}wgt \text{ until } s\text{”}) = 0 = wgt(s_0 \ldots s_k)$, in which case $wgt(\xi) = wgt(s_k)$ and the claim follows directly from assumption (†). Suppose now $s \neq s_k$. By (C2), we get:

$$\mathbb{E}_{C,s_{k+1}}(\text{“}wgt \text{ until } s\text{”}) \;=\; \mathbb{E}_{C,t}(\text{“}wgt \text{ until } s\text{”}) \;\text{ for all } t \in \mathit{Post}(s_k).$$

Hence:

$$\begin{aligned}
-\sum_{i=0}^{k-1} wgt(s_i) \quad &= \quad \mathbb{E}_{C,s_k}(\text{“}wgt \text{ until } s\text{”}) \\
&= \quad wgt(s_k) \;+\; \sum_{t \in \mathit{Post}(s_k)} P(s_k,t) \cdot \underbrace{\mathbb{E}_{C,t}(\text{“}wgt \text{ until } s\text{”})}_{= \mathbb{E}_{C,s_{k+1}}(\text{“}wgt \text{ until } s\text{”})} \\
&= \quad wgt(s_k) \;+ \mathbb{E}_{C,s_{k+1}}(\text{“}wgt \text{ until } s\text{”}) \cdot \underbrace{\sum_{t \in \mathit{Post}(s_k)} P(s_k,t)}_{=1} \\
&= \quad wgt(s_k) \;+\; \mathbb{E}_{C,s_{k+1}}(\text{“}wgt \text{ until } s\text{”})
\end{aligned}$$

We conclude:

$$-wgt(\xi) \;=\; -\sum_{i=0}^{k} wgt(s_i) \;=\; \mathbb{E}_{C,s_{k+1}}(\text{“}wgt \text{ until } s\text{”})$$

This completes the proof of the induction step.

We finally use statement (C5) to show that C is a 0-BSCC. Let ξ be a cycle in C and s an arbitrary state on ξ. Statement (C5) yields: $-wgt(\xi) \;=\; \mathbb{E}_{C,s}(\text{“}wgt \text{ until } s\text{”}) \;=\; 0$, which completes the proof. □

The goal is to apply the observation above on the relation between expected mean payoff and expected accumulated weights until reaching a target in Markov chains for checking the existence of 0-EC in strongly connected MDPs with $\mathbb{E}^{\max}_{\mathcal{M}}(\mathrm{MP}) = 0$. We start with the following observation stating that MD-schedulers with expected mean payoff 0 maximize the expected accumulated weight until reaching any state s of its BSCCs. Here, the maximum is taken over all MD-schedulers $\mathfrak{S}$ where s belongs to a BSCC of $\mathfrak{S}$.[2] More precisely:

Lemma B.12. *Let $\mathcal{M}$ be a strongly connected MDP with $\mathbb{E}^{\max}_{\mathcal{M}}(\mathrm{MP}) = 0$ and let $\mathfrak{S}$ be an MD-scheduler with $\mathbb{E}^{\mathfrak{S}}_{\mathcal{M},s}(\mathrm{MP}) = 0$ for all states s. Furthermore, let $\mathcal{B}$ be a BSCC of the Markov chain induced by $\mathfrak{S}$ and (s,α) a state-action pair in $\mathcal{B}$ (i.e., s is a state of $\mathcal{B}$ and $\alpha = \mathfrak{S}(s)$). Then:*

$$wgt(s,\alpha) \;+\; \sum_{t \in \mathcal{B}} P(s,\alpha,t) \cdot \mathbb{E}^{\mathfrak{S}}_{\mathcal{M},t}(\text{“}wgt \text{ until } s\text{”}) \;=\; 0. \tag{‡}$$

Moreover, whenever $\mathfrak{T}$ is an MD-scheduler for $\mathcal{M}$ with $\mathfrak{T}(s) = \alpha$ where s belongs to a BSCC of the Markov chain $C_{\mathfrak{T}}$ induced by $\mathfrak{T}$, then for all states t with $P(s,\alpha,t) > 0$

$$\mathbb{E}^{\mathfrak{S}}_{\mathcal{M},t}(\text{“}wgt \text{ until } s\text{”}) \;\geqslant\; \mathbb{E}^{\mathfrak{T}}_{\mathcal{M},t}(\text{“}wgt \text{ until } s\text{”}) \;.$$

Proof. The first part (statement (‡)) follows directly from Lemma B.3 applied to $\mathcal{B}$ viewed as a strongly connected Markov chain. For the second part, we assume that $\mathfrak{T}$ is some MD-scheduler with $\mathfrak{T}(s) = \alpha$ where s belongs to a BSCC $\mathcal{B}'$ of $\mathfrak{T}$. Suppose by contradiction that there is some state t with $P(s,\alpha,t) > 0$ and $\mathbb{E}^{\mathfrak{S}}_{\mathcal{M},t}(\text{“}wgt \text{ until } s\text{”}) < \mathbb{E}^{\mathfrak{T}}_{\mathcal{M},t}(\text{“}wgt \text{ until } s\text{”})$. Let $\mathfrak{V}$ denote the following scheduler operating in two modes: In its first mode, $\mathfrak{V}$ behaves as $\mathfrak{S}$. It switches to the second mode when entering t via the α-transition from s. In its second mode, $\mathfrak{V}$ behaves as $\mathfrak{T}$ until it visits s, in which case it switches back to its first mode where it behaves as $\mathfrak{S}$. Note that $\mathfrak{V}$ is not memoryless, but a finite-memory scheduler. Hence, the Markov chain $C_{\mathfrak{V}}$ induced by $\mathfrak{V}$ is finite as well. Lemma B.3 applied to the $C_{\mathfrak{V}}$ yields:

$$\mathbb{E}^{\mathfrak{V}}_{\mathcal{M},s}(\mathrm{MP}) \;=\; \frac{wgt(s,\alpha) \;+\; P(s,\alpha,t) \cdot \mathbb{E}^{\mathfrak{T}}_{\mathcal{E},t}(\text{“}wgt \text{ until } s\text{”}) \;+\; \sum_{u \neq t} P(s,\alpha,u) \cdot \mathbb{E}^{\mathfrak{S}}_{\mathcal{E},u}(\text{“}wgt \text{ until } s\text{”})}{1 \;+\; P(s,\alpha,t) \cdot \mathbb{E}^{\mathfrak{T}}_{\mathcal{E},t}(\text{“steps until } s\text{”}) \;+\; \sum_{u \neq t} P(s,\alpha,u) \cdot \mathbb{E}^{\mathfrak{S}}_{\mathcal{E},u}(\text{“steps until } s\text{”})} \;.$$

Using $\mathbb{E}^{\mathfrak{T}}_{\mathcal{E},t}(\text{“}wgt \text{ until } s\text{”}) > \mathbb{E}^{\mathfrak{S}}_{\mathcal{E},t}(\text{“}wgt \text{ until } s\text{”})$ and (‡), we get for the numerator of $\mathbb{E}^{\mathfrak{V}}_{\mathcal{M},s}(\mathrm{MP})$:

$$\begin{aligned}
& wgt(s,\alpha) + P(s,\alpha,t) \cdot \mathbb{E}^{\mathfrak{T}}_{\mathcal{E},t}(\text{“}wgt \text{ until } s\text{”}) + \sum_{u \neq t} P(s,\alpha,u) \cdot \mathbb{E}^{\mathfrak{S}}_{\mathcal{E},u}(\text{“}wgt \text{ until } s\text{”}) \\
> \;& wgt(s,\alpha) + P(s,\alpha,t) \cdot \mathbb{E}^{\mathfrak{T}}_{\mathcal{E},t}(\text{“}wgt \text{ until } s\text{”}) + \sum_{u \neq t} P(s,\alpha,u) \cdot \mathbb{E}^{\mathfrak{S}}_{\mathcal{E},u}(\text{“}wgt \text{ until } s\text{”}) \\
= \;& wgt(s,\alpha) + \sum_{u \in S} P(s,\alpha,u) \cdot \mathbb{E}^{\mathfrak{S}}_{\mathcal{E},u}(\text{“}wgt \text{ until } s\text{”}) \;=\; 0
\end{aligned}$$

This yields $\mathbb{E}^{\mathfrak{V}}_{\mathcal{M},s}(\mathrm{MP}) > 0$, which is impossible as $\mathbb{E}^{\mathfrak{V}}_{\mathcal{M},s}(\mathrm{MP}) \leqslant \mathbb{E}^{\max}_{\mathcal{M}}(\mathrm{MP}) = 0$. □

B.4.3 Algorithm to Check the Existence of 0-ECs

We show that given a strongly connected MDP $\mathcal{M}$ with $\mathbb{E}^{\max}_{\mathcal{M}}(\mathrm{MP}) = 0$, the task to decide the existence of 0-ECs is solvable by an algorithm that runs in time in polynomial in the size of the given MDP.

For this, we rely on the observation that the property of being a 0-EC does not depend on the precise transition probabilities, but only on the graph structure and the weights. The idea is now to modify the transition probabilities of a state-action pair (s,α) in a gambling BSCC of $\mathcal{M}$ such that the transformed MDP $\mathcal{M}'$ has the same graph structure and weights (thus, $\mathcal{M}$ and $\mathcal{M}'$ have the same 0-ECs) and $\mathcal{M}'$ enjoys the following property:

Whenever $\mathfrak{S}$ is an MD-scheduler for $\mathcal{M}$ and $\mathfrak{S}(s) = \alpha$ such that $\mathfrak{S}$ is gambling in $\mathcal{M}$, then $\mathbb{E}^{\mathfrak{S}}_{\mathcal{M}',s}(\mathrm{MP}) < 0$.

Thus, the gambling BSCCs of $\mathcal{M}$ containing the state-action pair (s,α) are no longer gambling in $\mathcal{M}'$. Hence, the only end components in $\mathcal{M}'$ with maximal mean payoff 0 are the 0-ECs. This then ensures that $\mathbb{E}^{\max}_{\mathcal{M}'}(\mathrm{MP}) = 0$ if and only if $\mathcal{M}$ (and $\mathcal{M}'$) have a 0-EC.

The algorithm for checking the existence of a 0-EC in a strongly connected MDP $\mathcal{M}$ with $\mathbb{E}^{\max}_{\mathcal{M}}(\mathrm{MP}) = 0$ proceeds as follows. It first runs a standard polynomial-time algorithm to compute an MD-scheduler $\mathfrak{S}$ with $\mathbb{E}^{\mathfrak{S}}_{\mathcal{M},s}(\mathrm{MP}) = 0$ for all states s. We may

[2] The supremum of the expectations of “wgt until s” under all schedulers is infinite if $\mathcal{M}$ has gambling end components that do not contain s. This is, however, irrelevant for our purposes as we are only interested in the maximal expectation of “wgt until s” when ranging over MD-schedulers under which the long-run frequency of s is positive.

assume w.l.o.g. that the Markov chain $\mathcal{C}_{\mathfrak{S}}$ induced by $\mathfrak{S}$ has a single BSCC $\mathcal{B}$. We then check in polynomial time whether $\mathcal{B}$ is a 0-BSCC. (For this, we can rely on Lemma 3.2 and check the nonexistence of positive cycles using standard graph algorithms.) If so, then $\mathcal{B}$ is a 0-EC of $\mathcal{M}$ and the algorithm terminates by returning $\mathcal{B}$. Otherwise, there exist states s, t, u in $\mathcal{B}$ such that $P(s,\alpha,t) > 0$, $P(s,\alpha,u) > 0$ where $\alpha = \mathfrak{S}(s)$ and (see Lemma B.11)

$$\mathbb{E}^{\mathfrak{S}}_{t}(\text{“}wgt \text{ until } s\text{”}) \;>\; \mathbb{E}^{\mathfrak{S}}_{u}(\text{“}wgt \text{ until } s\text{”}) \;.$$

Let now $\mathcal{M}'$ be the MDP resulting from $\mathcal{M}$ by changing the transition probabilities for the state-action pair (s,α) as follows. We pick a value $\delta > 0$ such that $P(s,\alpha,t) - \delta > 0$ and $P(s,\alpha,u) + \delta < 1$ and define:

$$P'(s,\alpha,t) \;=\; P(s,\alpha,t) - \delta, \qquad P'(s,\alpha,u) \;=\; P(s,\alpha,u) + \delta$$

and $P'(s,\alpha,s') = P(s,\alpha,s')$ for all other states $s' \in S \setminus \{t,u\}$. The transition probabilities for all other state-action pairs as well as the weight function remain unchanged. We then have $\mathbb{E}^{\mathfrak{S}}_{\mathcal{M}',s'}(\mathrm{MP}) < 0$ for all states s' in $\mathcal{B}$ and $\mathbb{E}^{\mathfrak{T}}_{\mathcal{M}',s'}(\mathrm{MP}) \leqslant \mathbb{E}^{\mathfrak{S}}_{\mathcal{M},s'}(\mathrm{MP})$ for all states s' and all MD-schedulers $\mathfrak{T}$ for $\mathcal{M}'$ (see Lemma B.13 below). In particular, $\mathbb{E}^{\max}_{\mathcal{M}'}(\mathrm{MP}) \leqslant 0$. We then call again an algorithm to compute an MD-scheduler $\mathfrak{S}'$ for $\mathcal{M}'$ that maximizes the expected mean payoff. If $\mathbb{E}^{\max}_{\mathcal{M}'}(\mathrm{MP}) < 0$ then the algorithm terminates with the answer “no, $\mathcal{M}$ has no 0-EC”. Otherwise, the algorithm repeats to modify the transition probabilities of a state-action pair (s',α') in some BSCC of $\mathfrak{S}'$, and so on.

The presented algorithm terminates after at most $|S| \cdot |Act|$ steps as the transition probabilities of each state-action pair are perturbed at most once (see Lemma B.13 below). The cost of iteration are dominated by the cost for computing an MD-scheduler $\mathfrak{S}$ maximizing the expected mean payoff and the values $\mathbb{E}^{\mathfrak{S}}_{\mathcal{M},s'}(\text{“}wgt \text{ until } s\text{'})$. Thus, the time complexity is polynomial in the size of $\mathcal{M}$.

Lemma B.13 (Soundness of the transformation)**.** *Let $\mathcal{M}$ be a strongly connected MDP with $\mathbb{E}^{\max}_{\mathcal{M}}(\mathrm{MP}) = 0$ and let $\mathfrak{S}$ and $\mathcal{M}'$ be as above, modifying the transition probabilities for (s,α). Then, for all MD-schedulers $\mathfrak{T}$ for $\mathcal{M}$ (and $\mathcal{M}'$):*

(a) If s belongs to a BSCC of $\mathcal{C}_{\mathfrak{T}}$ and $\mathfrak{T}(s) = \alpha$ then $\mathbb{E}^{\mathfrak{T}}_{\mathcal{M}',s}(\mathrm{MP}) < 0$.

(b) If $\mathcal{B}$ is a BSCC of $\mathfrak{T}$ such that either s does not belong to $\mathcal{B}$ or $\mathfrak{T}(s) \neq \alpha$ then $\mathbb{E}^{\mathfrak{T}}_{\mathcal{M}',s'}(\mathrm{MP}) = \mathbb{E}^{\mathfrak{T}}_{\mathcal{M},s'}(\mathrm{MP})$ for all states $s' \in \mathcal{B}$.

In particular, we have $\mathbb{E}^{\max}_{\mathcal{M}'}(\mathrm{MP}) \leqslant 0$.

Proof. Statement (b) is obvious, as in this case, $\mathcal{B}$ is not affected by the switch from $\mathcal{M}$ to $\mathcal{M}'$. For the proof of statement (a) we rely on the second part of Lemma B.12 yielding for all states s' with $P(s,\alpha,s') > 0$ that

$$\mathbb{E}^{\mathfrak{T}}_{\mathcal{M}',s'}(\text{“}wgt \text{ until } s\text{”}) \;=\; \mathbb{E}^{\mathfrak{T}}_{\mathcal{M},s'}(\text{“}wgt \text{ until } s\text{”}) \;\leqslant\; \mathbb{E}^{\mathfrak{S}}_{\mathcal{M},s'}(\text{“}wgt \text{ until } s\text{”}) \;.$$

But then:

$$\begin{aligned}
& wgt(s,\alpha) \;+\; \sum_{s'\in S} P'(s,\alpha,s') \cdot \mathbb{E}^{\mathfrak{T}}_{\mathcal{M}',s'}(\text{“}wgt \text{ until } s\text{”}) \\
\leqslant\;\; & wgt(s,\alpha) \;+\; \sum_{s'\in S\setminus\{t,u\}} P(s,\alpha,s') \cdot \mathbb{E}^{\mathfrak{S}}_{\mathcal{M},s'}(\text{“}wgt \text{ until } s\text{”}) \\
& \quad +\; (P(s,\alpha,t) - \delta) \cdot \mathbb{E}^{\mathfrak{S}}_{\mathcal{M},t}(\text{“}wgt \text{ until } s\text{”}) \\
& \quad +\; (P(s,\alpha,u) + \delta) \cdot \mathbb{E}^{\mathfrak{S}}_{\mathcal{M},u}(\text{“}wgt \text{ until } s\text{”}) \\
=\;\; & wgt(s,\alpha) \;+\; \sum_{s'\in S} P(s,\alpha,s') \cdot \mathbb{E}^{\mathfrak{S}}_{\mathcal{M}',s'}(\text{“}wgt \text{ until } s\text{”}) \\
& \quad -\; \delta \cdot \big(\mathbb{E}^{\mathfrak{S}}_{\mathcal{M},t}(\text{“}wgt \text{ until } s\text{”}) - \mathbb{E}^{\mathfrak{S}}_{\mathcal{M},u}(\text{“}wgt \text{ until } s\text{”}) \big) \\
<\;\; & wgt(s,\alpha) \;+\; \sum_{s'\in S} P(s,\alpha,s') \cdot \mathbb{E}^{\mathfrak{S}}_{\mathcal{M}',s'}(\text{“}wgt \text{ until } s\text{”}) \;=\; 0 \;.
\end{aligned}$$

Here, we use statement (‡) of Lemma B.12 and the facts that $\delta > 0$ and $\mathbb{E}^{\mathfrak{S}}_{\mathcal{M},t}(\text{“}wgt \text{ until } s\text{”}) > \mathbb{E}^{\mathfrak{S}}_{\mathcal{M},u}(\text{“}wgt \text{ until } s\text{”})$. Hence, by Lemma B.3 we obtain

$$\mathbb{E}^{\mathfrak{T}}_{\mathcal{M}'}(\mathrm{MP}) \;=\; \frac{wgt(s,\alpha) \;+\; \sum_{s'\in S} P'(s,\alpha,s')\cdot \mathbb{E}^{\mathfrak{T}}_{\mathcal{M}',s'}(\text{“}wgt \text{ until } s\text{”})}{1 \;+\; \sum_{s'\in S} P'(s,\alpha,s')\cdot \mathbb{E}^{\mathfrak{T}}_{\mathcal{M}',s'}(\text{“steps until } s\text{”})} \;<\; 0 \;.$$

□

B.4.4 Complexity of Checking the Existence of 0-ECs

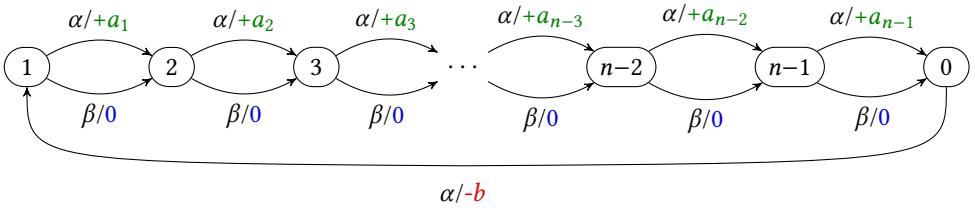


Figure 6. Reduction from subset sum for the NP-hardness of checking the existence of 0-ECs.

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.*

Proof. A polynomial-time procedure for checking the existence of a 0-EC in strongly connected MDPs with $\mathbb{E}^{\max}_{\mathcal{M}}(\mathrm{MP}) = 0$ has been presented in Section B.4.3.

We now show the NP-completeness of the general case. An NP-algorithm is obtained by nondeterministically guessing an MD-scheduler and checking whether the induced Markov chain has a 0-BSCC. NP-hardness can be easily obtained via a reduction from the subset sum problem: We are given a finite sequence of nonnegative integers $a_1, \ldots a_{n-1}, b$ and the task is to find a subset I of $\{1,\ldots,n{-}1\}$ with $\sum_{i\in I} a_i = b$. For this, we regard the MDP $\mathcal{M}$ with state space $\{0,1,\ldots,n{-}1\}$ illustrated by Figure 6. Each state $i \in \{1,\ldots,n{-}1\}$ has two actions α and β with $wgt(i,\alpha) = a_i$, $wgt(i,\beta) = 0$, as well as $P(i,\alpha,(i{+}1) \bmod n) = P(i,\beta,(i{+}1) \bmod n) = 1$. In state 0, only action α is enabled with $wgt(0,\alpha) = -b$ and $P(0,\alpha,1) = 1$. In all remaining cases, we have $P(\cdot) = 0$. Then, $\mathcal{M}$ is strongly connected and each subset I of $\{1,\ldots,n{-}1\}$ induces an end component $\mathcal{E}_I$ consisting of the state-action pairs $(0,\alpha)$ and (i,α) for $i \in I$ and (i,β) for $i \in \{1,\ldots,n{-}1\} \setminus I$. The BSCCs of MD-schedulers are exactly the end components $\mathcal{E}_I$ for $I \subseteq \{1,\ldots,n{-}1\}$. Moreover, $\mathcal{E}_I$ is a 0-EC iff $\sum_{i\in I} a_i = b$. Hence, $\mathcal{M}$ has a 0-EC iff there exists a subset I of $\{1,\ldots,n{-}1\}$ with $\sum_{i\in I} a_i = b$. □

B.5 Spider Construction and Weight-Divergence Algorithm

Recall that the purpose of the spider construction $\mathcal{M} \rightsquigarrow \mathrm{Spider}_{\mathcal{E}}(\mathcal{M})$ was to flatten a 0-BSCC $\mathcal{E}$ in the MDP $\mathcal{M}$. For a detailed presentation in this appendix, we recall the construction from the main paper:

As $\mathcal{E}$ is a 0-BSCC, for each state s in $\mathcal{E}$ there is a single state-action pair $(s,\alpha_s) \in \mathcal{E}$. Given two states s,t in $\mathcal{E}$, recall that $w(s,t)$ denotes the weight of every path from s to t in $\mathcal{E}$. The spider construction picks a reference state s_0 in $\mathcal{E}$. Then, the MDP $\mathcal{N} = \mathrm{Spider}_{\mathcal{E},s_0}(\mathcal{M})$ arises from $\mathcal{M}$ by performing the following steps for each state s that appears in $\mathcal{E}$:

- **(i)** Remove the state-action pair (s,α_s)
- **(ii)** In case $s \neq s_0$, add a new state-action pair (s,τ) with $P_{\mathcal{N}}(s,\tau,s_0) = 1$ and $wgt_{\mathcal{N}}(s,\tau) = w(s,s_0)$
- **(iii)** In case $s \neq s_0$, replace every state-action pair $(s,\beta) \in \mathcal{M}$ with $\beta \neq \alpha_s$ by (s_0,β). The transition probabilities and weights of these state-action pairs are given by $P_{\mathcal{N}}(s_0,\beta,u) = P_{\mathcal{M}}(s,\beta,u)$ for all states u and $wgt_{\mathcal{N}}(s_0,\beta) = w(s_0,s) + wgt_{\mathcal{M}}(s,\beta)$

Recall that we often write $\mathrm{Spider}_{\mathcal{E}}(\mathcal{M})$ rather than $\mathrm{Spider}_{\mathcal{E},s_0}(\mathcal{M})$ when the reference state s_0 is clear from the context or irrelevant, *e.g.*, in case the spider construction is used as a vehicle to reduce the MDP's number of state-action pairs where the actual structure of the arising graph is not of interest. As an example to illustrate how the choice of the reference state influences the graph structure of the MDP, let us return to the MDP of Example 3.6. When taking state s instead of state t as first reference state and then state u instead of t, we obtain the MDPs $\mathcal{M}_1^s = \mathrm{Spider}_{\mathcal{E},s}(\mathcal{M})$ and $\mathcal{M}_2^u = \mathrm{Spider}_{\mathcal{E},u}(\mathcal{M}_1^s)$ as depicted in Figure 7. Note that by changing the reference state during an iterative application of the spider construction, chains of τ transitions may arise.

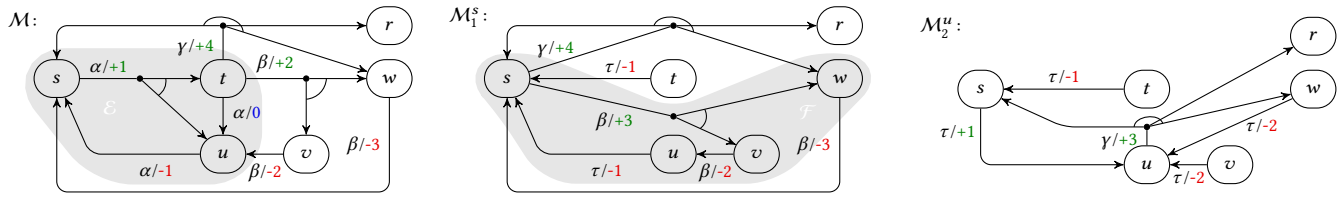


Figure 7. Spider constructions with s and u as reference states.

In the following, we suppose that in a preprocessing step the actions in $\mathcal{M}$ are renamed such that $Act_{\mathcal{M}}(s) \cap Act_{\mathcal{M}}(t) = \varnothing$ for all states s, t in $\mathcal{M}$ with $s \neq t$. This technical requirement will be used in upcoming proofs and can be achieved by simply renaming actions.

B.5.1 Properties of the Spider Construction

We now prove that the MDP $\mathsf{Spider}_{\mathcal{E}}(\mathcal{M})$ enjoys the properties as stated in Lemma 3.7. The proofs for statements (S1), (S2) will be established in Lemma B.14. Property (S4) will be shown in Lemma B.16. The equivalence of $\mathcal{M}$ and $\mathsf{Spider}_{\mathcal{E}}(\mathcal{M})$ as stated in (S3) will be a consequence of Lemma B.18.

Lemma B.14 (See (S1) and (S2) in Lemma 3.7)**.** *Let $\mathcal{M}$ be an MDP and $\mathcal{E}$ a 0-BSCC of $\mathcal{M}$. Then, the spider construction generates an MDP $\mathsf{Spider}_{\mathcal{E},s_0}(\mathcal{M})$ that satisfies the following properties:*

(S1) $\mathcal{M}$ and $\mathsf{Spider}_{\mathcal{E},s_0}(\mathcal{M})$ have the same state space and $\|\mathsf{Spider}_{\mathcal{E},s_0}(\mathcal{M})\| = \|\mathcal{M}\|-1$

(S2) If $\mathcal{M}$ is strongly connected and $\mathcal{E} \neq \mathcal{M}$ then $\mathsf{Spider}_{\mathcal{E},s_0}(\mathcal{M})$ has a single maximal end component $\mathcal{F}$, reachable from all states and containing the reference state s_0.

Proof. The first statement of (S1) is obvious: If m is the number of states in $\mathcal{E}$ then step **(i)** removes m state-action pairs, while step **(ii)** introduces $m-1$ new state-action pairs. Step **(iii)** has no effect on the number of state-action pairs. Hence, $\|\mathsf{Spider}_{\mathcal{E}}(\mathcal{M})\| = \|\mathcal{M}\|-1$.

To prove statement (S2), we suppose that $\mathcal{M} \neq \mathcal{E}$ and that $\mathcal{M}$ is strongly connected. Let T denote the set of states s in $\mathcal{E}$ such that $s \neq s_0$ and $P_{\mathcal{M}}(u, \alpha, s) > 0$ for some state-action pair $(u, \alpha) \in \mathcal{M} \setminus \mathcal{E}$. Let $\mathcal{F}$ denote the sub-MDP of $\mathsf{Spider}_{\mathcal{E}}(\mathcal{M})$ obtained by removing all state-action pairs (s, τ) where s is a state of $\mathcal{E}$ that is not contained in T. Thus, the state space of $\mathcal{F}$ consists of the states of $\mathcal{M}$ that do not belong to $\mathcal{E}$ and the states in $T \cup \{s_0\}$.

Obviously, $\mathcal{F}$ is reachable from all states and $\mathcal{F}$ contains the reference state s_0. We show that $\mathcal{F}$ is strongly connected. For this, we prove that all states s in $\mathcal{F}$ with $s \neq s_0$ are reachable from s_0 and can reach s_0.

- We first show that s_0 is reachable from each state s in $\mathsf{Spider}_{\mathcal{E}}(\mathcal{M})$. Let $\pi = t_0\, \beta_0\, t_1\, \beta_1 \ldots \beta_{n-1}\, t_n$ be a simple path in $\mathcal{M}$ from $t_0 = s$ to $t_n = s_0$, where "simple" means that $t_i \neq t_j$ for $0 \leqslant i < j \leqslant n$. We pick the smallest index $i \in \{0, 1, \ldots, n\}$ such that t_i belongs to $\mathcal{E}$. Then, $(t_j, \beta_j) \in \mathcal{F}$ for $0 \leqslant j < i$. Hence, if $t_i \neq s_0$ then $t_0\, \beta_0\, t_1\, \beta_1 \ldots \beta_{i-1}\, t_i\, \tau\, s_0$ is a path in $\mathcal{F}$ from $s = t_0$ to s_0. Likewise, if $t_i = s_0$ then $t_0\, \beta_0\, t_1\, \beta_1 \ldots \beta_{i-1}\, t_i$ is a path in $\mathcal{F}$ from $s = t_0$ to $s_0 = t_i$.
- We next show that each state s that is not contained in $\mathcal{E}$ is reachable from s_0 in $\mathcal{F}$. Let $\pi = t_0\, \beta_0\, t_1\, \beta_1 \ldots \beta_{n-1}\, t_n$ be a simple path in $\mathcal{M}$ from $t_0 = s_0$ to $t_n = s$. If none of the states $t_1, \ldots, t_n$ belongs to $\mathcal{E}$ then $(t_i, \beta_i) \in \mathcal{F}$ for all $0 \leqslant i < n$ and π is a path in $\mathcal{F}$ from s_0 to s. Suppose now that at least one of the states $t_1, \ldots, t_n$ is contained in $\mathcal{E}$. We pick the largest index $i \in \{1, \ldots, n\}$ where t_i is a state of $\mathcal{E}$. By assumption $s = t_n$ is not contained in $\mathcal{E}$. This yields $i < n$. Then, $(t_i, \beta_i) \notin \mathcal{E}$ and therefore $(s_0, \beta_i) \in \mathcal{F}$ due to step **(iii)**. As the states $t_{i+1}, \ldots, t_n$ do not belong to $\mathcal{E}$, we have $(t_j, \beta_j) \in \mathcal{F}$ for $i < j < n$. Thus, $s_0\, \beta_i\, t_{i+1}\, \beta_{i+1} \ldots \beta_{n-1}\, t_n$ is a path from s_0 to s in $\mathcal{F}$.
- It remains to show that each state $s \in T$ is reachable from s_0 in $\mathcal{F}$. By definition of T there is state-action pair (u, α) such that $P_{\mathcal{M}}(u, \alpha, s) > 0$ and $(u, \alpha) \in \mathcal{M} \setminus \mathcal{E}$. If u belongs to $\mathcal{E}$ then $(s_0, \alpha) \in \mathcal{F}$ and hence, $s_0\, \alpha\, s$ is a path in $\mathcal{F}$. If u does not belong to $\mathcal{E}$ then s is reachable from s_0 in $\mathcal{F}$ via a path of the form $\pi\, \alpha\, s$ where π is a path from s_0 to u in $\mathcal{F}$ (see above).

Thus, $\mathcal{F}$ is an end component of $\mathsf{Spider}_{\mathcal{E}}(\mathcal{M})$. As the states in $\mathcal{E}$ that are not contained in T do not have incoming edges in $\mathsf{Spider}_{\mathcal{E}}(\mathcal{M})$, these states are not contained in any end component of $\mathsf{Spider}_{\mathcal{E}}(\mathcal{M})$. Thus, $\mathcal{F}$ subsumes each other end component of $\mathsf{Spider}_{\mathcal{E}}(\mathcal{M})$. This shows that $\mathcal{F}$ is the unique maximal end component of $\mathsf{Spider}_{\mathcal{E}}(\mathcal{M})$. □

Remark B.15 (Result of the spider construction if $\mathcal{M}$ is a 0-BSCC)**.** Note that the requirement $\mathcal{M} \neq \mathcal{E}$ in (S2) is necessary as if $\mathcal{M}$ is a 0-BSCC then $\mathsf{Spider}_{\mathcal{M}}(\mathcal{M})$ is acyclic and consists of τ-transitions from all states s with $s \neq s_0$ to s_0.

Lemma B.16 (See (S4) in Lemma 3.7)**.** *Let $\mathcal{M}$ be an MDP and $\mathcal{E}$ a 0-BSCC of $\mathcal{M}$ that is contained in a maximal end component of $\mathcal{M}$ with maximal expected mean payoff 0. Then:*

(a) For each state s that is not contained in $\mathcal{E}$: s belongs to a 0-EC of $\mathcal{M}$ iff s belongs to a 0-EC of $\mathrm{Spider}_{\mathcal{E}}(\mathcal{M})$.
(b) 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 $\mathrm{Spider}_{\mathcal{E},s_0}(\mathcal{M})$.

Proof. We first consider a 0-EC $\mathcal{Z}$ of $\mathcal{M}$. The claim is obvious for $\mathcal{Z} = \mathcal{E}$. Hence, we may suppose $\mathcal{Z} \neq \mathcal{E}$. We now show that there is a 0-EC $\mathcal{F}$ of $\mathrm{Spider}_{\mathcal{E}}(\mathcal{M})$ that contains all states in $\mathcal{F}$ that are not contained in $\mathcal{E}$ and all actions α such that $(s,\alpha) \in \mathcal{Z} \setminus \mathcal{E}$.

If $\mathcal{Z}$ does not contain a state of $\mathcal{E}$, then $\mathcal{Z}$ is clearly a 0-EC of $\mathrm{Spider}_{\mathcal{E}}(\mathcal{M})$. If $\mathcal{Z}$ contains the reference state s_0, then the set $\mathcal{F}$ is a 0-EC of $\mathrm{Spider}_{\mathcal{E}}(\mathcal{M})$, consisting of the state-action pairs

- $(s,\alpha) \in \mathcal{Z}$ where s is a state not contained in $\mathcal{E}$,
- (s,τ) where s is a state of $\mathcal{E}$ with $s \neq s_0$ and $(s,\alpha_s) \in \mathcal{Z}$,
- (s_0,α) where s is a state of $\mathcal{E}$ (possibly $s = s_0$) with $\alpha \neq \alpha_s$ and $(s,\alpha) \in \mathcal{Z}$.

Otherwise, *i.e.*, $\mathcal{Z}$ does not contain s_0 but some state of $\mathcal{E}$, then $\mathcal{Z} \cup \mathcal{E}$ is a 0-EC of $\mathcal{M}$ (see Lemma B.10) and we can apply the argument before.

Vice versa, we show that each 0-EC $\mathcal{F}$ of $\mathrm{Spider}_{\mathcal{E}}(\mathcal{M})$ induces a 0-EC $\mathcal{Z}$ of $\mathcal{M}$ such that $\mathcal{Z}$ contains all states of $\mathcal{F}$ and all actions α of $\mathcal{M}$ where $(s_0,\alpha) \in \mathcal{F}$. If $\mathcal{F}$ is a 0-EC of $\mathrm{Spider}_{\mathcal{E}}(\mathcal{M})$ that does not contain s_0, then $\mathcal{F}$ is a 0-EC of $\mathcal{M}$. (We use here the fact that the states s in $\mathcal{E}$ with $s \neq s_0$ have a single τ-transition to s_0 in $\mathrm{Spider}_{\mathcal{E}}(\mathcal{M})$, but no other transition. Hence, each end component of $\mathrm{Spider}_{\mathcal{E}}(\mathcal{M})$ that contains s must also contain s_0.) If $\mathcal{F}$ is a 0-EC of $\mathrm{Spider}_{\mathcal{E}}(\mathcal{M})$ containing s_0, then the set consisting of all state-action pairs in $\mathcal{E}$ plus the state-action pairs (s,α_s) where s is a state in $\mathcal{F}$ and $(s_0,\alpha_s) \in \mathcal{F}$ is a 0-EC of $\mathcal{M}$. □

Remark B.17 (The values $w(s,t)$)**.** Recall from Section B.4.1 that for each maximal end component $\mathcal{G}$ of $\mathcal{M}$ with $\mathbb{E}^{\max}_{\mathcal{G}}(\mathrm{MP}) = 0$, each 0-EC of $\mathcal{G}$ is contained in a *maximal 0-ECs*, where maximality is understood with respect to the property "being a 0-EC". This implies the existence of integers $w(s,t) \in \mathbb{Z}$ for all states s,t that belong to the same maximal 0-EC of $\mathcal{M}$ that is contained in some maximal end component $\mathcal{G}$ of $\mathcal{M}$ with $\mathbb{E}^{\max}_{\mathcal{G}}(\mathrm{MP}) = 0$ such that the weight of all paths from s to t inside some 0-EC equals $w(s,t)$, no matter which 0-EC is chosen.

The transformation "0-EC $\mathcal{F}$ of $\mathrm{Spider}_{\mathcal{E}}(\mathcal{M}) \rightsquigarrow$ 0-EC $\mathcal{Z}$ of $\mathcal{M}$" explained in the proof of Lemma B.16 preserves the w-values. More precisely, if $\mathcal{F}$ is a 0-EC of $\mathrm{Spider}_{\mathcal{E}}(\mathcal{M})$ then $w(s,t)$ is the weight of each path from s to t in $\mathcal{F}$. Vice versa, if $\mathcal{Z}$ is a 0-EC of $\mathcal{M}$ and s, t are states in $\mathcal{Z}$ then $w(s,t)$ is the weight of each path from s to t in $\mathrm{Spider}_{\mathcal{E}}(\mathcal{M})$ that is built by τ-transitions and actions belonging to $\mathcal{Z}$. ■

We now turn to the proof of the scheduler transformations stated in (S3) of Lemma 3.7. Before doing so, let us recall the definition of the purge-function and of $\mathcal{E}$-invariant properties: Given a (finite or infinite) path $\varsigma = t_0\,\alpha_0\,t_1\,\alpha_1\ldots$ in $\mathcal{M}$, let $\mathit{purge}_{\mathcal{E}}(\varsigma) \in (S\times\mathbb{Z})^\omega \cup (S\times\mathbb{Z})^*\times S$ denote the sequence arising from ς by

(1) replacing each fragment $t_i\,\alpha_i\,\ldots\,\alpha_{j-1}\,t_j\,\alpha_j\,t_{j+1}$ of ς such that
- either $i=0$ or $(t_{i-1},\alpha_{i-1}) \notin \mathcal{E}$,
- $(t_j,\alpha_j) \notin \mathcal{E}$
- $(t_\ell,\alpha_\ell) \in \mathcal{E}$ for $\ell = i, i{+}1,\ldots,j$

with $t_i\,w\,t_{j+1}$ where $w = w(t_i,t_j) + \mathit{wgt}(t_j,\alpha_j)$ and

(2) replacing each action α_i in the resulting sequence with $\mathit{wgt}(s_i,\alpha_i)$.

Note that step (1) yields a (finite or infinite) sequence $t'_0\,c_0\,t'_1\,c_1\,\ldots$ where the $t'_0,t'_1,\ldots$ are states and $c_i \in \mathit{Act}\cup\mathbb{Z}$. Moreover, $c_i \in \mathbb{Z}$ implies $t'_i \in \mathcal{E}$, while $c_i \in \mathit{Act}$ implies that (t'_i,c_i) is a state-action pair of $\mathcal{M}$ that is not contained in $\mathcal{E}$ (but still $t'_i \in \mathcal{E}$ is possible). For example, if

$$\pi \;=\; t_0\,\beta_0\,t_1\,\beta_1\,t_2\,\gamma_2\,t_3\,\beta_3\,t_4\,\gamma_4\,t_5\,\gamma_5\,t_6\,\beta_6\,t_7$$

where all state-action pairs of the form (t_k,γ_k) belong to $\mathcal{E}$, while the state-action pairs of the form (t_l,β_l) do not, then after step (1) we obtain the sequence

$$t_0\,\beta_0\,t_1\,\beta_1\,t_2\,w_2\,t_4\,w_4\,t_7$$

where $w_2 = \underbrace{\mathit{wgt}(t_2,\gamma_2)}_{=w(t_2,t_3)} + \mathit{wgt}(t_3,\beta_3)$ and $w_4 = \underbrace{\mathit{wgt}(t_4,\gamma_4) + \mathit{wgt}(t_5,\gamma_5)}_{=w(t_4,t_6)} + \mathit{wgt}(t_6,\beta_3)$. This yields:

$$\mathit{purge}_{\mathcal{E}}(\pi) \;=\; t_0\,w_0\,t_1\,w_1\,t_2\,w_2\,t_4\,w_4\,t_7$$

where $w_0 = \mathit{wgt}(t_0,\beta_0)$ and $w_1 = \mathit{wgt}(t_1,\beta_1)$. Note that this implies t_2,t_3,t_4,t_5,t_6 to be states in $\mathcal{E}$, while states t_0,t_1 and t_7 might or might not be contained in $\mathcal{E}$.

A property φ is called $\mathcal{E}$-invariant if for all maximal paths ς the following conditions (I1) and (I2) hold:

(I1) If ς has an infinite suffix consisting of state-action pairs in $\mathcal{E}$ then $\varsigma \not\models \varphi$.
(I2) If $\varsigma \models \varphi$ and ς' is a maximal path with $\mathit{purge}_{\mathcal{E}}(\varsigma) = \mathit{purge}_{\mathcal{E}}(\varsigma')$ then $\varsigma' \models \varphi$.

Examples for $\mathcal{E}$-invariant properties are (positive or negative) weight-divergence or the pumping property, and so are properties of the form $\Diamond(t \wedge (\mathrm{wgt} \bowtie w))$ where t is a state not contained in $\mathcal{E}$, $\bowtie$ a comparison operator (*e.g.*, $=$ or $\geqslant$), and $w \in \mathbb{Z}$.

Recall that $\lim(\varsigma)$ denotes the set of state-action pairs that appear infinitely often in ς. In what follows, we write $\mathit{Limit}_{\mathcal{E}}$ to denote the set $\{\varsigma \in \mathit{IPaths} : \lim(\varsigma) = \mathcal{E}\}$.

Lemma B.18 (See (S3) in Lemma 3.7)**.** *Let $\mathcal{M}$ and $\mathcal{E}$ be as before. Then:*

(S3.1) For each scheduler $\mathfrak{T}$ for $\mathrm{Spider}_{\mathcal{E}}(\mathcal{M})$ there is a scheduler $\mathfrak{S}$ for $\mathcal{M}$ such that $\Pr^{\mathfrak{T}}_{\mathrm{Spider}_{\mathcal{E}}(\mathcal{M}),s}(\varphi) = \Pr^{\mathfrak{S}}_{\mathcal{M},s}(\varphi)$ for all states s in $\mathcal{M}$ and all $\mathcal{E}$-invariant properties φ. If $\mathfrak{T}$ is an MD-scheduler then $\mathfrak{S}$ can be chosen as an MD-scheduler.
(S3.2) For each scheduler $\mathfrak{S}$ for $\mathcal{M}$ there exists a (randomized) scheduler $\mathfrak{T}$ for $\mathrm{Spider}_{\mathcal{E}}(\mathcal{M})$ such that

$$\Pr^{\mathfrak{S}}_{\mathcal{M},s}(\varphi) \;\leqslant\; \Pr^{\mathfrak{T}}_{\mathrm{Spider}_{\mathcal{E}}(\mathcal{M}),s}(\varphi) \;\leqslant\; \Pr^{\mathfrak{S}}_{\mathcal{M},s}(\varphi) + \Pr^{\mathfrak{S}}_{\mathcal{M},s}(\mathit{Limit}_{\mathcal{E}})$$

for all states s and all $\mathcal{E}$-invariant properties φ.

Note that (S3.2) implies that $\Pr^{\mathfrak{S}}_{\mathcal{M},s}(\varphi) = \Pr^{\mathfrak{T}}_{\mathrm{Spider}_{\mathcal{E}}(\mathcal{M})}(\varphi)$ if $\Pr^{\mathfrak{S}}_{\mathcal{M},s}(\mathit{Limit}_{\mathcal{E}}) = 0$.

Proof. For statement (S3.1), we observe that $\mathcal{M}$ can mimick $\mathrm{Spider}_{\mathcal{E}}(\mathcal{M})$'s τ-transitions, followed by the state-action pair (s_0, β) with $\beta \in \mathit{Act}_{\mathcal{M}}(s)$ for some state s in $\mathcal{E}$ with $s \neq s_0$: First, the MD-scheduler that realizes $\mathcal{E}$ is simulated by choosing α_u for each state u in $\mathcal{E}$ until state s has been reached and then taking action β in state s. Note that $w(s, s_0) = -w(s_0, s)$ and therefore

$$\mathit{wgt}_{\mathrm{Spider}_{\mathcal{E}}(\mathcal{M})}(s, \tau) + \mathit{wgt}_{\mathrm{Spider}_{\mathcal{E}}(\mathcal{M})}(s_0, \beta) \;=\; w(s, s_0) + w(s_0, s) + \mathit{wgt}_{\mathcal{M}}(s, \beta) \;=\; \mathit{wgt}_{\mathcal{M}}(s, \beta)\ .$$

This yields a scheduler transformation "scheduler $\mathfrak{T}$ for $\mathrm{Spider}_{\mathcal{E}}(\mathcal{M}) \rightsquigarrow$ scheduler $\mathfrak{S}$ for $\mathcal{M}$" preserving the probabilities of all $\mathcal{E}$-invariant properties. Moreover, $\mathfrak{S}$ is MD if so is $\mathfrak{T}$.

The idea to provide scheduler transformations as stated in (S3.2) relies on the observation that $\mathrm{Spider}_{\mathcal{E}}(\mathcal{M})$ can mimic $\mathcal{M}$'s behavior inside $\mathcal{E}$, followed by a state-action pair (s, β), where s is a state in $\mathcal{E}$ with $s \neq s_0$ and $\beta \neq \alpha_s$, by taking the τ-transition from s to s_0, followed by the state-action pair (s_0, β).

Statement (S3.2) is obvious if $\mathcal{M} = \mathcal{E}$ in which case $\Pr^{\mathfrak{S}}_{\mathcal{M},s}(\psi) = 0$ for all $\mathcal{E}$-invariant properties ψ. (Recall that all paths with an infinite suffix in $\mathcal{E}$ violate ψ; see (I1) in the definition of $\mathcal{E}$-invariance.) Suppose now that $\mathcal{M} \neq \mathcal{E}$.

Let $\mathcal{H}$ be the MDP resulting from $\mathrm{Spider}_{\mathcal{E},s_0}(\mathcal{M})$ by adding a fresh state *final*, a fresh action name ι and a deterministic transition from the reference state s_0 of the spider construction to *final* and a deterministic self-loop at state *final*, both with action label ι and weight 0. Let $S' = S \cup \{\mathit{final}\}$. Obviously, $\mathcal{H}$ and $\mathcal{M}$ have the same traps.

We introduce a purge-function, called $\mathit{purge}_{\mathcal{H}}(\cdot)$, similar to the purge-function for paths in $\mathcal{M}$. Recall that the idea to define $\mathit{purge}_{\mathcal{E}}(\pi')$ for a path π' in $\mathcal{M}$ was to abstract away from the behavior inside $\mathcal{E}$ and just representing the state where $\mathcal{E}$ is entered and the action where $\mathcal{E}$ is left, and then replacing the action with the corresponding weight. Thus, for a path fragment of π' in which $\mathcal{E}$ is entered via state s and left via action $\alpha \in \mathit{Act}(t)$ (where t is a state of $\mathcal{E}$, while the state-action pair (t, α) does not belong to $\mathcal{E}$) leading to state u, the corresponding path fragment is replaced with swu where $w = w(s, t) + \mathit{wgt}_{\mathcal{M}}(t, \alpha)$. With the switch from $\mathcal{M}$ to $\mathrm{Spider}_{\mathcal{E},s_0}(\mathcal{M})$ resp. $\mathcal{H}$ the corresponding behavior of such a path fragment from s to u consists of two or one transition, namely $s\,\tau\,s_0\,\alpha\,u$ if $s \neq s_0$ or $s_0\,\alpha\,u$ if $s = s_0$. The idea is now to replace these path fragments with $s\,w'\,u$ where $w' = w(s, s_0) + \mathit{wgt}_{\mathcal{H}}(s_0, \alpha)$. Note that $\mathit{wgt}_{\mathcal{H}}(s_0, \alpha) = \mathit{wgt}_{\mathrm{Spider}_{\mathcal{E},s_0}(\mathcal{M})}(s_0, \alpha) = \mathit{wgt}(s_0, t) + \mathit{wgt}_{\mathcal{M}}(t, \alpha)$. As $w(s, s_0) + w(s_0, t) = w(s, t)$ we get $w = w'$.

The formal definition of $\mathit{purge}_{\mathcal{H}}(\pi)$ for the paths in $\mathcal{H}$ is as follows. Let $\pi = t_0\alpha_0t_1\alpha_1t_2\alpha_2\ldots$ be a (finite or infinite) path in $\mathcal{H}$. The sequence $\mathit{purge}_{\mathcal{H}}(\pi) \in (S' \times \mathbb{Z})^{\omega} \cup (S' \times \mathbb{Z})^* \times S'$ results from π as follows: First, each α_i with $\mathit{wgt}_{\mathcal{H}}(t_i, \alpha_i)$ is replaced, provided $t_i \neq s_0$ and (t_i, α_i) is a state-action pair of $\mathcal{M}$. Then, each path fragment of the form $t_i\,\tau\,t_{i+1}\,\alpha_{i+1}\,t_{i+2}$ where t_i belongs to $\mathcal{E}$, $t_i \neq s_0$, and t_{i+1} is the reference state s_0 of the spider construction is replaced with $t_i\,w\,t_{i+2}$ where $w = \mathit{wgt}_{\mathcal{H}}(t_i, \tau) + \mathit{wgt}_{\mathcal{H}}(s_0, \alpha_{i+1})$. Finally, each action α_i where $t_i = s_0$ is replaced with $\mathit{wgt}_{\mathcal{H}}(s_0, \alpha_i)$. Note that $\mathit{purge}_{\mathcal{H}}(\pi)$ is finite iff π is finite.

We now define the randomized scheduler $\mathfrak{U}$ for $\mathcal{H}$ that mimics the given scheduler $\mathfrak{S}$ for $\mathcal{M}$ as follows. Suppose $\pi = t_0\alpha_0t_1\alpha_1\ldots\alpha_{n-1}t_n$ is a finite path in $\mathcal{H}$ where t_n is not a trap and different from the auxiliary state *final*.

- If t_n is not a state of $\mathcal{E}$ and (t_n, α) a state-action pair of $\mathcal{M}$ (and $\mathcal{H}$) then $\mathfrak{U}(\pi)(\alpha)$ equals the conditional probability for $\mathfrak{S}$ to generate a path of the form $\pi'\alpha u$ where $\mathit{purge}_{\mathcal{E}}(\pi') = \mathit{purge}_{\mathcal{H}}(\pi)$ and u is an arbitrary α-successor of t_n under the condition that $\mathfrak{S}$ indeed schedules such a path π'. That is, if $t_0 = s$ and $\Pi_{\mathcal{M},s}$ denotes the set of all finite paths π' with

$first(\pi') = s$ and $purge_{\mathcal{E}}(\pi') = purge_{\mathcal{H}}(\pi)$ then

$$\mathfrak{U}(\pi)(\alpha) \;=\; \frac{\sum\limits_{\pi' \in \Pi_{\mathcal{M},s}} \sum\limits_{u \in S} \Pr^{\mathfrak{S}}_{\mathcal{M},s}(\pi' \alpha u)}{\sum\limits_{\pi' \in \Pi_{\mathcal{M},s}} \Pr^{\mathfrak{S}}_{\mathcal{M},s}(\pi')}$$

where we assume that $\Pi_{\mathcal{M},s}$ contains at least one $\mathfrak{S}$-path and use $\Pr^{\mathfrak{S}}_{\mathcal{M},s}(\pi')$ as a short-form notation for the probability for the cylinder set of π' under $\mathfrak{S}$.[3] If there is no $\mathfrak{S}$-path $\pi' \in \Pi_{\mathcal{M},s}$ then $\mathfrak{U}(\pi)$ is irrelevant for our purposes.

- Suppose now t_n is a state of $\mathcal{E}$, $t_n \neq s_0$ and either $n = 0$ or $n \geqslant 1$ and $\alpha_{n-1} \neq \tau$. Then, $\mathfrak{U}(\pi)(\tau) = 1$. Moreover, if α is an action in $Act_{\mathcal{H}}(s_0) \setminus \{\iota\}$ then $\mathfrak{U}(\pi\tau s_0)(\alpha)$ equals the conditional probability for $\mathfrak{S}$ to generate a path of the form π' with $purge_{\mathcal{E}}(\pi') = purge_{\mathcal{H}}(\pi\tau s_0 \alpha u)$ where u is an arbitrary α-successor of s_0 in $\mathcal{H}$. With the remaining probability, $\mathfrak{U}$ moves from s_0 to *final* via the fresh action ι. That is, $\mathfrak{U}(\pi\tau s_0)(\iota) = 1 - \sum_\alpha \mathfrak{U}(\pi\tau s_0)(\alpha)$. This value equals the conditional probability for $\mathfrak{S}$ to generate an infinite path $\varsigma = \pi''; \varsigma'$ where $purge_{\mathcal{E}}(\pi'') = purge_{\mathcal{H}}(\pi)$ and ς' is an infinite path in $\mathcal{E}$. In both cases, the condition is that $\mathfrak{S}$ generates at least one path π'' with $purge_{\mathcal{E}}(\pi'') = purge_{\mathcal{H}}(\pi)$. If no such path π'' exists then $\mathfrak{U}(\pi)$ is irrelevant for our purposes.
- The definition of $\mathfrak{U}(\pi)$ is analogous when $last(\pi) = s_0$ and either $n = 0$ or (t_{n-1}, α_{n-1}) is a state-action pair of $\mathcal{M}$.

Note that $\mathfrak{U}(\pi_1) = \mathfrak{U}(\pi_2)$ whenever π_1, π_2 are finite paths with $purge_{\mathcal{E}}(\pi_1) = purge_{\mathcal{E}}(\pi_2)$.

For simplicity, let us assume that $\mathcal{M}$ has no traps, in which case all maximal paths in $\mathcal{M}$ (and $\mathcal{H}$) are infinite. Given an $\mathcal{E}$-invariant property ψ for $\mathcal{M}$, let $purge(\psi)$ denote the set of all words $purge_{\mathcal{E}}(\varsigma)$ where ς is an infinite path in $\mathcal{M}$ with $\varsigma \models \psi$. (Recall that $\varsigma \not\models \psi$ for all infinite paths ς with $\lim(\varsigma) = \mathcal{E}$ by definition of $\mathcal{E}$-invariance.) Then, $purge(\psi) \subseteq (S \times \mathbb{Z})^\omega$ and $purge(\psi)$ can be viewed as a measurable subset of $(S' \times \mathbb{Z})^\omega$ where measurability is understood with respect to the sigma-algebra $\Sigma_{\mathcal{H}}$ generated by the cylinder sets spanned by the finite strings $(S' \times \mathbb{Z})^* \times S'$. (Recall that $S' = S \cup \{final\}$.)

The probability measure $\Pr^{\mathfrak{S}}_{\mathcal{M},s}$ on the maximal paths of $\mathcal{M}$ induces a probability measure $\mu^{\mathfrak{S}}_{\mathcal{M},s}$ on the sigma-algebra $\Sigma_{\mathcal{H}}$ over $(S' \times \mathbb{Z})^\omega$ using the embedding $e \colon IPaths_{\mathcal{M}} \to (S' \times \mathbb{Z})^\omega$ given by:

- $e(\varsigma) = purge_{\mathcal{E}}(\varsigma)$ if ς contains infinitely many actions not contained in $\mathcal{E}$ and
- $e(\varsigma) = purge_{\mathcal{E}}(\pi')\, 0\, final\, 0\, final \ldots$ if $\varsigma = \pi'; \varsigma'$ where ς' is an infinite paths consisting of state-action pairs in $\mathcal{E}$ and either π' consists of a single state or the last state-action pair of π' does not belong to $\mathcal{E}$.

Then, $\mu^{\mathfrak{S}}_{\mathcal{M},s}$ is the unique probability measure given by $\mu^{\mathfrak{S}}_{\mathcal{M},s}(Cyl(t_0\, w_0\, \ldots\, w_{n-1}\, t_n)) = \sum_{\pi'} \Pr^{\mathfrak{S}}_{\mathcal{M},s}(Cyl(\pi'))$ where π' ranges over all finite paths in $\mathcal{M}$ with $purge_{\mathcal{E}}(\pi') = t_0\, w_0\, \ldots\, w_{n-1}\, t_n$. Given a 0-EC-invariant property ψ, the image of the set of infinite ς in $\mathcal{M}$ satisfying ψ under e equals $purge(\psi)$. This yields $\Pr^{\mathfrak{S}}_{\mathcal{M},s}(\psi) = \mu^{\mathfrak{S}}_{\mathcal{M},s}(purge(\psi))$.

Likewise, scheduler $\mathfrak{U}$ for $\mathcal{H}$ induces a probability measure $\mu^{\mathfrak{U}}_{\mathcal{H},s}$ over this sigma-algebra such that $\Pr^{\mathfrak{U}}_{\mathcal{H},s}(\psi) = \mu^{\mathfrak{U}}_{\mathcal{H},s}(purge(\psi))$ for each 0-EC-invariant property ψ.

By construction, $\mu^{\mathfrak{S}}_{\mathcal{M},s}$ and $\mu^{\mathfrak{U}}_{\mathcal{H},s}$ agree on the cylinder sets of the finite strings in $(S' \times \mathbb{Z})^* \times S'$.

This can be shown by induction on the length of strings in $(S' \times \mathbb{Z})^* \times S'$. In the step of induction, we consider a string of the form $t_0\, w_0\, t_1\, w_1 \ldots w_{n-1}\, t_n\, w_n\, t_{n+1} \in (S' \times \mathbb{Z})^* \times S'$. Let $s = t_0$. Suppose for simplicity that t_n is a state in $\mathcal{M}$ that does not belong to $\mathcal{E}$. Let A denote the set of actions α where (t_n, α) is a state-action pair in $\mathcal{M}$ (and $\mathcal{H}$) and $wgt(t_n, \alpha) = w_n$. Let $\Pi_{\mathcal{M},s}$ denote the set of finite paths π' in $\mathcal{M}$ with $purge_{\mathcal{E}}(\pi') = t_0\, w_0\, t_1\, w_1 \ldots w_{n-1}\, t_n$. Likewise, we write $\Pi_{\mathcal{H},s}$ to denote the set of finite paths π in $\mathcal{H}$ with $purge_{\mathcal{H}}(\pi) = t_0\, w_0\, t_1\, w_1 \ldots w_{n-1}\, t_n$. By induction hypothesis we have

$$\sum_{\pi \in \Pi_{\mathcal{H},s}} \Pr^{\mathfrak{U}}_{\mathcal{H},s}(\pi) \;=\; \mu^{\mathfrak{U}}_{\mathcal{H},s}(t_0\, w_0\, t_1\, w_1 \ldots w_{n-1}\, t_n) \;=\; \mu^{\mathfrak{S}}_{\mathcal{M},s}(t_0\, w_0\, t_1\, w_1 \ldots w_{n-1}\, t_n) \;=\; \sum_{\pi' \in \Pi_{\mathcal{M},s}} \Pr^{\mathfrak{U}}_{\mathcal{H},s}(\pi')$$

[3] The cylinder set of a finite path π' denotes the set of maximal paths ς where π' is a prefix of ς.

Let p denote this value and suppose $p > 0$. Then:

$$\begin{aligned}
\mu^{\mathfrak{U}}_{\mathcal{H},s}(t_0\, w_0\, t_1\, w_1 \ldots w_{n-1}\, t_n\, w_n\, t_{n+1}) &= \sum_{\pi\in\Pi_{\mathcal{H},s}} \sum_{\beta\in A} \Pr^{\mathfrak{U}}_{\mathcal{H},s}(\pi)\cdot \mathfrak{U}(\pi)(\beta)\cdot P(t_n,\beta,t_{n+1}) \\
&= \sum_{\pi\in\Pi_{\mathcal{H},s}} \sum_{\beta\in A} \Pr^{\mathfrak{U}}_{\mathcal{H},s}(\pi)\cdot \frac{\sum\limits_{\pi'\in\Pi_{\mathcal{M},s}} \sum\limits_{u\in S} \Pr^{\mathfrak{S}}_{\mathcal{M},s}(\pi'\beta u)}{\sum\limits_{\pi'\in\Pi_{\mathcal{M},s}} \Pr^{\mathfrak{S}}_{\mathcal{M},s}(\pi')}\cdot P(t_n,\beta,t_{n+1}) \\
&= \sum_{\pi\in\Pi_{\mathcal{H},s}} \sum_{\beta\in A} \Pr^{\mathfrak{U}}_{\mathcal{H},s}(\pi)\cdot \frac{1}{p}\cdot \sum_{\pi'\in\Pi_{\mathcal{M},s}} \sum_{u\in S} \Pr^{\mathfrak{S}}_{\mathcal{M},s}(\pi'\beta u)\cdot P(t_n,\beta,t_{n+1}) \\
&= \frac{1}{p}\cdot \sum_{\beta\in A} P(t_n,\beta,t_{n+1})\cdot \sum_{\pi'\in\Pi_{\mathcal{M},s}} \sum_{u\in S} \Pr^{\mathfrak{S}}_{\mathcal{M},s}(\pi'\beta u)\cdot \sum_{\pi\in\Pi_{\mathcal{H},s}} \Pr^{\mathfrak{U}}_{\mathcal{H},s}(\pi) \\
&= \sum_{\beta\in A} P(t_n,\beta,t_{n+1})\cdot \sum_{\pi'\in\Pi_{\mathcal{M},s}} \Pr^{\mathfrak{S}}_{\mathcal{M},s}(\pi')\cdot \mathfrak{S}(\pi')(\beta)\cdot \sum_{u\in S} P(t_n,\beta,u) \\
&= \sum_{\pi'\in\Pi_{\mathcal{M},s}} \sum_{\beta\in A} \Pr^{\mathfrak{S}}_{\mathcal{M},s}(\pi')\cdot \mathfrak{S}(\pi')(\beta)\cdot P(t_n,\beta,t_{n+1}) \\
&= \mu^{\mathfrak{S}}_{\mathcal{M},s}(t_0\, w_0\, t_1\, w_1 \ldots w_{n-1}\, t_n\, w_n\, t_{n+1})
\end{aligned}$$

The calculation for the other cases is similar.

By Caratheodory's measure-extension theorem, $\Pr^{\mathfrak{S}}_{\mathcal{M},s}$ and $\Pr^{\mathfrak{U}}_{\mathcal{H},s}$ agree when viewed as measures over $(S'\times\mathbb{Z})^\omega$. For all $\mathcal{E}$-invariant properties ψ it holds $\psi = e^{-1}(\mathit{purge}_{\mathcal{E}}(\psi))$ and hence, we obtain

$$\Pr^{\mathfrak{S}}_{\mathcal{M},s}(\psi) \;=\; \mu^{\mathfrak{S}}_{\mathcal{M},s}(\mathit{purge}_{\mathcal{E}}(\psi)) \;=\; \mu^{\mathfrak{S}}_{\mathcal{H},s}(\mathit{purge}_{\mathcal{E}}(\psi)) \;=\; \Pr^{\mathfrak{U}}_{\mathcal{H},s}(\psi)\ .$$

We finally switch from scheduler $\mathfrak{U}$ for $\mathcal{H}$ to a scheduler $\mathfrak{T}$ for $\mathsf{Spider}_{\mathcal{E},s_0}(\mathcal{M})$. For this, we pick arbitrary actions α_s enabled in state s of $\mathcal{M}$ and define $\mathfrak{T}$ by $\mathfrak{T}(\pi)(\beta) = \mathfrak{U}(\pi)(\beta)$ for each action $\beta \neq \alpha_s$ that is enabled in s as a state of $\mathsf{Spider}_{\mathcal{E},s_0}(\mathcal{M})$ and $\mathfrak{T}(\pi)(\alpha_s) = \mathfrak{U}(\pi)(\alpha_s) + \mathfrak{U}(\pi)(\iota)$. We then have:

$$\Pr^{\mathfrak{T}}_{\mathsf{Spider}_{\mathcal{E},s_0}(\mathcal{M}),s}(\psi) \;\geqslant\; \Pr^{\mathfrak{U}}_{\mathcal{H},s}(\psi)$$

for all $\mathcal{E}$-invariant properties ψ. Clearly, $\Pr^{\mathfrak{U}}_{\mathcal{H},s}(\psi) + \Pr^{\mathfrak{S}}_{\mathcal{M},s}(\mathit{Limit}_{\mathcal{E}})$ is an upper bound for $\Pr^{\mathfrak{T}}_{\mathsf{Spider}_{\mathcal{E},s_0}(\mathcal{M}),s}(\psi)$. □

As a consequence of Lemma B.18 we get:

Corollary B.19. *For each $\mathcal{E}$-invariant property φ and each state s in $\mathcal{M}$, $\Pr^{\sup}_{\mathcal{M},s}(\varphi) = \Pr^{\sup}_{\mathsf{Spider}_{\mathcal{E}}(\mathcal{M}),s}(\varphi)$. Furthermore, the existence of a scheduler $\mathfrak{S}$ for $\mathcal{M}$ with $\Pr^{\sup}_{\mathcal{M},s}(\varphi) = \Pr^{\mathfrak{S}}_{\mathcal{M},s}(\varphi)$ implies the existence of a scheduler $\mathfrak{T}$ for $\mathsf{Spider}_{\mathcal{E}}(\mathcal{M})$ with $\Pr^{\sup}_{\mathcal{M},s}(\varphi) = \Pr^{\mathfrak{T}}_{\mathsf{Spider}_{\mathcal{E}}(\mathcal{M}),s}(\varphi)$, and vice versa.*

As weight-divergence and the gambling condition are $\mathcal{E}$-invariant properties, we obtain that the spider construction preserves weight-divergence and the pumping property as stated in Corollary 3.8. Moreover, we get:

Corollary B.20. *Suppose $\mathcal{M}$ is strongly connected and $\mathbb{E}^{\max}_{\mathcal{M}}(\mathrm{MP}) = 0$. If $\mathcal{E}$ is a 0-BSCC of $\mathcal{M}$ then either $\mathsf{Spider}_{\mathcal{E}}(\mathcal{M})$ has no maximal end component or $\mathbb{E}^{\max}_{\mathcal{F}}(\mathrm{MP}) \leqslant 0$ for the unique maximal end component $\mathcal{F}$ of $\mathsf{Spider}_{\mathcal{E}}(\mathcal{M})$.*

Proof. Suppose $\mathsf{Spider}_{\mathcal{E}}(\mathcal{M})$ has end components. Let $\mathcal{F}$ be the unique maximal end component of $\mathsf{Spider}_{\mathcal{E}}(\mathcal{M})$. But then $\mathbb{E}^{\max}_{\mathcal{F}}(\mathrm{MP}) \leqslant 0$ as otherwise $\mathcal{F}$ would be pumping (see Lemma 3.3), in which case $\mathcal{M}$ would be pumping (by Corollary 3.8). This, however, is impossible (again by Lemma 3.3) as $\mathbb{E}^{\max}_{\mathcal{M}}(\mathrm{MP}) = 0$. □

When the spider construction is applied to an MDP $\mathcal{M}$ that is not strongly connected, then $\mathsf{Spider}_{\mathcal{E}}(\mathcal{M})$ is obtained from $\mathcal{M}$ by replacing $\mathcal{F}$ with $\mathsf{Spider}_{\mathcal{E}}(\mathcal{F})$ where $\mathcal{F}$ is the unique maximal end component of $\mathcal{M}$ that contains the given 0-BSCC $\mathcal{E}$. Moreover, state-action pairs $(s,\alpha) \in \mathcal{M}\setminus\mathcal{F}$ with s being a state of $\mathcal{F}$ that is different from the reference state s_0 are replaced with (s_0,α) where $P_{\mathsf{Spider}_{\mathcal{E}}(\mathcal{M})}(s_0,\alpha,u) = P_{\mathcal{M}}(s,\alpha,u)$ for all states u and $\mathit{wgt}_{\mathsf{Spider}_{\mathcal{E}}(\mathcal{M})}(s_0,\alpha) = w(s_0,s) + \mathit{wgt}_{\mathcal{M}}(s,\alpha)$. Obviously, there is no end component $\mathcal{G}$ of $\mathsf{Spider}_{\mathcal{E}}(\mathcal{M})$ that subsumes $\mathsf{Spider}_{\mathcal{E}}(\mathcal{F})$ and $\mathcal{G} \neq \mathsf{Spider}_{\mathcal{E}}(\mathcal{F})$. Note that otherwise there would be a corresponding end component of $\mathcal{M}$ that strictly subsumes $\mathcal{F}$, which is impossible by the maximality of $\mathcal{F}$. Hence, by Corollary 3.8:

Corollary B.21 (Generalization of Corollary 3.8 for possibly not strongly connected MDPs)**.** *Let $\mathcal{M}$ be a (possibly not strongly connected) MDP and $\mathcal{E}$ a 0-BSCC of $\mathcal{M}$. Then, $\mathcal{M}$ has a weight-divergent (resp. pumping) end component iff* $\mathrm{Spider}_{\mathcal{E}}(\mathcal{M})$ *has a weight-divergent (resp. pumping) end component.*

B.5.2 Iterative Application of the Spider Construction

Let $\mathcal{M}$ be a (possibly not strongly connected) MDP such that $\mathbb{E}^{\max}_{\mathcal{F}}(\mathrm{MP}) \leqslant 0$ for all maximal end components $\mathcal{F}$ of $\mathcal{M}$. The iterative application of the spider construction generates a sequence $\mathcal{M}_0 = \mathcal{M}, \mathcal{M}_1, \ldots, \mathcal{M}_\ell = \mathcal{N}$ of MDPs with the same state space and where $\mathcal{M}_{i+1} = \mathrm{Spider}_{\mathcal{E}_i, s_{0,i}}(\mathcal{M}_i)$ arises from $\mathcal{M}_i$ by flattening some 0-BSCC $\mathcal{E}_i$ of $\mathcal{M}_i$ through the spider construction. More precisely, $\mathcal{E}_i$ is a 0-BSCC contained in a maximal end component $\mathcal{F}_i$ of $\mathcal{M}_i$ where $\mathbb{E}^{\max}_{\mathcal{F}_i}(\mathrm{MP}) = 0$ and $\mathcal{M}_{i+1}$ is obtained from $\mathcal{M}_i$ by

- replacing $\mathcal{F}_i$ with $\mathrm{Spider}_{\mathcal{E}_i, s_{0,i}}(\mathcal{F}_i)$ and
- replacing all state-action pairs $(s, \alpha) \in \mathcal{M}_i \setminus \mathcal{F}_i$ with $(s_{0,i}, \alpha)$, provided s is a state of $\mathcal{F}_i$ different from the reference state $s_{0,i}$.

The transition probabilities of the new state-action pairs $(s_{0,i}, \alpha)$ are the same as in $\mathcal{M}_i$, *i.e.*, $P_{\mathcal{M}_i}(s, \alpha, u) = P_{\mathcal{M}_{i+1}}(s_{0,i}, \alpha, u)$ for all states u, and the weight is given by $wgt_{\mathcal{M}_{i+1}}(s_{0,i}, \alpha) = w(s_{0,i}, s) + wgt_{\mathcal{M}_i}(s, \alpha)$. Here, $w(s,t)$ is the weight of some/each path s to t inside some 0-EC of $\mathcal{F}_i$ (see Remark B.17).

The algorithm terminates with the final MDP $\mathcal{M}_\ell$ if either there is no maximal end component $\mathcal{F}$ of $\mathcal{M}_\ell$ where $\mathbb{E}^{\max}_{\mathcal{F}}(\mathrm{MP}) = 0$ or for each maximal end component $\mathcal{F}$ of $\mathcal{M}_\ell$ with $\mathbb{E}^{\max}_{\mathcal{F}}(\mathrm{MP}) = 0$, the constructed MD-scheduler $\mathfrak{S}_{\mathcal{F}}$ for $\mathcal{F}$ maximizing the expected mean payoff has a gambling BSCC.

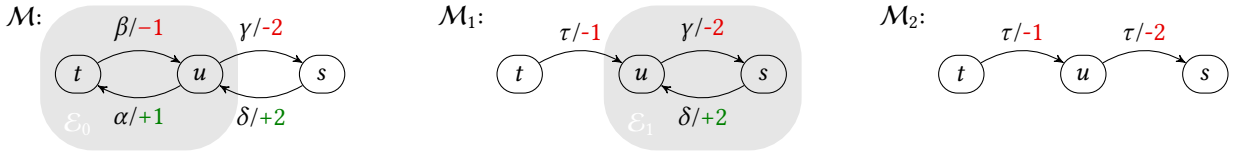


Figure 8. Iterative application of the spider construction.

Example B.22. The MDPs occurring within the following example are illustrated by Figure 8. Let $\mathcal{M}_0 = \mathcal{M}$ be the strongly connected MDP on the left of the figure. The iterative spider construction might first detect the 0-BSCC $\mathcal{E}_0 = \{(t, \beta), (u, \alpha)\}$. It then generates $\mathcal{M}_1 = \mathrm{Spider}_{\mathcal{E}_0, u}(\mathcal{M})$ shown in the center, where u is the reference state. Then, $\mathcal{E}_1 = \{(u, \gamma), (s, \delta)\}$ is the unique maximal end component of $\mathcal{M}_1$, and $\mathcal{E}_1$ is even a 0-BSCC of $\mathcal{M}_1$. The next iteration is $\mathcal{M}_2 = \mathrm{Spider}_{\mathcal{E}_1, s}(\mathcal{M}_1)$ shown on the right. As $\mathcal{M}_2$ does not have any end component, the iterative spider construction terminates with the MDP $\mathcal{M}_2$. ∎

We get by Lemma 3.7 and Corollary 3.8:

Lemma B.23 (Maximal end components of $\mathcal{M}_i$)**.** *For each $i \in \{1, \ldots, \ell\}$:*

(a) There is an injection ι that maps each maximal end component $\mathcal{F}$ of $\mathcal{M}_i$ to one of the maximal end components of the original MDP $\mathcal{M}$ such that the state space of $\mathcal{F}$ is contained in the state space of $\iota(\mathcal{F})$ and $\mathcal{F}$ is weight-divergent (resp. pumping) iff $\iota(\mathcal{F})$ is weight-divergent (resp. pumping).

(b) The states and actions that are contained in a 0-EC of $\mathcal{M}$ are exactly the states and non-τ actions that belong to one of the 0-BSCCs of $\mathcal{E}_1, \ldots, \mathcal{E}_{i-1}$ or are contained in a 0-EC of $\mathcal{M}_i$.

As the spider construction preserves the weight-divergence and pumping property (see Corollary 3.8), Lemma B.23 yields:

Corollary B.24. *If $\mathcal{M}$ has no weight-divergent end component then $\mathbb{E}^{\max}_{\mathcal{F}}(\mathrm{MP}) < 0$ for all end components $\mathcal{F}$ of the final MDP $\mathcal{M}_\ell$.*

By property (S2) of Lemma 3.7, the number of state-action pairs is strictly decreasing, *i.e.*, $\|\mathcal{M}_0\| > \|\mathcal{M}_1\| > \ldots > \|\mathcal{M}_\ell\|$. Hence, the number ℓ of recursive calls of the spider construction is bounded by $\|\mathcal{M}\|$.

Analogous to (S3) in Lemma 3.7 (see also Lemma B.18) we obtain the equivalence of $\mathcal{M}$ and the MDPs $\mathcal{M}_1, \ldots, \mathcal{M}_\ell$ with respect to the class of 0-EC-invariant properties. These are properties that are $\mathcal{E}$-invariant for each 0-EC $\mathcal{E}$ of $\mathcal{M}$.

Lemma B.25 (Equivalence of $\mathcal{M}$ and $\mathcal{M}_i$ w.r.t. 0-EC-invariant properties)**.** *The original MDP $\mathcal{M}$ and the MDP $\mathcal{M}_i$ for $i \in \{1, \ldots, \ell\}$ are equivalent in the following sense:*

(a) For each scheduler $\mathfrak{T}$ for $\mathcal{M}_i$ there is a scheduler $\mathfrak{S}$ for $\mathcal{M}$ such that $\Pr^{\mathfrak{T}}_{\mathcal{M}_i, s}(\varphi) = \Pr^{\mathfrak{S}}_{\mathcal{M}, s}(\varphi)$ for all states s in $\mathcal{M}$ and all 0-EC-invariant properties φ. If $\mathfrak{T}$ is an MD-scheduler then $\mathfrak{S}$ can be chosen as an MD-scheduler.

(b) For each scheduler $\mathfrak{S}$ for $\mathcal{M}$ there is a scheduler $\mathfrak{T}$ for $\mathcal{M}_i$ such that

$$\Pr^{\mathfrak{S}}_{\mathcal{M},s}(\varphi) \;\leqslant\; \Pr^{\mathfrak{T}}_{\mathcal{M}_i,s}(\varphi) \;\leqslant\; \Pr^{\mathfrak{S}}_{\mathcal{M},s}(\varphi) + \Pr^{\mathfrak{S}}_{\mathcal{M},s}\big\{\varsigma \in \mathit{IPaths} : \lim(\varsigma) \in \{\mathcal{E}_1,\ldots,\mathcal{E}_{i-1}\}\big\}$$

for all states s in $\mathcal{M}$ and all 0-EC-invariant properties φ. In particular:

$$\Pr^{\mathfrak{S}}_{\mathcal{M},s}(\varphi) = \Pr^{\mathfrak{T}}_{\mathcal{M}_i,s}(\varphi) \quad \textit{if } \Pr^{\mathfrak{S}}_{\mathcal{M},s}\{\varsigma \in \mathit{IPaths} : \lim(\varsigma) \textit{ is a 0-EC}\} = 0.$$

Thus, $\Pr^{\sup}_{\mathcal{M},s}(\varphi) = \Pr^{\sup}_{\mathcal{M}_i,s}(\varphi)$ for each 0-EC-invariant property φ and each state s in $\mathcal{M}$. Furthermore, the existence of a scheduler $\mathfrak{S}$ for $\mathcal{M}$ with $\Pr^{\sup}_{\mathcal{M},s}(\varphi) = \Pr^{\mathfrak{S}}_{\mathcal{M},s}(\varphi)$ implies the existence of a scheduler $\mathfrak{T}$ for $\mathcal{M}_i$ with $\Pr^{\sup}_{\mathcal{M},s}(\varphi) = \Pr^{\mathfrak{T}}_{\mathcal{M},s}(\varphi)$, and vice versa.

The proof follows from Lemma B.18 using an inductive argument.

Graph structure of the MDPs $\mathcal{M}_i$. With the renaming of the actions – prior to the application of the spider construction – to ensure that the action sets for each pair of distinct states are disjoint, the action label τ of transitions in $\mathcal{M}_{j-1}$ will be renamed when constructing $\mathcal{M}_i$ by applying the spider construction towards $\mathcal{M}_{i-1}$. This, however, is irrelevant for the following arguments and we will simply refer to them as τ-transitions.

Lemma B.26 (τ-transitions in the MDPs $\mathcal{M}_i$)**.** *For $i \in \{1,\ldots,\ell\}$:*

(a) The τ-transitions are always deterministic, i.e., have a single target state that will be reached with probability 1. The weight of a τ-transition from s to t in $\mathcal{M}_i$ is $w(s,t)$.
(b) Each state s has at most one τ-transition in $\mathcal{M}_i$, and if s has a τ-transition in $\mathcal{M}_i$ then no other action is enabled in s as a state of $\mathcal{M}_i$.
(c) The graph built by the τ-transitions in $\mathcal{M}_i$ is acyclic.

In particular, n is an upper bound for the total number of τ-transitions in each of the MDPs $\mathcal{M}_i$.

Proof. Statement (a) is clear by the spider construction. The proof for statement (b) is by induction on i. For $i = 1$ (basis of induction) this is clear by the spider construction. For the step of induction $i \Longrightarrow i{+}1$, we use the fact that if s does not belong to the selected 0-BSCC $\mathcal{E}_i$ of the unique maximal end component $\mathcal{F}_i$ of $\mathcal{M}_i$ then $(s,\alpha) \in \mathcal{M}_i$ iff $(s,\alpha) \in \mathcal{M}_{i+1}$. (Recall that $\mathcal{M}_{i+1}$ arises from $\mathcal{M}_i$ by replacing $\mathcal{F}_i$ with $\mathrm{Spider}_{\mathcal{E}_i}(\mathcal{F}_i)$ for some 0-BSCC $\mathcal{E}_i$ of $\mathcal{F}_i$ and replacing all state-action pairs $(s,\alpha) \in \mathcal{M}_i \setminus \mathcal{E}_i$ where s is a state of $\mathcal{F}_i$ different from the reference state $s_{0,i}$ with $(s_{0,i},\alpha)$.) If s is contained in $\mathcal{E}_i$, but not the reference state $s_{0,i}$ of the spider construction, then the spider construction replaces all state-action pairs $(s,\alpha) \in \mathcal{M}_i \setminus \mathcal{E}_i$ with $(s_{0,i},\alpha)$, discards the unique state-action pair $(s,\alpha) \in \mathcal{E}_i$ and creates a τ-transition from s to $s_{0,i}$. Thus, if s belongs to $\mathcal{E}_i$ and $(s,\tau) \in \mathcal{M}_i$ then $(s,\tau) \in \mathcal{E}_i$ (as s has no other actions in $\mathcal{M}_i$) and (s,τ) will be discarded. In particular, the reference state $s_{0,i}$ does not get "new" τ-transitions, *i.e.*,

$$\text{State } t \text{ is a } \tau\text{-successor of } s_{0,i} \text{ in } \mathcal{M}_i \text{ iff } t \text{ is a } \tau\text{-successor of } s_{0,i} \text{ in } \mathcal{M}_{i+1}. \tag{*}$$

This yields the claim for (b).

Statement (c) follows by induction on i. The claim is obvious for $i = 0$. In the step of induction, we assume that the MDPs $\mathcal{M}_0,\ldots,\mathcal{M}_i$ do not have any τ-cycle, *i.e.*, a cycle built by τ-transitions. Suppose by contradiction that $\mathcal{M}_{i+1} = \mathrm{Spider}_{\mathcal{E}}(\mathcal{M}_i)$ has a τ-cycle $\xi = t_0\,\tau\,t_1\,\tau\ldots\tau\,t_n$. At least one of the states t_i must be contained in the 0-BSCC $\mathcal{E}_i$ of $\mathcal{M}_i$ as otherwise ξ would be a τ-cycle in $\mathcal{M}_i$. All states s in $\mathcal{E}_i$ that are different from the reference state $s_{0,i}$ have a τ-transition to $s_{0,i}$ in $\mathcal{M}_{i+1}$ and no other action is enabled in s as a state of $\mathcal{M}_{i+1}$. Hence, with some state of $\mathcal{E}_i$ also the reference state $s_{0,i}$ must be contained in ξ. Now let $s_{0,i} = t_0$ without loss of generality. But then the transition from $t_0 = s_{0,i}$ to t_1 must have been generated in an earlier application of the spider construction (see (*) above). That is, $t_1 = s_{0,i_1}$ for some $i_1 < i$. We repeat this argument and get natural numbers $i_1,\ldots,i_n$ with $0 \leqslant i_n < i_{n-1} < \ldots < i_1$ such that $t_m = s_{0,i_m}$. But then ξ is a τ-cycle in $\mathcal{M}_i$. Contradiction. □

Towards establishing the polynomial-time complexity of the iterative application of the spider construction, let us discuss the sizes of the MDPs $\mathcal{M}_i$. Although a single application $\mathcal{M}_i \rightsquigarrow \mathcal{M}_{i+1}$ of the spider construction can be done in polynomial time depending of the size of $\mathcal{M}_i$, this is not as obvious for the sequence $\mathcal{M}_0 \rightsquigarrow \ldots \rightsquigarrow \mathcal{M}_l$: As the number l of applications required depends on the input $\mathcal{M}_0$, the size of $\mathcal{M}_l$ could be exponential in the size of $\mathcal{M}_0$. Fortunately, this is not the case, as the following lemma shows.

Lemma B.27 (Size of the MDPs $\mathcal{M}_i$)**.** *Notations as before. The size of $\mathcal{M}_i$ is polynomially bounded by the size of $\mathcal{M}$.*

Proof. Recall that the size of an MDP has been defined as the number of states plus the total sum of the logarithmic lengths of the weights of all state-action pairs and the transition probabilities. Let S be the state space of $\mathcal{M}$ (and $\mathcal{M}_i$), and let $n = |S|$ denote the number of states in $\mathcal{M}$. Furthermore, let $w^{\max} = \{|wgt(s,\alpha)| : s \in S, \alpha \in Act_{\mathcal{M}}(s)\}$.

The state-action pairs of $\mathcal{M}_i$ have the form (s,β) where β is an action of $\mathcal{M}$ or stand for a τ-transition. All τ-actions are deterministic (*i.e.*, probability 1 for a single target state and 0 for all other states). The weight of a τ-transition from s to t in

$\mathcal{M}_i$ is $w(s,t)$, which is the weight of all paths from s to t in each 0-EC containing s and t. The latter statement is a consequence of the fact that the union of 0-ECs in MDPs with maximal expected mean payoff 0 is again a 0-EC. See Lemma B.10. Moreover, $|w(s,t)| \leqslant (n-1) \cdot w^{\max}$.

The total number of τ-transitions in $\mathcal{M}_i$ is bounded by n (see Lemma B.26). Hence, the total logarithmic length of the weights of all τ-transitions in $\mathcal{M}_i$ is polynomially bounded by the size of the original MDP $\mathcal{M}$. Likewise, for the state-action pairs (s,β) where β is an action of $\mathcal{M}$, say $\beta \in Act_{\mathcal{M}}(t)$, the logarithmic length of the transition probabilities in $\mathcal{M}_i$ are the same as in $\mathcal{M}$ as we have $P_{\mathcal{M}_i}(s,\beta,u) = P_{\mathcal{M}}(s,\beta,u)$. Moreover, we have $wgt_{\mathcal{M}_i}(s,\beta) = w(s,t) + wgt_{\mathcal{M}}(t,\beta)$. But then again the total logarithmic length of the weights for these state-action pairs in $\mathcal{M}_i$ is polynomially bounded by the size of $\mathcal{M}$. □

Lemma B.28. *If α is an action of $\mathcal{M}$ that does not belong to a 0-EC then α is an action of each of the MDPs $\mathcal{M}_i$.*

Proof. By induction on i. □

As a consequence of part (b) of Lemma B.23 and Lemma B.28 we get:

Remark B.29 (Actions in the final MDP $\mathcal{N}$)**.** Let $\mathcal{N} = \mathcal{M}_\ell$ be the MDP that has been generated by the iterative application of the spider construction to an MDP $\mathcal{M}$ that has no weight-divergent end component. Recall from Lemma B.26 that each state s has at most one τ-transition in $\mathcal{N}$, and if so, then no other action is enabled in s as a state of $\mathcal{N}$.

The actions in $\mathcal{N}$ are either actions of $\mathcal{M}$ that do not belong to a 0-EC of $\mathcal{M}$ or τ. In more detail, this means the following.

- If s is a state that does not belong to a 0-EC of $\mathcal{M}$ then $(s,\alpha) \in \mathcal{M}$ iff $(s,\alpha) \in \mathcal{M}_i$ with the same weight and the same transition probabilities.
- Suppose now that s is a state that belongs to some 0-EC of $\mathcal{M}$. Then, for each state-action pair $(s,\alpha) \in \mathcal{N}$:
  - Either $\alpha = \tau$, in which case $P_{\mathcal{N}}(s,\alpha,t) = 1$ and $wgt(s,\alpha) = w(s,t)$ for some state t that belongs to the same maximal 0-EC as s,
  - or there is a state t that belongs to the same maximal 0-EC $\mathcal{E}$ of $\mathcal{M}$ and $(t,\alpha) \in \mathcal{M}$. In this case, $P_{\mathcal{M}}(t,\alpha,u) = P_{\mathcal{N}}(s,\alpha,u)$ for all states u and $wgt_{\mathcal{N}}(s,\alpha) = w(s,t) + wgt_{\mathcal{M}}(t,\alpha)$ and (t,α) does not belong to any 0-EC of $\mathcal{M}$.

In particular, $\mathcal{N}$ does not contain any action of $\mathcal{M}$ that belongs to a 0-EC of $\mathcal{M}$. ■

Lemma B.30. *Let $\mathcal{Z}$ be a maximal 0-EC of $\mathcal{M}$ and $i \in \{0,1,\ldots,\ell\}$. Let T_i denote the set states t that belong to $\mathcal{Z}$ such that either $t = s_{0,j}$ for the largest index $j \in \{0,1,\ldots,i-1\}$ where all states of $\mathcal{E}_j$ are contained in $\mathcal{Z}$ or $\mathcal{M}_i$ contains a state-action pair (t,α) where α is an action of $\mathcal{M}$. Then, for each state s in $\mathcal{Z}$ and each state $t \in T$ there is a path $\pi = t_0\,\alpha_0\,t_1\,\alpha_1 \ldots \alpha_{m-1}\,t_m$ from $t_0 = s$ to $t = t_m$ in $\mathcal{M}_i$ such that for each $j \in \{0,1,\ldots,m-1\}$ either $\alpha_j = \tau$ or α_j is an action of $\mathcal{Z}$. Moreover, the weight of each such path is $w(s,t)$.*

Proof. We first observe that if s is a state of $\mathcal{Z}$ and α an action of $\mathcal{M}$ such that $(s,\alpha) \in \mathcal{M}_i \setminus \mathcal{Z}$ then either s is the reference state $s_{0,j}$ of $\mathcal{E}_j$ for some $j < i$ such that s is not contained in $\mathcal{E}_{j+1} \cup \ldots \mathcal{E}_{i-1}$ or there is some action of $\mathcal{Z}$ such that $(s,\beta) \in \mathcal{M}_i$. This is a consequence of (S4) in Lemma 3.7 (see also Lemma B.16). Hence, it suffices to consider for the case where $\mathcal{M}$ is a 0-EC, in which case $\mathcal{Z} = \mathcal{M}$. The claim then follows by induction on i. The basis of induction $i = 0$ is trivial and the step of induction follows from statement (S2) of Lemma B.14. □

Corollary B.31. *If $\mathcal{M}$ is a 0-EC then the final MDP $\mathcal{N} = \mathcal{M}_\ell$ generated by the weight-divergence algorithm can be viewed as an acyclic graph built by τ-transitions with a single trap state that is reachable from all other states.*

Lemma B.32 (Properties of the final MDP)**.** *If $\mathcal{M} = \mathcal{M}_0$ is not weight-divergent then the final MDP $\mathcal{N} = \mathcal{M}_\ell$ generated through the iterative application of the spider construction enjoys the following properties:*

(a) Let π be a path in $\mathcal{N}$ from s to t built by τ-transitions. Then, $wgt_{\mathcal{N}}(\pi) = w(s,t)$.

(b) For each maximal 0-EC $\mathcal{Z}$ of $\mathcal{M}$ there is a state $t_{\mathcal{Z}}$ such that:
- *For each state s in $\mathcal{Z}$ there is a path π_s from s to $t_{\mathcal{Z}}$ in $\mathcal{N}$ built by τ-transitions. Moreover, π_s is a prefix of each maximal path ς in $\mathcal{N}$ with $s = first(\varsigma)$.*
- *If (s,α) is a state-action pair in $\mathcal{N}$ where s belongs to $\mathcal{Z}$ and α is an action of $\mathcal{M}$ then $s = t_{\mathcal{Z}}$.*
- *Whenever s is a state of $\mathcal{Z}$ and (s,α) a state-action pair of $\mathcal{M}$ that does not belong any 0-EC then $(t_{\mathcal{Z}},\alpha)$ is a state-action pair of $\mathcal{N}$.*

(c) If $\mathcal{Z}$ is a maximal 0-EC in $\mathcal{M}$ and $\mathcal{N}$ does not contain a state-action pair (t,α) where t belongs to $\mathcal{Z}$ and α is an action of $\mathcal{M}$ then $\mathcal{Z} = \mathcal{M}$.

Proof. Statement (a) is clear from Lemma B.26. Statement (b) follows from Lemma B.30 and the observation that all actions in $\mathcal{N}$ are either actions of $\mathcal{M}$ or τ (see Remark B.29). This yields that whenever s and t belong to the same maximal 0-EC $\mathcal{Z}$ of $\mathcal{M}$ and $\mathcal{N}$ contains a state-action pair (t,α) where α is an action of $\mathcal{M}$ then t is reachable from s in $\mathcal{N}$ via τ-transitions. As $\mathcal{N}$ has no 0-EC, each maximal 0-EC of $\mathcal{M}$ can contain only one such a state t. Statement (c) follows from Lemma B.28. □

B.5.3 Soundness of the Weight-Divergence Algorithm

We are now ready to prove the soundness of the weight-divergence algorithm for a given strongly connected MDP $\mathcal{M}$ presented in Section 3.2. We rephrase now Theorem 3.9 to make the connection between $\mathcal{M}$ and $\mathcal{N}$ obtained from the algorithm explicit to check weight-divergence.

Theorem B.33. *The algorithm for checking weight-divergence of a strongly connected MDP $\mathcal{M}$ runs in time polynomial in the size of $\mathcal{M}$. If $\mathcal{M}$ is weight-divergent then it either finds a pumping or gambling MD-scheduler. If $\mathcal{M}$ is not weight-divergent, then it generates a new MDP $\mathcal{N}$ such that*

(W1) $\mathcal{N}$ and $\mathcal{M}$ have the same state space and $\|\mathcal{N}\| \leqslant \|\mathcal{M}\|$.
(W2) $\mathcal{N}$ has at most one maximal end component, and if so, $\mathbb{E}^{\max}_{\mathcal{F}}(\mathrm{MP}) < 0$ for the unique maximal end component $\mathcal{F}$ of $\mathcal{N}$ and $\mathcal{F}$ is reachable from all states in $\mathcal{N}$.
(W3) $\mathcal{N}$ and $\mathcal{M}$ are equivalent with respect to the class of 0-EC-invariant properties in the sense that the statement of Lemma B.25 holds.

Proof. The weight-divergence algorithm generates a sequence $\mathcal{M}_0 = \mathcal{M}, \mathcal{M}_1, \ldots, \mathcal{M}_\ell = \mathcal{N}$ of MDPs as stated at the beginning of Section B.5.2. Hence, each of the $\mathcal{M}_i$'s for $i < \ell$ has a unique maximal end component $\mathcal{F}_i$, $\mathbb{E}^{\max}_{\mathcal{F}_i}(\mathrm{MP}) = 0$ and $\mathcal{M}_{i+1} = \mathsf{Spider}_{\mathcal{E}_i}(\mathcal{M}_i)$ where $\mathcal{E}_i$ is a 0-BSCC of $\mathcal{F}_i$. Then, $\mathcal{M}_0, \ldots, \mathcal{M}_\ell$ have the same state space and $\|\mathcal{M}_{i+1}\| = \|\mathcal{M}_i\|-1$. This yields (W1) and $\ell \leqslant \|\mathcal{M}\|$. By Lemma B.23, for each $i \in \{0, 1, \ldots, \ell\}$ we have that $\mathcal{M}$ is weight-divergent iff $\mathcal{M}$ is weight-divergent. Lemma B.25 yields the equivalence of $\mathcal{M}$ and $\mathcal{M}_i$ as stated in (W3).

In case that $\mathcal{N} = \mathcal{M}_\ell$ has end components, the reachability of the unique maximal end component $\mathcal{F}$ of $\mathcal{N}$ follows from the observation that the reference state $s_{0,i}$ is accessible from all states in $\mathcal{M}_{i+1} = \mathsf{Spider}_{\mathcal{E}_i, s_{0,i}}(\mathcal{M}_i)$ (by induction on i).

For $i = 0, 1, \ldots, \ell$, the weight-divergence algorithm computes $\mathbb{E}^{\max}_{\mathcal{F}_i}(\mathrm{MP})$ and an MD-scheduler $\mathfrak{T}$ maximizing the expected mean payoff. The case $\mathbb{E}^{\max}_{\mathcal{F}_i}(\mathrm{MP}) \neq 0$ is only possible if $i = \ell$ as we then have:

- If $\mathbb{E}^{\max}_{\mathcal{F}_i}(\mathrm{MP}) < 0$ then all schedulers for $\mathcal{M}_i$ are negatively pumping (Lemma B.9 with all weights multiplied by -1), and hence, $\mathcal{M}_i$ and $\mathcal{M}$ are not positively weight-divergent. In this case, the final MDP $\mathcal{M}_i$ enjoys the properties (W1), (W2) and (W3).
- If $\mathbb{E}^{\max}_{\mathcal{F}_i}(\mathrm{MP}) > 0$ then $\mathcal{M}$ is pumping (Lemma 3.3) and therefore positively weight-divergent. In this case, $\mathfrak{T}$ is a pumping MD-scheduler for $\mathcal{M}_i$. We now can rely on the scheduler transformation presented in part (a) of Lemma B.25 to obtain a pumping MD-scheduler $\mathfrak{S}$ for $\mathcal{M}$.

Suppose now $\mathbb{E}^{\max}_{\mathcal{F}_i}(\mathrm{MP}) = 0$. If $\mathfrak{T}$ has a gambling BSCC then $\mathcal{M}$ is gambling and therefore positively weight-divergent. In this case, we can again rely on the scheduler transformation presented in part (a) of Lemma B.25 to obtain a gambling MD-scheduler $\mathfrak{S}$ for $\mathcal{M}$.

Otherwise, each BSCC of the Markov chain induced by $\mathfrak{T}$ is a 0-BSCC (Lemma 3.2). In this case, $i < \ell$ and $\mathcal{E}_i$ is one of the 0-BSCCs of $\mathfrak{T}$. The weight-divergence algorithm generates the MDP $\mathcal{M}_{i+1} = \mathsf{Spider}_{\mathcal{E}_i}(\mathcal{M}_i)$. If $\mathcal{F}_i = \mathcal{E}_i$ then $\ell = i+1$ and $\mathcal{M}_i$ (and therefore $\mathcal{M}$) are not weight-divergent and the final MDP $\mathcal{M}_{i+1}$ has no end components. Thus, the conditions (W1), (W2) and (W3) are fulfilled. Suppose now that $\mathcal{F}_i \neq \mathcal{E}_i$. In this case, the procedure will be repeated with the MDP $\mathcal{M}_{i+1}$.

Lemma B.27 yields that the size of the MDPs $\mathcal{M}_i$ is polynomially bounded in the size of the original MDP $\mathcal{M}$. Thus, the cost per iteration (the computation of the maximal end component $\mathcal{F}_i$ of $\mathcal{M}_i$ and its maximal expected mean payoff as well as the new MDP $\mathcal{M}_{i+1} = \mathsf{Spider}_{\mathcal{E}_i}(\mathcal{M}_i)$) are polynomially bounded in the size of $\mathcal{M}$. This yields a polynomial-time bound for the weight-divergence algorithm. □

Statement (W2) in Theorem B.33 implies that if the final MDP $\mathcal{N}$ has end components then $\mathcal{N}$ is *universally negatively pumping* in the sense that $\limsup_{n\to\infty} wgt(pref(\varsigma, n)) = -\infty$ holds for almost all paths ς under each scheduler for $\mathcal{N}$.

B.6 Universal Negative Weight-Divergence and Boundedness

This section provides the proof for Theorem 3.14 stating that for strongly connected MDPs with maximal mean payoff 0, the absence of 0-ECs is equivalent to the universal negative weight-divergence property:

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.*

The proof for Theorem 3.14 is split in the proof of each statement (a) and (b). The essential argument for the proof is that if $\mathcal{M}$ has schedulers where the set of paths that are bounded from below then $\mathcal{M}$ has positive measure then $\mathcal{M}$ must have a 0-EC.

Let us first recall the definition of boundedness from below and introduce related notions. Let $L, U \in \mathbb{Z} \cup \{\pm\infty\}$ with $L \leqslant U$. An infinite path ς is said to be (L,U)-bounded iff

$$\overset{\infty}{\forall}\, n \in \mathbb{N}.\ L \ \leqslant\ wgt(pref(\varsigma, n)) \ \leqslant\ U$$

where $\overset{\infty}{\forall}$ means "for all, but finitely many". Thus:

$$\varsigma \text{ is } (L,+\infty)\text{-bounded} \quad \text{iff} \quad \liminf_{n\to\infty} wgt(pref(\varsigma,n)) \geqslant L \ .$$

An infinite path ς is *bounded from below* if there is some integer L such that ς is $(L,+\infty)$-bounded. Clearly, this is equivalent to the requirement that $\liminf_{n\to\infty} wgt(pref(\varsigma,n)) \in \mathbb{Z}$.

Scheduler $\mathfrak{S}$ for a strongly connected MDP $\mathcal{M}$ is said to be *almost-surely (L,U)-bounded* from state s, or briefly *(L,U)-bounded* from s, if for almost all infinite $\mathfrak{S}$-path ς starting in s are (L,U)-bounded. Scheduler $\mathfrak{S}$ is said to be *probably (L,U)-bounded* from state s if

$$\Pr^{\mathfrak{S}}_{\mathcal{M},s}\big\{\ \varsigma \in \mathit{IPaths}\ :\ \overset{\infty}{\forall}\, n \in \mathbb{N}.\ L \leqslant wgt(pref(\varsigma,n)) \leqslant U\ \big\}\ >\ 0$$

We say $\mathfrak{S}$ is *probably bounded* (resp. *almost-surely bounded*) from s if there exist $L, U \in \mathbb{Z}$ with $L \leqslant U$ such that $\mathfrak{S}$ is probably (resp. almost-surely) (L,U)-bounded from s. Likewise, $\mathfrak{S}$ is called *probably bounded from below* (resp. *almost-surely bounded from below* or briefly *bounded from below*) from state s if there exists an integer L such that $\mathfrak{S}$ is probably (resp. almost-surely) $(L,+\infty)$-bounded from s_0.

Lemma B.34. *If $\mathfrak{S}$ is a scheduler that is probably bounded from below from some state s then there is some integer L such that $\mathfrak{S}$ is probably $(L,+\infty)$-bounded from s.*

Proof. Let Π denote the set of all $\mathfrak{S}$-paths ς from s that are bounded from below. Likewise, for $L \in \mathbb{Z}$, let Π_L denote the set of all $\mathfrak{S}$-paths ς from s that are $(L,+\infty)$-bounded. Then, $\Pi = \bigcup_{L\in\mathbb{Z}} \Pi_L$. Hence, $\Pr^{\mathfrak{S}}_{\mathcal{M},s}(\Pi) > 0$ iff there exists $L \in \mathbb{Z}$ with $\Pr^{\mathfrak{S}}_{\mathcal{M},s}(\Pi_L) > 0$. □

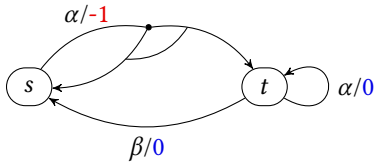


Figure 9. Lemma B.34 does not hold for almost-sure boundedness

Note that the analogous statement of Lemma B.34 does not hold for almost-sure boundedness. An example is the strongly connected MDP $\mathcal{M}$ depicted in Figure 9 consisting of the state-action pairs (s,α), (t,α) and (t,β). Then, the MD-scheduler $\mathfrak{S}$ that choses α for states s and t is almost-surely bounded from below, but there is no integer L such that $\mathcal{M}$ has an almost-surely $(L,+\infty)$-bounded scheduler.

Obviously, if $\mathcal{M}$ is strongly connected then the existence of a probably bounded scheduler does not depend on the starting state s. To see this, we suppose that $\mathcal{M}$ has a probably (L,U)-bounded scheduler from s. For each state t in $\mathcal{M}$ we pick a finite path π_t from t to s. Then, for each state t, $\mathcal{M}$ has a probably (L_t,U_t)-bounded scheduler from t where $L_t = L + wgt(\pi_t)$ and $U_t = U + wgt(\pi_t)$. The analogous statement does not hold for almost-surely bounded schedulers. For example, considering again the MDP $\mathcal{M}$ illustrated by Figure 9. Then $\mathcal{M}$ has an almost-surely $(0,0)$-bounded scheduler from t, but there is no scheduler that is almost-surely bounded from below from state s.

Lemma B.35 (From probably to almost-surely boundedness)**.** *Let $\mathcal{M}$ be a strongly connected MDP with $\mathbb{E}^{\max}_{\mathcal{M}}(\mathrm{MP}) = 0$ and let s_0 be a state in $\mathcal{M}$, $L \in \mathbb{Z}$ with $U \in \mathbb{Z} \cup \{+\infty\}$, and $\mathfrak{S}$ a scheduler that is probably (L,U)-bounded from s_0. Then, there exists an MD-scheduler $\mathfrak{T}$ such that the Markov chain $\mathcal{C}_{\mathfrak{T}}$ induced by $\mathfrak{T}$ contains a 0-BSCC $\mathcal{B}$. In particular, $\mathfrak{T}$ is bounded from below.*

Proof. Recall that for an infinite paths $\varsigma = s_0\,\alpha_0\,s_1\,\alpha_1\ldots$, the limit of ς, denoted $\lim(\varsigma)$, is the set of all state-action pairs (s,α) such that $(s,\alpha) = (s_n,\alpha_n)$ for infinitely many indices $n \in \mathbb{N}$. Given an end component $\mathcal{E}$ of $\mathcal{M}$, we define:

$$\Pi_{\mathcal{E}} \ =\ \big\{\, \varsigma \in \mathit{IPaths}\ :\ \lim(\varsigma) = \mathcal{E}\ \text{ and }\ \overset{\infty}{\forall}\, n \in \mathbb{N}.\ L \leqslant wgt(pref(\varsigma,n)) \leqslant U\,\big\}\ .$$

As $\mathfrak{S}$ is probably (L,U)-bounded, there exists an end component $\mathcal{E}$ of $\mathcal{M}$ such that $\Pr^{\mathfrak{S}}_{\mathcal{M},s_0}(\Pi_{\mathcal{E}})$ is positive.

For each state s in $\mathcal{E}$ and each integer $k \in \{\ell \in \mathbb{Z} : L \leqslant \ell \leqslant U\}$, let $\Pi_{s,k}$ denote the following set:

$$\Pi_{s,k} \;=\; \big\{\, \varsigma \in \Pi_{\mathcal{E}} \;:\overset{\infty}{\exists}\, n \in \mathbb{N}.\; (\, \varsigma[n] = s \;\wedge\; wgt(pref(\varsigma, n)) = k \,) \,\big\}$$

where $\varsigma[n] = s_n$ denotes the $(n{+}1)$-st state of ς.

For each state s in $\mathcal{E}$, let K_s denote the set of integers $k \in \{\ell \in \mathbb{Z} : L \leqslant \ell \leqslant U\}$ such that $\Pr^{\mathfrak{S}}_{\mathcal{M},s_0}(\Pi_{s,k}) > 0$. Then, K_s is nonempty, as $\Pi_{\mathcal{E}}$ agrees with the union of the sets $\Pi_{s,k}$ when k ranges over all integers ℓ with $L \leqslant \ell \leqslant U$. Let $k_s = \min K_s$ and $\Pi_s = \Pi_{s,k_s}$. Thus, $\Pr^{\mathfrak{S}}_{\mathcal{M},s_0}(\Pi_s) > 0$.

For each state-action pair (s,α) in $\mathcal{E}$, let $\Pi_{s,\alpha}$ denote the set of all paths $\varsigma \in \Pi_s$ such that the following condition holds:

$$\overset{\infty}{\exists}\, n \in \mathbb{N}.\; (\, \varsigma[n] = s \;\wedge\; wgt(pref(\varsigma, n)) = k_s \;\wedge\; \mathfrak{S}(pref(\varsigma, n)) = \alpha \,) \;\;.$$

The above condition assumes that $\mathfrak{S}$ is deterministic. If $\mathfrak{S}$ is randomized then we replace the condition "$\mathfrak{S}(pref(\varsigma,n)) = \alpha$" with "$\mathfrak{S}(pref(\varsigma,n))(\alpha) \geqslant 1/|Act(s)|$". As there are only finitely many actions α with $(s,\alpha) \in \mathcal{E}$ and Π_s is the union of the sets $\Pi_{s,\alpha}$ there is some action $\alpha_s \in Act_{\mathcal{E}}(s)$ with $\Pr^{\mathfrak{S}}_{\mathcal{M},s_0}(\Pi_{s,\alpha_s}) > 0$. But then

$$L \;\leqslant\; k_s + wgt(s,\alpha_s) \;\leqslant\; U$$

and for each state t with $P(s,\alpha_s,t) > 0$ we have that t belongs to $\mathcal{E}$ and

$$k_t \;\leqslant\; k_s + wgt(s,\alpha_s) \;\;.$$

Let $R = L + \max_{s\in\mathcal{E}} k_s$. We now consider the MD-scheduler $\mathfrak{T}$ that schedules α_s for each state s in $\mathcal{E}$ and satisfies $\Pr^{\mathfrak{T}}_{\mathcal{M},u}(\Diamond\mathcal{E}) = 1$ for all states u in $\mathcal{M}$ that do not belong to $\mathcal{E}$. (Such a scheduler exists as $\mathcal{M}$ is strongly connected.)

By induction on the length $|\pi|$ of finite $\mathfrak{T}$-paths starting in some state of $\mathcal{E}$, we obtain $wgt(\pi) \geqslant L - R + k_{last(\pi)}$ if π is a finite $\mathfrak{T}$-path with $first(\pi) \in \mathcal{E}$.

- Basis of induction: If $|\pi| = 0$, say $\pi = s \in \mathcal{E}$, then $wgt(\pi) = 0 \geqslant L - R + k_s$ as $R \geqslant L + k_s$ by the choice of R.
- Step of induction: If π is a path of length $n{+}1$ and its last transition is $s \xrightarrow{\alpha_s} t$ then we apply the induction hypothesis to its prefix of length n and obtain:

$$\begin{aligned} wgt(\pi) \;&=\; wgt(pref(\pi,n)) \;+\; wgt(s,\alpha_s) \\ &\geqslant\; (L - R + k_s) \;+\; wgt(s,\alpha_s) \\ &\geqslant\; L - R \;+\; \underbrace{(k_s + wgt(s,\alpha_s))}_{\geqslant k_t} \;\geqslant\; L - R + k_t \end{aligned}$$

In particular, $wgt(\pi) \geqslant L - R + \min_{s\in\mathcal{E}} k_s \overset{\text{def}}{=} L^*$ for all finite $\mathfrak{T}$-paths starting in some state of $\mathcal{E}$.

Let now $\mathcal{B}$ be a BSCC of $\mathfrak{T}$. Then, $\mathcal{B}$ is a sub-component of $\mathcal{E}$. As the weight of all finite paths in $\mathcal{B}$ is bounded by L^* from below, $\mathcal{B}$ does not have negative cycles. Hence, $\mathbb{E}_{\mathcal{B}}(\mathrm{MP}) \geqslant 0$. On the other hand, $\mathbb{E}_{\mathcal{B}}(\mathrm{MP}) \leqslant \mathbb{E}^{\max}_{\mathcal{M}}(\mathrm{MP}) = 0$. This yields that $\mathbb{E}_{\mathcal{B}}(\mathrm{MP}) = 0$ and that $\mathcal{B}$ is a 0-BSCC (Lemma 3.2). □

The following lemma restates part (a) of Theorem 3.14.

Lemma B.36. *Let $\mathcal{M}$ be a strongly connected MDP with $\mathbb{E}^{\max}_{\mathcal{M}}(\mathrm{MP}) = 0$. Then, the following statements are equivalent*

(a) $\mathcal{M}$ has a 0-EC.
(b) There exists an MD-scheduler $\mathfrak{S}$ such that the Markov chain induced by $\mathfrak{S}$ contains a 0-BSCC.
(c) $\mathcal{M}$ has a scheduler that is almost-surely bounded from below from each state.
(d) There exist integers L, U such that $\mathcal{M}$ has a probably (L,U)-bounded scheduler from some state.
(e) There exists an integer L such that $\mathcal{M}$ has a probably $(L,+\infty)$-bounded scheduler from some state.

Proof. We first show the equivalence of (a) and (b). The implication "(b) $\Longrightarrow$ (a)" is trivial. For the proof of "(a) $\Longrightarrow$ (b)", we suppose we are given a 0-EC $\mathcal{E}$ of $\mathcal{M}$. We pick an MD-scheduler $\mathfrak{S}$ for $\mathcal{M}$ such that the state-action pairs $(s,\mathfrak{S}(s))$ belong to $\mathcal{E}$ whenever s is a state of $\mathcal{E}$ and $\Pr^{\mathfrak{S}}_{\mathcal{M},s}(\Diamond\mathcal{E}) = 1$ for all states s in $\mathcal{M}$ with $s \notin \mathcal{E}$. Then, each BSCC of the Markov chain induced for $\mathfrak{S}$ is a 0-BSCC. The equivalence of statements (c), (d) and (e) is a consequence of Lemma B.35, which shows "(e) $\Longrightarrow$ (c)", while "(c) $\Longrightarrow$ (d)" and "(d) $\Longrightarrow$ (e)" are trivial. We finally check the equivalence of statements (a)/(b) and (c)/(d). The implication "(b) $\Longrightarrow$ (d)" is obvious, while "(c) $\Longrightarrow$ (b)" has been shown in the proof of Lemma B.35. □

The next lemma is part (b) in Theorem 3.14.

Lemma B.37. *Let $\mathcal{M}$ be a strongly connected MDP with $\mathbb{E}^{\max}_{\mathcal{M}}(\mathrm{MP}) = 0$. Then, the following statements are equivalent:*

(a) $\mathcal{M}$ has no 0-EC.
(b) $\mathcal{M}$ has no scheduler that is bounded from below.
(c) Each scheduler for $\mathcal{M}$ is negatively weight-divergent.

Proof. The implications "(c) $\Longrightarrow$ (b)" and "(b) $\Longrightarrow$ (c)" are trivial as schedulers realizing a 0-EC $\mathcal{E}$ are bounded from below and as no negatively weight-divergent scheduler is bounded from below. To prove "(a) $\Longrightarrow$ (c)" we suppose that $\mathcal{M}$ has no 0-EC. We suppose by contraction that $\mathcal{M}$ has a scheduler $\mathfrak{S}$ that is not negatively weight-divergent. That is, there is a state s such that

$$\Pr^{\mathfrak{S}}_{\mathcal{E},s}\big\{\, \varsigma \in \mathit{IPaths} \;:\; \liminf_{n\to\infty} \; \mathit{wgt}(\mathit{pref}(\varsigma,n)) \;>\; -\infty \big\} \;>\; 0 \;.$$

But then, there exists $L \in \mathbb{Z}$ such that $\mathfrak{S}$ is probably $(L,+\infty)$-bounded from s (see Lemma B.34). We derive, by Lemma B.36 that $\mathcal{M}$ has a 0-EC, a contradiction. □

B.7 Checking the Gambling Property

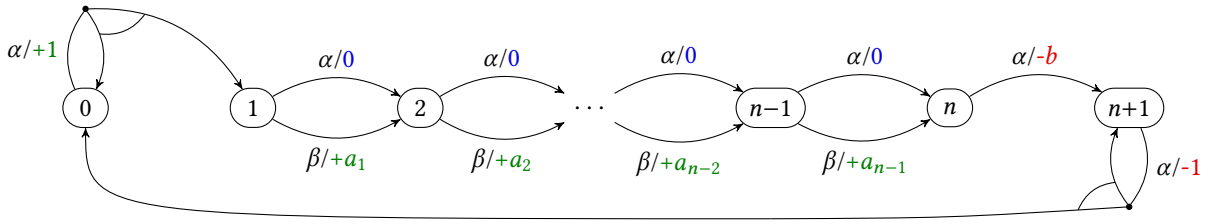


Figure 10. Reduction from subset sum for the NP-hardness of checking the gambling property.

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.*

Proof. We first observe that if $\mathbb{E}^{\max}_{\mathcal{M}}(\mathrm{MP}) = 0$, then an end component is gambling iff it is weight-divergent. Hence, statement (a) follows from Theorem 3.9.

We show the NP-completeness in the general case (statement (b)). A nondeterministic polynomially time-bounded algorithm is obtained by first guessing nondeterministically an MD-scheduler $\mathfrak{S}$ and then checking deterministically whether $\mathbb{E}^{\mathfrak{S}}_{\mathcal{M}}(\mathrm{MP}) = 0$ and the Markov chain induced by $\mathfrak{S}$ has a positive cycle in some its BSCCs. NP-hardness is achieved by a polynomial reduction from the subset sum problem, which takes as input a finite nonempty sequence $a_1, a_2, \ldots, a_{n-1}, b$ of positive integers and asks for a subset I of $\{1,\ldots,n-1\}$ such that $\sum_{i\in I} a_i = b$. Let $\mathcal{M}$ be the MDP illustrated by Figure 10. Clearly, $\mathcal{M}$ is strongly connected. There is a one-to-one correspondence between the MD-schedulers for $\mathcal{M}$ and the subsets I of $\{1,\ldots,n-1\}$. Given $I \subseteq \{1,\ldots,n-1\}$, we define $\mathfrak{S}_I$ as the MD-scheduler that picks β for the states $i \in I$ and action α for all other states. Then, the Markov chain $\mathcal{C}_I$ induced by $\mathfrak{S}_I$ consists of a single BSCC, and $\mathfrak{S}_I$ is gambling iff $\mathbb{E}^{\mathfrak{S}_I}_{\mathcal{M}}(\mathrm{MP}) = 0$ iff $\sum_{i\in I} a_i = b$. Thus, $\mathcal{M}$ has a gambling MD-scheduler iff there exists $I \subseteq \{1,\ldots,n-1\}$ with $\sum_{i\in I} a_i = b$. □

B.8 Computing the set *ZeroEC* and the Recurrence Values

We now provide the proof for Lemma 3.13. So, we suppose that we are given a strongly connected MDP $\mathcal{M}$ with $\mathbb{E}^{\max}_{\mathcal{M}}(\mathrm{MP}) = 0$. Section B.8.1 explains how to compute the maximal 0-ECs. In Section B.8.2, we will explain how to compute the recurrence values of the states belonging to a 0-EC.

B.8.1 Computing the Maximal 0-ECs

We now turn to the computation of the maximal 0-ECs. (Recall the notion of maximal 0-ECs from Section B.4.1.) Let *ZeroEC* denote the set of all states that belong to some 0-EC. Thus, *ZeroEC* is the union of the state spaces of all maximal 0-ECs.

Lemma B.38 (First part of Lemma 3.13)**.** *If $\mathcal{M}$ is strongly connected and $\mathbb{E}^{\max}_{\mathcal{M}}(\mathrm{MP}) = 0$ then the maximal 0-ECs (and the set ZeroEC) are computable in polynomial time.*

Proof. To compute the maximal 0-ECs we present an algorithm that identifies all state-action pairs (s,α) that are contained in some 0-EC. For this, we combine the polynomial-time algorithm to check the existence of 0-BSCCs (Section B.4.3) and the polynomial-time algorithm to flatten 0-BSCCs (spider construction).

Obviously, all state-action pairs (s,α) with $P(s,\alpha,s) = 1$ and $\mathit{wgt}(s,\alpha) = 0$ constitute a 0-EC. In the sequel, we suppose that such "trivial" state-action pairs have been removed from $\mathcal{M}$. We first rename the actions in $\mathcal{M}$ to ensure that $\mathit{Act}_{\mathcal{M}}(s) \neq$

$Act_{\mathcal{M}}(s')$ for all states s, s' with $s \neq s'$. That is, for each action name α in $\mathcal{M}$ there is a unique state-action pair $(s, \alpha) \in \mathcal{M}$. Thus, it suffices to identify all actions that belong to some 0-EC.

The algorithm works as follows. We first run the polynomial-time algorithm to check the existence of a 0-BSCC (see Section B.4.3). If a 0-BSCC $\mathcal{B}$ of $\mathcal{M}$ is found then we apply the spider construction to transform the original $\mathcal{M}$ into an equivalent MDP $\mathcal{N} = \mathrm{Spider}_{\mathcal{B}}(\mathcal{M})$ with the same state space and where $\mathcal{B}$ has been flattened. Recall that $\mathcal{N}$ contains fewer state-action pairs than $\mathcal{M}$ (see (S1) in Lemma 3.7). By property (S4) stated in Lemma 3.7 (see also Lemma B.16) we get that *ZeroEC* equals the union of the state space of $\mathcal{B}$ and the set of states belonging to some 0-EC of $\mathcal{N}$, and the analogous statement for the actions that are contained in some 0-EC of $\mathcal{M}$ resp. $\mathcal{N}$. (Of course, the extra τ-actions of $\mathcal{N}$ from all states s in $\mathcal{B} \setminus \{s_0\}$ to s_0 have to be ignored.) Hence, we can repeat the same procedure to the new MDP $\mathcal{N}$. In this way we encounter all states and actions that belong to a 0-EC. The number of iterations is bounded by the number of state-action pairs that belong to some 0-EC of $\mathcal{M}$. Thus, due to Lemma B.27, the time complexity is polynomial in the size of $\mathcal{M}$. □

B.8.2 Recurrence Values in Maximal 0-ECs

Given a strongly connected MDP $\mathcal{M}$ with $\mathbb{E}^{\max}_{\mathcal{M}}(\mathrm{MP}) = 0$ and a state s that belongs to some maximal 0-EC $\mathcal{Z}$ of $\mathcal{M}$, the long-run weight and the recurrence value of state s are defined by:

$$\begin{aligned} lgr(s) &= \max\left\{ w \in \mathbb{Z} \,:\, \exists \mathfrak{S}.\ \Pr^{\mathfrak{S}}_{\mathcal{Z},s}\big(\Diamond\Box(\mathrm{wgt} \geqslant w)\big) = 1 \right\} \\ rec(s) &= \max\left\{ w \in \mathbb{Z} \,:\, \exists \mathfrak{S}.\ \Pr^{\mathfrak{S}}_{\mathcal{Z},s}\big(\Box(\mathrm{wgt} \geqslant w) \wedge \Box\Diamond s\big) = 1 \right\} \end{aligned}$$

where the existential quantifier $\exists\mathfrak{S}$ ranges over the schedulers for $\mathcal{Z}$ (rather than $\mathcal{M}$).

Whenever s and t are states that belong to the same maximal 0-EC then $lgr(s) = w(s,t) + lgr(t)$. On the other hand, $rec(s) \leqslant 0$ for all states s in a 0-EC and $rec(s) \neq w(s,t) + rec(t)$ is possible if s and t belong to the same maximal 0-EC.

Example B.39. Consider the MDP $\mathcal{M}$ with the following deterministic (*i.e.*, with probability 1) transitions:

$$s \xrightarrow{3} t \xrightarrow{-2} u \xrightarrow{-1} s \qquad \text{and} \qquad v \xrightarrow{4} s \xrightarrow{-4} v$$

For simplicity, we dropped here the action names and attached the weights to the transitions. Then, $\mathcal{M}$ constitutes a maximal 0-EC with $rec(s) = lgr(s) = 0$, $rec(t) = lgr(t) = -3$, $rec(u) = lgr(u) = -1$, while $rec(v) = 0$ and $lgr(v) = 4$. ■

Lemma B.40 (Second part of Lemma 3.13)**.** *If $\mathcal{M}$ is strongly connected and $\mathbb{E}^{\max}_{\mathcal{M}}(\mathrm{MP}) = 0$ then the long-run weights $lgr(s)$ and the recurrence values $rec(s)$ for the states $s \in ZeroEC$ are computable in polynomial time.*

Proof. Let $\mathcal{Z}$ be a maximal 0-EC of $\mathcal{M}$. As before, $w(s,t) = wgt(\pi)$ for some/each path from s to t in $\mathcal{Z}$. Then, $w(s,s) = 0$ and $w(s,t) = w(s,u) + w(u,t)$ for all states $s, t, u \in \mathcal{E}$. Obviously, the values $w(s,t)$ for all states $s, t \in \mathcal{Z}$, can be obtained in polynomial time, *e.g.*, using a BFS or DFS from every state. Let

$$W \;=\; \big\{\, w(s,t) \,:\, s,t \text{ are states in } \mathcal{Z} \,\big\} \;\subseteq\; \mathbb{Z}\ .$$

Note that W contains at most n^2 elements when n is the number of states in $\mathcal{Z}$. Moreover, the absolute value of the elements in W is bounded by $(n-1)\cdot \max_{s,\alpha} |wgt(s,\alpha)|$ where (s,α) ranges over all state-action pairs in $\mathcal{Z}$. Thus, the logarithmic lengths of the elements in W is polynomially bounded in the size of $\mathcal{Z}$. Recall that the size of an MDP is defined as the number of states plus the total sum of the logarithmic lengths of its transition probabilities and weights.

Let s be a state in $\mathcal{Z}$. Let $w_1, w_2, \ldots, w_k$ an enumeration of the elements in $\{w(s,t) : t \in \mathcal{Z}\}$ such that $w_1 > w_2 > \ldots > w_k$. For $j = 1, 2, \ldots, k$, let

$$T_{s,j} \;=\; \big\{\, t \,:\, t \text{ is a state in } \mathcal{Z} \text{ with } w(s,t) \geqslant w_j \,\big\}\ .$$

For each state $t \in T_{s,j}$, let

$$Act_{s,j}(t) \;=\; \big\{\, \beta \in Act_{\mathcal{Z}}(t) \,:\, Post(t,\beta) \subseteq T_{s,j} \,\big\}$$

where $Post(t,\beta)$ denotes the set of states u with $P(t,\beta,u) > 0$. Consider the sub-MDP $\mathcal{Z}_{s,j}$ consisting of all state-action pairs $(t,\beta) \in \mathcal{Z}$ with $t \in T_{s,j}$ and $\beta \in Act_{s,j}(t)$. Let $MEC_{s,j}$ be the set of states in $\mathcal{Z}_{s,j}$ that are contained in some maximal end component of $\mathcal{Z}_{s,j}$.

Some comments are in order. First, as $w(s,s) = 0$ we have $s \notin T_{s,j}$ iff w_j is positive. In particular, $s \in MEC_{s,j}$ is only possible if $w_j \leqslant 0$. Second, for each path π in $\mathcal{Z}$ with $first(\pi) = s$ and $last(\pi) = t \in T_{s,j}$ we have $wgt(\pi) = w(s,t) \geqslant w_j$. Third, the sets $Act_{s,j}(t)$ can be empty, in which case t is a trap in $\mathcal{Z}_{s,j}$.

Let j be the smallest index in $\{1, \ldots, k\}$ such that $\Pr^{\max}_{\mathcal{Z},s}(\Diamond MEC_{s,j}) = 1$. Note that $\mathcal{Z}_{s,k} = \mathcal{Z}$ and $s \in MEC_{s,k}$, and therefore $\Pr^{\max}_{\mathcal{Z},s}(\Diamond MEC_{s,k}) = 1$. This ensures the existence of such an index j. We now show that $lgr(s) = w_j$.

- To prove $lgr(s) \leqslant w_j$, we pick a scheduler $\mathfrak{S}$ for $\mathcal{Z}$ with $\Pr^{\mathfrak{S}}_{\mathcal{Z},s}\big(\Diamond\Box(\mathrm{wgt} \geqslant lgr(s))\big) = 1$. Let m be the largest value such that $w_m \geqslant lgr(s)$. Then, $\Pr^{\mathfrak{S}}_{\mathcal{Z},s}(\Diamond MEC_{s,m}) = 1$. But then $j \leqslant m$, and therefore $w_j \geqslant w_m \geqslant lgr(s)$.
- To see why $lgr(s) \geqslant w_j$, we pick an MD-scheduler $\mathfrak{S}$ for $\mathcal{Z}$ such that $\Pr^{\mathfrak{S}}_{\mathcal{Z},s}(\Diamond MEC_{s,j}) = 1$. But then $wgt(\pi) = w(s,t) \geqslant w_j$ for all $\mathfrak{S}$-paths π from s to a state t that belongs to a BSCC of $\mathfrak{S}$. Hence, $\Pr^{\mathfrak{S}}_{\mathcal{Z},s}(\Diamond\Box(\mathrm{wgt} \geqslant w_j)) = 1$. By the definition of $lgr(s)$, we get $lgr(s) \geqslant w_j$.

With similar arguments, we get $rec(s) = w_i$ where i is the smallest index such that $s \in MEC_{s,i}$. Clearly, these indices i and j can be computed in polynomial time using standard algorithms to compute the maximal end components in MDPs and maximal reachability probabilities. Note that $k \leqslant |W| \leqslant n^2$ where n is the number of states in $\mathcal{Z}$. □

Example B.41. Let us revisit the MDP of Example B.39. The values $w(x,y)$ are as follows:

	s	t	u	v
s	0	3	1	−4
t	−3	0	−2	−7
u	−1	2	0	−5
v	4	7	5	0

Hence, $W = \{7,5,4,3,2,1,0,-1,-2,-3,-4,-5,-7\}$.

For instance, for state s, we consider the values in the row for s: $w_1 = 3$, $w_2 = 1$, $w_3 = 0$ and $w_4 = -4$, and look for the smallest index $j \in \{1,\ldots,4\}$ such that $\Pr^{max}_{\mathcal{Z},s}(\Diamond MEC_{s,j}) = 1$. The MDP $\mathcal{Z}_{s,1}$ consists of state t, which is a trap in $\mathcal{Z}_{s,1}$. Hence, $MEC_{s,1} = \varnothing$. The MDP $\mathcal{Z}_{s,2}$ consists of the transition $t \to u$ and u is a trap in $\mathcal{Z}_{s,2}$. Again, we have $MEC_{s,2} = \varnothing$. The MDP $\mathcal{Z}_{s,3}$ consists of the cycle $s \to t \to u \to s$. Hence, $MEC_{s,2} = \{s,t,u\}$. This yields $lgr(s) = rec(s) = w_3 = 0$.

Let us look now for state v. Here, we deal with the values $w_1 = 7$, $w_2 = 5$, $w_3 = 4$, and $w_4 = 0$. We have $MEC_{v,1} = \varnothing$ as the MDP $\mathcal{Z}_{v,1}$ consists of the trap state t. The MDP $\mathcal{Z}_{v,2}$ consists of states t and u. Both are traps in $\mathcal{Z}_{v,2}$. Hence, $MEC_{v,2} = \varnothing$. The MDP $\mathcal{Z}_{v,3}$ consists of the cycle $s \to t \to u \to v$ plus the trap states v. Hence, $MEC_{v,3} = \{s,t,u\}$ and $lgr(v) = w_3 = 4$. The MDP $\mathcal{Z}_{v,4}$ equals $\mathcal{M}$. Therefore, $v \in MEC_{v,4} = \{s,t,u,v\}$, which yields $rec(v) = w_4 = 0$. ■

The proof of Lemma B.40 also shows the existence of MD-schedulers that achieve the long-run weights and recurrence values.

Corollary B.42. *For each maximal 0-EC $\mathcal{Z}$ there exist MD-schedulers $\mathfrak{S}$ and $\mathfrak{T}$ such that $\Pr^{\mathfrak{S}}_{\mathcal{Z},s}(\Diamond\Box\big(wgt \geqslant lgr(s)\big)) = 1$ and $\Pr^{\mathfrak{T}}_{\mathcal{Z},s}\big(\Box(wgt \geqslant rec(s)) \wedge \Box\Diamond s\big) = 1$ for each state s in $\mathcal{Z}$.*

Remark B.43 (Minimal credits in energy-MDPs)**.** Let $\mathcal{M}$ be an MDP, F a set of states in $\mathcal{M}$ and s a state in $\mathcal{M}$. The value

$$mincredit_{\mathcal{M}}(s,F) \;=\; \min\big\{w \in \mathbb{Z} \;:\; \exists\mathfrak{S}.\ \Pr^{\mathfrak{S}}_{\mathcal{Z},s}\big(\,\Box(\mathrm{wgt}{+}w \geqslant 0) \;\wedge\; \Box\Diamond F\,\big) = 1\big\}$$

is the minimal initial weight budget required in state s to ensure that the accumulated weight is always nonnegative and that the Büchi condition $\Box\Diamond F$ holds almost surely (under some scheduler). Following the literature on energy-MDPs with Büchi objectives, this value is called the minimal credit for state s in $\mathcal{M}$. It relates to the recurrence values as follows. If $\mathcal{Z}$ is a maximal 0-EC of $\mathcal{M}$ and s a state in $\mathcal{Z}$ then

$$rec(s) \;=\; -mincredit_{\mathcal{Z}}(s,\{s\})\ .$$

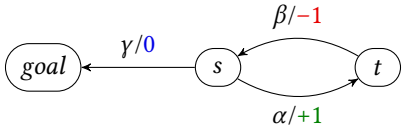
Pseudo-polynomial algorithms for computing the minimal credits for all states in an arbitrary MDPs $\mathcal{M}$ are known from the literature on energy games and energy-MDPs with Büchi objectives [8] and [16, 21]. If $\mathcal{M}$ is a strongly connected MDP with $wgt(\xi) = 0$ for all cycles in $\mathcal{M}$ then $mincredit_{\mathcal{M}}(s,F)$ can be computed in polynomial time by an algorithm similar to the one for the long-run weights and recurrence values. Using the notations introduced in the proof of Lemma B.40, $mincredit_{\mathcal{M}}(s,F) = -w_j$ where j is the largest index such that $\Pr^{\max}_{\mathcal{Z}_{s,j},s}(\Diamond MEC^F_{s,j}) = 1$ where $MEC^F_{s,j}$ denotes the set of all states in $\mathcal{Z}_{s,j}$ that belong to a maximal end component $\mathcal{E}$ of $\mathcal{Z}_{s,j}$ where $\mathcal{E}$ contains at least one state in F.

Thus, our results show that these algorithms can be improved for MDPs that constitute a 0-EC as then the values $mincredit_{\mathcal{M}}(s,F)$ are computable in polynomial time. This is not an interesting instance of energy-MDPs. However, the efficient computability of the recurrence values of states belonging to 0-ECs will be crucial for the weight-bounded repeated reachability problems (see Section D.5). ■

C Proofs of Section 4

We will interpret the assumption (BT) as follows: if $\Pr^{\mathfrak{S}}_{\mathcal{M},s}(\Diamond goal) < 1$ then the set of pumping $\mathfrak{S}$-paths from state s has positive measure under $\mathfrak{S}$. If (BT) holds then $\mathbb{E}^{\inf}_{\mathcal{M},s_{init}}(\Diamond goal)$ is achieved by an MD-scheduler and such an MD-scheduler is computable in polynomial time using linear-programming techniques [4].

Example C.1 (Incompleteness of condition (BT))**.** Consider the following MDP



Then, $\mathbb{E}^{\mathfrak{S}}_{\mathcal{M},s}(\Diamond goal) = 0$, $\mathbb{E}^{\mathfrak{S}}_{\mathcal{M},t}(\Diamond goal) = -1$ for each proper scheduler $\mathfrak{S}$. However, $\mathcal{M}$ violates condition (BT) as the expected total weight of the improper MD-scheduler $\mathfrak{S}$ with $\mathfrak{S}(s) = \alpha$ is not $+\infty$. ∎

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, $\mathbb{E}^{\inf}_{\mathcal{M},s_{init}}(\Diamond goal)$ 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 implication "$\Longrightarrow$" is shown in Lemma C.2, and the last statement in Lemma C.3. The proof of the implication "$\Longleftarrow$" in Lemma 4.1 will be presented afterwards together with an explanation how the iterative application of the spider construction to flatten 0-ECs can be used to generate a new MDP that satisfies condition (BT) and has the same minimal expected accumulated weight.

Intuitively, if there are negatively weight-divergent components, then one can define a family of schedulers that decrease the weight arbitrarily low before moving to the goal state, hence showing that the minimal expectation is $-\infty$. We prove this formally in the following lemma.

Lemma C.2 (Implication "$\Longrightarrow$" of Lemma 4.1)**.** *Let $\mathcal{M}$ be as in Lemma 4.1. If $\mathcal{M}$ has negatively weight-divergent end components then $\mathbb{E}^{\inf}_{\mathcal{M},s_{init}}(\Diamond goal) = -\infty$.*

Proof. Assume thus that $\mathcal{M}$ has a negatively weight-divergent end component, and let us show that $\mathbb{E}^{\inf}_{\mathcal{M},s_{init}}(\Diamond goal) = -\infty$ by exhibiting a sequence $(\mathfrak{S}_R)_{R\in\mathbb{N}}$ of proper schedulers for $\mathcal{M}$ with $\inf_{R\in\mathbb{N}} \mathbb{E}^{\mathfrak{S}_R}_{\mathcal{M},s_{init}}(\Diamond goal) = -\infty$.

Let $\mathcal{E}$ be a negatively weight-divergent end component of $\mathcal{M}$. To define $\mathfrak{S}_R$, we combine three natural schedulers. Let first $\mathfrak{S}$ be an MD-scheduler for $\mathcal{M}$ such that $\mathfrak{S}$ is proper, *i.e.*, for every $s \in S$, $\Pr^{\mathfrak{S}}_{\mathcal{M},s}(\Diamond goal) = 1$. Let then $\mathfrak{T}$ be an MD-scheduler that reaches $\mathcal{E}$ with positive probability and agrees with $\mathfrak{S}$ in states from which $\mathcal{E}$ is no longer reachable: $\Pr^{\mathfrak{T}}_{\mathcal{M},s_{init}}(\Diamond\mathcal{E}) > 0$, and $\mathfrak{T}(t) = \mathfrak{S}(t)$ for each state $t \in T$ where $T = \{t \in S : \Pr^{\mathfrak{T}}_{\mathcal{M},t}(\Diamond\mathcal{E}) = 0\}$. Therefore, for every $t \in T$, $\Pr^{\mathfrak{S}}_{\mathcal{M},t}(\Diamond goal) = 1$. Last, let $\mathfrak{V}$ be a negatively weight-divergent scheduler for $\mathcal{E}$.

From $\mathfrak{S}$, $\mathfrak{T}$, $\mathfrak{V}$ and $R \in \mathbb{N}$, we define $\mathfrak{S}_R$ as follows. Initially, $\mathfrak{S}_R$ mimics $\mathfrak{T}$ until reaching a state, say u, belonging to $\mathcal{E}$ or T.

- If $u \in \mathcal{E}$, and the accumulated weight is r, then $\mathfrak{S}_R$ switches mode and simulates $\mathfrak{V}$ until the accumulated weight is at most $-R$ (this happens almost surely because $\mathcal{E}$ is negatively weight-divergent); then $\mathfrak{S}_R$ again switches mode and behaves as $\mathfrak{S}$ until reaching *goal*.
- If $u \in T$, then $\mathfrak{S}_R$ switches mode and behaves as $\mathfrak{S}$ until reaching *goal*.

We claim that $\inf_{R\in\mathbb{N}} \mathbb{E}^{\mathfrak{S}_R}_{\mathcal{M},s_{init}}(\Diamond goal) = -\infty$. To prove it, we provide an upper bound to $\mathbb{E}^{\mathfrak{S}_R}_{\mathcal{M},s_{init}}(\Diamond goal)$ for fixed $R \in \mathbb{N}$. For each state $u \in S$, we define p_u as the probability under $\mathfrak{T}$ to reach u before traversing $\mathcal{E}$. Let $E = \max_{s\in\mathcal{E}} \mathbb{E}^{\mathfrak{S}}_{\mathcal{M},s}(\Diamond goal)$ be the maximum of the expected accumulated weights until reaching *goal*, taken over all paths starting in a state of $\mathcal{E}$. Then, the expected accumulated weight until reaching *goal* under $\mathfrak{S}_R$ is

$$\begin{aligned}\mathbb{E}^{\mathfrak{S}_R}_{\mathcal{M},s_{init}}(\Diamond goal) &\leqslant \sum_{s\in\mathcal{E}} p_s \cdot (-R+E) \;+\; \sum_{t\in T} p_t \cdot \mathbb{E}^{\mathfrak{S}}_{\mathcal{M},t}(\Diamond goal)\\ &= \Pr^{\mathfrak{T}}_{\mathcal{M},s_{init}}(\Diamond\mathcal{E}) \cdot (-R+E) \;+\; \sum_{t\in T} p_u \cdot \mathbb{E}^{\mathfrak{S}}_{\mathcal{M},t}(\Diamond goal)\ .\end{aligned}$$

From $\Pr^{\mathfrak{T}}_{\mathcal{M},s_{init}}(\Diamond\mathcal{E}) > 0$, and the fact that E and the sum over T are constants, we derive the desired limit:

$$\inf_{R\in\mathbb{N}} \mathbb{E}^{\mathfrak{S}_R}_{\mathcal{M},s_{init}}(\Diamond goal) \;=\; -\infty\ .$$

This completes the proof of Lemma C.2. □

Lemma C.3 (Last statement of Lemma 4.1)**.** *If $\mathcal{M}$ has no negatively weight-divergent end component then condition* (BT) *holds if and only if $\mathcal{M}$ has no 0-ECs.*

Proof. "$\Longrightarrow$" is trivial as the expected total weight of each scheduler that realizes a 0-EC is bounded. To prove "$\Longleftarrow$" we suppose that $\mathcal{M}$ has no negatively weight-divergent end component and no 0-EC. But then $\mathbb{E}^{\inf}_{\mathcal{E}}(\mathrm{MP}) > 0$ for each end component $\mathcal{E}$ of $\mathcal{M}$. By Lemma B.9, all end components of $\mathcal{M}$ are universally pumping. Hence, condition (BT) holds. □

To complete the proof of Lemma 4.1 we need to show the finiteness of $\mathbb{E}^{\inf}_{\mathcal{M},s_{init}}(\oplus\, goal)$ if $\mathcal{M}$ has no negatively weight-divergent end components.

For a better fit of the notations used in the previous section, we switch here to the maximal expected accumulated weight until reaching the goal state:

$$\mathbb{E}^{\max}_{\mathcal{M},s}(\oplus\, goal) \;=\; \sup\big\{\mathbb{E}^{\mathfrak{S}}_{\mathcal{M},s}(\oplus\, goal) \;:\; \mathfrak{S} \text{ is a proper scheduler}\big\}.$$

The aim is to show that $\mathbb{E}^{\max}_{\mathcal{M},s_{init}}(\oplus\, goal)$ is finite if $\mathcal{M}$ has no positively weight-divergent end components. The corresponding result for the minimal expected accumulated weight until reaching the goal state is then obtained by multiplying all weights by -1.

Rephrased for maximal expected weights, the assumption of Bertsekas and Tsitsiklis [4] asserts that the expected total weight of each improper scheduler is $-\infty$. More precisely:

$$\text{If } \Pr^{\mathfrak{S}}_{\mathcal{M},s}(\Diamond goal) < 1 \text{ then } \Pr^{\mathfrak{S}}_{\mathcal{M},s}\{\varsigma \in \mathit{IPaths} : \varsigma \text{ is negatively pumping}\} > 0. \tag{BTmax}$$

Lemma C.4 ([4])**.** *Under assumption* (BTmax), *$\mathbb{E}^{\max}_{\mathcal{M},s}(\oplus\, goal)$ is finite, and there is an MD-scheduler $\mathfrak{S}$ with $\mathbb{E}^{\max}_{\mathcal{M},s}(\oplus\, goal) = \mathbb{E}^{\mathfrak{S}}_{\mathcal{M},s}(\oplus\, goal)$ for all states s.*

Rephrased for maximal expectations, Lemma C.2 states that $\mathbb{E}^{\inf}_{\mathcal{M},s_{init}}(\oplus\, goal) = +\infty$ if $\mathcal{M}$ has positively weight-divergent end components. Likewise, Lemma C.3 yields that if $\mathcal{M}$ has no positively weight-divergent end component then (BTmax) is equivalent to the nonexistence of 0-ECs.

Lemma C.5. *If $\mathcal{M}$ is an MDP with $\mathbb{E}^{\max}_{\mathcal{E}}(\mathrm{MP}) < 0$ for each end component $\mathcal{E}$ of $\mathcal{M}$ then $\mathcal{M}$ satisfies condition* (BTmax).

Proof. If $\Pr^{\mathfrak{S}}_{\mathcal{M},s}(\Diamond goal) < 1$ then there exists an end component $\mathcal{E}$ of $\mathcal{M}$ such that $\Pr^{\mathfrak{S}}_{\mathcal{M},s}\{\varsigma \in \mathit{IPaths} : \lim(\varsigma) = \mathcal{E}\} > 0$. (Recall that $\lim(\varsigma)$ denotes the set of state-action pairs that are taken infinitely often in ς.) As $\mathbb{E}^{\max}_{\mathcal{E}}(\mathrm{MP}) < 0$, all schedulers for $\mathcal{E}$ are negatively pumping. Hence, the set of negatively pumping $\mathfrak{S}$-paths starting in s has positive measure. □

Construction of the new MDP $\mathcal{N}$. Suppose now that $\mathcal{M}$ is an MDP that has no positively weight-divergent end component. We now generate from $\mathcal{M}$ a new MDP $\mathcal{N}$ with the same state space that satisfies (BTmax) and is equivalent to $\mathcal{M}$ with respect to the maximal expected accumulated weight until reaching the goal state. For this, we apply the iterative spider construction to $\mathcal{M}$ of Section B.5.2. The resulting MDP $\mathcal{N}$ has the following properties:

Lemma C.6. *Let $\mathcal{M}$ be an MDP that has no weight-divergent end component, and let $\mathcal{N}$ be the MDP resulting from $\mathcal{M}$ by flattening the 0-ECs using the iterative spider construction of Section B.5.2. Then:*

- *(E1) $\|\mathcal{N}\| \leqslant \|\mathcal{M}\|$ and the size of $\mathcal{N}$ is polynomially bounded by the size of $\mathcal{M}$.*
- *(E2) $\mathcal{N}$ satisfies condition* (BTmax).
- *(E3) For proper scheduler $\mathfrak{T}$ for $\mathcal{N}$ there is a proper scheduler $\mathfrak{S}$ for $\mathcal{M}$ with $\mathbb{E}^{\mathfrak{S}}_{\mathcal{M},s}(\oplus\, goal) = \mathbb{E}^{\mathfrak{T}}_{\mathcal{N},s}(\oplus\, goal)$ for all states s. If $\mathfrak{T}$ is an MD-scheduler, then $\mathfrak{S}$ can be chosen as an MD-scheduler.*
- *(E4) For each proper scheduler $\mathfrak{S}$ for $\mathcal{M}$ there is a proper scheduler $\mathfrak{T}$ for $\mathcal{N}$ with $\mathbb{E}^{\mathfrak{S}}_{\mathcal{M},s}(\oplus\, goal) = \mathbb{E}^{\mathfrak{T}}_{\mathcal{N},s}(\oplus\, goal)$ for all states s.*

Proof. Statement (E1) follows from property (S1) of Lemma 3.7 and Lemma B.27. Statement (E2) is a consequence of (W2) in Theorem B.33 which yields that $\mathbb{E}^{\max}_{\mathcal{E}}(\mathrm{MP}) < 0$ for all end components $\mathcal{E}$ of $\mathcal{N}$. By Lemma C.5, we obtain that $\mathcal{N}$ satisfies (BTmax).

We now turn to the proof of the scheduler transformations as stated in (E3) and (E4). For this, we can rely on the equivalence of $\mathcal{M}$ and $\mathcal{N}$ with respect to the class of all 0-EC-invariant properties as stated in Lemma B.25. To apply Lemma B.25 we use the following facts:

- Whenever $\mathfrak{S}$ is a proper scheduler for $\mathcal{M}$ then $\Pr^{\mathfrak{S}}_{\mathcal{M},s}(\mathit{Limit}_{\mathcal{E}}) = 0$ for each end component $\mathcal{E}$ of $\mathcal{M}$. (As before, $\mathit{Limit}_{\mathcal{E}} = \{\varsigma \in \mathit{IPaths} : \lim(\varsigma) = \mathcal{E}\}$.)
- For each $K \in \mathbb{Z}$, the property $\psi_K = \Diamond(\mathit{goal} \wedge (\mathrm{wgt} = K))$ is measurable and 0-EC-invariant. (Recall that *goal* is a trap. Therefore, there is no end component containing *goal*.)

- For the proper schedulers $\mathfrak{S}$ for $\mathcal{M}$ resp. proper schedulers $\mathfrak{T}$ for $\mathcal{N}$ we have:

$$\mathbb{E}^{\mathfrak{S}}_{\mathcal{M},s}(\oplus\, goal) \;=\; \sum_{K=-\infty}^{+\infty} K \cdot \Pr^{\mathfrak{S}}_{\mathcal{M},s}(\psi_K) \qquad\qquad \mathbb{E}^{\mathfrak{T}}_{\mathcal{N},s}(\oplus\, goal) \;=\; \sum_{K=-\infty}^{+\infty} K \cdot \Pr^{\mathfrak{T}}_{\mathcal{N},s}(\psi_K)\ .$$

Thus, statement (E3) follows directly from part (a) of Lemma B.25, while (E4) follows from part (b) of Lemma B.25. □

By (E2) in Lemma C.6 and Lemma C.4, $\mathbb{E}^{\max}_{\mathcal{N},s}(\oplus\, goal)$ is finite for all states s and $\mathcal{N}$ has a proper MD-scheduler that maximizes the expected accumulated weight from each state. Moreover, we have:

Lemma C.7. *Let $\mathcal{M}$ and $\mathcal{N}$ be as in Lemma C.6. We have $\mathbb{E}^{\max}_{\mathcal{N},s}(\oplus\, goal) = \mathbb{E}^{\max}_{\mathcal{M},s}(\oplus\, goal)$ for each state s, and $\mathcal{M}$ has an MD-scheduler $\mathfrak{S}$ with $\mathbb{E}^{\mathfrak{S}}_{\mathcal{M},s}(\oplus\, goal) = \mathbb{E}^{\max}_{\mathcal{M},s}(\oplus\, goal)$.*

Proof. Follows from Lemma C.4 and (E3) and (E4) in Lemma C.6. □

Corollary C.8 (Implication "$\Longleftarrow$" of Lemma 4.1 for maximal expectations)**.** *If $\mathcal{M}$ has no positively weight-divergent end component then $\mathbb{E}^{\max}_{\mathcal{N},s}(\oplus\, goal)$ is finite for all states s. Moreover, one can construct in polynomial time an MDP $\mathcal{N}$ with the same state space such that $\mathcal{N}$ satisfies condition* (BTmax) *and $\mathbb{E}^{\max}_{\mathcal{M},s}(\oplus\, goal) = \mathbb{E}^{\max}_{\mathcal{N},s}(\oplus\, goal)$ for all states s.*

By multiplying all weights by -1, we obtain that $\mathbb{E}^{\inf}_{\mathcal{N},s}(\oplus\, goal)$ is finite for all states s in an MDP $\mathcal{M}$ without negatively weight-divergent end components (assuming the existence of proper schedulers). This completes the proof of Lemma 4.1.

D Proofs of Section 5

We prove here the statements of Section 5. Our results for weight-bounded reachability and Büchi constraints are summarized in Figure 11 where $\mathcal{M} = (S, Act, P, wgt)$ is an MDP, s_{init} a state of $\mathcal{M}$, T and F are set of states in $\mathcal{M}$. Moreover, $K \in \mathbb{Z}$ and $K_t \in \mathbb{Z} \cup \{-\infty\}$ for $t \in T$.

solvable in polynomial time		in NP ∩ coNP, solvable in pseudo-polynomial time hard for non-stochastic two-player mean-payoff games	
$\mathrm{DWR}^{\exists,>0}$	$\exists \mathfrak{S}.\ \Pr^{\mathfrak{S}}_{\mathcal{M},s_{init}}\big(\bigvee_{t\in T} \Diamond(t \wedge (\mathrm{wgt} \geqslant K_t))\big) > 0\,?$	$\mathrm{DWR}^{\exists,=1}$	$\exists \mathfrak{S}.\ \Pr^{\mathfrak{S}}_{\mathcal{M},s_{init}}\big(\bigvee_{t\in T} \Diamond(t \wedge (\mathrm{wgt} \geqslant K_t))\big) = 1\,?$
$\mathrm{DWR}^{\forall,=1}$	$\forall \mathfrak{S}.\ \Pr^{\mathfrak{S}}_{\mathcal{M},s_{init}}\big(\bigvee_{t\in T} \Diamond(t \wedge (\mathrm{wgt} \geqslant K_t))\big) = 1\,?$	$\mathrm{DWR}^{\forall,>0}$	$\forall \mathfrak{S}.\ \Pr^{\mathfrak{S}}_{\mathcal{M},s_{init}}\big(\bigvee_{t\in T} \Diamond(t \wedge (\mathrm{wgt} \geqslant K_t))\big) = 1\,?$
$\mathrm{WB}^{\exists,>0}$	$\exists \mathfrak{S}.\ \Pr^{\mathfrak{S}}_{\mathcal{M},s_{init}}\big(\Box\Diamond(\mathrm{wgt} \geqslant K) \wedge \Box\Diamond F\big) > 0\,?$	$\mathrm{WB}^{\exists,=1}$	$\exists \mathfrak{S}.\ \Pr^{\mathfrak{S}}_{\mathcal{M},s_{init}}\big(\Box\Diamond(\mathrm{wgt} \geqslant K) \wedge \Box\Diamond F\big) = 1\,?$
$\mathrm{WB}^{\forall,=1}$	$\forall \mathfrak{S}.\ \Pr^{\mathfrak{S}}_{\mathcal{M},s_{init}}\big(\Box\Diamond(\mathrm{wgt} \geqslant K) \wedge \Box\Diamond F\big) = 1\,?$	$\mathrm{WB}^{\forall,>0}$	$\forall \mathfrak{S}.\ \Pr^{\mathfrak{S}}_{\mathcal{M},s_{init}}\big(\Box\Diamond(\mathrm{wgt} \geqslant K) \wedge \Box\Diamond F\big) > 0\,?$

Figure 11. Results for weight-bounded reachability and Büchi constraints

We start with the four cases of disjunctive weight-bounded reachability properties (Sections D.1, D.2, D.3 and D.4) and then consider weight-bounded Büchi properties in Section D.5. For weight constraints not to be trivial, we safely assume that $T \setminus T^*$ is nonempty where $T^* = \{t \in T : K_t = -\infty\}$.

D.1 Positive Reachability Under Some Scheduler

Theorem D.1. *Problem $DWR^{\exists,>0}$ belongs to P and the value $K^{\exists,>0}_{\mathcal{M},s}$ is computable in polynomial time.*

Proof. The existence of a scheduler satisfying a DWR-property φ with positive probability is equivalent to the existence of a path from the initial state s to one of the targets $t \in T$ with accumulated weight at least K_t. To decide the latter in polynomial time, one can rely on shortest-path algorithms for weighted graphs, such as the Bellman-Ford algorithm. More precisely, consider the weighted graph obtained from $\mathcal{M}$ by ignoring action names and probabilities, and switching the weight function from wgt to $-wgt$. Then $K^{\exists,>0}_{\mathcal{M},s}$ is the weight of a shortest path from s to $goal$. To decide $\mathrm{DWR}^{\exists,>0}$, we apply a shortest-path algorithm for each $t \in T$ with source s and target t, and compare the obtained value with K_t. □

D.2 Positive Reachability Under All schedulers

Lemma D.2. *The problem $DWR^{\forall,>0}$ for general MDPs can be reduced in polynomial time to the case where $\mathcal{M}$ has a single goal state which is a trap.*

Proof. Let $\varphi(T, (K_t)_{t\in T})$ denote the weight-bounded reachability constraint $\bigvee_{t\in T} \Diamond(t \wedge (\mathrm{wgt} \geqslant K_t))$ and let us first show that one can assume that $K_t > -\infty$ for all $t \in T$. In fact, for all states $t \in T^* = \{t \mid K_t = -\infty\}$ (which can be assumed to be trap states), one can add one action α with $P(t, \alpha, t) = \frac{1}{2}$ and $P(t, \alpha, t') = \frac{1}{2}$ for some arbitrary $t' \in T \setminus T^*$ with $wgt(t, \alpha) = +1$. Then, for any scheduler, if state t is reached, then with positive probability so will be t' with weight at least $K_{t'}$. The converse is also true: if $\varphi(T \setminus T^*, (K_t)_{t\in T\setminus T^*})$ holds with positive probability for all schedulers for the new MDP, so does $\varphi(T, (K_t)_{t\in T})$ on the original one.

Let us now show how to make sure all target states are trap states by defining $\mathcal{M}'$. For each non-trap $t \in T$, we add a new trap goal state g_t in $\mathcal{M}'$. Moreover, for each state-action pair (t, α) in $\mathcal{M}$, the new MDP $\mathcal{M}'$ has a fresh state $g_{t,\alpha}$. Let G denote the set consisting of all states $t \in T$ that are traps in $\mathcal{M}$ and all states g_t where $t \in T$ is not a trap in $\mathcal{M}$ and $K_{g_t} = K_t$. The new MDP $\mathcal{M}'$ is obtained from $\mathcal{M}$ by adding deterministic transitions from $g_{t,\alpha}$ to g_t with action label τ and weight $-wgt(t, \alpha)$ and by modifying the transition probabilities of each state-action pair (t, α) where $t \in T$ is not a trap in $\mathcal{M}$ is as follows: $P_{\mathcal{M}'}(t, \alpha, s) = \frac{1}{2}P_{\mathcal{M}}(t, \alpha, s)$ for all $s \in S$, and $P_{\mathcal{M}'}(t, \alpha, g_t) = \frac{1}{2}$. Thus, whenever $\mathcal{M}'$ visits t it moves to g_t with probability $\frac{1}{2}$, otherwise $\mathcal{M}'$ continues as in $\mathcal{M}$. Now, any scheduler $\mathfrak{S}$ for $\mathcal{M}'$ with $\Pr^{\mathfrak{S}}_{\mathcal{M}', s_{init}}\big(\varphi(G, (K_g)_{g\in G})\big) > 0$ also satisfies $\Pr^{\mathfrak{S}}_{\mathcal{M}, s_{init}}\big(\varphi(T, (K_t)_{t\in T})\big) > 0$ as any $\mathfrak{S}$-path reaching g_t in $\mathcal{M}'$ has a prefix ending in t with the same accumulated weight. Conversely, if scheduler $\mathfrak{S}$ for $\mathcal{M}$ satisfies $\Pr^{\mathfrak{S}}_{\mathcal{M}}\big(\varphi(T, (K_t)_{t\in T})\big) > 0$, then there exists a $\mathfrak{S}$-path π from s_{init} that ends in some state $t \in T$ with accumulated weight at least K_t. If t is a trap in $\mathcal{M}$ then t is also a goal state in $\mathcal{M}'$. Otherwise, *i.e.*, if t is not a trap in $\mathcal{M}$, then there is a $\mathfrak{S}$-path $\pi' = \pi\, \alpha\, g_{t,\alpha}\, \tau\, g_t$ in $\mathcal{M}'$ with $wgt(\pi') = wgt(\pi) + wgt(t, \alpha) - wgt(t, \alpha) = wgt(\pi) \geqslant K_t = K_{g_t}$. Here, α is any action that $\mathfrak{S}$ schedules with positive probability for the input path π. This shows $\Pr^{\mathfrak{S}}_{\mathcal{M}', s_{init}}\big(\varphi(G, (K_g)_{g\in G})\big) > 0$.

Suppose now that all goal states $t \in T$ are trap states in $\mathcal{M}$. It is now easy to reduce them to a single trap state. In fact, the MDP $\mathcal{M}$ can be modified by adding a fresh goal state g, and from each $t \in T$, a single action that deterministically leads to g with weight $-K_t$. If $\mathcal{M}''$ denotes this new MDP, then satisfying $\varphi(T, (K_t)_{t\in T})$ in $\mathcal{M}$ is equivalent to satisfying $\varphi(g, 0)$ in $\mathcal{M}''$. □

The following two lemmas establish the complexity of the problem $\mathrm{DWR}^{\forall,>0}$ (the first part of Theorem 5.2). The algorithm for the computation of the values will be given afterwards.

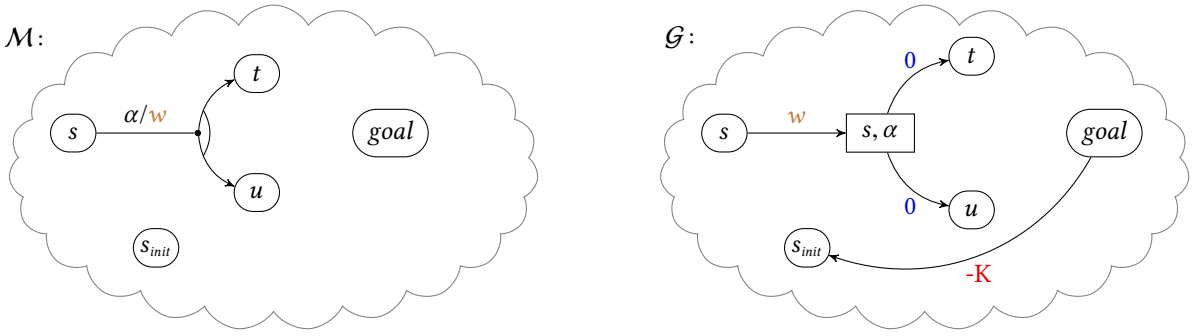


Figure 12. Construction of a two-player game $\mathcal{G}$ from an MPD $\mathcal{M}$.

Lemma D.3. *Let $\mathcal{M}$ be an MDP, goal $\in \mathcal{M}$ a trap state, and $K \in \mathbb{Z}$. Checking whether for all schedulers $\mathfrak{S}$, $\Pr^{\mathfrak{S}}_{\mathcal{M}, s_{init}}\big(\Diamond(goal \wedge (wgt \geqslant K))\big) > 0$ reduces to the resolution of a two-player mean-payoff Büchi game. The problem $DWR^{\forall,>0}$ is thus in NP ∩ coNP.*

Proof. From $\mathcal{M}$, we construct a two-player game $\mathcal{G}$ intuitively as follows: player 1 is responsible for choosing the actions, and player 2 resolves the probabilistic choices; moreover, from state *goal*, controlled by player 1, we add an action leading back to the initial state with weight $-K$. Formally, $\mathcal{G}$ has set of vertices $V = S \cup S \times Act$, partitionned into $V_1 = S$ and $V_2 = S \times Act$, for each player. For every state $s \in S$ in the MDP $\mathcal{M}$ and every action α enabled in s there exists a transition in $\mathcal{G}$ from $s \in V_1$ to $(s, \alpha) \in V_2$ with weight $wgt(s, \alpha)$. Now, for all states $s, t \in S$ in the MDP and actions α satisfying $P(s, \alpha, t) > 0$ there exists a transition in $\mathcal{G}$ from $(s, \alpha) \in V_2$ to $t \in V_1$ with weight 0. Finally, there is a transition from *goal* to the initial state s_{init} with weight $-K$. This transformation is represented in Figure 12.

In the sequel, σ denotes a (pure) strategy for player 1 and τ a (pure) strategy for player 2, and we write $\mathsf{Play}_{\mathcal{G}}(\sigma,\tau)$ for the play in $\mathcal{G}$ yield by σ and τ. The above transformation satifies

$$\begin{aligned}\forall \mathfrak{S},\ \Pr^{\mathfrak{S}}_{\mathcal{M},s_{init}}(\Diamond(goal \wedge (\mathrm{wgt} \geqslant K))) > 0 &\iff \forall\sigma\exists\tau,\ \mathsf{Play}_{\mathcal{G},s_{init}}(\sigma,\tau) \models (\neg goal)\mathbf{U}(goal \wedge (\mathrm{wgt} \geqslant K))\\ &\iff \forall\sigma\exists\tau,\ \mathsf{Play}_{\mathcal{G},s_{init}}(\sigma,\tau) \models \Diamond(goal \wedge (\mathrm{wgt} \geqslant K))\\ &\iff \forall\sigma\exists\tau,\ \mathsf{Play}_{\mathcal{G},s_{init}}(\sigma,\tau) \models (\Box\Diamond goal \wedge \mathrm{MP} \geq 0)\\ &\iff \exists\tau\forall\sigma,\ \mathsf{Play}_{\mathcal{G},s_{init}}(\sigma,\tau) \models (\Box\Diamond goal \wedge \mathrm{MP} \geq 0)\end{aligned}$$

The first equivalence is immediate from the transformation since the positive probability of the eventually property corresponds to the existence of a path in the MDP.

For the second equivalence, the left-to-right implication is obvious; let us prove the right-to-left one. Let σ be a strategy for player 1, and fix τ a strategy for player 2, such that $\mathsf{Play}_{\mathcal{G},s_{init}}(\sigma,\tau) \models \Diamond(goal \wedge (\mathrm{wgt} \geqslant K))$. The play $\mathsf{Play}_{\mathcal{G},s_{init}}(\sigma,\tau)$ until *goal* is reached with accumulated weight at least K, can be decomposed into factors from s_{init} to *goal*, alternated with transitions from *goal* to s_{init}: $\pi_1(goal,\alpha,s_{init})\pi_2\cdots(goal,\alpha,s_{init})\pi_m$ where the π_i's do not visit *goal*. We let K_i be the accumulated weight along π_i. Then the accumulated weight of this prefix play is $\sum_{i=1}^{m-1}(K_i - K) + K_m$, and by assumption, it is greater than K. We derive that $\sum_{i=1}^{m} K_i \geq m \cdot K$, and thus there exists i with $K_i \geq K$. This fragment thus satisfies the property $(\neg goal)\mathbf{U}(goal \wedge (\mathrm{wgt} \geqslant K))$. To conclude, it suffices to observe that the strategy σ can be arbitrary on each of these fragments.

The third equivalence is relatively simple. First of all, from left to right, given a strategy σ for player 1, we aim at building a strategy τ' for player 2 ensuring $(\Box\Diamond goal \wedge \mathrm{MP} \geq 0)$. To do so, the idea is to apply the counterstrategy τ until *goal* is reached with accumulated weight at least K; then τ' takes the α transition from *goal* to s_{init} with weight $-K$, so that the accumulated weight is nonnegative; and we iterate the reasoning from s_{init} again. Doing so, τ' guarantees infinitely many visits to *goal* with accumulated weight at least K, and infinitely many visits to s_{init} with nonnegative accumulated weight. The mean-payoff of $\mathsf{Play}_{\mathcal{G},s_{init}}(\sigma,\tau)$ is thus nonnegative.

The last equivalence is a consequence of the determinacy of two-player turn-based games with mean-payoff and Büchi objectives, a consequence of Martin's general determinacy theorem [22]. Mean-payoff Büchi games are even finite-memory determined [8].

The complexity of the problem $\mathsf{DWR}^{\forall,>0}$ then follows directly, as determining the winner in a turn-based game with mean-payoff Büchi winning condition is in $\mathsf{NP} \cap \mathsf{coNP}$ [8]. □

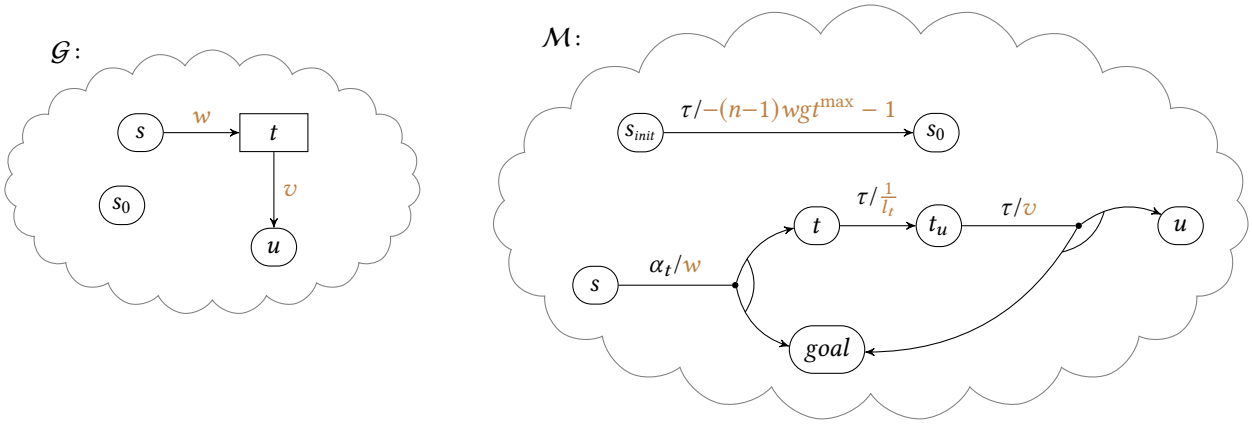


Figure 13. Construction of an MDP $\mathcal{M}$ from a two-player game $\mathcal{G}$.

Lemma D.4. *The problem $DWR^{\forall,>0}$ is hard for (non-stochastic) two-player mean-payoff games.*

Proof. We now prove the lower bound, that is, checking whether player 1 of a (non-stochastic) mean-payoff game has a winning strategy is polynomially reducible to the complement of $\mathsf{DWR}^{\forall,>0}$.

More precisely, we provide a polynomial reduction to the problem to decide whether $\Pr^{\mathfrak{S}}_{\mathcal{M},s_{init}}\big(\Diamond(goal \wedge (\mathrm{wgt} \geqslant 0))\big) = 0$ holds for all schedulers $\mathfrak{S}$ for a given MDP $\mathcal{M}$ with distinguished states *goal* and s_{init}.

Consider a mean-payoff game $\mathcal{G}$ with starting state s_0. Let $\mathcal{M}$ be the MDP obtained from $\mathcal{G}$ by performing the following steps (see also Figure 13).

- Add a new initial state s_{init} and a trap state *goal*.
- For each player-1 state s and edge $s \xrightarrow{w} t$ in $\mathcal{G}$, state s in $\mathcal{M}$ has an enabled action α_t with $P(s,\alpha_t,t) = P(s,\alpha_t,goal) = \frac{1}{2}$ and $wgt(s,\alpha_t) = w$.
- For each player-2 state s in $\mathcal{G}$, we add states s and s_t for all successors t of s to $\mathcal{M}$. State s in $\mathcal{M}$ has a single enabled action τ with $P(s,\tau,s_t) = \frac{1}{\ell_s}$ where ℓ_s denotes the number of successors of s in $\mathcal{G}$ and where t ranges over all successors of s in $\mathcal{G}$. The states s_t have a single enabled action τ with $P(s_t,\tau,goal) = P(s_t,\tau,t) = \frac{1}{2}$ and $wgt(s_t,\tau)$ equals the weight of the edge from s to t in $\mathcal{G}$.
- State s_{init} has a single action with $P(s_{init},\tau,s_0) = 1$ and $wgt(s_{init},\tau) = -(n-1)wgt^{\max} - 1$ where $wgt^{\max}$ is the maximal weight attached to the edges in $\mathcal{G}$ and n is the number of states in $\mathcal{G}$. (We suppose $wgt^{\max} > 0$. If this is not the case we put $wgt(s_{init},\tau) = -1$.)

Then, $\mathcal{M}$ is contracting in the sense $\Pr^{\min}_{\mathcal{M},s}(\Diamond goal) = 1$. In particular, $\mathcal{M}$ has no end components. We have for all schedulers $\mathfrak{S}$ for $\mathcal{M}$ that $\Pr^{\mathfrak{S}}_{\mathcal{M},s_{init}}\big(\Diamond goal \wedge (\mathrm{wgt} < 0)\big) = 1$ iff $\Pr^{\mathfrak{S}}_{\mathcal{M},s_{init}}\big(\Diamond goal \wedge (\mathrm{wgt} \geqslant 0)\big) = 0$. Moreover, there is a one-to-one correspondence between the schedulers for $\mathcal{M}$ and the strategies for player 1 in $\mathcal{G}$.

If $\mathfrak{S}$ is an MD-strategy for player 1 in $\mathcal{G}$ such that the mean payoff of all $\mathfrak{S}$-plays is nonpositive, then $\mathfrak{S}$ has no positive cycles and

$$\Pr^{\mathfrak{S}}_{\mathcal{M},s_{init}}\big(\,\Diamond goal \wedge (\mathrm{wgt} < 0)\,\big) \;=\; 1.$$

In fact $wgt(\pi) < -wgt(s_{init},s_0)$ for all simple $\mathfrak{S}$-paths π starting in state s_0, and since there are no positive cycles under $\mathfrak{S}$, any non-simple path has also negative weight. That is, $wgt(\pi) < 0$ for all $\mathfrak{S}$-paths starting in s_{init}.

Conversely, if $\mathfrak{S}$ is a scheduler for $\mathcal{M}$ with $\Pr^{\mathfrak{S}}_{\mathcal{M},s_{init}}\big(\Diamond(goal \wedge (\mathrm{wgt} < 0))\big) = 1$ then there is an MD-scheduler $\mathfrak{T}$ for $\mathcal{M}$ with $\Pr^{\mathfrak{T}}_{\mathcal{M},s_{init}}\big(\Diamond(goal \wedge (\mathrm{wgt} < 0))\big) = 1$ (see Lemma D.19) and the Markov chain induced by $\mathfrak{T}$ has no positive cycles. Thus, the mean payoff of all $\mathfrak{T}$-plays starting is nonpositive. □

Finally, we explain how to compute $K^{\forall,>0}_{\mathcal{M},s}$ in pseudo-polynomial time. Note that this implies that the decision problem is solvable in pseudo-polynomial time as well.

We may assume w.l.o.g. that $T \setminus T^*$ is a singleton (following the argumentation provided in Lemma D.18 later on). That is, T contains a single trap state with finite K_t. Additionally, we make the following assumption **(A)**:

(A) $\Pr^{\min}_{\mathcal{M},s_{init}}(\Diamond T) > 0$.

This assumption is justified as $\Pr^{\min}_{\mathcal{M},s_{init}}(\Diamond T) = 0$ implies $\Pr^{\mathfrak{S}}_{\mathcal{M},s_{init}}(\varphi) = 0$ for some scheduler $\mathfrak{S}$ regardless the value of K_t.

Preprocessing. In what follows, let $T = \{goal\}$ and suppose that $\mathcal{M}$ satisfies assumption **(A)**. We now define a new MDP $\mathcal{N}$ that arises from $\mathcal{M}$ by adding a fresh action symbol τ and a new trap state *fail* and by performing the following steps:

1. Collapse all states s with $\Pr^{\min}_{\mathcal{M},s}(\Diamond T) = 0$ to the single trap state *fail*.
2. Remove all states s with $s_{init} \not\models \exists\Diamond s$.

As all states s that belong to some end component $\mathcal{E}$ of $\mathcal{M}$ are collapsed to *fail* (see step 1), the MDP $\mathcal{N}$ has no end component. Hence, under all schedulers for $\mathcal{N}$, almost surely one of its two trap states *fail* or *goal* will be reached:

$$\Pr^{\min}_{\mathcal{N},s}\big(\,\Diamond(goal \vee fail)\,\big) \;=\; 1 \quad \text{for all states } s \text{ of } \mathcal{N}\ .$$

Assumption **(A)** yields $\Pr^{\min}_{\mathcal{N},s_{init}}\big(\,\Diamond(goal)\,\big) \;>\; 0$.

For $K \in \mathbb{Z}$, let

$$\psi_K \;=\; \Diamond(goal \wedge (\mathrm{wgt} \geqslant K))\ .$$

Problem $\mathrm{DWR}^{\forall,>0}$ rephrased for $\mathcal{N}$ asks whether $\Pr^{\mathfrak{S}}_{\mathcal{N},s_{init}}(\psi_K) > 0$ for all schedulers $\mathfrak{S}$ for $\mathcal{N}$ where $K \in \mathbb{Z}$ is fixed. The corresponding optimization problem asks to compute for the states s in $\mathcal{N}$ the values

$$K^{\forall,>0}_{\mathcal{N},s} \;=\; \sup\big\{\, K \mid \forall\mathfrak{S}.\, \Pr^{\mathfrak{S}}_{\mathcal{N},s}(\psi_K) > 0 \,\big\}\ .$$

We have $K^{\forall,>0}_{\mathcal{M},s} = K^{\forall,>0}_{\mathcal{N},s}$ for all states s in $\mathcal{M}$ with $\Pr^{\min}_{\mathcal{M},s}(\Diamond goal) > 0$, while $K^{\forall,>0}_{\mathcal{M},s} = -\infty$ if $\Pr^{\min}_{\mathcal{M},s}(\Diamond T) = 0$.

Assumptions after the preprocessing. We now have the following assumptions

- **(C1)** $\mathcal{M}$ has no end components and two traps states *goal* and *fail*.
- **(C2)** $\Pr^{\min}_{\mathcal{M},s}\big(\Diamond(goal \vee fail)\big) = 1$ for all states s of $\mathcal{M}$.
- **(C3)** $\Pr^{\min}_{\mathcal{M},s}\big(\Diamond(goal)\big) > 0$ for all states s of $\mathcal{M}$ with $s \neq fail$.
- **(C4)** All states in $\mathcal{M}$ are reachable from s_{init}.

The values for the trap states are trivial as we have $K^{\forall,>0}_{\mathcal{M},fail} = -\infty$ and $K^{\forall,>0}_{\mathcal{M},goal} = 0$.

Lemma D.5. *If $\mathcal{M}$ satisfies the above assumptions **(C1)** to **(C3)**, then $K^{\forall,>0}_{\mathcal{M},s} \in \mathbb{Z} \cup \{+\infty\}$ for all non-trap states s in $\mathcal{M}$.*

Proof. Let s be a non-trap state in $\mathcal{M}$. Let E_s denote the maximal conditional expected number of steps for reaching *goal* from s in $\mathcal{M}$, under the condition $\Diamond(goal)$. By assumption **(C3)** and the results of [19], E_s is finite for all states s in $\mathcal{M}$. Let $k_s = \lceil E_s \rceil$. Then, for each scheduler $\mathfrak{S}$ there is at least one path from s to *goal* of length at most k_s, which yields $K^{\forall,>0}_{\mathcal{M},s} \in \mathbb{Z} \cup \{\infty\}$. □

Lemma D.6. *Assumptions and notations as before. For each $K \in \mathbb{Z}$ and each state s in $\mathcal{N}$:*

$$\Pr^{\mathfrak{S}}_{\mathcal{M},s}(\varphi_K) > 0 \text{ for all schedulers } \mathfrak{S} \text{ for } \mathcal{M} \quad \textit{iff} \quad \Pr^{\mathfrak{S}}_{\mathcal{N},s}(\varphi_K) > 0 \text{ for all schedulers } \mathfrak{S} \text{ for } \mathcal{N} \ .$$

Thus, the value $K^{\forall,>0}_{\mathcal{M},s}$ is the maximal value $K \in \mathbb{Z} \cup \{\infty\}$ such that $\Pr^{\mathfrak{S}}_{\mathcal{N},s}(\varphi_K) > 0$ for all schedulers $\mathfrak{S}$ for $\mathcal{N}$.

Corollary D.7. $K^{\forall,>0}_{\mathcal{M},s} = K^{\forall,>0}_{\mathcal{N},s}$ *for all states s in $\mathcal{N}$.*

Given a scheduler $\mathfrak{S}$ for $\mathcal{N}$ we define

$$K^0_{\mathfrak{S},s} \quad = \quad \sup\left\{ K \in \mathbb{Z} \,:\, \Pr^{\mathfrak{S}}_{\mathcal{N},s}(\varphi_K) > 0 \right\} \ .$$

If $\mathfrak{S}$ is an MD-scheduler then let $\mathcal{N}_{\mathfrak{S}}$ denote the Markov chain induced by $\mathfrak{S}$.

Lemma D.8. *Let $\mathfrak{S}$ be an MD-scheduler for $\mathcal{N}$. Then, for all non-trap states s in $\mathcal{N}$:*

(a) $K^0_{\mathfrak{S},s} = +\infty$ iff $\mathcal{N}_{\mathfrak{S}}$ has a positive cycle that is reachable from s.
(b) If $\mathcal{N}_{\mathfrak{S}}$ does not contain any positive cycle that is reachable from s, then

$$K^0_{\mathfrak{S},s} \quad = \quad \max\left\{ wgt(\pi) \,:\, \pi \text{ is a path from } s \text{ to } goal \text{ in } \mathcal{N}_{\mathfrak{S}} \right\} \ .$$

The values $K^0_{\mathfrak{S},s}$ are computable in polynomial time.

Proof. Statements (a) and (b) are obvious. To check the existence of positive cycles and to compute the values $K^0_{\mathfrak{S},s}$ in (b) we can apply standard shortest-path algorithms to the weighted graph that arises from $\mathcal{N}_{\mathfrak{S}}$ by ignoring the transition probabilities and multiplying all weights with -1. □

Let S be the state space of $\mathcal{N}$ without state *fail*. If (s,α) is a state-action pair in $\mathcal{N}$ then $Post(s,\alpha) = \{t \in S \cup \{fail\} : P(s,\alpha,t) > 0\}$.

Lemma D.9. *For each state $s \in S$ we have that $K^{\forall,>0}_{\mathcal{N},s} \in \mathbb{Z}$ iff there is at least one MD-scheduler $\mathfrak{S}$ for $\mathcal{N}$ such that the Markov chain $\mathcal{N}_{\mathfrak{S}}$ induced by $\mathfrak{S}$ has no positive cycle that is reachable from s. In this case, $K^{\forall,>0}_{\mathcal{N},s} = \min_{\mathfrak{S}} K^0_{\mathfrak{S},s}$ where the minimum ranges over all MD-schedulers $\mathfrak{S}$ for $\mathcal{N}$.*

Proof. "$\Longleftarrow$": If there is an MD-scheduler $\mathfrak{S}$ without positive cycles, then $K^{\forall,>0}_{\mathcal{N},s}$ is bounded from above by the maximal weight of the $\mathfrak{S}$-paths from s to *goal*. This value is finite.

"$\Longrightarrow$": Suppose $K \stackrel{\text{def}}{=} K^{\forall,>0}_{\mathcal{N},s} \in \mathbb{Z}$. Because K is a maximum, there is some scheduler $\mathfrak{S}$ such that $wgt(\pi) \leqslant K$ for all $\mathfrak{S}$-paths from s to *goal*. But then:

$$\Pr^{\mathfrak{S}}_{\mathcal{N},s}\big(\Diamond fail \ \vee \ \Diamond(goal \wedge (\mathrm{wgt} \leqslant K)) \big) \quad = \quad 1 \ .$$

As $\mathcal{N}$ has no end components, $\mathcal{N}$ has no (positively or negatively) weight-divergent scheduler. Hence, we may apply Lemma D.19 to obtain the existence of an MD-scheduler $\mathfrak{T}$ such that

$$\Pr^{\mathfrak{T}}_{\mathcal{N},s}\big(\Diamond fail \ \vee \ \Diamond(goal \wedge (\mathrm{wgt} \leqslant K)) \big) \quad = \quad 1 \ .$$

But then the weight of all $\mathfrak{T}$-paths from s to *goal* is bounded by K. Lemma D.8 yields that $\mathfrak{T}$ has no positive cycle that is reachable from s. The last part is obvious from Lemma D.8. □

Note that the previous lemma is sufficient to derive an exponential-time algorithm to compute the values: one can enumerate all MD-schedulers and pick the one with the best value. In the remaining of this section, we will show how to compute these values in pseudo-polynomial time.

Lemma D.10. *Let $S_\infty = \{s \in S : K^{\forall,>0}_{\mathcal{N},s} = \infty\}$. Then for each state $s \in S$ the following statements are equivalent:*

(a) $s \in S_\infty$
(b) For each $w \in \mathbb{Z}$ and each scheduler $\mathfrak{S}$ there is an $\mathfrak{S}$-path π from s to goal with $wgt(\pi) \geqslant w$.
(c) $\Pr^{\min}_{\mathcal{N},s}(\Diamond S_\infty) > 0$

Proof. "(a) $\iff$ (b)" and "(a) $\implies$ (c)" are trivial. To prove "(c) $\implies$ (b)" we suppose $\Pr^{\min}_{\mathcal{N},s}(\Diamond S_\infty) > 0$. Let $w \in \mathbb{Z}$ and $\mathfrak{S}$ be a scheduler. As $\Pr^{\mathfrak{S}}_{\mathcal{N},s}(\Diamond S_\infty) > 0$ there is a state $t \in S_\infty$ and an $\mathfrak{S}$-path π' from s to t. Let $w' = wgt(\pi')$. We now consider the residual scheduler $\mathfrak{S}' = \mathfrak{S}{\uparrow}\pi'$. As (a) and (b) are equivalent and $t \in S_\infty$, there is an $\mathfrak{S}'$-path π'' from t to *goal* with $wgt(\pi'') \geqslant w - w'$. But then $\pi \stackrel{\text{def}}{=} \pi'; \pi''$ is an $\mathfrak{S}$-path from s to *goal* with

$$wgt(\pi) \;=\; wgt(\pi') + wgt(\pi'') \;\geqslant\; w' \,+\, (w - w') \;=\; w \;.$$

This completes the proof of Lemma D.10. □

Remark D.11 (Reduction to mean-payoff games)**.** Checking whether $K^{\forall,>0}_{\mathcal{M},s} = +\infty$ is polynomially reducible to non-stochastic two-player mean-payoff games. For this, we regard MDPs as non-stochastic two-player games (action player against probabilistic player). The objective of the action player is to ensure that the mean payoff is nonpositive. Then, "all MD-schedulers of an MDP have a positive cycle" is equivalent to "there is no winning strategy for the action player". Thus, Lemma D.9 yields a polynomial reduction to the complement of non-stochastic mean-payoff games with threshold 0.

The previous remark allows us to compute S_∞ in pseudo-polynomial time since mean-payoff games can be solved in pseudo-polynomial time. In the rest of this section, we will assume that S_∞ is given, and show how to compute the values in polynomial time. The overall complexity will thus be pseudo-polynomial time.

Computing the values in $\mathcal{N}$. Suppose we have an oracle to compute S_∞. Let

$$S_{\text{fin}} \;\stackrel{\text{def}}{=}\; S \setminus S_\infty \;=\; \left\{\, s \in S \;:\; K^{\forall,>0}_{\mathcal{N},v} \in \mathbb{Z} \,\right\} \;.$$

For each state $s \in S_{\text{fin}}$ we define $Act_{\text{fin}}(s)$ as the set of actions $\alpha \in Act(s)$ such that $P(s,\alpha,s') > 0$ implies $s' \in S_{\text{fin}}$. Note that $Act_{\text{fin}}(s)$ is nonempty if $s \in S_{\text{fin}} \setminus \{goal\}$. We have $\Pr^{\min}_{\mathcal{N},s}(\Diamond S_\infty) = 0$ for all states $s \in S_{\text{fin}}$.

The values $K^{\forall,>0}_{\mathcal{N},s}$ for the states $s \in S_{\text{fin}} \setminus \{goal\}$ satisfy the following equation:

$$K^{\forall,>0}_{\mathcal{N},s} \;=\; \min\left\{\, K_{s,\alpha} \;:\; \alpha \in Act_{\text{fin}}(s) \,\right\}$$

where for $(s,\alpha) \in \mathcal{N}$

$$K_{s,\alpha} \;=\; wgt(s,\alpha) \;+\; \max\left\{K^{\forall,>0}_{\mathcal{N},v} \;:\; v \in Post(s,\alpha) \setminus \{fail\}\right\} \;.$$

Recall that $K^{\forall,>0}_{\mathcal{N},goal} = 0$.

We now provide a polynomial-time algorithm for the computation of the values $K^{\forall,>0}_{\mathcal{N},s}$ for $s \in S_{\text{fin}}$. Let $n = |S_{\text{fin}}|$ denote the number of states in S_{fin}.

Initialization. Let $K^{(j)}_{goal} = 0$ for $j = 0, 1, \ldots, n{-}1$. For all states $s \in S_{\text{fin}} \setminus \{goal\}$ we start with $K^{(0)}_s \;=\; -\infty$.

Iteration. For $j = 1, \ldots, n{-}1$ we compute the following values for all states $s \in S \setminus \{goal\}$ and all actions $\alpha \in Act_{\text{fin}}(s)$:

$$K^{(j)}_{s,\alpha} \;=\; wgt(s,\alpha) \;+\; \max_{t \in Post(s,\alpha)} K^{(j-1)}_t \qquad \text{and} \qquad K^{(j)}_s \;=\; \min\left\{\, K^{(j)}_{s,\alpha} \;:\; \alpha \in Act_{\text{fin}}(s) \,\right\}.$$

Lemma D.12 (Soundness)**.** *The above algorithm correctly computes the values $K^{(n-1)}_s = K^{\forall,>0}_{\mathcal{N},s}$ for all states $s \in S_{fin}$.*

Proof. Let $\mathcal{N}_{\text{fin}}$ be the largest sub-MDP of $\mathcal{N}$ that does not contain any state of S_∞. That is, the state space of $\mathcal{N}_{\text{fin}}$ is $S_{\text{fin}} \cup \{fail\}$ and $\mathcal{N}_{\text{fin}}$ results from $\mathcal{N}$ by removing the states $t \in S_\infty$ and all state-action pairs (s,α) with $P(s,\alpha,t) > 0$ for some $t \in S_\infty$. Thus, the action set of each state $s \in S_{\text{fin}}$ is $Act_{\text{fin}}(s)$. Then, $\mathcal{N}_{\text{fin}}$ has no positive cycle (Lemma D.9).

By induction on j, we get for all states $s \in S_{\text{fin}}$ and all actions $\alpha \in Act_{\text{fin}}(s)$:

$$K^{(j)}_{s,\alpha} \;\leqslant\; K^{(j+1)}_{s,\alpha} \;\leqslant\; K_{s,\alpha} \quad \text{and} \quad K^{(j)}_s \;\leqslant\; K^{(j+1)}_s \;\leqslant\; K^{\forall,>0}_{\mathcal{N},s} \;.$$

Given a scheduler $\mathfrak{S}$ for $\mathcal{N}_{\text{fin}}$, let

$$maxwgt^{(j)}_s[\mathfrak{S}] \;=\; \max\left\{\; wgt(\pi) \;:\; \pi \text{ is a } \mathfrak{S}\text{-path from } s \text{ to } goal \text{ with } |\pi| \leqslant j \;\right\} \;.$$

Then, by induction on j we get:

$$K^{(j)}_s \;=\; \min_{\mathfrak{S}}\; maxwgt^{(j)}_s[\mathfrak{S}]$$

where $\mathfrak{S}$ ranges over all schedulers for $\mathcal{N}_{\text{fin}}$. Moreover, there exists a scheduler $\mathfrak{S}_j$ for $\mathcal{N}_{\text{fin}}$ such that $K^{(j)}_s = maxwgt^{(j)}_s[\mathfrak{S}_j]$.

Let now $\mathfrak{S} = \mathfrak{S}_{n-1}$. As $\mathcal{N}_{\text{fin}}$ has no positive cycles, we have: If π is a finite path of length at least n, then $wgt(\pi) \leqslant wgt(\pi')$ where π' results from π by removing all cycles. Thus, $maxwgt^{(n-1)}_s[\mathfrak{S}]$ is the maximal weight of a $\mathfrak{S}$-path from s to *goal*. But then $K^{(n-1)}_s = K^{\forall,>0}_{\mathcal{N},v}$ for all states $s \in S_{\text{fin}}$. □

Corollary D.13. *The values $K^{\forall,>0}_{\mathcal{M},s}$ for the states $s \in S_{fin}$ are computable in polynomial time, assuming an oracle to compute the set S_∞.*

By Lemma D.9, if $\mathcal{M}$ has no positive cycles then S_∞ is empty. Hence:

Corollary D.14 (Complexity of $\mathrm{DWR}^{\forall,>0}$ for MDPs without positive cycles)**.** *If $\mathcal{M}$ has no positive cycles then problem $DWR^{\forall,>0}$ is in P and the values $K^{\forall,>0}_{\mathcal{M},s}$ are computable in polynomial time.*

For the general case we have:

Theorem D.15 (Complexity of $\mathrm{DWR}^{\forall,>0}$)**.** *The decision problem $DWR^{\forall,>0}$ is in coNP and the values $K^{\forall,>0}_{\mathcal{M},s}$ for the states s in $\mathcal{M}$ are computable in pseudo-polynomial time.*

Proof. To prove membership to coNP we rely on the statements of Lemma D.9, which yields that the answer to question $\mathrm{DWR}^{\forall,>0}$ is "no" iff there is an MD-scheduler $\mathfrak{S}$ for $\mathcal{N}$ such that the Markov chain induced by $\mathfrak{S}$ contains no positive cycle and $K^0_{\mathfrak{S},s_{init}} < K$. So, a nondeterministic polynomially time-bounded algorithm for the complement of $\mathrm{DWR}^{\forall,>0}$ is obtained by guessing an MD-scheduler for $\mathcal{N}$, computing the value $K^0_{\mathfrak{S},s_{init}}$ in polynomial time (see Lemma D.8) and finally checking whether $K^0_{\mathfrak{S},s_{init}} < K$.

To compute the values $K^{\forall,>0}_{\mathcal{M},s}$ in pseudo-polynomial time, we compute S_∞ in pseudo-polynomial time by Remark D.11, and apply the above algorithm to compute the values $K^{\forall,>0}_{\mathcal{M},s}$ for the states $s \in S_{\mathrm{fin}}$. □

D.3 Almost-Sure Reachability Under All Schedulers

In this section, we prove the following theorem.

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.*

From $\mathcal{M}$ we construct a weighted directed graph $G = (V, \to, wgt)$. The set of vertices is $V = \{s \in S : \Pr^{\min}_{\mathcal{M},s}(\Diamond T^*) < 1\}$. There is an edge in G from s to s' iff there exists an action $\alpha \in Act$ with $P(s,\alpha,s') > 0$. The weight associated with edge $s \to s'$ is the minimum among actions that can lead from s to s': $wgt(s \to s') = \min\{wgt(s,\alpha) \mid P(s,\alpha,s') > 0\}$. Finally, for $s \in V$, G_s denotes the subgraph of G reachable from s.

In the case where all states in T are traps, Theorem 5.3 derives from the following characterization of positive instances of $\mathrm{DWR}^{\forall,=1}$:

Lemma D.16. *Let φ be a DWR-property with all states in T being traps. Then $\Pr^{\min}_{\mathcal{M},s}(\varphi) = 1$ iff the following two conditions hold:*

(i) $\Pr^{\min}_{\mathcal{M},s}(\Diamond T) = 1$, *and*

(ii) if $\Pr^{\min}_{\mathcal{M},s}(\Diamond T^*) < 1$, *then the weighted graph G_s does not contain any negative cycle, and for each path π in G_s that starts in s and ends in some $t \in T \setminus T^*$ we have $wgt(\pi) \geqslant K_t$.*

In this case and if $T \setminus T^* = \{goal\}$ is a singleton, then $K^{\forall,=1}_{\mathcal{M},s}$ is the minimal weight of a path from s to goal in G_s.

Proof. "$\Longrightarrow$": $\Pr^{\min}_{\mathcal{M},s}(\Diamond T) = 1$ is clearly a necessary condition for $\Pr^{\min}_{\mathcal{M},s}(\varphi) = 1$. Assume now that $\Pr^{\min}_{\mathcal{M},s}(\Diamond T^*) < 1$ and that either G_s contains a negative cycle, or there is a path π from s to some $t \in T \setminus T^*$ with $wgt(\pi) < K_t$. In both cases there is a scheduler $\mathfrak{S}$ such that $\Pr^{\mathfrak{S}}_{\mathcal{M},s}(\Diamond(t \wedge \mathrm{wgt} < K_t)) > 0$ and hence $\Pr^{\min}_{\mathcal{M},s}(\varphi) < 1$.

"$\Longleftarrow$": $\Pr^{\min}_{\mathcal{M},s}(\Diamond T^*) = 1$ clearly implies $\Pr^{\min}_{\mathcal{M},s}(\varphi) = 1$. Assume now that $\Pr^{\min}_{\mathcal{M},s}(\Diamond T) = 1$, $\Pr^{\min}_{\mathcal{M},s}(\Diamond T^*) < 1$, G_s does not contain any negative cycle, and for each path π in G_s that starts in s and ends in some $t \in T \setminus T^*$ we have $wgt(\pi) \geq K_t$. Then under all schedulers and for every path from s to a target state $t \in T \setminus T^*$ the accumulated weight necessarily is at least K_t. We thus derive $\Pr^{\min}_{\mathcal{M},s}(\varphi) = 1$. □

Condition (i) from the characterization of Lemma D.16 is a classical verification question for MDP and can be solved in P. For condition (ii), the weighted graph can be constructed in polynomial time, and using standard shortest-path algorithms in weighted graphs one can check for the nonexistence of a negative cycle and compute the minimal weight of paths from s to a target state $t \in T \setminus T^*$. Thus, in case all T-states are traps, $\mathrm{DWR}^{\forall,=1}$ can be solved in polynomial time.

We now address the general case. Intuitively, this case is harder since a target state $t \in T \setminus T^*$ might be visited several times before t is actually visited with the constraint $\mathrm{wgt} \geqslant K_t$. However, let us explain how to reduce the general case to the case where all states in T are traps. To ease the presentation, we consider a simple DWR-property of the form $\Diamond(goal \wedge (\mathrm{wgt} \geqslant K))$.

Clearly, $\Pr^{\min}_{\mathcal{M},s}(\Diamond goal) = 1$ is a necessary condition for $\Pr^{\min}_{\mathcal{M},s}(\varphi) = 1$. We thus check first whether $\Pr^{\min}_{\mathcal{M},s}(\Diamond goal) = 1$ holds. Then, without loss of generality, we assume that all states are reachable from the initial state s, and that $\Pr^{\min}_{\mathcal{M},t}(\Diamond goal) = 1$ for

all states t. Under these assumptions, all end components of $\mathcal{M}$ must contain *goal*. If $\mathcal{M}$ has no end components then *goal* is a trap, and we are back to the special case (Lemma D.16). Suppose now that *goal* is not a trap. Then, $\mathcal{M}$ has a unique maximal end component $\mathcal{E}$ and $\mathcal{E}$ contains *goal*.

- If $\mathbb{E}^{\min}_{\mathcal{E}}(\mathrm{MP}) > 0$ then all end components of $\mathcal{M}$ are pumping. Hence, $\Pr^{\min}_{\mathcal{M},s}\big(\Diamond(goal \wedge (\mathrm{wgt} \geqslant K))\big) = 1$ for all $K \in \mathbb{Z}$, and therefore $K^{\forall,=1}_{\mathcal{M},s} = +\infty$.
- If $\mathbb{E}^{\min}_{\mathcal{E}}(\mathrm{MP}) < 0$ then $\mathcal{M}$ has negatively weight-divergent end components. In this case, $\Pr^{\mathfrak{S}}_{\mathcal{E},goal}\big(\bigcirc \Diamond(goal \wedge (\mathrm{wgt} < 0))\big) > 0$ where $\mathfrak{S}$ is an MD-scheduler for $\mathcal{E}$ with $\mathbb{E}^{\mathfrak{S}}_{\mathcal{E}}(\mathrm{MP}) = \mathbb{E}^{\min}_{\mathcal{E}}(\mathrm{MP}) < 0$. (This follows from the quotient representation of the expected mean payoff in the Markov chain induced by $\mathfrak{S}$, see Lemma B.3.) Hence, $K^{\forall,=1}_{\mathcal{M},s} = K^{\forall,=1}_{\mathcal{M}',s}$ where $\mathcal{M}'$ is the MDP resulting from $\mathcal{M}$ when turning *goal* into a trap (*i.e.*, removing all state-action pairs $(goal, \alpha)$). For $\mathcal{M}'$ we can then rely on Lemma D.16.
- Suppose now that $\mathbb{E}^{\min}_{\mathcal{E}}(\mathrm{MP}) = 0$.
  - If $\mathcal{E}$ does not contain any 0-EC then $\mathcal{E}$ is universally weight-divergent, *i.e.*, all end components of $\mathcal{E}$ are positively weight-divergent. This is a consequence of Theorem 3.14 applied to the MDP resulting from $\mathcal{M}$ by multiplying all weights with -1. Hence, in this case we have $K^{\forall,=1}_{\mathcal{M},s} = +\infty$.
  - Suppose now that $\mathcal{E}$ contains at least one 0-EC. In this case, we can treat *goal* as a trap and rely on Lemma D.16. In fact, let $\mathcal{F}$ be a 0-EC. If $goal \notin \mathcal{F}$, then from *goal* under some scheduler, $\mathcal{F}$ is reached with positive probability and the run remains almost surely in $\mathcal{F}$. So if the first visit to *goal* does not satisfy the weight constraint, no further visits might be possible under some schedulers. If $goal \in \mathcal{F}$, for some schedulers that remain in $\mathcal{F}$ the accumulated weight will be identical at each visit to *goal* since $\mathcal{F}$ is a 0-EC.

To check which of the above cases applies we can use standard polynomial-time algorithms to compute the minimal expected mean payoff in strongly connected MDPs and the algorithms to check the existence of 0-ECs presented in Section 3.3 (see also Section B.4.3). Recall from part (b) of Theorem 3.14 and Corollary 3.15 that the latter can also be used to check universal weight-divergence. Putting things together, this proves Theorem 5.3.

We finish this section with a very similar result applied to the particular special case of Markov chains, which will be useful in the next section.

Lemma D.17. *Let φ be a DWR-property and s a state of a Markov chain C. Then $\Pr_{C,s}(\varphi) = 1$ if and only if:*

(i) $\Pr_{C,s}(\Diamond T) = 1$ *and*

(ii) for each $t \in T \setminus T^*$, there is no path π from s to t containing a state that belongs to a negative cycle; moreover the minimal weight of a path from s to t is at least K_t.

Observe that conditions (i) and (ii) in Lemma D.17 can be checked in polynomial time: in particular (ii) reduces to standard shortest-path algorithms in weighted graphs. As a consequence, for Markov chains we obtain that $\Pr_{C,s}\big(\Diamond T^* \vee \bigvee_{t \in T \setminus T^*} t \wedge (\mathrm{wgt} \geqslant K_t)\big) = 1$ can be decided in polynomial time.

D.4 Almost-Sure Reachability Under Some Scheduler

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

To establish the upper complexity bound of Theorem 5.5, we first justify that we can assume without loss of generality that $\mathcal{M}$ has certain properties.

Lemma D.18. *The problem $DWR^{\exists,=1}$ for general MDPs can be reduced in polynomial time to the case where instances of $DWR^{\exists,=1}$ enjoy the following properties:*

(A1) *$T^* = \{good\}$ and $T \setminus T^* = \{goal\}$ with good and goal are both traps.*

(A2) *For all states $s \in S \setminus T^*$, $\Pr^{\max}_{\mathcal{M},s}(\Diamond T^*) < 1$.*

(A3) *For all states $s \in S$, $\Pr^{\max}_{\mathcal{M},s}(\Diamond T) = 1$.*

Proof. Let $\mathcal{M}$ be an MDP, and $\varphi = \bigvee_{t \in T} \Diamond(t \wedge (\mathrm{wgt} \geqslant K_t))$. We start by proving that **(A1)** is not a real restriction. From $\mathcal{M}$, we build $\mathcal{M}'$ that extends $\mathcal{M}$ with copies t' of the states $t \in T^*$, an additional state *goal*, and new state-action pairs (t, τ) such that: in case $t \in T^*$, $P_{\mathcal{M}'}(t, \tau, t') = 1$ and $wgt_{\mathcal{M}'}(t, \tau) = 0$; and for $t \in T \setminus T^*$, $P_{\mathcal{M}'}(t, \tau, goal) = 1$ and $wgt_{\mathcal{M}'}(t, \tau) = -K_t$. The new states (*goal* and each t' for $t \in T^*$) are traps. We then let

$$\varphi' \;=\; \Diamond(goal \wedge (\mathrm{wgt} \geqslant 0) \vee \bigvee_{t \in T^*} \Diamond(t' \wedge (\mathrm{wgt} \geqslant K_t)) \;.$$

This construction in particular ensures $\Pr^{\max}_{\mathcal{M},s_{init}}(\varphi) = 1$ iff $\Pr^{\max}_{\mathcal{M}',s_{init}}(\varphi') = 1$.

To justify assumption **(A2)**, observe that $K^{\exists,=1}_{\mathcal{M},s} = +\infty$ for all states s with $\Pr^{\max}_{\mathcal{M},s}(\Diamond T^*) = 1$, so that $\mathsf{DWR}^{\exists,=1}$ is trivial for such states. Moreover, setting $\widetilde{T}^* = \{s \in S : \Pr^{\max}_{\mathcal{M},s}(\Diamond T^*) = 1\}$ and

$$\widetilde{\varphi}_K \quad = \quad \Diamond\widetilde{T}^* \vee \Diamond(\mathit{goal} \wedge (\mathrm{wgt} \geqslant K))$$

we have that for all states s in $\mathcal{M}$, $\Pr^{\max}_{\mathcal{M},s}(\varphi_K) \quad = \quad \Pr^{\max}_{\mathcal{M},s}(\widetilde{\varphi}_K)$. We can thus safely assume $\widetilde{T}^* = T^*$.

Finally, it is no loss of generality to assume **(A3)** because $\Pr^{\max}_{\mathcal{M},s}(\Diamond T) < 1$ implies $\Pr^{\max}_{\mathcal{M},s}(\varphi_K) < 1$ for all $K \in \mathbb{Z}$ and therefore $K^{\exists,=1}_{\mathcal{M},s} = -\infty$. Transforming $\mathcal{M}$ into the largest sub-MDP $\widetilde{\mathcal{M}}$ where the state space is $\widetilde{S} = \{s \in S : \Pr^{\max}_{\mathcal{M},s}(\Diamond T) = 1\}$ we get

$$\forall K \in \mathbb{Z},\ \forall s \in \widetilde{S},\ \Pr^{\max}_{\mathcal{M},s}(\varphi_K) = 1 \Longleftrightarrow \Pr^{\max}_{\widetilde{\mathcal{M}},s}(\varphi_K) = 1 \quad .$$

Therefore, for all states $s \in \widetilde{S}$, $K^{\exists,=1}_{\mathcal{M},s} = K^{\exists,=1}_{\widetilde{\mathcal{M}},s}$. □

In the rest of this section, we hence assume $\varphi = \Diamond \mathit{good} \vee \Diamond(\mathit{goal} \wedge \mathrm{wgt} \geqslant K)$ for trap states *good* and *goal*.

Case of MDPs Without Positively Weight-Divergent End Components

We first address the special case where $\mathcal{M}$ has no positively weight-divergent end component. Thanks to the spider construction from Section 3.1, we may assume that $\mathcal{M}$ has no 0-EC. As a consequence, for all end components $\mathcal{E}$ of $\mathcal{M}$, $\mathbb{E}^{\max}_{\mathcal{E}}(\mathrm{MP}) \leqslant 0$. Let us prove that MD-schedulers are sufficient for $\mathsf{DWR}^{\exists,=1}$, assuming $\mathcal{M}$ has no positively weight-divergent end component.

Lemma D.19 (MD-scheduler suffice if no weight-divergent EC)**.** *Let $\mathcal{M}$ be an MDP such that $\mathcal{M}$ has no positively weight-divergent end component. Let $\varphi = \Diamond \mathit{good} \vee \Diamond(\mathit{goal} \wedge \mathit{wgt} \geqslant K)$ where good and goal are traps. If there exists a scheduler $\mathfrak{S}$ with $\Pr^{\mathfrak{S}}_{\mathcal{M},s_{init}}(\varphi) = 1$, then there exists an MD-scheduler $\mathfrak{T}$ with $\Pr^{\mathfrak{T}}_{\mathcal{M},s_{init}}(\varphi) = 1$.*

Proof. Suppose we are given a scheduler $\mathfrak{S}$ with $\Pr^{\mathfrak{S}}_{\mathcal{M},s_{init}}(\varphi) = 1$. First, we consider the case where $\varphi = \Diamond(\mathit{goal} \wedge \mathrm{wgt} \geqslant K)$, *i.e.*, $T^* = \varnothing$. Later we will explain how to adapt the proof for the case where $T = \{\mathit{goal}, \mathit{good}\}$ with $K_{\mathit{good}} = -\infty$.

Let $\sim$ denote the following equivalence relation on the state space of $\mathcal{M}$:

$$s \sim t \qquad \text{iff} \qquad s \text{ and } t \text{ belong to the same maximal end component of } \mathcal{M} \quad .$$

A state-action pair (s,α) is called a *progress move* if $\alpha \in \mathit{Act}(s)$ and $P(s,\alpha,t) > 0$ for at least one state t with $s \nsim t$. Note that if state s does not belong to any end component then all state-action pairs (s,α) with $\alpha \in \mathit{Act}(s)$ are progress moves. Moreover, if (s,α) is a progress move then there is at least one state t such that $\Pr^{\max}_{\mathcal{M},t}(\Diamond s) < 1$.

Let X denote the set of state-weight pairs $(s,w) \in (S \setminus \{\mathit{goal}\}) \times \mathbb{Z}$ such that there is some $\mathfrak{S}$-path π from s_{init} to s with $\mathit{wgt}(\pi) = w$ and $s \notin T$. Let $X(s) = \{w \in \mathbb{Z} : (s,w) \in X\}$.

Claim 1: If $X(s) \neq \varnothing$ then $\min X(s)$ exists.

To prove the Claim, suppose by contradiction that $\inf X(s) = -\infty$. By assumption, for each $R \in \mathbb{N}$ there exists an $\mathfrak{S}$-path π_R from s_{init} to s with $\mathit{wgt}(\pi_R) \leqslant -R$. Let $\mathcal{N}$ be the MDP that extends $\mathcal{M}$ by a trap state goal' and the state-action pair (goal,τ) with $\mathit{wgt}(\mathit{goal},\tau) = -K$ and $P(\mathit{goal},\tau,\mathit{goal}') = 1$. Obviously, the residual scheduler $\mathfrak{S}{\restriction}\pi_R$ can be extended to a scheduler $\mathfrak{B}_R$ for $\mathcal{N}$ such that

$$\Pr^{\mathfrak{B}_R}_{\mathcal{N},s}\big(\Diamond(\mathit{goal}' \wedge (\mathrm{wgt} \geqslant R))\big) \quad = \quad 1 \quad .$$

This holds for every $R \in \mathbb{N}$, thus $\mathcal{N}$ has a positively weight-divergent end component $\mathcal{E}$; a contradiction. This completes the proof of Claim 1.

Let $U = \{s \in S : X(s) \neq \varnothing\}$. For $s \in U$ we define $w_s = \min X(s) \in \mathbb{Z}$. Let $A(s)$ denote the set of actions $\alpha \in \mathit{Act}(s)$ such that $\mathfrak{S}(\pi) = \alpha$ for at least one $\mathfrak{S}$-path π from s_{init} to s with $\mathit{wgt}(\pi) = w_s$.

Let now Y_0 be the set of all states $s \in U$ such that $A(s)$ contains at least one action α_s where (s,α_s) is a progress move. We then inductively define Y_{i+1} as the set of states $s \in U \setminus (Y_0 \cup \ldots \cup Y_i)$ where $A(s)$ contains at least one action α_s such that $P(s,\alpha_s,t) > 0$ for at least one state $t \in Y_i$. Clearly, there is some j such that Y_{j+1} is empty. Let $Y \quad = \quad Y_0 \cup Y_1 \cup \ldots \cup Y_j$. Then, $Y \subseteq U$.

Claim 2: $Y = U$.

To prove this claim, we again suppose by contradiction that Y is a proper subset of U. Consider the sub-MDP $\widetilde{\mathcal{M}}$ of $\mathcal{M}$ induced by the state-action pairs (s,α) with $s \in U \setminus Y$ and $\alpha \in A(s)$. Note that $s \in U \setminus Y$ and $\alpha \in A(s)$ implies $t \in U \setminus Y$ for all

states t with $P(s,\alpha,t) > 0$. The residual schedulers $\mathfrak{S}{\restriction}\pi$ for the paths π from s_{init} to s with $wgt(\pi) = w_s$ and $s \in U \setminus Y$ can be viewed as schedulers for $\widetilde{\mathcal{M}}$. Hence:

$$\Pr^{\mathfrak{S}{\restriction}\pi}_{\widetilde{\mathcal{M}},s}\big(\,\Box(U \setminus Y)\,\big) \;=\; 1 \;.$$

As $goal \notin U$ (recall that U consists of non-trap states, whereas $goal$ is a trap) we get $\Pr^{\mathfrak{S}}_{\mathcal{M},s_{init}}(\varphi) < 1$, a contradiction, so that $U = Y$.

To conclude, for the states $s \in U$ we can now pick some action $\alpha_s \in A(s)$ such that either $s \in Y_0$ and (s,α_s) is a progress move or $s \in Y_{i+1}$ and $P(s,\alpha_s,t) > 0$ for some state $t \in Y_i$. Let $\mathfrak{T}$ be a memoryless scheduler such that $\mathfrak{T}(s) = \alpha_s$ for $s \in U$ (and defined arbitrary from other states). Then, $\Pr^{\mathfrak{T}}_{\mathcal{M},s_{init}}(\varphi) = 1$ and hence, for the case $T^* = \emptyset$ we are done with the proof of Lemma D.19.

Suppose now that $T^* = \{good\}$, and $\varphi = \Diamond good \vee \Diamond(goal \wedge \mathsf{wgt} \geq K)$. We pick an MD-scheduler $\mathfrak{U}$ such that $\Pr^{\mathfrak{U}}_{\mathcal{M},s}(\Diamond good) = p_s$ where for each state s, $p_s = \Pr^{\max}_{\mathcal{M},s}(\Diamond good)$. We may assume w.l.o.g. that $\mathfrak{S}$ behaves as $\mathfrak{U}$ whenever a state s with $p_s = 1$ has been reached.

The definition of the sets X, $X(s)$ is as before. In Claim 1, we show that $\min X(s)$ exists for all states s where $p_s < 1$. The argument is again by contradiction using paths π_R as above. As $\Pr^{\mathfrak{S}{\restriction}\pi_R}_{\mathcal{M},s}(\Diamond good) \leqslant p_s < 1$ we have:

$$\Pr^{\mathfrak{S}{\restriction}\pi_R}_{\mathcal{M},s}\Big(\bigvee_{t\in T\setminus T^*}(\Diamond(t \wedge (\mathsf{wgt} \geqslant K_t)))\Big) \;\geqslant\; 1 - p_s \;.$$

The MDP $\mathcal{N} = \mathcal{N}_s$ is defined as above, with a τ-transition from $goal$ to the fresh state $goal'$, while $good$ has a τ-transition to s, with weight small enough to ensure that $\mathcal{N}$ has no positively weight-divergent end component.

> Let us briefly explain how to find the value $wgt(good,\tau) \in \mathbb{Z}$ so that the constructed MDP $\mathcal{N}$ has no positively weight-divergent end component. Suppose $\mathcal{N}$ has an end component that contains $good$. Let $\mathcal{E}$ be the maximal end component of $\mathcal{N}$ that contains $good$, and view it as an MDP. Obviously, we have $\Pr^{\max}_{\mathcal{E},s}(\Diamond good) = 1$ and $\mathcal{E}$ is a sub-MDP of $\mathcal{M}$. (Notice that none of the new state-action pairs (t,τ) for $t \in \{goal, good\}$ is contained in $\mathcal{E}$.) As $\mathcal{M}$ has no positively weight-divergent end component, so does $\mathcal{E}$. Thus, the maximal expected accumulated weight until reaching $good$ is finite. Let $E = \mathbb{E}^{\max}_{\mathcal{E},s}(\text{“}wgt \text{ until } good\text{”})$. We then may define $wgt(t,\tau)$ as any value that is smaller than $-E$. This ensures that $\mathbb{E}^{\mathfrak{U}}_{\mathcal{E},s}(\mathrm{MP}) < 0$ for each MD-scheduler $\mathfrak{U}$ for $\mathcal{E}$ with a single BSCC $\mathcal{B}$ and $(good,\tau) \in \mathcal{B}$ (Lemma B.3). But then $\mathbb{E}^{\max}_{\mathcal{E}}(\mathrm{MP}) < 0$. Hence, $\mathcal{E}$ is not positively weight-divergent (Corollary 3.4).

Scheduler $\mathfrak{V}_R$ for $\mathcal{N}$ is now defined as follows. Starting in state s, $\mathfrak{V}_R$ first behaves as $\mathfrak{S}{\restriction}\pi_R$ until reaching a state $t \in \{good, goal\}$.

- If $t = goal$ then $\mathfrak{V}_R$ schedules τ and moves to $goal'$.
- If $t = good$ then $\mathfrak{V}_R$ takes the τ-transition back to state s. If R' is the weight of the (complete) path that $\mathfrak{V}_R$ has generated then $\mathfrak{V}_R$ behaves now as $\mathfrak{S}{\restriction}\pi_{R'}$ until reaching again a state $t \in \{good, goal\}$.

As the probablities to generate a path from s to $good$ under all residual schedulers $\mathfrak{S}{\restriction}\pi_R$ is at least $1-p_s$ we obtain:

$$\Pr^{\mathfrak{V}_R}_{\mathcal{M},s}\big(\,\Diamond(goal' \wedge (\mathsf{wgt} \geqslant R))\,\big) \;=\; 1 \;.$$

This is impossible as $\mathcal{N}$ has no positively weight-divergent end component. The remaining argument and proof of Claim 2 is the same as for the case $T^* = \varnothing$. This concludes the proof of Lemma D.19 in the general case. □

As a consequence of Lemma D.19, one can decide in NP the problem $\mathsf{DWR}^{\exists,=1}$ by guessing an MD-scheduler $\mathfrak{S}$ and checking that it ensures $\Pr^{\mathfrak{S}}(\Diamond good \vee \Diamond(goal \wedge \mathsf{wgt} \geqslant K)) = 1$. This would also yield an EXPTIME-algorithm to compute the values $K^{\exists,=1}_{\mathcal{M},s}$. We now give a better alternative in terms of complexity, both to answer the decision problem $\mathsf{DWR}^{\exists,=1}$ and to compute the values $K^{\exists,=1}_{\mathcal{M},s}$, by using a reduction to mean-payoff games.

Theorem D.20. *If $\mathcal{M}$ has no weight-divergent end components then the problem $\mathsf{DWR}^{\exists,=1}$ is in NP ∩ coNP. The values $K^{\exists,=1}_{\mathcal{M},s}$ can be computed in pseudo-polynomial time.*

Proof. The complexity upper bound for the decision problem is established through a polynomial-time reduction to mean-payoff games, illustrated on Figure 14.

From $\mathcal{M}$, we build a two-player mean-payoff game $\mathcal{G}$, in which player 1 simulates the choices of the scheduler, player 2 is responsible for the probabilistic choices, and the weight of a transition by player 1 coincides with the weight in $\mathcal{M}$. Moreover, $\mathcal{G}$ is extended with two transitions: a self-loop on $good$ with weight 0, and an edge from $goal$ to s_{init} with weight $-K$.

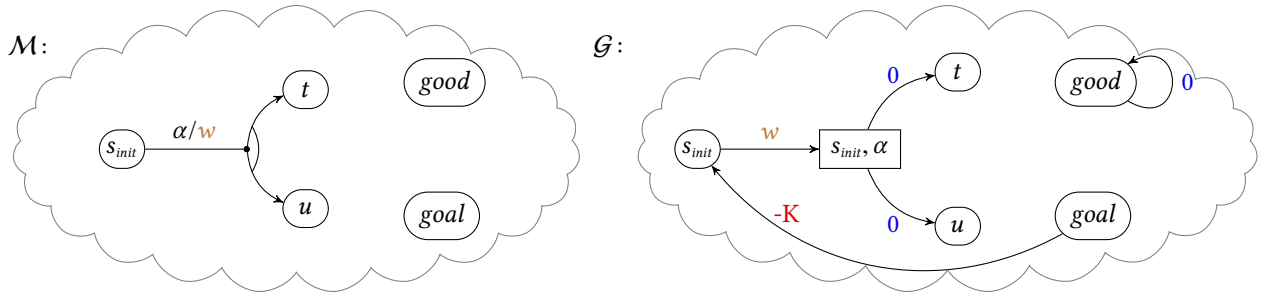


Figure 14. Reduction of $\mathrm{DWR}^{\exists,=1}$ to mean-payoff games, assuming no weight-divergent EC.

The construction ensures

$$\exists \mathfrak{S},\ \Pr^{\mathfrak{S}}_{\mathcal{M},s_{init}}(\Diamond good \vee \Diamond goal \wedge \mathrm{wgt} \geqslant K) = 1 \iff \exists \sigma,\ \forall \tau,\ \mathrm{Play}_{\mathcal{G}}(\sigma,\tau) \models \mathrm{MP} \geq 0\ .$$

($\Longrightarrow$) To prove the left-to-right direction, we pick a scheduler $\mathfrak{S}$ that satisfies $\Pr^{\mathfrak{S}}_{\mathcal{M},s_{init}}(\Diamond good \vee \Diamond goal \wedge \mathrm{wgt} \geqslant K) = 1$. By Lemma D.19, we may assume that $\mathfrak{S}$ is an MD-scheduler. It trivially induces an MD-strategy σ in $\mathcal{G}$, which mimics $\mathfrak{S}$ in all states except *goal*, and moves back to s_{init} from *goal*.
Thanks to Lemma D.17, along all paths from s_{init} to *goal* in the Markov chain C induced by $\mathfrak{S}$, no state belongs to a negative cycle of C. From the hypothesis that for all $s \neq good$, $\Pr^{\max}_{\mathcal{M},s}(\Diamond good) < 1$, we derive that all states (except *good*) have a path to *goal* in C. Therefore, C has no negative cycle on the way to *goal*. Moreover, the weight $-K$ of the transition from *goal* to s_{init} was chosen so that all cycles around s_{init} have nonnegative weight. Thus, for all strategy τ of player 2, the play induced by σ and τ has nonnegative mean payoff.

($\Longleftarrow$) For the other direction, we let σ be an MD-strategy winning for player 1 in $\mathcal{G}$. Here it also trivially defines a scheduler $\mathfrak{S}$ for $\mathcal{M}$. We first show that $\mathfrak{S}$ ensures $\Pr^{\mathfrak{S}}_{\mathcal{M},s_{init}}(\Diamond good \vee \Diamond goal) = 1$, by contradiction. Then, there exists an EC $\mathcal{E}$ such that $\Pr^{\mathfrak{S}}_{\mathcal{M},s_{init}}(\Diamond\Box\mathcal{E}) > 0$. Since all EC of $\mathcal{M}$ have negative expected mean payoff, this holds in particular for $\mathcal{E}$, and thus $\mathfrak{S}$ induces at least a path with negative mean-payoff. Now, assume that the weight constraint in *goal* is not met, *i.e.*, there is an $\mathfrak{S}$-path reaching *goal* with accumulated weight at most $K-1$. Iterations of this path followed by the transition from *goal* to s_{init} yields a play in $\mathcal{G}$ under σ which had negative mean-payoff, a contradiction with the fact that σ is winning. Thus, all $\mathfrak{S}$-paths to *goal* reach it with accumulated weight at least K. All in all, $\mathfrak{S}$ ensures that φ holds almost surely.

For the computation of the values, $K^{\exists,=1}_{\mathcal{M},s} = \infty$ if $\Pr^{\max}_{\mathcal{M},s}(\Diamond good) = 1$, and $K^{\exists,=1}_{\mathcal{M},s} = -\infty$ if $\Pr^{\max}_{\mathcal{M},s}(\Diamond(good \vee goal)) < 1$. Otherwise, for each state $s \in S$ we have that $K^{\exists,=1}_{\mathcal{M},s} \in \{-\infty\} \cup [|S| \cdot W_{\min}, |S| \cdot W_{\max}]$, where $W_{\min}$ is the minimal weight and $W_{\max}$ is the maximal weight that appears in $\mathcal{M}$. In fact, MD-schedulers suffice to achieve a given threshold K, and an MD-scheduler $\mathfrak{S}$ satisfying $\Pr^{\mathfrak{S}}_{\mathcal{M},s}\big(\Diamond good \vee \Diamond(goal \wedge \mathrm{wgt} \geqslant K)\big) = 1$ cannot induce any negative cycle in $\mathcal{M}$. Thus, any threshold K that can be ensured by $\mathfrak{S}$ must be equal to or greater than the weight of a simple shortest path from s to *goal*, which is at least $|S| \cdot W_{\min}$. Conversely, under any MD-scheduler $\mathfrak{S}$, *goal* (which is a trap state) is reachable following a simple path so a threshold K with $\Pr^{\mathfrak{S}}_{\mathcal{M},s}\big(\Diamond good \vee \Diamond(goal \wedge \mathrm{wgt} \geqslant K)\big) = 1$ cannot be larger than $|S| \cdot W_{\max}$. To compute the values, we can thus run a binary search in this interval with $\log(|S| \cdot W)$ calls to a pseudo-polynomial mean-payoff solver, where $W = W_{\max} - W_{\min}$. The binary search either determines a finite value, or it returns that the value must be less than $|S| \cdot W_{\min}$ in which case the value is $-\infty$. □

To prepare the proof of the general case (when $\mathcal{M}$ may have weight-divergent end components), we provide a characterization of the different cases that arise for $K^{\exists,=1}_{\mathcal{M},s}$: whether the value is $-\infty$, finite, or $+\infty$. To do so, we introduce a weighted directed graph associated with $\mathcal{M}$ and an MD-scheduler $\mathfrak{S}$. Let $G^{\mathfrak{S}}_s$ denote the weighted directed graph where the vertex set consists of all states u in $\mathcal{M}$ that belong to $\mathfrak{S}$-path from s to *goal*. The edge relation in $G^{\mathfrak{S}}_s$ is given by $u \to u'$ if $P_{\mathcal{M}}(u, \mathfrak{S}(u), u') > 0$. The weight of the edge $u \to u'$ is $wgt_{\mathcal{M}}(u, \mathfrak{S}(u))$.

Lemma D.21. *Let $\mathcal{M}$ be an MDP with no positively weight-divergent end components that satisfies assumptions **(A1)**, **(A2)** and **(A3)** and let s be a state of $\mathcal{M}$. Then:*

(a) $K^{\exists,=1}_{\mathcal{M},s} = +\infty$ iff $s = good$

(b) $K^{\exists,=1}_{\mathcal{M},s} \in \mathbb{Z}$ iff $s \neq good$ and there exists an MD-scheduler $\mathfrak{S}$ such that the graph $G^{\mathfrak{S}}_s$ does not contain any negative cycle.

(c) $K^{\exists,=1}_{\mathcal{M},s} = -\infty$ iff $s \neq good$ and there is no MD-scheduler $\mathfrak{S}$ satisfying the condition stated in (b).

Proof. Statement (a) is trivial. Statements (b) and (c) are easy consequences of Lemma D.19 and Lemma D.17. □

General Case

We now consider the general case where $\mathcal{M}$ might have positively weight-divergent end components, still assuming **(A1)**, **(A2)** and **(A3)** to hold.

We first observe that states that belong to the same positively weight-divergent end component have the same truth values for $\mathsf{DWR}^{\exists,=1}$ and the same values for the corresponding optimization problem. More precisely:

Lemma D.22 (Same values for states in weight-divergent ECs)**.** *Let $\mathcal{E}$ be a positively weight-divergent end component of $\mathcal{M}$. Then:*

(a) For all states $s, s' \in \mathcal{E}$, $\Pr^{\max}_{\mathcal{M},s}(\varphi) = 1$ iff $\Pr^{\max}_{\mathcal{M},s'}(\varphi) = 1$ and $\Pr^{\max}_{\mathcal{M},s}(\varphi) > 0$ iff $\Pr^{\max}_{\mathcal{M},s'}(\varphi) > 0$.

(b) There exists $K^{\exists,=1}_{\mathcal{M},\mathcal{E}} \in \{+\infty, -\infty\}$ such that for all states $s \in \mathcal{E}$ we have $K^{\exists,=1}_{\mathcal{M},s} = K^{\exists,=1}_{\mathcal{M},\mathcal{E}}$.

(c) If $K^{\exists,=1}_{\mathcal{M},\mathcal{E}} = +\infty$ and s is a state with $\Pr^{\max}_{\mathcal{M},s}(\Diamond\mathcal{E}) = 1$, then $K^{\exists,=1}_{\mathcal{M},s} = +\infty$.

Proof. Given two states $s, s' \in \mathcal{E}$ and a natural number $R \in \mathbb{N}$, since $\mathcal{E}$ is positively weight-divergent there exists a scheduler $\mathfrak{B}_R$ for $\mathcal{E}$ such that $\Pr^{\mathfrak{B}_R}_{\mathcal{E},s}\big(\Diamond(s' \wedge (\mathrm{wgt} \geqslant R))\big) = 1$. Hence, for each scheduler $\mathfrak{S}$ for $\mathcal{M}$ and each $K \in \mathbb{Z}$ there is a scheduler $\mathfrak{S}_R$ for $\mathcal{M}$ such that $\Pr^{\mathfrak{S}}_{\mathcal{M},s}(\varphi_K) = \Pr^{\mathfrak{S}_R}_{\mathcal{M},s'}(\varphi_{K+R})$. With $R = 0$, we obtain statement (a) and $K^{\exists,=1}_{\mathcal{M},s} = K^{\exists,=1}_{\mathcal{M},s'}$. With $s = s'$ and letting R tend to ∞, we obtain that the value $K^{\exists,=1}_{\mathcal{M},s}$ cannot be finite. Thus, $K^{\exists,=1}_{\mathcal{M},s}, K^{\exists,>0}_{\mathcal{M},s} \in \{\pm\infty\}$. This yields statement (b).

To prove statement (c), we pick a family of schedulers $(\mathfrak{S}_K)_{K\in\mathbb{Z}}$ such that $\Pr^{\mathfrak{S}_K}_{\mathcal{M},u}(\varphi_K) = 1$ for all states u in $\mathcal{E}$ and all integers K. Let now s be a state such that $\Pr^{\mathfrak{S}}_{\mathcal{M},s}(\Diamond\mathcal{E}) = 1$. Given $R \in \mathbb{Z}$, we now compose $\mathfrak{S}$ and the schedulers $\mathfrak{S}_K$ to obtain a scheduler $\mathfrak{T}_R$ with $\Pr^{\mathfrak{T}_R}_{\mathcal{M},s}(\varphi_R) = 1$. $\mathfrak{T}_R$ first behaves as $\mathfrak{S}$ until reaching a state in $\mathcal{E}$. As soon as some state u of $\mathcal{E}$ is reached, say along a path π from s to u with $wgt(\pi) = w$, $\mathfrak{T}$ switches mode and behaves as $\mathfrak{S}_{R-w}$ from then on. □

For solving the general case, let us now explain how to use the particular case of MDPs without positively weight-divergent end components. From $\mathcal{M}$ we construct another $\mathcal{N}$ by intuitively replacing each maximal weight-divergent end component $\mathcal{E}$ with an entry state $\mathcal{E}_{in}$ and an exit state $\mathcal{E}_{out}$ and a transition from $\mathcal{E}_{in}$ to $\mathcal{E}_{out}$ with "sufficiently high" weight. $\mathcal{N}$ has no positively weight-divergent end components, however the values $K^{\exists,=1}_{\mathcal{N},s}$ are only a lower bound on $K^{\exists,=1}_{\mathcal{M},s}$. In particular, it may happen that $K^{\exists,=1}_{\mathcal{N},s} = -\infty$ and $K^{\exists,=1}_{\mathcal{M},s} = +\infty$. To remedy this problem, we will see how to identify end components with value $+\infty$. This is performed by a fixed-point computation of the "good" end components. All details of these steps are provided in the remainder of this section.

To formally define $\mathcal{N}$ we introduce some notations. Let $\mathcal{E}_1, \ldots, \mathcal{E}_k$ be the maximal end components of $\mathcal{M}$ that are positively weight-divergent. Recall that these can be computed in polynomial time by first computing the maximal end components of $\mathcal{M}$ and then checking whether each of them is weight-divergent thanks to Theorem 3.9 from Section 3.2.

Let *WDMEC* consist of all states in $\mathcal{M}$ that are contained in one of the weight-divergent maximal end components $\mathcal{E}_1, \ldots, \mathcal{E}_k$. Since *good* and *goal* are traps, the maximal end components of $\mathcal{M}$ do not contain them, and $\{good, goal\} \cap WDMEC = \varnothing$. For simplicity, we assume that the action sets $Act_{\mathcal{M}}(s)$ are pairwise disjoint, and we write $wgt^{\min}_{\mathcal{M}} \in \mathbb{Z}$ for the minimal weight assigned to some state-action pair in $\mathcal{M}$. We have now all ingredients to precisely define $\mathcal{N}$.

state space $S_{\mathcal{N}} = \big(S_{\mathcal{M}} \setminus WDMEC\big) \cup \big\{\, \mathcal{E}_{in}, \mathcal{E}_{out} \;:\; \mathcal{E} \in \{\mathcal{E}_1, \ldots, \mathcal{E}_k\} \,\big\}$;
action set $Act_{\mathcal{N}} = Act \cup \{\tau\}$, with $\tau \notin Act$ a fresh action symbol;
transitions

- If (s, α) is a state-action pair in $\mathcal{M}$ where $s \in S_{\mathcal{M}} \setminus WDMEC$, then (s, α) is a state-action pair of $\mathcal{N}$ with the same weight; moreover for each state $s' \in S_{\mathcal{M}} \setminus WDMEC$, $P_{\mathcal{N}}(s, \alpha, s') = P_{\mathcal{M}}(s, \alpha, s')$ and for each end component $\mathcal{E} \in \{\mathcal{E}_1, \ldots, \mathcal{E}_k\}$ we define $P_{\mathcal{N}}(s, \alpha, \mathcal{E}_{in}) = P_{\mathcal{M}}(s, \alpha, \mathcal{E})$.[4]
- Each state-action pair (s, α) in $\mathcal{M}$ where $s \in \mathcal{E} \subseteq WDMEC$and such that for some state s' outside $\mathcal{E}$, $P_{\mathcal{M}}(s, \alpha, s') > 0$ is turned into a state-action pair $(\mathcal{E}_{out}, \alpha)$ in $\mathcal{N}$; the weight of the state-action pair $(\mathcal{E}, \alpha)$ in $\mathcal{N}$ coincides with the weight of (s, α) in $\mathcal{M}$; moreover the transition probabilities are defined as follows: for each state $s' \in S_{\mathcal{M}} \setminus WDMEC$ we set $P_{\mathcal{N}}(\mathcal{E}_{out}, \alpha, s') = \frac{P_{\mathcal{M}}(s,\alpha,s')}{1-P_{\mathcal{M}}(s,\alpha,\mathcal{E})}$, for each maximal weight-divergent end component $\mathcal{F} \neq \mathcal{E}$ we set $P_{\mathcal{N}}(\mathcal{E}_{out}, \alpha, \mathcal{F}_{in}) = \sum_{s'\in\mathcal{F}} \frac{P_{\mathcal{M}}(s,\alpha,s')}{1-P_{\mathcal{M}}(s,\alpha,\mathcal{E})}$, and $P_{\mathcal{N}}(\mathcal{E}, \alpha, \mathcal{E}) = 0$.
- Last, for each maximal weight-divergent end component $\mathcal{E}$ in $\mathcal{M}$, $\mathcal{N}$ contains a state-action pair $(\mathcal{E}_{in}, \tau)$ for some fresh action symbol τ with $P_{\mathcal{N}}(\mathcal{E}_{in}, \tau, \mathcal{E}_{out}) = 1$ and $wgt_{\mathcal{N}}(\mathcal{E}_{in}, \tau) = \omega$, with $\omega = \max\{\, 0, \; -(|S_{\mathcal{N}}|-1) \cdot wgt^{\min}_{\mathcal{M}} \,\}$.

[4]The notation $P_{\mathcal{M}}(s, \alpha, \mathcal{E})$ stands for the probability from s to reach any state of $\mathcal{E}$.

An example of transformation from $\mathcal{M}$ to $\mathcal{N}$ is provided by Example 5.4 in the core of the paper.

The representation of all states belonging to the same maximal weight-divergent end component $\mathcal{E}$ by the states $\mathcal{E}_{in}$ and $\mathcal{E}_{out}$ is motivated by part (a) of Lemma D.22, which expresses that states in the same WDMEC have the same truth value for $\mathrm{DWR}^{\exists,=1}$. Intuitively, the τ-transition from $\mathcal{E}_{in}$ to $\mathcal{E}_{out}$ serves to mimic all state-action pairs (s,α) with $s \in \mathcal{E}$ and $P_{\mathcal{M}}(s,\alpha,\mathcal{E}) = 1$.

Lemma D.23 (Simple properties of the new MDP)**.** *$\mathcal{N}$ satisfies assumptions **(A1)**, **(A2)** and **(A3)**. Moreover:*

(a) None of the states $\mathcal{E}_{in}$, $\mathcal{E}_{out}$ for $\mathcal{E} \in \{\mathcal{E}_1,\ldots,\mathcal{E}_k\}$ belong to an end component of $\mathcal{N}$.

(b) $\mathcal{N}$ has no positively weight-divergent end component.

(c) None of the states $\mathcal{E}_{in}$, $\mathcal{E}_{out}$ for $\mathcal{E} \in \{\mathcal{E}_1,\ldots,\mathcal{E}_k\}$ belongs to a simple negative cycle in $\mathcal{N}$.

Proof. Given that $\mathcal{M}$ satisfies assumptions **(A1)**, **(A2)** and **(A3)**, it is easy to see that $\mathcal{N}$ satisfies these assumptions, too.

Statements (a) and (b) follow from the fact that each end component $\mathcal{E}$ of $\mathcal{N}$ is an end component of $\mathcal{M}$ that is not positively weight-divergent. Under the assumption that $wgt^{\min}_{\mathcal{M}} \geqslant 0$, statement (c) is trivial, since in this case all weights in $\mathcal{M}$ are nonnegative, and also in $\mathcal{N}$ because all τ transitions then have weight $\omega = 0$.

Assume now that $wgt^{\min}_{\mathcal{M}} < 0$, in which case the weight of τ-transitions satisfies $\omega = -(|S_{\mathcal{N}}|-1)\cdot wgt^{\min}_{\mathcal{M}} < 0$. Let ξ be a simple cycle in $\mathcal{N}$ that contains one of the states $\mathcal{E}_{in}$ or $\mathcal{E}_{out}$ for some $\mathcal{E} \in \{\mathcal{E}_1,\ldots,\mathcal{E}_k\}$. Necessarily, ξ then also contains the τ-transition from $\mathcal{E}_{in}$ to $\mathcal{E}_{out}$ (otherwise ξ could not enter $\mathcal{E}_{out}$ and could not leave $\mathcal{E}_{in}$).

Regarding ξ as a sequence of state-action pairs, we let ρ denote the sequence $(s_1,\alpha_1)\ldots(s_m,\alpha_m)$ of state-action pairs in $\mathcal{N}$ that results from ξ by removing the state-action pairs $(\mathcal{F}_{in},\tau)$ for $\mathcal{F} \in \{\mathcal{E}_1,\ldots,\mathcal{E}_k\}$. In particular, $\{s_1,\ldots,s_m\} \subseteq S_{\mathcal{M}} \setminus \mathit{WDMEC}$ and all state-action pairs (s_i,α_i) in ρ belong to $\mathcal{M}$. Since ξ is a simple cycle, we derive $m \leqslant |S_{\mathcal{N}}|-2$. As a consequence, we get

$$wgt_{\mathcal{N}}(\xi) \geqslant \sum_{i=1}^{m} wgt_{\mathcal{M}}(s_i,\alpha_i) + wgt_{\mathcal{N}}(\mathcal{E}_{in},\tau) \geqslant m\cdot wgt^{\min}_{\mathcal{M}} + wgt_{\mathcal{N}}(\mathcal{E}_{in},\tau) \geqslant (N-2)\cdot wgt^{\min}_{\mathcal{M}} - (N-1)\cdot wgt^{\min}_{\mathcal{M}} = -wgt^{\min}_{\mathcal{M}} > 0 \ .$$

Thus ξ is not a negative cycle, and the proof of statement (c) is complete. □

To establish a relation between values in $\mathcal{M}$ and in $\mathcal{N}$, we first assign states of $\mathcal{M}$ a corresponding state in $\mathcal{N}$. For $s \in \mathcal{M}$, $s_{\mathcal{N}}$ is defined as s if $s \in S_{\mathcal{M}} \setminus \mathit{WDMEC}$, and otherwise $s = \mathcal{E}_{out}$ if s belongs to the MEC $\mathcal{E}$ that is positively weight-divergent. Furthermore, we define $\mathit{WDMEC}_{\mathcal{N}} = \{\mathcal{E}_{in},\mathcal{E}_{out} : \mathcal{E} \in \{\mathcal{E}_1,\ldots,\mathcal{E}_k\}\}$, as the set of entry and exit states in $\mathcal{N}$.

The values for states in $\mathcal{N}$ is defined analogously to values in $\mathcal{M}$: for $u \in \mathcal{N}$, $K^{\exists,=1}_{\mathcal{N},u}$ is the supremum over all integers K such that $\Pr^{\max}_{\mathcal{N},u}(\varphi_K) = 1$. On the one hand recall from Lemma D.22 that for each $\mathcal{E} \in \{\mathcal{E}_1,\ldots,\mathcal{E}_k\}$ all states in $\mathcal{E}$ share the same value $K^{\exists,=1}_{\mathcal{M},\mathcal{E}} \in \{+\infty,-\infty\}$. On the other hand, as $\mathcal{N}$ satisfies assumptions **(A1)**, **(A2)** and **(A3)**, has no positively weight-divergent end components, and (see Lemma D.23) we can apply Lemma D.21, we obtain that $K^{\exists,=1}_{\mathcal{N},\mathcal{E}_{out}} \in \{-\infty\} \cup \mathbb{Z}$. Thus, we cannot expect that $K^{\exists,=1}_{\mathcal{M},s}$ and $K^{\exists,=1}_{\mathcal{N},s_{\mathcal{N}}}$ agree. Nevertheless, we will prove that the values $K^{\exists,=1}_{\mathcal{M},s}$ can be derived from the values $K^{\exists,=1}_{\mathcal{N},s_{\mathcal{N}}}$.

We start with the following observation that relates the values in $\mathcal{M}$ and $\mathcal{N}$ where we switch to a different objective for $\mathcal{N}$.

Lemma D.24. *Let s and $K \in \mathbb{Z}$ be such that $\Pr^{\max}_{\mathcal{M},s}(\varphi_K) = 1$. Then, for $T^*_{\mathcal{N}} = T^* \cup \{\mathcal{E}_{out} : K^{\exists,=1}_{\mathcal{M},\mathcal{E}} = +\infty\}$ it holds*

$$\Pr^{\max}_{\mathcal{N},s_{\mathcal{N}}}\big(\Diamond T^*_{\mathcal{N}} \vee \Diamond(goal \wedge (wgt \geqslant K))\big) = 1 \ .$$

Proof. If the given state s of $\mathcal{M}$ belongs to a maximal weight-divergent end component of $\mathcal{M}$, then $s_{\mathcal{N}} \in T^*_{\mathcal{N}}$ and the claim is obvious. Suppose now that $s \notin \mathit{WDMEC}$, in which case s is also a state of $\mathcal{N}$.

We pick an MD-scheduler $\mathfrak{U}$ for $\mathcal{N}$ that maximizes the probabilities to reach T^* from every state in $\mathcal{N}$. Let $V = \{v \in S_{\mathcal{N}} : \Pr^{\max}_{\mathcal{N},v}(\Diamond T^*) = 1\}$. Clearly, we have $\Pr^{\mathfrak{S}}_{\mathcal{N},v}(\Diamond T^*) = 1$ for every state $v \in V$. Consider now any scheduler $\mathfrak{T}$ for $\mathcal{N}$ such that:

- $\mathfrak{T}(\pi) = \mathfrak{S}(\pi)$ for each finite path π in $\mathcal{N}$ with $\mathit{last}(\pi) \notin V$ consisting of states that do not belong to $\mathit{WDMEC}_{\mathcal{N}}$.
- $\mathfrak{T}(\pi) = \mathfrak{U}(\mathit{last}(\pi))$ if $\mathit{last}(\pi) \in V$.

(The behavior of $\mathfrak{T}$ for the paths π that contain a state in $\mathit{WDMEC}_{\mathcal{N}}$ is irrelevant.)

Obviously, $\Pr^{\mathfrak{T}}_{\mathcal{N},s_{\mathcal{N}}}(\Diamond\mathcal{E}_{in}) > 0$ implies $\Pr^{\mathfrak{S}}_{\mathcal{M},s}(\Diamond\mathcal{E}) > 0$. But then there is an $\mathfrak{S}$-path π from s to some state u in $\mathcal{E}$. For the residual scheduler, we have $\Pr^{\mathfrak{S}\uparrow\pi}_{\mathcal{M},u}\big(\varphi_{K-wgt(\pi)}\big) = 1$. In particular, $\Pr^{\max}_{\mathcal{M},u}(\varphi_{K-wgt(\pi)}) = 1$. Therefore, $K^{\exists,=1}_{\mathcal{M},\mathcal{E}} = +\infty$ by part (b) of Lemma D.22. But then $\mathcal{E}_{out} \in T^*_{\mathcal{N}}$. Moreover, each $\mathfrak{T}$-path π from $s_{\mathcal{N}}$ to *goal* that does not enter a state in $\mathit{WDMEC}_{\mathcal{N}}$ is a $\mathfrak{S}$-path from s to *goal*. Hence, $wgt(\pi) \geqslant K$. □

To simplify our notations, we write $K^{\exists,=1}_{\mathcal{N},\mathcal{E}}$ rather than $K^{\exists,=1}_{\mathcal{N},\mathcal{E}_{out}}$ for a given $\mathcal{E} \in \{\mathcal{E}_1,\ldots,\mathcal{E}_k\}$. Recall that by assumption **(A2)** we have $K^{\exists,=1}_{\mathcal{N},\mathcal{E}_{out}} \in \mathbb{Z} \cup \{-\infty\}$. The idea is now to identify the end components $\mathcal{E} \in \{\mathcal{E}_1,\ldots,\mathcal{E}_k\}$ of $\mathcal{M}$ where $K^{\exists,=1}_{\mathcal{M},\mathcal{E}} = +\infty$. For this, we observe:

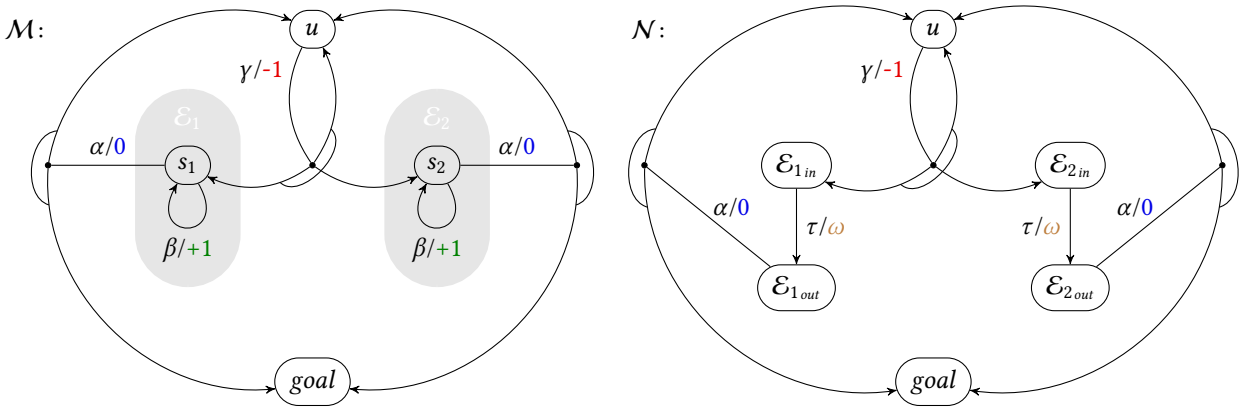


Figure 15. We have $K^{\exists,=1}_{\mathcal{M},s_1} = K^{\exists,=1}_{\mathcal{M},s_1} = +\infty$ while $K^{\exists,=1}_{\mathcal{N},\mathcal{E}_{1\,out}} = K^{\exists,=1}_{\mathcal{N},\mathcal{E}_{2\,out}} = -\infty$.

Lemma D.25. *For each scheduler $\mathfrak{T}$ for $\mathcal{N}$ and each $K \in \mathbb{Z}$, there is a scheduler $\mathfrak{S}$ for $\mathcal{M}$ such that for all states s in $\mathcal{M}$:*

$$\Pr^{\mathfrak{T}}_{\mathcal{N},s_{\mathcal{N}}}(\varphi_K) \;\leqslant\; \Pr^{\mathfrak{S}}_{\mathcal{M},s}(\varphi_K) \;.$$

Proof. The proof is an easy verification. It relies on the fact that any scheduler $\mathfrak{T}$ for $\mathcal{N}$ naturally induces a scheduler $\mathfrak{S}$ for $\mathcal{M}$ that mimics the behavior of $\mathfrak{T}$ and uses a weight-divergent scheduler for the behavior inside the end components $\mathcal{E} \in \{\mathcal{E}_1, \ldots, \mathcal{E}_k\}$ to ensure that the accumulated weight inside $\mathcal{E}$ is at least ω. □

As a consequence, values in $\mathcal{M}$ are at least as large as values in $\mathcal{N}$.

Corollary D.26 (Values in $\mathcal{N}$ are lower bounds for the values in $\mathcal{M}$)**.**

(a) For every state s of $\mathcal{M}$, $K^{\exists,=1}_{\mathcal{N},s} \leqslant K^{\exists,=1}_{\mathcal{M},s}$.

(b) If $\mathcal{E} \in \{\mathcal{E}_1, \ldots, \mathcal{E}_k\}$ and $K^{\exists,=1}_{\mathcal{N},\mathcal{E}_{out}} \in \mathbb{Z}$ then $K^{\exists,=1}_{\mathcal{M},\mathcal{E}} = +\infty$.

Still there can be end components $\mathcal{E} \in \{\mathcal{E}_1, \ldots, \mathcal{E}_k\}$ with $K^{\exists,=1}_{\mathcal{M},\mathcal{E}} = +\infty$ while $K^{\exists,=1}_{\mathcal{N},\mathcal{E}_{out}} = -\infty$. The following example illustrates this phenomenon.

Example D.27. Let $\mathcal{M}$ be the MDP depicted in Figure 15 on the left. Then, $\mathcal{M}$ has two maximal end components $\mathcal{E}_1$, $\mathcal{E}_2$ where $\mathcal{E}_1$ consists of the state-action pair (s_1, β) and $\mathcal{E}_2$ of the state-action pair (s_2, β). State u does not belong to any end component. Hence, $\mathit{WDMEC} = \{s_1, s_2\}$. We have $K^{\exists,=1}_{\mathcal{M},s_1} = K^{\exists,=1}_{\mathcal{M},s_1} = +\infty$. The new MDP $\mathcal{N}$ illustrated by Figure 15 on the right can be seen as a Markov chain. As $\mathit{wgt}_{\mathcal{N}}(u,\gamma) = -1$, the five states $\mathcal{E}_{1\,in}, \mathcal{E}_{1\,out}, \mathcal{E}_{2\,in}, \mathcal{E}_{2\,out}$ and u consistute a strongly connected component of $\mathcal{N}$ that contains a negative cycle. Hence, $\Pr_{\mathcal{N},\mathcal{E}_{1\,out}}(\varphi_K) = \Pr_{\mathcal{N},\mathcal{E}_{2\,out}}(\varphi_K) = 0$ for each K, and therefore $K^{\exists,=1}_{\mathcal{N},\mathcal{E}_{1\,out}} = K^{\exists,=1}_{\mathcal{N},\mathcal{E}_{2\,out}} = -\infty$.

To detect end components $\mathcal{E}$ with with $K^{\exists,=1}_{\mathcal{M},\mathcal{E}} = +\infty$ and $K^{\exists,=1}_{\mathcal{N},\mathcal{E}_{out}} = -\infty$, we introduce the notation of *good states* and *good end components*.

Definition D.28 (Good states and end components)**.** If $X \subseteq \{\mathcal{E}_1, \ldots, \mathcal{E}_k\}$ then we define $X_{in} = \{\mathcal{E}_{in} : \mathcal{E} \in X\}$ and

$$\varphi_K[X] \;=\; \Diamond(T^* \cup X_{in}) \;\vee\; \Diamond(\mathit{goal} \wedge (\mathrm{wgt} \geqslant K)) \;.$$

The set of *good end components GoodEC* is the largest subset X of $\{\mathcal{E}_1, \ldots, \mathcal{E}_k\}$ such that:

$$\forall \mathcal{E} \in X \; \exists K \in \mathbb{Z} \text{ s.t. } \Pr^{\max}_{\mathcal{N},\mathcal{E}_{out}}\big(\varphi_K[X]\big) \;=\; 1 \;. \tag{Good}$$

The set of *good states* is defined as $\mathit{Good} = \bigcup_{\mathcal{E} \in \mathit{GoodEC}} S_{\mathcal{E}}$ where $S_{\mathcal{E}}$ denotes the set of states of end component $\mathcal{E}$.

Remark D.29 (Greatest fixed-point characterization of the *GoodEC*)**.** *By definition, GoodEC is the greatest fixed point of the operator $\Omega\colon 2^{\{\mathcal{E}_1,\ldots,\mathcal{E}_k\}} \to 2^{\{\mathcal{E}_1,\ldots,\mathcal{E}_k\}}$ that maps a given $X \subseteq \{\mathcal{E}_1, \ldots, \mathcal{E}_k\}$ to*

$$\Omega(X) \;=\; \big\{\, \mathcal{E} \in \{\mathcal{E}_1, \ldots, \mathcal{E}_k\} \;:\; \exists K \in \mathbb{Z} \text{ s.t. } \Pr^{\max}_{\mathcal{N},\mathcal{N}_{out}}\big(\varphi_K[X]\big) = 1 \,\big\} \;.$$

The operator Ω is monotonic, thus Tarski's fixed-point theorem ensures the existence of a greatest fixed point that can be obtained as the limit of the sequence $X_0 = \{\mathcal{E}_1, \ldots, \mathcal{E}_k\}$, $X_{i+1} = \Omega(X_i)$ for $i \geqslant 0$.

First we prove that good states have value $+\infty$ in $\mathcal{M}$.

Lemma D.30. *If $s \in Good$ then $K^{\exists,=1}_{\mathcal{M},s} = +\infty$.*

Proof. Obviously, there exists $K \in \mathbb{Z}$ such that $\Pr^{\max}_{\mathcal{N},\mathcal{E}_{out}}\big(\varphi_K[GoodEC]\big) = 1$ for all $\mathcal{E} \in GoodEC$. Using Lemma D.19 one can show that there is an MD-scheduler $\mathfrak{T}$ for $\mathcal{N}$ such that

$$\Pr^{\mathfrak{T}}_{\mathcal{N},\mathcal{E}_{out}}\big(\varphi_K[GoodEC]\big) = 1 \quad \text{for all } \mathcal{E} \in GoodEC \ .$$

Scheduler $\mathfrak{T}$ enjoys the following properties:

(1) $\Pr^{\mathfrak{T}}_{\mathcal{N},\mathcal{E}_{out}}(\Diamond \mathcal{E}_{in}) < 1$ for all $\mathcal{E} \in \{\mathcal{E}_1,\ldots,\mathcal{E}_k\}$.
This follows from the observation that the exit and entry states of the end components $\mathcal{E} \in \{\mathcal{E}_1,\ldots,\mathcal{E}_k\}$ do not belong to any end component of $\mathcal{N}$ (Lemma D.23).

(2) If $\mathcal{E},\mathcal{F} \in \{\mathcal{E}_1,\ldots,\mathcal{E}_k\}$ and $\mathcal{E}$ is good then $\Pr^{\mathfrak{T}}_{\mathcal{N},\mathcal{E}_{out}}((\neg Good_{\mathcal{N}})\,\mathbf{U}\,\mathcal{F}_{in}) > 0$ implies $\mathcal{F}$ is good.
Suppose by contradiction that $\mathcal{F}$ is not good. Let $X = GoodEC \cup \{\mathcal{F}\}$. We pick a $\mathfrak{T}$-path π from $\mathcal{E}_{out}$ to $\mathcal{F}_{in}$ that does not contain a state in $Good_{\mathcal{N}}$, the set of good states of $\mathcal{N}$. Then, $\pi' = \pi\,\tau\,\mathcal{F}_{out}$ is a $\mathfrak{T}$-path, too, and $\Pr^{\mathfrak{T}}_{\mathcal{N},\mathcal{F}_{out}}\big(\varphi_L[GoodEC]\big) = 1$ where $L = K - wgt_{\mathcal{N}}(\pi')$. With $H = \min\{K,L\}$ we get $\Pr^{\mathfrak{T}}_{\mathcal{N},\mathcal{E}_{out}}\big(\varphi_H[X]\big) = 1$ for all $\mathcal{E} \in X$. Thus, X is a fixed point of Ω. This contradicts that $GoodEC$ is the greatest fixed point of Ω.

(3) $\Pr^{\mathfrak{T}}_{\mathcal{N},\mathcal{E}_{out}}\big(\Diamond(T \cup Good_{\mathcal{N}})\big) = 1$ as $\Pr^{\mathfrak{T}}_{\mathcal{N},\mathcal{E}_{out}}\big(\varphi_K[GoodEC]\big) = 1$.

To prove $K^{\exists,=1}_{\mathcal{M},s} = +\infty$ we pick some $R \in \mathbb{Z}$ and design a scheduler $\mathfrak{S} = \mathfrak{S}_R$ for $\mathcal{M}$ such that $\Pr^{\mathfrak{S}}_{\mathcal{M},s}(\varphi_R) = 1$ for all good states s. The behavior of $\mathfrak{S}$ is as follows.

- For all input paths π ending in a state u of $S_{\mathcal{M}} \setminus WDMEC_{\mathcal{N}} \subseteq S_{\mathcal{N}} \setminus Good_{\mathcal{N}}$, scheduler $\mathfrak{S}$ behaves as $\mathfrak{T}$, *i.e.*, $\mathfrak{S}(\pi) = \mathfrak{T}(u)$.
- For all input paths π that end in the entry state $\mathcal{E}_{in}$ of some end component $\mathcal{E} \in GoodEC$, scheduler $\mathfrak{S}$ uses a weight-divergent scheduler for $\mathcal{E}$ until it has generated a path π' where the total weight is at least $R-K$ and where $\mathfrak{T}(\mathcal{E}_{out})$ is an action of state $last(\pi')$. Scheduler $\mathfrak{S}$ then schedules action $\mathfrak{T}(\mathcal{E}_{out})$ for π'.
- The behavior of $\mathfrak{S}$ for input paths π where $last(\pi)$ is the entry state $\mathcal{E}_{in}$ of some non-good end component is irrelevant.

Using properties (1), (2) and (3) of $\mathfrak{T}$, we get that none of the $\mathfrak{S}$-paths starting in a good state will visit a non-good end component and that $\Pr^{\mathfrak{S}}_{\mathcal{M},s}(\varphi_R) = 1$ for all good states s. □

Conversely, we establish that states belonging to a weight-divergent end component and with value $+\infty$ are good states.

Lemma D.31. *If $s \in WDMEC$ and $K^{\exists,=1}_{\mathcal{M},s} = +\infty$ then $s \in Good$.*

Proof. Recall that all states that belong to the same end component $\mathcal{E} \in \{\mathcal{E}_1,\ldots,\mathcal{E}_k\}$ have the same value $K^{\exists,=1}_{\mathcal{M},\mathcal{E}} = \{+\infty,-\infty\}$.

Set X Let X denote the set of end components $\mathcal{E} \in \{\mathcal{E}_1,\ldots,\mathcal{E}_k\}$ such that $K^{\exists,=1}_{\mathcal{M},\mathcal{E}} = +\infty$. It suffices to show that X is a fixed point of Ω, as then $X \subseteq GoodEC$ will follow, and therefore all states belonging to an end component $\mathcal{E} \in X$ are good.

Progress moves As in the proof of Lemma D.19, page 45, we use the notion of progress moves. This time, we need the term progress move for maximal end components. Given a maximal end component $\mathcal{E}$, we refer to a state-action pair (u,α) with $u \in \mathcal{E}$ as a *progress move* for $\mathcal{E}$ if there is some state v with $P_{\mathcal{M}}(u,\alpha,v) > 0$ and v does not belong to $\mathcal{E}$. By assumption **(A3)**, each maximal end component of $\mathcal{M}$ has a progress move. Moreover, whenever $\mathfrak{S}$ is a scheduler for $\mathcal{M}$ and s a state in $\mathcal{M}$ such that $\Pr^{\mathfrak{S}}_{\mathcal{M},s}(\Diamond T) = 1$ then for each $\mathfrak{S}$-path π from s to some maximal end component $\mathcal{E}$ there is an $\mathfrak{S}$-path π' that extends π and where $(last(\pi'),\mathfrak{S}(\pi'))$ is a progress move of $\mathcal{E}$.

Schedulers $\mathfrak{S}_{\mathcal{E}}$ For each $\mathcal{E} \in X$, there is some $R \in \mathbb{Z}$ and a scheduler $\mathfrak{S}_{\mathcal{E}}$ enjoying the following properties:

(P1) $\Pr^{\mathfrak{S}_{\mathcal{E}}}_{\mathcal{M},s}(\varphi_R) = 1$ for all states s in $\mathcal{E}$

(P2) There is a progress move $(u_{\mathcal{E}},\alpha_{\mathcal{E}})$ of $\mathcal{E}$ such that $\mathfrak{S}_{\mathcal{E}}(\pi) = \alpha_{\mathcal{E}}$ for each $\mathfrak{S}$-path π starting in some state of $\mathcal{E}$ with $wgt_{\mathcal{M}}(\pi) \geqslant 0$ and $last(\pi) = u_{\mathcal{E}}$.

(P3) Whenever π is a $\mathfrak{S}_{\mathcal{E}}$-path consisting of states and state-action pairs in $\mathcal{E}$ such that $\mathfrak{S}_{\mathcal{E}}(\pi)$ is a progress move then $wgt_{\mathcal{M}}(\pi) \geqslant 0$, $last(\pi) = u_{\mathcal{E}}$ and $\mathfrak{S}_{\mathcal{E}}(\pi) = \alpha_{\mathcal{E}}$.

Conditions **(P2)** and **(P3)** can be ensured by using a weight-divergent scheduler until having accumulated enough weight and the current state in $u_{\mathcal{E}}$.

Property **(P1)** implies:

(P4) All states that are reachable from $\mathcal{E}$ via an $\mathfrak{S}_{\mathcal{E}}$-path are either not contained in *WDMEC* or belong to an end component $\mathcal{E} \in X$.

(P5) Whenever π is an $\mathfrak{S}_{\mathcal{E}}$-path from $\mathcal{E}$ to *goal* where only the first state is contained in *WDMEC*, then π can be seen as a path in $\mathcal{N}$ starting in $\mathcal{E}_{out}$ and $wgt_{\mathcal{N}}(\pi) \geqslant R$.

(P6) $\Pr^{\mathfrak{S}_\mathcal{E}}_{\mathcal{M},u_\mathcal{E}}(\Diamond\mathcal{E}) < 1$

Let now $\mathfrak{T}_\mathcal{E}$ be a scheduler for $\mathcal{N}$ that schedules $\alpha_\mathcal{E}$ for state $\mathcal{E}_{out}$ and behaves as $\mathfrak{S}$ afterwards (when identifying $u_\mathcal{E}$ with $\mathcal{E}_{in}$) until reachaing a trap state $t \in T$ or the entry state of an end component $\mathcal{F} \in \{\mathcal{E}_1,\ldots,\mathcal{E}_k\}$. In the latter case, $\mathcal{F} \in X$ (by **(P4)**) and the behavior of $\mathfrak{T}_\mathcal{E}$ after having reached $\mathcal{F}_{in}$ is irrelevant for our purposes. By **(P5)** and **(P6)** we then have

$$\Pr^{\mathfrak{T}_\mathcal{E}}_{\mathcal{N},\mathcal{E}_{out}}(\varphi_R[X]) = 1 \ .$$

This shows that X is indeed a fixed point of Ω.

As a consequence, all states in X are good states, showing the desired result. □

From Lemmas D.30 and D.31 we obtain a characterization of good states and good end components.

Corollary D.32. *Good* $= \{s \in \mathit{WDMEC} : K^{\exists,=1}_{\mathcal{M},s} = +\infty\}$ *and* $\mathit{GoodEC} = \{\mathcal{E} \in \{\mathcal{E}_1,\ldots,\mathcal{E}_k\} : K^{\exists,=1}_{\mathcal{M},\mathcal{E}} = +\infty\}$.

Finally we characterize states of $\mathcal{M}$ with value $+\infty$:

Lemma D.33. *Let s be a state in $\mathcal{M}$. Then $K^{\exists,=1}_{\mathcal{M},s} = +\infty$ iff $\Pr^{\max}_{\mathcal{M},s}(\Diamond(T^* \cup \mathit{Good})) = 1$.*

Proof. The implication "$\Longleftarrow$" is an easy verification. The task is to provide schedulers $\mathfrak{T}_K$ with $\Pr^{\mathfrak{S}_K}_{\mathcal{M},s}(\varphi_K) = 1$ for each $K \in \mathbb{Z}$. The idea is to combine an MD-scheduler $\mathfrak{S}$ satisfying $\Pr^{\mathfrak{S}}_{\mathcal{M},s}(\Diamond(T^* \cup \mathit{Good})) = 1$ with schedulers of a family $(\mathfrak{S}_R)_{R\in\mathbb{Z}}$ where $\Pr^{\mathfrak{S}_R}_{\mathcal{M},u}(\varphi_R) = 1$ for each state $u \in \mathit{Good}$ and each $R \in \mathbb{Z}$. For this, scheduler $\mathfrak{T}_K$ first behaves as $\mathfrak{S}$ until reaching the target state in $\mathit{good} \in T^*$ or a good state. In the latter case, if w is the weight that has been accumulated so far, $\mathfrak{T}_K$ behaves as $\mathfrak{S}_{K-w}$ after having reached a good state.

To prove "$\Longrightarrow$", we pick a state s of $\mathcal{M}$ such that $K^{\exists,=1}_{\mathcal{M},s} = +\infty$. The claim is trivial if $s \in \{\mathit{good}\} \cup \mathit{Good}$. Consider now the case where $s \notin \{\mathit{good}\} \cup \mathit{Good}$. Suppose by contradiction that $\Pr^{\max}_{\mathcal{M},s}(\Diamond(\mathit{good} \cup \mathit{Good})) < 1$. Then, for each scheduler $\mathfrak{S}$ with $\Pr^{\mathfrak{S}}_{\mathcal{M},s}(\Diamond T) = 1$ we have $\Pr^{\mathfrak{S}}_{\mathcal{M},s}((\neg \mathit{Good})\,\mathbf{U}\,\mathit{goal}) > 0$.

Let $\mathcal{M}'$ be the MDP that results from $\mathcal{M}$ by (i) removing all state-action pairs (u,α) with $u \in \mathit{WDMEC}$ and (ii) adding the state-action pairs (u,τ) with $P(u,\tau,s) = 1$ and $\mathit{wgt}(u,\tau) = 0$ for $u \in \mathit{Good} \cup \{\mathit{good}\}$. Then, the end components of $\mathcal{M}'$ are exactly the end components of $\mathcal{M}$ that do not contain any state in WDMEC. In particular, $\mathcal{M}'$ has no (positively) weight-divergent end components. This is because s and the states $u \in \mathit{Good} \cup \{\mathit{good}\}$ cannot belong to an end component of $\mathcal{M}'$ since $\Pr^{\max}_{\mathcal{M}',s}(\Diamond(\{\mathit{good}\} \cup \mathit{Good})) = \Pr^{\max}_{\mathcal{M},s}(\Diamond(\{\mathit{good}\} \cup \mathit{Good})) < 1$.

We now consider a sequence of schedulers $(\mathfrak{S}_K)_{K\in\mathbb{N}}$ with $\Pr^{\mathfrak{S}_K}_{\mathcal{M},s}(\varphi_K) = 1$ for all $K \in \mathbb{N}$. None of the states that are reachable from s via a $\mathfrak{S}_K$-path belongs to $\mathit{WDMEC} \setminus \mathit{Good}$ as no $\mathfrak{S}_K$-path from s to goal can traverse a state u where $K^{\mathrm{E1}}_{\mathcal{M},u} = -\infty$.

Given $R \in \mathbb{Z}$, we design a scheduler $\mathfrak{T}_R$ for $\mathcal{M}'$ as follows. Given an input path π for $\mathfrak{T}$ where π does not contain a τ-transition from a state $u \in \mathit{Good} \cup \{\mathit{good}\}$ then $\mathfrak{T}_R$ behaves as $\mathfrak{S}_R$. Otherwise, π has the form $\pi_1;\pi_2$ where π_1 is a path from s to s where the last transition is a τ-transition from some state $u \in \mathit{Good} \cup \{\mathit{good}\}$ to s and π_2 is a path from s that does not contain such a τ-transition. In this case, we define $w = \mathit{wgt}(\pi_1)$ and $\mathfrak{T}_R$ behaves for π in the same way as $\mathfrak{S}_{R-w}$ behaves for the path π_2. We then have $\Pr^{\mathfrak{T}_R}_{\mathcal{M}',s}(\Diamond(\mathit{goal} \wedge (\mathrm{wgt} \geqslant R))) = 1$. In particular, $\mathbb{E}^{\mathfrak{T}_R}_{\mathcal{M}',s}(\Diamond\!\!\!\!-\, \mathit{goal}) \geqslant R$. As this holds for each $R \in \mathbb{Z}$, we obtain: $\mathbb{E}^{\sup}_{\mathcal{M}',s}(\Diamond\!\!\!\!-\, \mathit{goal}) = +\infty$. This is impossible by Lemma 4.1 (rephrased for maximal expected accumulated weights) as $\mathcal{M}'$ does not have positively weight-divergent end components. □

We can finally relate values in $\mathcal{M}$ and in $\mathcal{N}$.

Definition D.34 (Values for the states in $\mathcal{N}$)**.** Given a state u in $\mathcal{N}$ the *value of u in $\mathcal{N}$* is

$$K_{\mathcal{N},u} \;=\; \sup\,\{\, K \in \mathbb{Z} \;:\; \Pr^{\max}_{\mathcal{N},u}(\varphi_K[\mathit{GoodEC}]) = 1 \,\} \ .$$

Lemma D.35. *For each state s in $\mathcal{M}$, $K^{\exists,=1}_{\mathcal{M},s} = K_{\mathcal{N},s_\mathcal{N}}$.*

Proof. We have $K^{\exists,=1}_{\mathcal{M},s} \leqslant K_{\mathcal{N},s_\mathcal{N}}$ by Lemma D.24 and Corollary D.32. To prove $K^{\exists,=1}_{\mathcal{M},s} \geqslant K_{\mathcal{N},s_\mathcal{N}}$ we show that $\Pr^{\max}_{\mathcal{N},s_\mathcal{N}}(\varphi_K[\mathit{GoodEC}]) = 1$ implies $\Pr^{\max}_{\mathcal{M},s}(\varphi_K) = 1$. For this, pick a scheduler $\mathfrak{T}$ for $\mathcal{N}$ with $\Pr^{\mathfrak{T}}_{\mathcal{N},s_\mathcal{N}}(\varphi_K[\mathit{GoodEC}]) = 1$ and a family $(\mathfrak{S}_R)_{R\in\mathbb{Z}}$ of schedulers for $\mathcal{M}$ such that $\Pr^{\mathfrak{S}_R}_{\mathcal{M},u}(\varphi_R) = 1$ for all good states $u \in \mathit{Good}$. Let now $\mathfrak{S}$ be the following scheduler for $\mathcal{M}$ that mimics $\mathfrak{T}$ until reaching a good end component $\mathcal{E}$. If w is the weight that has been accumulated so far then $\mathfrak{T}$ behaves as $\mathfrak{S}_{K-w}$ from then on. We then have $\Pr^{\mathfrak{T}}_{\mathcal{M},s}(\varphi_K) = 1$. □

Computation of the Values $K^{\exists,=1}_{\mathcal{M},s}$ in the General Case

Relying on Lemma D.35, the values $K^{\exists,=1}_{\mathcal{M},s}$ for each s can be computed as follows.

1. Construct from $\mathcal{M}$ the MDP $\mathcal{N}$.
2. Compute the set *GoodEC*, iteratively starting with $X_0 = \{\mathcal{E}_1, \ldots, \mathcal{E}_k\}$, and with
$$X_{i+1} \;=\; \left\{ \mathcal{E} \in X_i \;:\; \exists K \in \mathbb{Z} \text{ s.t. } \Pr^{\max}_{\mathcal{N},\mathcal{E}_{out}}(\varphi_K[X_i]) = 1 \right\} .$$
When the sequence converges, the obtained set is *GoodEC*.
Since $\mathcal{N}$ has no weight-divergent end components, for the computation of the sets $X_1, X_2, \ldots$ we rely on the techniques for MDPs with this restriction, presented at the beginning of this section as a particular case.
3. Compute the values $K_{\mathcal{N},s}$ (see Definition D.34), again using the techniques for MDPs without weight-divergent end components.

For the third step, we can switch from $\mathcal{N}$ to the sub-MDP that arises by removing the entry and exit states for the end components $\mathcal{E} \in \{\mathcal{E}_1, \ldots, \mathcal{E}_k\} \setminus \mathit{GoodEC}$. These are the maximal weight-divergent end components of $\mathcal{M}$ where $K^{\exists,=1}_{\mathcal{M},\mathcal{E}} = -\infty$. Furthermore, for $\mathcal{E} \in \mathit{GoodEC}$, state $\mathcal{E}_{out}$ can be turned into a trap.

Let us analyse the complexity of the above procedure. The number of iterations in the second step is bounded by the number k of maximal weight-divergent end components. The values $K^{\exists,=1}_{\mathcal{M},s}$ can be computed in polynomial time, assuming an oracle that determines the values $K^{\exists,=1}_{\mathcal{N},s}$ for MDPs without weight-divergent end components. Recall (see Theorem D.20) that for MDPs without weight-divergent end components, the decision problem lies in $\mathsf{NP} \cap \mathsf{coNP}$, and the values are computable in pseudo-polynomial time. Since $\mathsf{P}^{\mathsf{NP}\cap\mathsf{coNP}}$ agrees with $\mathsf{NP} \cap \mathsf{coNP}$ [20], we conclude:

Theorem D.36. *The decision problem $DWR^{\exists,=1}$ belongs to NP ∩ coNP. The values $K^{\exists,=1}_{\mathcal{M},s}$ can be computed in pseudo-polynomial time.*

Mean-Payoff Game Hardness

We now prove a lower complexity bound for the $\mathsf{DWR}^{\exists,=1}$ decision problem.

Lemma D.37. *$DWR^{\exists,=1}$ is mean-payoff game hard.*

Proof. To establish mean-payoff hardness, we describe a polynomial-time reduction from the problem to decide whether player 1 of a (non-stochastic) two-player mean-payoff game has a winning strategy from a given game location s_{init}. This problem is known to be in $\mathsf{NP} \cap \mathsf{coNP}$ (even $\mathsf{UP} \cap \mathsf{coUP}$), but not known to be in P.

Let $\mathcal{G} = (V, V_1, V_2, E, wgt)$ be a two-player mean-payoff game where V is a finite set of game locations, disjointly partitioned into V_1 and V_2. The set V_i stands for the set of game locations where player i has to move. $E \subseteq V_1 \times V_2 \cup V_2 \times V_1$ is the edge relation, where we suppose that E is total in the sense that each location has at least one outgoing edge, and $wgt\colon E_1 \to \mathbb{Z}$ is the weight function[5] where $E_1 = E \cap (V_1 \times V)$. The objective of player 1 is to ensure that the mean payoff of all plays is nonnegative. More precisely, we consider the problem where we are given $\mathcal{G}$ and a distinguished starting location $v_0 \in V$ and where the task is to decide whether player 1 has a strategy $\mathfrak{S}$ such that the mean payoff of all $\mathfrak{S}$-plays from v_0 is nonnegative.

Let now $\mathcal{M}$ be the following MDP, as illustrated on Figure 16. The state space is $S_{\mathcal{M}} = V \cup \{s_{init}, goal\}$ where s_{init} is the initial state of $\mathcal{M}$. The action set is $Act = V \cup \{\alpha, \tau\}$. The transition probabilities and weights are defined as follows. In s_{init}, actions τ and α are enabled with $P(s_{init}, \tau, v_0) = 1$ and $P(s_{init}, \alpha, s_{init}) = 1$ and $wgt(s_{init}, \tau) = 0$, $wgt(s_{init}, \alpha) = 1$. If $(v_1, v) \in E$ where $v_1 \in V_1$ then $P(v_1, v, v) = 1$. The weight of the state-action pair (v_1, v) is the weight of the edge (v_1, v) in $\mathcal{G}$. In all other cases, $P(v_1, \cdot) = 0$. (That is, only the actions $v \in V$ where (v_1, v) is an edge in $\mathcal{G}$ are enabled in state v_1 of $\mathcal{M}$.) Let now $v_2 \in V_2$ and let $Post(v_2) = \{v \in V : (v_2, v) \in E\}$. The only enabled action in v_2 is τ. The transition probabilities are given by $P(v_2, \tau, goal) = \frac{1}{2}$ and $P(v_2, \tau, v) = \frac{1}{2k}$ where $k = |Post(v_2)|$. The weight of the state-action pair (v_2, τ) is 0. State *goal* is a trap in $\mathcal{M}$.

As the edge relation E of $\mathcal{G}$ is total, *goal* is the only trap state of $\mathcal{M}$. By construction, we have $\Pr^{\min}_{\mathcal{M},s}(\Diamond goal) = 1$ for all states s in $\mathcal{M}$ with $s \neq s_{init}$. This also implies that $\mathcal{M}$ has a single end component $\mathcal{E}$ consisting of the state-action pair (s_{init}, α).

Obviously, $\mathcal{M}$ can be constructed in polynomial time from $\mathcal{G}$. We let
$$\varphi \;=\; \Diamond(goal \wedge (\mathrm{wgt} \geqslant 0))$$
that is, $T = \{goal\}$ and $K_{goal} = 0$. We claim that player 1 has a winning strategy $\mathfrak{S}$ in $\mathcal{G}$ iff there exists a scheduler $\mathfrak{S}$ in $\mathcal{M}$ with $\Pr^{\max}_{\mathcal{M},s_{init}}(\varphi) = 1$.

[5]Note that the general case, where players do not strictly alternate, and weights are also attached to moves of player 2 can easily be reduced to game structures of this form.

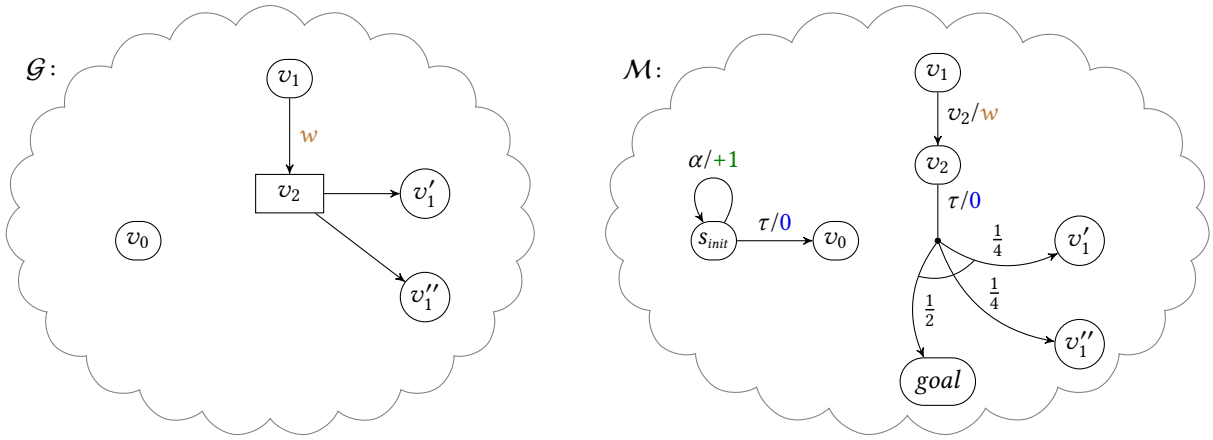


Figure 16. Reduction from mean-payoff games to $\mathrm{DWR}^{\exists,=1}$.

Suppose first that player 1 has a winning strategy σ in the mean-payoff game. Without loss of generality, this winning strategy can be assumed to be an MD-strategy. Let $\mathcal{G}'$ be the graph structure induced by σ restricted to the states that are reachable from v_0 along finite σ-plays. As σ is winning, $\mathcal{G}'$ has no negative cycle. Let w be the minimal weight of a path π from v_0 to *goal* in $\mathcal{G}'$, and let $k = \max\{0, -w\}$. Consider now the following scheduler $\mathfrak{S}$ for $\mathcal{M}$. It schedules k-times action α in state s_{init}, moves to state v_0 via the τ-transition afterwards and behaves as σ from then on. In particular $\mathfrak{S}$ is a finite-memory scheduler. The underlying graph of the Markov chain C induced by $\mathfrak{S}$ (restricted to the states reachable from s_{init}) agrees with $\mathcal{G}'$ extended by an initial phase

$$\underbrace{s_{init} \xrightarrow{\alpha} s_{init} \xrightarrow{\alpha} \ldots \xrightarrow{\alpha} s_{init}}_{k \text{ transitions}} \xrightarrow{\tau} v_0 \ .$$

As $\mathcal{G}'$, also C has no negative cycles. Moreover, $wgt(\pi) \geqslant k + w \geqslant 0$ for all paths π in C from s_{init} to *goal*. Hence, $\Pr^{\mathfrak{S}}_{\mathcal{M}, s_{init}}(\varphi) = 1$.

Vice versa, suppose $\mathfrak{S}$ is a scheduler for $\mathcal{M}$ such that $\Pr^{\mathfrak{S}}_{\mathcal{M}, s_{init}}(\varphi) = 1$. In particular, $\Pr^{\mathfrak{S}}_{\mathcal{M}, s_{init}}(\Diamond goal) = 1$. Thus, $\mathfrak{S}$ schedules τ for s_{init} after having generated a path of the form $\pi = s_{init}\, \alpha\, s_{init} \alpha s_{init} \alpha \ldots \alpha\, s_{init}$. With $k = |\pi|$ we have $wgt(\pi) = k$. The residual scheduler $\mathfrak{T} = \mathfrak{S}{\uparrow}\pi\, \tau\, v_0$ can be viewed as a strategy for player 1 in the game structure $\mathcal{G}$. As

$$\Pr^{\mathfrak{T}}_{\mathcal{M}, v_0}(\Diamond(goal \wedge (\mathrm{wgt} \geqslant -k))) \ = \ 1 \ ,$$

all $\mathfrak{T}$-paths starting in v_0 and ending in *goal* have weight at least $-k$. The sub-MDP $\mathcal{M}'$ resulting from $\mathcal{M}$ by removing the state-action pair (s_{init}, α) has no end components. As $\mathfrak{T}$ can be viewed as a scheduler for $\mathcal{M}'$, we can rely on Lemma D.19, which ensures the existence of an MD-scheduler $\mathfrak{U}$ with

$$\Pr^{\mathfrak{U}}_{\mathcal{M}, v_0}(\Diamond(goal \wedge (\mathrm{wgt} \geqslant -k))) \ = \ 1 \ .$$

Lemma D.16 yields that the Markov chain induced by $\mathfrak{U}$ has no negative cycle. Hence, $\mathfrak{U}$ is a winning strategy for player 1 in the game $\mathcal{G}$. □

D.5 Weight-Bounded Büchi Constraints

We now provide the proofs for Section 5.2. Throughout this section, we suppose that $\mathcal{M} = (S, Act, P, wgt)$ is an MDP without traps, which ensures that all maximal paths are infinite. Furthermore, let s_{init} be a state in $\mathcal{M}$, F a set of states in $\mathcal{M}$, and $K \in \mathbb{Z}$.

Lemma D.38. *Problems $WB^{\exists,=1}$ and $WB^{\forall,>0}$ are hard for non-stochastic two-player mean-payoff games.*

Proof. As the corresponding result has been established for the DWR problems $\mathrm{DWR}^{\exists,=1}$ and $\mathrm{DWR}^{\forall,>0}$ (see Lemma D.37 and D.4), it suffices to provide prolynomial reductions from them. Let $\mathcal{M}$ be an MDP and $\varphi(T, (K_t)_{t \in T})$ be a disjunctive weight-bounded reachability constraint. Let $\mathcal{M}'$ be the MDP resulting from $\mathcal{M}$ by (i) discarding all state-action pairs $(t, \alpha) \in \mathcal{M}$ with $t \in T$, (ii) adding state-action pairs (t, τ) with $P_{\mathcal{M}'}(t, \tau, t) = 1$ and $wgt_{\mathcal{M}'}(t, \tau) = 0$ for all states $t \in T$ and all traps t of $\mathcal{M}$. Then, $(\mathcal{M}, s_{init})$ satisfies $\mathrm{DWR}^{\exists,=1}$ iff $(\mathcal{M}, s_{init})$ satisfies $\mathrm{WB}^{\exists,=1}$, and $(\mathcal{M}, s_{init})$ satisfies $\mathrm{DWR}^{\forall,>0}$ iff $(\mathcal{M}, s_{init})$ satisfies $\mathrm{WB}^{\forall,>0}$. □

D.5.1 The Existential Problems $\mathrm{WB}^{\exists,=1}$ and $\mathrm{WB}^{\exists,>0}$

We introduce notations for sets of states belonging to given end components. Let $PumpEC_F$ is the set of states that belong to a pumping end component $\mathcal{E}$ that contain at least one state in F. Likewise, $GambEC_F$ is the set of states that belong to a gambling end component containing at least one F-state. We write $ZeroEC_F$ for the set of states that belong to a 0-EC $\mathcal{Z}$ such that $\mathcal{Z}$ contains at least one F-state and $\mathcal{Z}$ is a sub-component of an MEC $\mathcal{E}$ with $\mathbb{E}^{\max}_{\mathcal{E}}(\mathrm{MP}) = 0$. Then, $WDMEC_F = PumpEC_F \cup GambEC_F$ consists of the states that belong to some weight-divergent end component containing at least one F-state. $PumpEC_F \cup ZeroEC_F$ is the set of states that are contained in some end component intersecting with F that has a scheduler where the accumulated weight is bounded from below almost surely.

Lemma D.39. *$PumpEC_F$, $WDMEC_F$, $PumpEC_F \cup ZeroEC_F$ and $WDMEC_F \cup ZeroEC_F$ are computable in polynomial time. Moreover, there exist a scheduler $\mathfrak{D}$ such that:*

- $\Pr^{\mathfrak{D}}_{\mathcal{M},s}(\Box\Diamond F) = 1$ *for all* $s \in WDMEC_F \cup ZeroEC_F$,
- $\mathfrak{D}$ *is pumping from all states* $s \in PumpEC_F$,
- $\mathfrak{D}$ *is gambling from all states* $s \in GambEC_F$, *and*
- *from all states* $t \in ZeroEC_F$, $\mathfrak{D}$ *realizes a 0-EC* $\mathcal{E}$ *with* $\mathcal{E} \cap F \neq \varnothing$. *In particular,* $\Pr^{\mathfrak{D}}_{\mathcal{M},t}(\Box\Diamond(wgt \geqslant 0) \wedge \Box\Diamond F) = 1$ *for all* $t \in ZeroEC_F$.

Proof. $PumpEC_F$ is the union of the state spaces of the MEC $\mathcal{E}$ of $\mathcal{M}$ that contain at least one F-state and that enjoy the property $\mathbb{E}^{\max}_{\mathcal{E}}(\mathrm{MP}) > 0$ (see Lemma 3.3). This yields the polynomial-time computability of $PumpEC_F$.

To compute $WDMEC_F$ we can rely on the fact that each weight-divergent end component is contained in an MEC that is weight-divergent. Thus, we can apply standard techniques to compute the MECs of $\mathcal{M}$. For each of the MECs $\mathcal{E}$, we check whether $\mathcal{E}$ contains at least one F-state, and if so, we check whether $\mathcal{E}$ is weight-divergent using the polynomial-time weight-divergence algorithm presented in Section 3.2. Then, $WDMEC_F$ arises by the union of the state spaces of these MECs.

The set $PumpEC_F \cup ZeroEC_F$ is the set of all states s that belong to a maximal end component $\mathcal{E}$ with $\mathcal{E} \cap F \neq \varnothing$ and such either $\mathcal{E}$ is pumping or $\mathbb{E}^{\max}_{\mathcal{E}}(\mathrm{MP}) = 0$ and s belongs to a maximal 0-EC $\mathcal{Z}$ of $\mathcal{E}$ with $\mathcal{Z} \cap F \neq \varnothing$. Thus, to compute $PumpEC_F \cup ZeroEC_F$ in polynomial time we can rely on the pumping criterion presented in Lemma 3.3 and the techniques of Lemma 3.13 to determine maximal 0-ECs of in strongly connected MDPs with maximal expected mean payoff 0. The statement about the recurrence values $rec(s)$ is immediate from Lemma 3.13.

It remains to explain how to obtain scheduler $\mathfrak{D}$. For each pumping resp. gambling MEC $\mathcal{E}$ containing at least one F-state, we pick a pumping resp. gambling scheduler $\mathfrak{B}_{\mathcal{E}}$ and an MD-scheduler $\mathfrak{U}_{\mathcal{E}}$ for $\mathcal{E}$ such that $\Pr^{\mathfrak{U}_{\mathcal{E}}}_{\mathcal{E},s}(\Diamond F) = 1$ for all states s in $\mathcal{E}$. Obviously, $\mathfrak{U}_{\mathcal{E}}$ and $\mathfrak{B}_{\mathcal{E}}$ can be combined to obtain a (possibly infinite-memory) weight-divergent scheduler $\mathfrak{D}_{\mathcal{E}}$ for $\mathcal{E}$ with $\Pr^{\mathfrak{D}_{\mathcal{E}}}_{\mathcal{E},s}(\Box\Diamond F) = 1$ for all states s in $\mathcal{E}$. Composing the schedulers $\mathfrak{D}_{\mathcal{E}}$ with schedulers that realize maximal 0-ECs contained in MECs with maximal expected mean payoff 0 yields a scheduler $\mathfrak{D}$ as stated in the lemma. □

We now show that problems $\mathrm{WB}^{\exists,=1}$ and $\mathrm{WB}^{\exists,>0}$ are polynomially reducible to $\mathrm{DWR}^{\exists,=1}$ and $\mathrm{DWR}^{\exists,>0}$, respectively (see Lemma D.40 and D.41 below). In both cases we use the disjunctive weight-bounded reachability constraint $\varphi = \varphi(T, (K_t)_{t\in T})$ where $T = WDMEC_F \cup ZeroEC_F$ and $K_t = -\infty$ for $t \in WDMEC_F$ and $K_t = K$ for $t \in ZeroEC_F$. That is, $T^* = WDMEC_F$ and

$$\varphi \;=\; \Diamond WDMEC_F \vee \bigvee_{t\in ZeroEC_F} \Diamond(t \wedge (\mathrm{wgt} \geqslant K)) \;.$$

Given an end component $\mathcal{E}$, let $Limit_{\mathcal{E}}$ denote the set of infinite paths ς such that the limit of ς equals $\mathcal{E}$. Recall that the limit of an infinite path ς is the set of all state-action pairs (s,α) that occur infinitely often in ς. In what follows, we often use de Alfaro's Theorem [10] stating that under each scheduler, the probability of the paths in $\bigcup_{\mathcal{E}} Limit_{\mathcal{E}}$ equals 1 when $\mathcal{E}$ ranges over all (possibly non-maximal) end components.

Lemma D.40. *Let φ be as above. Then* $\Pr^{\max}_{\mathcal{M},s_{init}}\big(\Box\Diamond(wgt \geqslant K) \wedge \Box\Diamond F\big) = 1$ *iff* $\Pr^{\max}_{\mathcal{M},s_{init}}(\varphi) = 1$.

Proof. The implication "$\Longleftarrow$" is an easy verification. Given a scheduler $\mathfrak{S}$ with $\Pr^{\mathfrak{S}}_{\mathcal{M},s_{init}}(\varphi) = 1$, combine $\mathfrak{S}$ with the scheduler $\mathfrak{D}$ of Lemma D.39 to obtain a new scheduler $\mathfrak{T}$ with $\Pr^{\mathfrak{T}}_{\mathcal{M},s_{init}}\big(\Box\Diamond(\mathrm{wgt} \geqslant K) \wedge \Box\Diamond F\big) = 1$.

To prove "$\Longrightarrow$", we suppose that we are given a scheduler $\mathfrak{T}$ for $\mathcal{M}$ with $\Pr^{\mathfrak{T}}_{\mathcal{M},s_{init}}\big(\Box\Diamond(\mathrm{wgt} \geqslant K) \wedge \Box\Diamond F\big) = 1$. Then, $\Pr^{\mathfrak{T}}_{\mathcal{M},s_{init}}(Limit_{\mathcal{E}}) > 0$ implies $\mathbb{E}^{\max}_{\mathcal{E}}(\mathrm{MP}) \geqslant 0$ for each end component $\mathcal{E}$. Thus, $\mathcal{E}$ is either weight-divergent or a 0-EC. As almost all $\mathfrak{T}$-paths satisfy $\Box\Diamond F$ almost surely, $\mathcal{E}$ must contain at least one F-state. This yields

$$\Pr^{\mathfrak{T}}_{\mathcal{M},s_{init}}\big(\Diamond(WDMEC_F \cup ZeroEC_F)\big) \;=\; 1\,.$$

Moreover, if $\mathcal{E}$ is a 0-EC with $\Pr^{\mathfrak{T}}_{\mathcal{M},s_{init}}(Limit_{\mathcal{E}}) > 0$ and $\mathcal{E}$ is not a sub-component of a weight-divergent MEC then all states of $\mathcal{E}$ belong to $ZeroEC_F$ and almost all $\mathfrak{T}$-paths $\varsigma \in Limit_{\mathcal{E}}$ have infinitely many prefixes π with $wgt(\pi) \geqslant K$. In particular, these paths ς satisfy the formula $\bigvee_{t \in ZeroEC_F} \Diamond(t \wedge (\mathrm{wgt} \geqslant K))$. This yields $\Pr^{\mathfrak{T}}_{\mathcal{M},s_{init}}(\varphi) = 1$. □

Lemma D.41. *Let φ be as in Lemma D.40. Then* $\Pr^{\max}_{\mathcal{M},s_{init}}\big(\Box\Diamond(wgt \geqslant K) \wedge \Box\Diamond F\big) > 0$ *iff* $\Pr^{\max}_{\mathcal{M},s_{init}}(\varphi) > 0$.

Proof. "$\Longrightarrow$": If $\mathfrak{S}$ is a scheduler for $\mathcal{M}$ with $\Pr^{\mathfrak{S}}_{\mathcal{M},s_{init}}(\varphi) > 0$, then there is a finite $\mathfrak{S}$-path π from s_{init} such that either $last(\pi) \in WDMEC_F$ or $last(\pi) \in ZeroEC_F$ and $wgt(\pi) \geqslant K$. Let now $\mathfrak{T}$ be any scheduler for $\mathcal{M}$ such that $\mathfrak{T}{\uparrow}\pi = \mathfrak{D}$ where $\mathfrak{D}$ is as in Lemma D.39. Then, $\Pr^{\mathfrak{T}}_{\mathcal{M},s_{init}}(\Box\Diamond(\mathrm{wgt} \geqslant K) \wedge \Box\Diamond F) > 0$.

"$\Longleftarrow$": We suppose we are given a scheduler $\mathfrak{T}$ for $\mathcal{M}$ such that $\Pr^{\mathfrak{T}}_{\mathcal{M},s_{init}}\big(\Box\Diamond(\mathrm{wgt} \geqslant K) \wedge \Box\Diamond F\big)$ is positive. There is an end component $\mathcal{E}$ such that $\Pr^{\mathfrak{T}}_{\mathcal{M},s_{init}}(\Lambda_{\mathcal{E}}) > 0$ where $\Lambda_{\mathcal{E}}$ denotes the set of infinite paths $\varsigma \in Limit_{\mathcal{E}}$ with $\varsigma \models (\mathrm{wgt} \geqslant K) \wedge \Box\Diamond F$. But then, $\mathcal{E}$ contains an F-state and is probably bounded from below. Thus, $\mathbb{E}^{\max}_{\mathcal{E}}(\mathrm{MP})$ is nonnegative.

If $\mathcal{E}$ is weight-divergent then all states of $\mathcal{E}$ are contained in $WDMEC_F$, in which case $\Pr^{\mathfrak{S}}_{\mathcal{M},s_{init}}(\Diamond\, WDMEC_F) > 0$ and therefore $\Pr^{\mathfrak{S}}_{\mathcal{M},s_{init}}(\varphi) > 0$.

Let us now consider the case where $\mathcal{E}$ is not weight-divergent. Then, $\mathbb{E}^{\max}_{\mathcal{E}}(\mathrm{MP}) = 0$. Corollary D.42 yields that $\mathcal{E}$ is a 0-EC. But then all states in $\mathcal{E}$ are conatined in $ZeroEC_F$. Hence, there is some state $t \in ZeroEC$ with $\Pr^{\mathfrak{S}}_{\mathcal{M},s_{init}}\big(\Diamond(t \wedge (\mathrm{wgt} \geqslant K))\big) > 0$. But then $\Pr^{\mathfrak{S}}_{\mathcal{M},s_{init}}(\varphi) > 0$. □

D.5.2 The Universal Problems $\mathrm{WB}^{\forall,=1}$ and $\mathrm{WB}^{\forall,>0}$

We now consider the universal variants $\mathrm{WB}^{\forall,=1}$ and $\mathrm{WB}^{\forall,>0}$ and show that they are solvable using algorithms for the following two coBüchi problems:

$\mathrm{WcoB}^{\exists,>0}$: does there exist a scheduler $\mathfrak{S}$ s.t. $\Pr^{\mathfrak{S}}_{\mathcal{M},s_{init}}\big(\Diamond\Box(\mathrm{wgt} \geqslant K)\big) > 0$?

$\mathrm{WcoB}^{\exists,=1}$: does there exist a scheduler $\mathfrak{S}$ s.t. $\Pr^{\mathfrak{S}}_{\mathcal{M},s_{init}}\big(\Diamond\Box(\mathrm{wgt} \geqslant K)\big) = 1$?

Property (S3) of the spider construction and the corresponding statement for the iterative application of the spider construction (Lemma B.25) will serve as a useful vehicle to prove the following two lemmas, which again will be used to reduce $\mathrm{WcoB}^{\exists,>0}$ to $\mathrm{DWR}^{\exists,>0}$ and $\mathrm{WcoB}^{\exists,=1}$ to $\mathrm{DWR}^{\exists,=1}$(see Lemma D.45 below).

Lemma D.42. *Let $\mathcal{M}$ be a strongly connected MDP where $\mathbb{E}^{\max}_{\mathcal{M}}(\mathrm{MP}) = 0$ and where $\mathcal{M}$ is not weight-divergent. If there exists a scheduler $\mathfrak{S}$ and some $K \in \mathbb{Z}$ such that $\Pr^{\mathfrak{S}}_{\mathcal{M},s}\{\varsigma \in IPaths : \varsigma \models \Box\Diamond(wgt \geqslant K) \wedge \lim(\varsigma) = \mathcal{M}\}$ is positive, then $\mathcal{M}$ is a 0-EC.*

Proof. Suppose by contradiction that $\mathcal{M}$ is not a 0-EC. Let $\mathcal{N}$ be the MDP resulting from applying the weight-divergence algorithm to $\mathcal{M}$ (iterative application of the spider construction). The set of infinite paths ς with $\limsup_{n\to\infty} wgt(pref(\varsigma,n)) > -\infty$ and where $\lim(\varsigma)$ contains at least one state-action pair that is not contained in a 0-EC constitutes a 0-EC-invariant property with positive measure under $\mathfrak{S}$ from state s. Hence, Lemma B.25 yields the existence of a scheduler $\mathfrak{T}$ for $\mathcal{N}$ such that the set of infinite paths ς in $\mathcal{N}$ with $\limsup_{n\to\infty} wgt(pref(\varsigma,n)) > -\infty$ has positive measure from s. This, however, is impossible as $\mathbb{E}^{\max}_{\mathcal{F}}(\mathrm{MP}) < 0$ for all end components $\mathcal{F}$ of $\mathcal{N}$ (see Theorem 3.9), which yields that $\mathcal{N}$ is universally negatively pumping. □

As a consequence of Lemma D.42 we get the following corollary which can be seen as an add-on for Lemma 3.14 (but will not be used for the following considerations on weight-bounded Büchi conditions).

Corollary D.43. *Let $\mathcal{M}$ be a strongly connected MDP where $\mathbb{E}^{\max}_{\mathcal{M}}(\mathrm{MP}) = 0$ and where $\mathcal{M}$ is not positively weight-divergent. Then, $\mathcal{M}$ has no 0-ECs if and only if $\mathcal{M}$ is universally negatively pumping.*

Proof. The implication "$\Longleftarrow$" is trivial as 0-ECs are obviously not negatively pumping. We now prove "$\Longrightarrow$". For this, we suppose that $\mathcal{M}$ is not universally negativey pumping and show that $\mathcal{M}$ has at least one 0-EC. Being not universally negatively pumping implies the existence of a scheduler $\mathfrak{S}$ and a state s such that:

$$\Pr^{\mathfrak{S}}_{\mathcal{M},s}\Big\{\varsigma \in IPaths : \liminf_{n\to\infty} wgt(pref(\varsigma,n)) \;>\; -\infty\Big\} \;>\; 0 \;.$$

But then there is some $K \in \mathbb{Z}$ and an end component $\mathcal{E}$ such that

$$\Pr^{\mathfrak{S}}_{\mathcal{M},s}\big\{\varsigma \in IPaths : \varsigma \models \Box\Diamond(\mathrm{wgt} \geqslant K) \wedge \lim(\varsigma) = \mathcal{E}\big\} \;>\; 0 \;.$$

As $\mathbb{E}^{\max}_{\mathcal{M}}(\mathrm{MP}) = 0$ we have $\mathbb{E}^{\max}_{\mathcal{E}}(\mathrm{MP}) \leqslant 0$. As $\mathbb{E}^{\max}_{\mathcal{E}}(\mathrm{MP}) < 0$ would imply that $\mathcal{E}$ is universally negatively pumping we get $\mathbb{E}^{\max}_{\mathcal{E}}(\mathrm{MP}) = 0$. Lemma D.42 applied to $\mathcal{E}$ yields that $\mathcal{E}$ is a 0-EC. □

Lemma D.44. *Let $\mathcal{M}$ be a strongly connected MDP where $\mathbb{E}^{\max}_{\mathcal{M}}(\mathrm{MP}) = 0$. If there exists a scheduler $\mathfrak{S}$ and some $K \in \mathbb{Z}$ such that $\Pr^{\mathfrak{S}}_{\mathcal{M},s}\{\varsigma \in \mathit{IPaths} : \varsigma \models \Diamond\Box(\mathit{wgt} \geqslant K) \wedge \lim(\varsigma) = \mathcal{M}\}$ is positive then $\mathcal{M}$ is a 0-EC.*

Proof. The argument is similar as for Lemma D.42. Suppose by contradiction that $\mathcal{M}$ is not a 0-EC. Let $\mathcal{N}$ denote the MDP resulting from $\mathcal{M}$ by successively applying the spider construction to flatten all 0-ECs of $\mathcal{M}$. For this we can rely on the algorithm to compute all maximal 0-ECs of $\mathcal{M}$ presented in Section B.8.1 (see also Lemma 3.13) and then successively apply the spider construction to the BSCCs of the maximal 0-ECs. The final MDP $\mathcal{N}$ has no 0-EC, but might still contain gambling end components. In any case, all end components of $\mathcal{N}$ are negatively weight-divergent. The set of infinite paths ς with $\varsigma \models \Diamond\Box(\mathit{wgt} \geqslant K)$ where $\lim(\varsigma)$ contains at least one state-action pair that does not belong to a 0-EC consitutes a measurable 0-EC-invariant property. The probability for this property under $\mathfrak{S}$ from state s is positive (by assumption and as $\mathcal{M}$ is not a 0-EC). Thanks to Lemma B.25, scheduler $\mathfrak{S}$ for $\mathcal{M}$ can be transformed into a scheduler $\mathfrak{T}$ for $\mathcal{N}$ such that $\Diamond\Box(\mathit{wgt} \geqslant K)$ holds with positive probability. But this is impossible as all end components of $\mathcal{N}$ are negatively weight-divergent. □

Lemma D.45. *Problem $\mathrm{WcoB}^{\exists,>0}$ is polynomially reducible to $\mathrm{DWR}^{\exists,>0}$, while $\mathrm{WcoB}^{\exists,=1}$ is polynomially reducible to $\mathrm{DWR}^{\exists,=1}$.*

Proof. Let $\mathit{PumpEC} = \mathit{PumpEC}_S$ be the set of states that are contained in some pumping end component and let $\mathit{ZeroEC} = \mathit{ZeroEC}_S$ denote the set of states that belong to a maximal 0-EC $\mathcal{Z}$ where $\mathcal{Z}$ is a sub-component of a maximal end component of $\mathcal{M}$ with maximal expected mean payoff 0. Using the results of Section 3 (see also Lemma D.39), *PumpEC* and *ZeroEC* are computable in polynomial time. Recall from Section B.8.2 that for each state $t \in \mathit{ZeroEC}$ the recurrence value $\mathit{rec}(t)$ defined as the maximal value $w \in \mathbb{Z}$ such that $\Pr^{\max}_{\mathcal{Z},s}\big(\,\Box(\mathrm{wgt} \geqslant w) \wedge \Box\Diamond t\,\big) = 1$ is computable in polynomial time where $\mathcal{Z}$ denotes the unique maximal 0-EC that contains t. Let now $\mathfrak{U}$ be a scheduler such that:

- $\mathfrak{U}$ is pumping from each state $t \in \mathit{PumpEC}$,
- $\Pr^{\mathfrak{U}}_{\mathcal{M},s}(\,\Box(\mathrm{wgt} \geqslant \mathit{rec}(t)) \wedge \Box\Diamond t\,) = 1$ for each state $t \in \mathit{ZeroEC}$.

Let φ be the disjunctive weight-bounded reachability constraint $\varphi = \varphi(T, (K_t)_{t\in T})$ where $T = \mathit{PumpEC} \cup \mathit{ZeroEC}$ and $K_t = -\infty$ for $t \in \mathit{PumpEC}$ and $K_t = K - \mathit{rec}(t)$ for $t \in \mathit{ZeroEC} \setminus \mathit{PumpEC}$. We now show:

(a) $\Pr^{\max}_{\mathcal{M},s_{init}}\big(\Diamond\Box(\mathrm{wgt} \geqslant K)\big) > 0$ iff $\Pr^{\max}_{\mathcal{M},s_{init}}(\varphi) > 0$
(b) $\Pr^{\max}_{\mathcal{M},s_{init}}\big(\Diamond\Box(\mathrm{wgt} \geqslant K)\big) = 1$ iff $\Pr^{\max}_{\mathcal{M},s_{init}}(\varphi) = 1$

Clearly, statement (a) implies the polynomial reducibility of $\mathrm{WcoB}^{\exists,>0}$ to $\mathrm{DWR}^{\exists,>0}$, while the polynomial reducibility of $\mathrm{WcoB}^{\exists,=1}$ to $\mathrm{DWR}^{\exists,=1}$ follows from statement (b).

Proof of statement (a). The implication is "$\Longleftarrow$" is obvious any scheduler $\mathfrak{T}$ with $\Pr^{\mathfrak{T}}_{\mathcal{M},s_{init}}(\varphi) > 0$ can be combined with the scheduler $\mathfrak{U}$ above to obtain a new scheduler $\mathfrak{S}$ with $\Pr^{\mathfrak{S}}_{\mathcal{M},s_{init}}\big(\Diamond\Box(\mathrm{wgt} \geqslant K)\big) > 0$.

To prove "$\Longrightarrow$", we suppose there is a scheduler $\mathfrak{T}$ with $\Pr^{\mathfrak{T}}_{\mathcal{M},s_{init}}\big(\Diamond\Box(\mathrm{wgt} \geqslant K)\big) > 0$. There exists an end component $\mathcal{E}$ such that

$$\Pr^{\mathfrak{T}}_{\mathcal{M},s_{init}}\big\{\,\varsigma \in \mathit{Limit}_{\mathcal{E}} \;:\; \varsigma \models \Diamond\Box(\mathrm{wgt} \geqslant K)\,\big\} \;>\; 0\ .$$

If $\mathcal{E}$ contains some state that belongs to a pumping end component (this covers the case where $\mathcal{E}$ itself is pumping) then $\mathcal{E}$ contains at least one state in *PumpEC*, which obviously yields $\Pr^{\mathfrak{T}}_{\mathcal{M},s_{init}}(\varphi) > 0$. Suppose now that none of the states in $\mathcal{E}$ belongs to a pumping end component. Obviously we then have $\mathbb{E}^{\max}_{\mathcal{E}}(\mathrm{MP}) = 0$. Thanks to Lemma D.44 we get that $\mathcal{E}$ is a 0-EC.

Pick some state t in $\mathcal{E}$. The presented algorithm for computing the recurrence values (see Section B.8.2) shows that $\mathit{rec}(t) \geqslant w_{\mathcal{E}}(t)$ where $w_{\mathcal{E}}(t) = \min\{w(t,s) : s \in \mathcal{E}\}$. (As before, $w(t,s)$ is the weight of all paths from t to s in the maximal 0-EC $\mathcal{Z}$ that subsumes $\mathcal{E}$.) Thus, if $\varsigma \in \mathit{Limit}_{\mathcal{E}}$ with $\varsigma \models \Diamond\Box(\mathrm{wgt} \geqslant K)$ then there exists an integer $L \geqslant K$ such that ς contains infinitely many finite prefixes π with $\mathit{last}(\pi) = t$ and $\mathit{wgt}(\pi) = L$. We then have $L + w(t,s) \geqslant K$ for all states s in $\mathcal{E}$. Thus, $L + w_{\mathcal{E}}(t) \geqslant K$ and therefore

$$\mathit{rec}(t) \;\;\geqslant\;\; w_{\mathcal{E}}(t) \;\;\geqslant\;\; K - L\ .$$

We get $L \geqslant K - \mathit{rec}(t) = K_t$. This yields $\varsigma \models \Diamond(t \wedge (\mathrm{wgt} \geqslant K_t))$. But then $\Pr^{\mathfrak{T}}_{\mathcal{M},s_{init}}(\varphi) > 0$.

Proof of statement (b). For the implication is "$\Longleftarrow$" we suppose that we are given a scheduler $\mathfrak{T}$ with $\Pr^{\max}_{\mathcal{M},s_{init}}(\varphi) = 1$. Composing $\mathfrak{T}$ with the above scheduler $\mathfrak{U}$ we obtain a new scheduler $\mathfrak{S}$ with $\Pr^{\mathfrak{S}}_{\mathcal{M},s_{init}}\big(\Diamond\Box(\mathrm{wgt} \geqslant K)\big) = 1$.

To prove "$\Longrightarrow$" we suppose that we are given a scheduler $\mathfrak{S}$ with $\Pr^{\mathfrak{S}}_{\mathcal{M},s_{init}}(\Diamond\Box(\mathrm{wgt} \geqslant K)) = 1$. But then each end component $\mathcal{E}$ where $\Pr^{\mathfrak{S}}_{\mathcal{M},s_{init}}\{\varsigma \in \mathit{Limit}_{\mathcal{E}} : \varsigma \models \Diamond\Box(\mathrm{wgt} \geqslant K)\} > 0$ has nonnegative maximal expected mean payoff. If $\mathbb{E}^{\max}_{\mathcal{E}}(\mathrm{MP}) > 0$ then $\mathcal{E}$ is pumping and all states of $\mathcal{E}$ belong to *PumpEC*. If $\mathbb{E}^{\max}_{\mathcal{E}}(\mathrm{MP}) = 0$ then Lemma D.44 implies that $\mathcal{E}$ is a 0-EC. As in the proof of statement (a) we can pick an arbitrary state t of $\mathcal{E}$ and show each $\mathfrak{S}$-paths $\varsigma \in \mathit{Limit}_{\mathcal{E}}$ with $\varsigma \models \Diamond\Box(\mathrm{wgt} \geqslant K)\} > 0$ has infinitely many prefixes π with $\mathit{last}(\pi) = t$ and $\mathit{wgt}(\pi) \geqslant K_t$. This yields $\Pr^{\mathfrak{S}}_{\mathcal{M},s_{init}}(\varphi) = 1$. □

Combining Lemma D.45 with the results on DWR-problems yields:

Corollary D.46. *$\mathrm{WcoB}^{\exists,>0}$ is solvable in polynomial time, while $\mathrm{WcoB}^{\exists,=1}$ is in NP ∩ coNP and solvable in pseudo-polynomial time.*

We now return to weight-bounded Büchi constraints and show how $\mathrm{WB}^{\forall,=1}$ and $\mathrm{WB}^{\forall,>0}$ are solvable using algorithms for (the complements of) the coBüchi problems $\mathrm{WcoB}^{\exists,>0}$ and $\mathrm{WcoB}^{\exists,=1}$, respectively.

Lemma D.47. *Problem $\mathrm{WB}^{\forall,=1}$ is solvable in polynomial time.*

Proof. Let $\mathcal{M}^-$ denote the MDP resulting from $\mathcal{M}$ by multiplying all weights with -1 and let $L = -(K-1)$.

$$\begin{aligned}
(\mathcal{M}, s_{init}) \text{ satisfies } \mathrm{WB}^{\forall,=1} \quad &\text{iff} \quad \Pr^{\min}_{\mathcal{M},s_{init}}(\Box\Diamond F) = 1 \text{ and } \forall \mathfrak{S}.\ \Pr^{\mathfrak{S}}_{\mathcal{M},s_{init}}\big(\Box\Diamond(\mathrm{wgt} \geqslant K)\big) = 1 \\
&\text{iff} \quad \Pr^{\min}_{\mathcal{M},s_{init}}(\Box\Diamond F) = 1 \text{ and } \neg\exists \mathfrak{S}.\ \Pr^{\mathfrak{S}}_{\mathcal{M},s_{init}}\big(\Diamond\Box(\mathrm{wgt} < K)\big) > 0 \\
&\text{iff} \quad \Pr^{\min}_{\mathcal{M},s_{init}}(\Box\Diamond F) = 1 \text{ and } \neg\exists \mathfrak{S}.\ \Pr^{\mathfrak{S}}_{\mathcal{M}^-,s_{init}}\big(\Diamond\Box(\mathrm{wgt} \geqslant L)\big) > 0
\end{aligned}$$

Thus, $\mathrm{WB}^{\forall,=1}$ is solvable using known polynomial-time algorithms to check whether $\Pr^{\min}_{\mathcal{M},s}(\Box\Diamond F) = 1$ and an algorithm for the complement of $\mathrm{WcoB}^{\exists,>0}$. As $\mathrm{WcoB}^{\exists,>0}$ is solvable in polynomial time (Corollary D.46), so is $\mathrm{WB}^{\forall,=1}$. □

Lemma D.48. *$\mathrm{WB}^{\forall,>0}$ is in NP ∩ coNP and solvable in pseudo-polynomial time.*

Proof. As before, let $\mathcal{M}^-$ denote the MDP resulting from $\mathcal{M}$ by multiplying all weights with -1 and let $L = -(K-1)$.

$$\begin{aligned}
&(\mathcal{M}, s) \text{ satisfies } \mathrm{WB}^{\forall,>0} \\
\text{iff} \quad &\forall \mathfrak{S}.\ \Pr^{\mathfrak{S}}_{\mathcal{M},s}\big(\Box\Diamond(\mathrm{wgt} \geqslant K) \ \wedge\ \Box\Diamond F\big) > 0 \\
\text{iff} \quad &\neg\exists \mathfrak{S}.\ \Pr^{\mathfrak{S}}_{\mathcal{M},s}\big(\Diamond\Box(\mathrm{wgt} < K) \ \vee\ \Diamond\Box\neg F\big) = 1 \\
\text{iff} \quad &\neg\exists \mathfrak{S}.\ \Pr^{\mathfrak{S}}_{\mathcal{M}^-,s}\big(\Diamond\Box(\mathrm{wgt} \geqslant L) \ \vee\ \Diamond\Box\neg F\big) = 1
\end{aligned}$$

Let G denote the union of all states belonging to a (possibly non-maximal) end component $\mathcal{E}$ of $\mathcal{M}$ (or $\mathcal{M}^-$) such that $\mathcal{E} \cap F = \varnothing$. The set G is computable in polynomial time using standard techniques. Let now $\mathcal{N}$ be the MDP resulting from $\mathcal{M}^-$ by collapsing all states in G into a fresh trap state g and adding a state-action pair (g, α) where $P_{\mathcal{N}}(g, \alpha, g) = 1$ and $wgt_{\mathcal{N}}(g, \alpha) = 1$. Then:

$$\exists \mathfrak{S}.\ \Pr^{\mathfrak{S}}_{\mathcal{M}^-,s}\big(\Diamond\Box(\mathrm{wgt} \geqslant L) \ \vee\ \Diamond\Box\neg F\big) = 1 \quad\Longleftrightarrow\quad \exists \mathfrak{S}.\ \Pr^{\mathfrak{S}}_{\mathcal{N},s}\big(\Diamond\Box(\mathrm{wgt} \geqslant L)\big) = 1\ .$$

This yields $(\mathcal{M}, s)$ satisfies $\mathrm{WB}^{\forall,>0}$ iff $(\mathcal{N}, s)$ does not satisfy $\mathrm{WcoB}^{\exists,=1}$ for the weight bound L. Thus, the claim follows from Corollary D.46. □

D.5.3 Optimal Values for Weight-Bounded Büchi Constraints

The values of the optimization problems $\mathrm{WB}^{\exists,>0}$, $\mathrm{WB}^{\exists,=1}$, $\mathrm{WB}^{\forall,>0}$ and $\mathrm{WB}^{\forall,=1}$ are computable using the above reductions to the DWR problems and the algorithms presented for the optimization variants for $\mathrm{DWR}^{\exists,=1}$ and $\mathrm{DWR}^{\exists,>0}$. Thus,

$$\begin{aligned}
B^{\exists,>0}_{\mathcal{M},s} &= \max\big\{K \in \mathbb{Z} \ :\ \exists \mathfrak{S}.\ \Pr^{\mathfrak{S}}_{\mathcal{M},s}\big(\Box\Diamond(\mathrm{wgt} \geqslant K) \wedge \Box\Diamond F\big) > 0\big\} \\
B^{\forall,=1}_{\mathcal{M},s} &= \max\big\{K \in \mathbb{Z} \ :\ \forall \mathfrak{S}.\ \Pr^{\mathfrak{S}}_{\mathcal{M},s}\big(\Diamond\Box(\mathrm{wgt} \geqslant K) \wedge \Box\Diamond F\big) = 1\big\}
\end{aligned}$$

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