On Skolem-Hardness and Saturation Points in Markov Decision Processes

Jakob Piribauer [ID]
Technische Universität Dresden, Germany
jakob.piribauer@tu-dresden.de

Christel Baier [ID]
Technische Universität Dresden, Germany
christel.baier@tu-dresden.de

—— **Abstract** ————————————————————————

The Skolem problem and the related Positivity problem for linear recurrence sequences are outstanding number-theoretic problems whose decidability has been open for many decades. In this paper, the inherent mathematical difficulty of a series of optimization problems on Markov decision processes (MDPs) is shown by a reduction from the Positivity problem to the associated decision problems which establishes that the problems are also at least as hard as the Skolem problem as an immediate consequence. The optimization problems under consideration are two non-classical variants of the stochastic shortest path problem (SSPP) in terms of expected partial or conditional accumulated weights, the optimization of the conditional value-at-risk for accumulated weights, and two problems addressing the long-run satisfaction of path properties, namely the optimization of long-run probabilities of regular co-safety properties and the model-checking problem of the logic frequency-LTL. To prove the Positivity- and hence Skolem-hardness for the latter two problems, a new auxiliary path measure, called weighted long-run frequency, is introduced and the Positivity-hardness of the corresponding decision problem is shown as an intermediate step. For the partial and conditional SSPP on MDPs with non-negative weights and for the optimization of long-run probabilities of constrained reachability properties ($a \,U\, b$), solutions are known that rely on the identification of a bound on the accumulated weight or the number of consecutive visits to certain sates, called a saturation point, from which on optimal schedulers behave memoryless. In this paper, it is shown that also the optimization of the conditional value-at-risk for the classical SSPP and of weighted long-run frequencies on MDPs with non-negative weights can be solved in pseudo-polynomial time exploiting the existence of a saturation point. As a consequence, one obtains the decidability of the qualitative model-checking problem of a frequency-LTL formula that is not included in the fragments with known solutions.

1   Introduction

Markov decision processes (MDPs) (see, e.g., [40]) constitute one of the most prominent classes of operational models combining randomization and non-determinism and are widely used in verification, articifical intelligence, robotics and operations research. Consequently, a vast landscape of optimization problems on MDPs has been studied. The task usually is to find a strategy resolving the non-deterministic choices, called a *scheduler*, such that a certain objective quantity is optimized or to decide whether the optimal value exceeds a given rational threshold (*threshold problem*).

Stochastic shortest path problems (SSPPs) are one important type of such optimization problems on MDPs equipped with weights. These problems ask for a scheduler maximizing or minimizing the expected accumulated weight before reaching a designated goal state. In the classical setting, only schedulers reaching the goal almost surely are taken into consideration. This *classical SSPP* is known to be solvable in polynomial time using graph-based algorithms and linear-programming techniques [10, 20, 3]. For various purposes, the requirement that the goal has to be reached almost surely, however, is not appropriate. This applies, e.g., to work on the semantics of probabilistic programs when no guarantee on almost sure termination can be given [25, 30, 9, 15, 36], to the analysis of the behavior of fault-tolerant systems in error scenarios which occur with low probability, or to the trade-off analysis when combinations of utility and cost constraints can be achieved with positive probability, but not almost surely (see, e.g., [5]). This motivates a switch to non-classical variants of the SSPP: The *conditional SSPP* [8] asks for a scheduler optimizing the conditional expected accumulated weight before reaching the goal under the condition that the goal will indeed be reached and the *partial SSPP* [16, 38] assigns weight 0 to all executions not reaching the goal. Both variants increase the algorithmic difficulties. In the special case of MDPs with non-negative weights, exponential-time algorithms for the partial and conditional SSPP exploit the monotonicity of accumulated weights and rely on the existence of a *saturation point* (a bound for the accumulated weight) from which on optimal schedulers behave memorylessly. Apart from a PSPACE lower bound and approximation algorithms [38], no algorithms are known for solving the partial or conditional SSPP in integer-weighted MDPs.

Conditional expectations also play a crucial role in risk management: The *conditional value-at-risk* is an established risk measure quanitfying the expected loss in bad cases [45, 1]. Given a probability value p, the *value-at-risk* of a random variable X is defined as the worst p-quantile. Quantile queries on the distribution of path lengths have been studied in [44]. The conditional value-at-risk is the expectation of X under the condition that the outcome is worse than the value-at-risk. For MDPs, the conditional value-at-risk has been studied for mean-payoffs and for weighted reachability where on each run only once a terminal weight is collected when a target state is reached [31]. In this paper, we consider the conditional value-at-risk for the more general accumulated weight before reaching the goal, i.e. for the classical SSPP. To the best of our knowledge, this problem has not been studied.

Other typical optimization problems arise in the context of verification, asking for worst-case schedulers that minimize or maximize the probability of a given path property. While such problems are well-understood, e.g., for properties given by linear temporal logic (LTL)-formulas or non-deterministic Büchi-automata [19], there has been increasing interest in ways to quantify the degree to which a property is satisfied not only by the probability (see [28]). Approaches in this direction include the work on robust satisfaction of temporal specifications [32, 43], coverage semantics [17], robustness distances [13], and the more general model-measurement semantics [29] among others. Furthermore, this has lead to different

notions quantifying to which degree a property is satisfied in the long-run: *Frequency-LTL* has been introduced in [23, 24] as an extension of LTL by a frequency modality that makes assertions on the portion of time (or relative frequency of positions in paths) where a given event holds. While [23, 24] presents model-checking algorithms for Markov chains and arbitrary frequency-LTL formulas, the presented model checking algorithms for MDPs are restricted to fragments of frequency-LTL. We address the model checking problem for frequency-LTL formulas not contained in these fragments. Further, the concept of *long-run probabilities* [4] has been introduced for reasoning about the probabilities of path properties when the system is in equilibrium and can, e.g., be useful to formalize refined notions of long-run availability. In [4], a pseudo-polynomial time algorithm that exploits the existence of a saturation point for the computation of optimal long-run probabilities of constrained reachability properties ($a \cup b$) is provided. Here, we study long-run probabilities of general regular co-safety properties.

Contributions.   The main contribution of the paper is to provide evidence for the mathematical difficulty of the series of decision problems described above in terms of a reduction from the *Positivity problem* of linear recurrence sequences. The Positivity problem is closely related to the *Skolem problem*, a prominent number-theoretic decision problem for linear recurrence sequences, and the decidability of both problems has been open for many decades (see, e.g., [27]). As it is well-known that the Skolem problem is reducible to the Positivity problem, the provided reductions establish that the investigated decision problems are also at least as hard as the Skolem problem. In the middle column of Table 1, these Skolem-hardness results are listed:

🟨 **Table 1** Overview of the results.

optimization problem on MDPs	threshold problem Positivity- and hence Skolem-hard for	exponential-time algorithm using a saturation point for
partial SSPP (1)	weights in $\mathbb{Z}$, *Thm. 3*	weights in $\mathbb{N}$ [16] (PSPACE-hard, *Prop. 15*)
conditional SSPP (2)	weights in $\mathbb{Z}$, *Thm. 5*	weights in $\mathbb{N}$ [8] (PSPACE-hard [8])
conditional value-at-risk for the classical SSPP (3)	weights in $\mathbb{Z}$, *Thm. 6*	weights in $\mathbb{N}$, *Thm. 12*
long-run probability (4)	regular co-safety properties, *Thm. 9*	constrained reachability $a \cup b$ [4] (NP-hard [4])
model checking of frequency-LTL (5)	$\mathrm{Pr}^{\max}_{\mathcal{M}}(G^{>\vartheta}_{\inf}(\varphi)) = 1?$ for an LTL-formula φ, *Thm. 11*	$\mathrm{Pr}^{\max}_{\mathcal{M}}(G^{>\vartheta}_{\inf}(a \cup b)) = 1?$ *Cor. 14*

To obtain these results, we construct an MDP-gadget in which a linear recurrence relation can be encoded. Together with different gadgets encoding initial values of a linear recurrence sequence, we use this gadget to establish Positivity-hardness for problems (1)-(3). Afterwards, we introduce a notion of *weighted long-run frequency* for constrained reachability properties that can be seen as a generalization of classical limit-average weights and serves here as a technical vehicle to provide a connective link to long-run probabilities and the model-checking problem of frequency-LTL. The Positivity-hardness for problems (4) and (5) is obtained via the Positivity-hardness of the threshold problem for weighted long-run frequencies by showing how to encode integer weights in terms of the satisfaction of a fixed co-safety property. The

Positivity-hardness of (4) and (5) is somehow surprising: The non-probabilistic variant (4) is shown to be decidable in [4], while our results show that Positivity-hardness of (4) holds even for a simple fixed co-safety property given by a very small counter-free non-deterministic finite automaton. Likewise, Positivity-hardness of (5) is established already for the restriction to the almost-sure satisfaction problem of a simple fixed frequency-LTL formula.

For special cases of some of the problems studied here it is known that optimal values can be computed in exponential time exploiting a saturation point. We extend this picture by showing analogous results for problems (3) and (5) (see Table 1). In particular, we provide a simple exponential time algorithm for the computation of the optimal conditional value-at-risk for the classical SSPP. Further, we pinpoint where the Positivity-hardness of the model checking problem of frequency-LTL arises: We observe that the techniques of [4] allow to solve the qualitative model-checking problem for a frequency-LTL formula with only one constrained reachability ($a \cup b$) property under a frequency-globally modality. Our Positivity-hardness result for model checking frequency-LTL uses an only slightly more complicated fixed formula where a Boolean combination of atomic propositions and constrained reachability properties occurs in the scope of the frequency-globally modality. In particular, the Positivity-hardness does not require deeper nesting of temporal operators.

Related work.   Besides the above cited work that presents algorithms for special cases of the investigated problems, closest to our work is [2] where Skolem-hardness for decision problems for Markov chains have been established. The problems are to decide whether for given states s, t and rational number p, there is a positive integer n such that the probability to reach t from s in n steps equals p and the model checking problem for a probabilistic variant of monadic logic and a variant of LTL that treats Markov chains as linear transformers of probability distributions. These decision problems are of quite different nature than the problems studied here, and so are the reductions from the Skolem problem. In this context also the results of [18] and [34] are remarkable as they show the decidability (subject to Schanuel's conjecture) of reachability problems in continuous linear dynamical systems and continuous-time MDPs, respectively, as instances of the continuous Skolem problem.

A class of problems related to SSPPs concerns the optimization of probabilities for weight-bounded reachability properties and also exhibits increasing algorithmic difficulty (for an overview see [41]): For non-negative weights, schedulers optimizing the probability for reaching a target while the accumulated weight stays below a given bound are computable in pseudo-polynomial time and the corresponding probability-threshold problem is in P for qualitative probability thresholds (">0" or "=1") and PSPACE-hard in the general case [44, 26]. For integer weights even in finite-state Markov chains, the probabilities for a weight-bounded reachability property can be irrational. Still, decidability for analogous problems for integer-weighted MDPs have been established for certain cases. Examples are pseudo-polynomial algorithms for qualitative threshold problems in integer-weighted MDPs [14, 12, 35, 3] or an exponential-time algorithm and a PSPACE lower bound for the almost-sure termination problem in one-counter MDPs [11].

Switching to more expressive models typically leads to the undecidability of infinite-horizon verification problems. This applies, e.g., to recursive MDPs [21], MDPs with two or more weight functions [7, 42] or partially observable MDPs [33, 6]. However, we are not aware of natural decision problems for standard (finite-state) MDPs with a single weight function and single objective that are known to be undecidable.

2 Preliminaries

We give basic definitions and present our notation (for more details see, e.g., [40]). We then formally define the quantitative objectives studied in this paper.

Notations for Markov decision processes. A *Markov decision process* (MDP) is a tuple $\mathcal{M} = (S, Act, P, s_{init}, wgt, \mathsf{AP}, L)$ where S is a finite set of states, Act a finite set of actions, $s_{init} \in S$ the initial state, $P \colon S \times Act \times S \to [0, 1] \cap \mathbb{Q}$ is the transition probability function, $wgt \colon S \times Act \to \mathbb{Z}$ the weight function, AP a finite set of atomic propositions, and $L \colon S \to 2^{\mathsf{AP}}$ a labeling function. If not needed, we might drop the weight function or the labeling. We require that $\sum_{t \in S} P(s, \alpha, t) \in \{0, 1\}$ for all $(s, \alpha) \in S \times Act$. We say that action α is enabled in state s iff $\sum_{t \in S} P(s, \alpha, t) = 1$. We assume that for all states s there is an enabled action and that all states are reachable from s_{init}. We call a state absorbing if there is only one enabled action, returning to the state with probability 1 and weight 0. The paths of $\mathcal{M}$ are finite or infinite sequences $s_0 \, \alpha_0 \, s_1 \, \alpha_1 \ldots$ where states and actions alternate such that $P(s_i, \alpha_i, s_{i+1}) > 0$ for all $i \geq 0$. For $\pi = s_0 \, \alpha_0 \, s_1 \, \alpha_1 \ldots \alpha_{k-1} \, s_k$, $wgt(\pi) = wgt(s_0, \alpha_0) + \ldots + wgt(s_{k-1}, \alpha_{k-1})$ denotes the accumulated weight of π, $P(\pi) = P(s_0, \alpha_0, s_1) \cdot \ldots \cdot P(s_{k-1}, \alpha_{k-1}, s_k)$ its probability, and $last(\pi) = s_k$ its last state. Further, we also write π to denote the word $L(s_0), L(s_1), \ldots$. The *size* of $\mathcal{M}$ is the sum of the number of states plus the total sum of the logarithmic lengths of the non-zero probability values $P(s, \alpha, s')$ as fractions of co-prime integers and the weight values $wgt(s, \alpha)$. An end component of $\mathcal{M}$ is a strongly connected sub-MDP.

Scheduler. A *scheduler* for $\mathcal{M}$ is a function $\mathfrak{S}$ that assigns to each finite path π a probability distribution over $Act(last(\pi))$. If there is a finite set X of memory modes and a memory update function $U : S \times Act \times S \times X \to X$ such that the choice of $\mathfrak{S}$ only depends on the current state after a finite path and the memory mode obtained from updating the memory mode according to U in each step, we say that $\mathfrak{S}$ is a finite-memory scheduler. If the choice depends only on the current state, we say that $\mathfrak{S}$ is memoryless. A scheduler $\mathfrak{S}$ is called deterministic if $\mathfrak{S}(\pi)$ is a Dirac distribution for each path π. Given a scheduler $\mathfrak{S}$, $\zeta = s_0 \, \alpha_0 \, s_1 \, \alpha_1 \ldots$ is a $\mathfrak{S}$-path iff ζ is a path and $\mathfrak{S}(s_0 \, \alpha_0 \ldots \alpha_{k-1} \, s_k)(\alpha_k) > 0$ for all $k \geq 0$.

Probability measure. We write $\mathrm{Pr}^{\mathfrak{S}}_{\mathcal{M},s}$ or briefly $\mathrm{Pr}^{\mathfrak{S}}_s$ to denote the probability measure induced by $\mathfrak{S}$ and s. For details, see [40]. We will use LTL-like formulas to denote measurable sets of paths. Given a measurable set ψ of infinite paths, we define $\mathrm{Pr}^{\min}_{\mathcal{M},s}(\psi) = \inf_{\mathfrak{S}} \mathrm{Pr}^{\mathfrak{S}}_{\mathcal{M},s}(\psi)$ and $\mathrm{Pr}^{\max}_{\mathcal{M},s}(\psi) = \sup_{\mathfrak{S}} \mathrm{Pr}^{\mathfrak{S}}_{\mathcal{M},s}(\psi)$ where $\mathfrak{S}$ ranges over all schedulers for $\mathcal{M}$. For a random variable X defined on infinte paths in $\mathcal{M}$, we denote the expected value of X under the probability measure induced by a scheduler $\mathfrak{S}$ and state s by $\mathbb{E}^{\mathfrak{S}}_{\mathcal{M},s}(X)$. Furthermore, for a measurable set of paths ψ with positive probability, $\mathbb{E}^{\mathfrak{S}}_{\mathcal{M},s}(X|\psi)$ denotes the conditional expectation of X under ψ. If $s = s_{init}$, we sometimes drop the subscript s.

Partial and conditional SSPP. Let $\mathcal{M}$ be an MDP with an absorbing state *goal*. On infinite paths ζ, we define the random variable $\oplus goal(\zeta)$ to be $wgt(\zeta)$ if $\zeta \vDash \Diamond goal$, and to be 0 otherwise. The *partial expectation* $PE^{\mathfrak{S}}_{\mathcal{M},s}$ of a scheduler $\mathfrak{S}$ is defined as $\mathbb{E}^{\mathfrak{S}}_{\mathcal{M},s}(\oplus goal)$. The maximal partial expectation is $PE^{\max}_{\mathcal{M},s} = \sup_{\mathfrak{S}} PE^{\mathfrak{S}}_{\mathcal{M},s}$. The *conditional expectation* $CE^{\mathfrak{S}}_{\mathcal{M},s}$ is defined as the conditional expected value $\mathbb{E}^{\mathfrak{S}}_{\mathcal{M},s}(\oplus goal|\Diamond goal)$ for all schedulers reaching *goal* with positive probability, and the maximal conditional expectations is $CE^{\max}_{\mathcal{M},s} = \sup_{\mathfrak{S}} CE^{\mathfrak{S}}_{\mathcal{M},s}$ where $\mathfrak{S}$ ranges over all schedulers $\mathfrak{S}$ with $\mathrm{Pr}^{\mathfrak{S}}_{\mathcal{M},s}(\Diamond goal) > 0$. The *partial SSPP* asks for the maximal partial expectations and the *conditional SSPP* for the maximal conditional expectation. These problems were first considered in [16] and [8]. For more details see [8, 38].

Conditional value-at-risk. Given an MDP $\mathcal{M}$ with a scheduler $\mathfrak{S}$, a random variable X defined on runs of the MDP with values in $\mathbb{R}$ and a value $p \in [0, 1]$, we define the *value-at-risk* as $VaR_p^{\mathfrak{S}}(X) = \sup\{r \in \mathbb{R} | \Pr_{\mathcal{M}}^{\mathfrak{S}}(X \leq r) \leq p\}$. So, the value-at-risk is the point at which the cumulative distribution function of X reaches or exceeds p. Denote $VaR_p^{\mathfrak{S}}(X)$ by v. The *conditional value-at-risk* is now the expectation of X under the condition that the outcome belongs to the p worst outcomes. Following the treatment of random variables that are not continuous in general in [31], we define the conditional value-at-risk as follows:

$$CVaR_p^{\mathfrak{S}}(X) = 1/p(\Pr_{\mathcal{M}}^{\mathfrak{S}}(X < v) \cdot \mathbb{E}_{\mathcal{M}}^{\mathfrak{S}}(X | X < v) + (p - \Pr_{\mathcal{M}}^{\mathfrak{S}}(X < v)) \cdot v).$$

Outcomes of X which are less than v are treated differently to outcomes equal to v as it is possible that the outcome v has positive probability and we only want to account exactly for the p worst outcomes. Hence, we take only $p - \Pr_{\mathcal{M}}^{\mathfrak{S}}(X < v)$ of the outcomes which are exactly v into account as well.

Threshold problems for the conditional value-at-risk in weighted MDPs have been studied in [31] for two random variables: the mean-payoff and weighted reachability where a set of final states is equipped with terminal weights obtained when reaching these states while all other transitions have weight 0. In this paper, we will address the conditional value-at-risk for the accumulated weight before reaching *goal* in MDPs with an absorbing state *goal*: Define $\Diamond\!\!\!\Diamond goal(\zeta)$ to be $wgt(\zeta)$ if $\zeta \vDash \Diamond goal$ and leave it undefined otherwise. The optimization of the expectation of $\Diamond\!\!\!\Diamond goal$ is known as the *classical SSPP*. Note that the expectation of this random variable is only defined under schedulers reaching *goal* with probability 1.

Long-run probability. Let $\mathcal{M}$ be an MDP with states labeled by atomic propositions from AP. Let φ be a path property, i.e., a measurable set of paths. The *long-run probability* for φ of a path ζ under a scheduler $\mathfrak{S}$ is $lrp_\varphi^{\mathfrak{S}}(\zeta) = \liminf_{n \to \infty} \frac{1}{n+1} \cdot \sum_{i=0}^{n} \Pr_{\mathcal{M}, \zeta[i]}^{\mathfrak{S}\uparrow\zeta[0...i]}(\varphi)$. Here, $\zeta[0...i]$ denotes the prefix from position 0 to i of ζ, $\zeta[i]$ denotes the state after i steps, and $\mathfrak{S}\uparrow\zeta[0...i]$ denotes the residual scheduler defined by $\mathfrak{S}\uparrow\zeta[0...i](\pi) = \mathfrak{S}(\zeta[0...i] \circ \pi)$ for all finite paths π starting in $\zeta[i]$. The long-run probability of φ under scheduler $\mathfrak{S}$ is $\mathbb{LP}_{\mathcal{M}}^{\mathfrak{S}}(\varphi) = \mathbb{E}_{\mathcal{M}}^{\mathfrak{S}}(lrp_\varphi^{\mathfrak{S}})$. The maximal long-run probability for φ is $\mathbb{LP}_{\mathcal{M}}^{\max}(\varphi) = \sup_{\mathfrak{S}} \mathbb{E}_{\mathcal{M}}^{\mathfrak{S}}(lrp_\varphi^{\mathfrak{S}})$. This notion was introduced in [4]. In this paper, we are interested in two kinds of path properties: Constrained reachability, $a \cup b$, where a and b are atomic propositions and the more general regular co-safety properties given by a finite non-deterministic automaton (NFA) $\mathcal{A}$ accepting "good" prefixes of a run. For a co-safety property given by an NFA $\mathcal{A}$, we also write $\mathbb{LP}_{\mathcal{M}}^{\max}(\mathcal{A})$.

3   Skolem-hardness

The *Skolem problem* and the closely related *Positivity problem* are outstanding problems in the fields of number theory and theoretical computer science (see, e.g., [27, 37]). Their decidability has been open for many decades. We call a problem to which the Skolem problem is reducible *Skolem-hard*. This is a hardness result in the sense that a decision procedure would imply a major breakthrough by settling the decidability of the Skolem problem and it shows that a problem possesses an inherent mathematical difficulty.

Skolem problem. Given a natural number $k \geq 2$, and rationals α_i and β_j with $1 \leq i \leq k$ and $0 \leq j \leq k - 1$, let $(u_n)_{n \geq 0}$ be defined by the initial values $u_0 = \beta_0, \ldots, u_{k-1} = \beta_{k-1}$ and the linear recurrence relation $u_{n+k} = \alpha_1 u_{n+k-1} + \cdots + \alpha_k u_n$ for all $n \geq 0$. The Skolem problem is to decide whether there is an $n \in \mathbb{N}$ with $u_n = 0$.

A closely related problem is the *Positivity problem*. It asks whether $u_n \geq 0$ for all n. It is folklore that the Skolem problem is polynomial-time reducible to the positivity problem (see, e.g., [22]). We will use the Positivity problem for our reductions leading to the main result:

▶ **Main result** (Theorems 3, 5, 6, 9, 11). *The Positivity problem and hence the Skolem problem are polynomial-time reducible to the threshold problems for the partial and conditional SSPP, the conditional value-at-risk in the classical SSPP, and long-run probabilities of regular co-safety properties, as well as to the qualitative model checking problem of frequency-LTL.*

For this purpose, we will construct an MDP gadget depicted in Figure 1a that encodes a linear recurrence relation in terms of the optimal values of different quantitative objectives. For the different problems, we then provide gadgets encoding the initial values of a linear recurrence sequence. We can plug these gadgets together to obtain an MDP and a scheduler $\mathfrak{S}$ such that $\mathfrak{S}$ maximizes the respective objective iff the linear recurrence sequence has no negative member. By computing the optimal values under $\mathfrak{S}$ in the MDPs – which turn out to be rational – we provide reductions from the positivity problem to the respective threshold problems with strict inequality (see also Remark 4).

3.1   Partial and Conditional SSPP

Given a linear recurrence sequence, we construct an MDP in which the sequence is encoded in terms of optimal partial expectations. So let k be a natural number and let $(u_n)_{n\geq 0}$ be the linear recurrence sequence given by rationals α_i for $1 \leq i \leq k$ and β_j for $0 \leq j \leq k-1$ as above. As $u_{n+k} = \alpha_1 u_{n+k-1} + \cdots + \alpha_k u_n$ for all n, we see that for any positive $\lambda \in \mathbb{Q}$ the sequence $(v_n)_{n\geq 0}$ defined by $v_n = \lambda^{n+1} u_n$ satisfies $v_{n+k} = \lambda^1 \alpha_1 v_{n+k-1} + \cdots + \lambda^k \alpha_k v_n$ for all n. Furthermore, v_n is non-negative if and only if u_n is. W.l.o.g., we hence can assume that $\sum_i |\alpha_i| < \frac{1}{4}$ and that $0 \leq \beta_j < \frac{1}{4k^{2k+2}}$ for all j (for details, see the extended version [39]).

Now, we construct an MDP-gadget with an example depicted in Figure 1a. This gadget contains states *goal*, s, and t, as well as $s_1, \ldots, s_k$ and $t_1, \ldots, t_k$. In state t, an action γ is enabled which has weight 0 and leads to state t_i with probability α_i if $\alpha_i > 0$ and to state s_i with probability $|\alpha_i|$ if $\alpha_i < 0$ for all i. The remaining probability leads to *goal*. From each state t_i, there is an action leading to t with weight $-i$. The action δ enabled in s as well as the actions leading from states s_i to s are constructed in the same way. This gadget will be integrated into a larger MDP where there are no other outgoing edges from states $s_1, \ldots, s_k, t_1, \ldots, t_k$. Now, for each state q and each integer w, let $e(q, w)$ be the optimal partial expectation when starting in state q with accumulated weight w. Further, let $d(w) = e(t, w) - e(s, w)$. The simple proof of the following lemma can be found in [39] and uses that optimal partial expectations satisfy that $e(q, w) = \sum_r P(q, \alpha, r)e(r, w+wgt(q, \alpha))$ if an optimal scheduler chooses action α in state q when the accumulated weight is w.

▶ **Lemma 1.** *Let $w \in \mathbb{Z}$. If an optimal scheduler chooses action γ in t and δ in s if the accumulated weight is w, then $d(w) = \alpha_1 d(w-1) + \cdots + \alpha_k d(w-k)$.*

Now we construct a gadget that encodes the initial values $\beta_0, \ldots, \beta_{k-1}$. The gadget is depicted in Figure 1b and contains states t, s, *goal*, and *fail*. For each $0 \leq j \leq k - 1$, it additionally contains states x_j and y_j. In state x_j, there is one action enabled that leads to *goal* with probability $\frac{1}{2k^{2(k-j)}} + \beta_j$ and to *fail* otherwise. From state y_j, *goal* is reached with probability $\frac{1}{2k^{2(k-j)}}$ and *fail* otherwise. In state t, there is an action γ_j leading to x_j with weight $+k - j$ for each $0 \leq j \leq k - 1$. Likewise, in state s there is an action δ_j leading to y_j with weight $k-j$ for each $0 \leq j \leq k - 1$. We now glue together the two gadgets at states s, t, and *goal*. The cumbersome choices of probability values lead to the following lemma via straight-forward computations presented in [39].
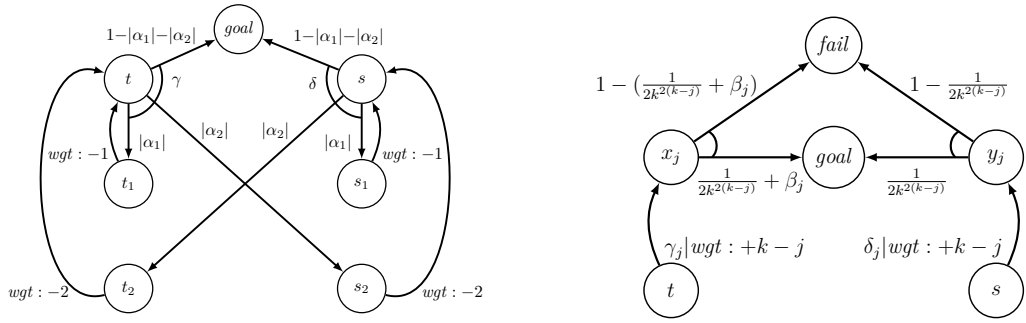
(a) In the depicted example, the recurrence depth is 2, $\alpha_1 > 0$, and $\alpha_2 < 0$.

(b) The gadget contains the depicted states and actions for each $0 \leq j \leq k - 1$.

■ **Figure 1** The gadget (a) encoding the linear recurrence relation in all reductions and (b) encoding the intial values in the reduction to the partial SSPP.

▶ **Lemma 2.** *Let* $0 \leq j \leq k - 1$. *Starting with weight* $-(k-1)+j$ *in state* t *or* s, *action* γ_j *and* δ_j *maximize the partial expectation. For positive starting weight,* γ *and* δ *are optimal.*

Comparing action γ_j and δ_j for starting weight $-(k-1)+j$, we conclude that the difference between optimal values $d(-(k-1)+j)$ is equal to β_j, for $0 \leq j \leq k-1$, and hence $d(-(k-1)+n) = u_n$ for all n. Finally, we equip the MDP with a simple initial gadget (see [39]): From the initial state s_{init}, one action with weight $+1$ is enabled. This action leads to a state c with probability $\frac{1}{2}$ and loops back to s_{init} with probability $\frac{1}{2}$. In c, the decision between action τ leading to state t and action σ leading to state s has to be made. So for any $n > 0$, state c is reached with accumulated weight n with positive probability. An optimal scheduler now has to decide whether the partial expectation when starting with weight n is better in state s or t: Action τ is optimal in c for accumulated weight w if and only if $d(w) \geq 0$. Further, the scheduler $\mathfrak{S}$ always choosing τ in c and actions $\gamma, \gamma_0, \ldots, \gamma_{k-1}, \delta, \ldots$ as described in Lemma 2 is optimal iff the given linear recurrence sequence is non-negative. We can compute the partial expectation of scheduler $\mathfrak{S}$ in the constructed MDP. The partial expectation turns out to be a rational. Hence, using this partial expectation as the threshold ϑ, we obtain the first main result. The technical proof computing the value of $\mathfrak{S}$ in the constructed MDP is given in the extended version [39].

▶ **Theorem 3.** *The Positivity problem is polynomial-time reducible to the following problem: Given an MDP* $\mathcal{M}$ *and a rational* ϑ, *decide whether* $PE_{\mathcal{M}}^{\max} > \vartheta$.

▶ **Remark 4.** There is no obvious way to adjust the construction such that the Skolem-hardness of the question whether $PE_{\mathcal{M}}^{\max} \geq \vartheta$ would follow. One attempt would be to provide an ε such that $PE_{\mathcal{M}}^{\max} > \vartheta$ iff $PE_{\mathcal{M}}^{\max} \geq \vartheta + \varepsilon$. This, however, probably requires a bound on the position at which the given linear recurrence sequence first becomes negative. But this question lies at the core of the positivity and the Skolem problem. All Skolem-hardness results in this paper hence concern only threshold problems with strict inequality.

The Skolem-hardness of the threshold problem for the conditional SSPP is obtained by a simple reduction showing that the threshold problems of the partial SSPP is polynomial-time reducible to the threshold problem of the conditional SSPP (see [39]).

▶ **Theorem 5.** *The Positivity problem is reducible in polynomial time to the following problem: Given an MDP* $\mathcal{M}$ *and a rational* ϑ, *decide whether* $CE_{\mathcal{M}}^{\max} > \vartheta$.
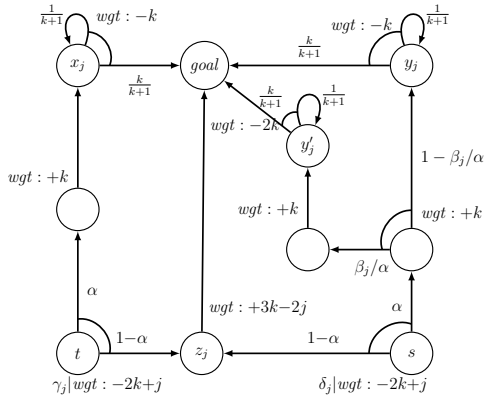
(a) The gadget contains the depicted states and actions for each $0 \leq j \leq k-1$. $\alpha = \sum_{1=i}^{k} |\alpha_i|$.
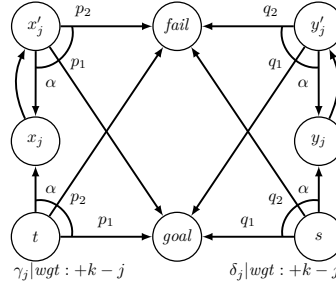
(b) The gadget contains the depicted states and actions for each $0 \leq j \leq k-1$. The probabilities are: $p_1 = (1-\alpha)(\frac{1}{2k^{2(k-j)}} + \beta_j)$, $p_2 = (1-\alpha)(1-(\frac{1}{2k^{2(k-j)}} + \beta_j))$, $q_1 = (1-\alpha)\frac{1}{2k^{2(k-j)}}$, $q_2 = (1-\alpha)(1-\frac{1}{2k^{2(k-j)}})$. All actions except for γ_j and δ_j have weight 0.

Figure 2 The gadgets encoding initial values for (a) the conditional value-at-risk for the classical SSPP and (b) weighted long-run frequencies.

3.2 Conditional value-at-risk for the classical SSPP

We reuse the gadget depicted in Figure 1a to prove the following result:

▶ **Theorem 6.** *The Positivity problem is polynomial-time reducible to the following problem: Given an MDP $\mathcal{M}$ and rationals ϑ and $p \in (0,1)$, decide whether $CVaR_p^{\max}(\diamond goal) > \vartheta$.*

We begin by the following consideration: Given an MDP $\mathcal{M}$ with initial state s_{init}, we construct a new MDP $\mathcal{N}$. We add a new initial state s'_{init}. In s'_{init}, there is only one action with weight 0 enabled leading to s_{init} with probability $\frac{1}{3}$ and to *goal* with probability $\frac{2}{3}$. So, at least two thirds of the paths accumulate weight 0 before reaching the goal. Hence, we can already say that $VaR_{1/2}^{\mathfrak{S}}(\diamond goal) = 0$ in $\mathcal{N}$ under any scheduler $\mathfrak{S}$. Note that schedulers for $\mathcal{M}$ can be seen as schedulers for $\mathcal{N}$ and vice versa. This considerably simplifies the computation of the conditional value-at-risk in $\mathcal{N}$. Define the random variable $\ominus goal(\zeta)$ to be $\diamond goal(\zeta)$ if $\diamond goal \leq 0$ and to be 0 otherwise. Now, the conditional value-at-risk for the probability value 1/2 under a scheduler $\mathfrak{S}$ in $\mathcal{N}$ is given by $CVaR_{1/2}^{\mathfrak{S}}(\diamond goal) = 2 \cdot \mathbb{E}_{\mathcal{N},s_{init}}^{\mathfrak{S}}(\ominus goal) = \frac{2}{3} \cdot \mathbb{E}_{\mathcal{M},s_{init}}^{\mathfrak{S}}(\ominus goal)$. So, the result follows from the following lemma:

▶ **Lemma 7.** *The Positivity problem is polynomial-time reducible to the following problem: Given an MDP $\mathcal{M}$ and a rational ϑ, decide whether $\mathbb{E}_{\mathcal{M},s_{init}}^{\max}(\ominus goal) > \vartheta$.*

We adjust the MDP used for the Skolem-hardness proof for the partial SSPP. So, let k be a natural number, $\alpha_1, \ldots, \alpha_k$ be rational coefficients of a linear recurrence sequence, and $\beta_0, \ldots, \beta_{k-1} \geq 0$ the rational initial values. W.l.o.g. we again assume these values to be small, namely: $\sum_{1 \leq i \leq k} |\alpha_i| \leq \frac{1}{5(k+1)}$ and for all j, $\beta_j \leq \frac{1}{3}\alpha$ where $\alpha = \sum_{1 \leq i \leq k} |\alpha_i|$.

The first important observation is that the optimal expectation of $\ominus goal$ for different starting states and starting weights behaves very similar to optimal partial expectations: For each state q and each integer w, let $e(q,w)$ be the optimal expectation of $\ominus goal$ when starting in state q with accumulated weight w. If an optimal scheduler chooses α when in q with accumulated weight w, then $e(q,w) = \sum_{r \in S} P(q, \alpha, r) \cdot e(r, w + wgt(q, \alpha))$. Reusing the MDP-gadget depicted in 1a, we observe that if we again let $d(w) = e(t,w) - e(s,w)$, the following holds as before: For any $w \in \mathbb{Z}$, if an optimal scheduler chooses action γ in t and δ in s if the accumulated weight is w, then $d(w) = \alpha_1 d(w-1) + \cdots + \alpha_k d(w-k)$.

Now, we construct a new gadget that encodes the initial values of a linear recurrence sequence. The new gadget is depicted in Figure 2a. Besides the actions γ_j and δ_j for $0 \leq j \leq k-1$ there are no non-deterministic choices. Again, we glue together the two gadgets in states s, t, and $goal$. The main idea is that for non-negative starting weights in state s or t actions γ_j and δ_j lead to a larger expected tail loss than actions γ and δ. For $0 \leq j \leq k-1$ and an accumulated weight $-k+j$ in state t or s, the actions γ_j and δ_j are, however, optimal for maximizing the expectation of $\lozenge goal$ sinve the goal is reached with non-negative weights with high probability under these actions (details in [39]). The difference of optimal values satisfies $e(t, -k+j) - e(s, -k+j) = \beta_j$ for $0 \leq j \leq k-1$ again. Finally, we add the same initial component as in the previous section and see that the scheduler $\mathfrak{S}$ always choosing τ in state c is optimal iff the linear recurrence sequence stays non-negative. As the expectation of $\lozenge goal$ under $\mathfrak{S}$ is again a rational (see [39]), this finishes the proof analogously to the previous section.

3.3 Long-run probability and frequency-LTL

In order to transfer the Skolem-hardness results to long-run probabilities and frequency-LTL, we introduce the auxiliary notion of *weighted long-run frequency*. Let $\mathcal{M}$ be an MDP with a weight function $wgt : S \times Act \rightarrow \mathbb{Z}$ and two disjoint sets of states $Goal, Fail \subseteq S$. On an infinite paths $\pi = s_0, \alpha_0, s_1, \ldots$, we define the random variable *wlf* as follows:

$$wlf(\pi) = \liminf_{n \to \infty} \frac{1}{n+1} \sum_{i=0}^{n} wgt(s_i, \alpha_i) \cdot \mathbb{1}_{\pi[i\ldots] \models \neg Fail\,\mathrm{U}\,Goal}$$

where $\mathbb{1}_{\pi[i\ldots] \models \neg Fail\,\mathrm{U}\,Goal}$ is 1 if the suffix $\pi[i \ldots] = s_i, \alpha_i, s_{i+1}, \ldots$ satisfies $\neg Fail\,\mathrm{U}\,Goal$, and 0 otherwise. Given a scheduler $\mathfrak{S}$, we define the weighted long-run frequency $WLF_{\mathcal{M}}^{\mathfrak{S}} = \mathbb{E}_{\mathcal{M}}^{\mathfrak{S}}(wlf)$ and $WLF_{\mathcal{M}}^{\max} = \sup_{\mathfrak{S}} WLF_{\mathcal{M}}^{\mathfrak{S}}$. This can be seen as a long-run average version of partial expectations. Weights are only received if afterwards $Goal$ is visited before $Fail$ and we measure the average weight received per step according to this rule. Note that we only consider the path property $\neg Fail\,\mathrm{U}\,Goal$ in this paper and hence do not include this property in our notation and terminology. An illustrating example can be found in [39].

We modifiy the MDP that was constructed in Section 3.1 for the Skolem-hardness of the partial SSPP. We replace the gadget encoding the initial values with the gadget depicted in Figure 2b. This gadget differs from the gadget used for partial expectations only in the expected time it takes to reach $goal$ or $fail$ under γ_j or δ_j. It is constructed in a way such that the expected time to reach $goal$ or $fail$ from s_{init} does not depend on the scheduler. Finally, we add a transition leading back to the initial state from $goal$ and $fail$. An optimal scheduler for weighted long-run frequencies in the constructed MDP $\mathcal{K}$ now just has to maximize the partial expectation leading to the Skolem-hardness result (for more details see [39]).

▶ **Theorem 8.** *The Positivity problem is polynomial-time reducible to the following problem: Given an MDP $\mathcal{M}$ and a rational ϑ, decide whether $WLF_{\mathcal{M}}^{\max} > \vartheta$.*

This result now serves as a tool to establish analogous results for long-run probabilities. The key idea is to encode integer weights via a labelling of states and to use a simple regular co-safety property to mimic the reception of weights in weighted long-run frequencies.

▶ **Theorem 9.** *The Positivity problem is polynomial-time reducible to the following problem: Given an MDP $\mathcal{M}$, an NFA $\mathcal{A}$, and a rational ϑ, decide whether $\mathbb{LP}_{\mathcal{M}}^{\max}(\mathcal{A}) > \vartheta$.*

In the sequel, we consider weighted states instead of weighted state-action pairs. Further, we assume that the weights are only -1, 0, and $+1$. This assumption leads to a pseudo-polynomial blow-up in the general case. The weights in the MDP $\mathcal{K}$ constructed for Theorem
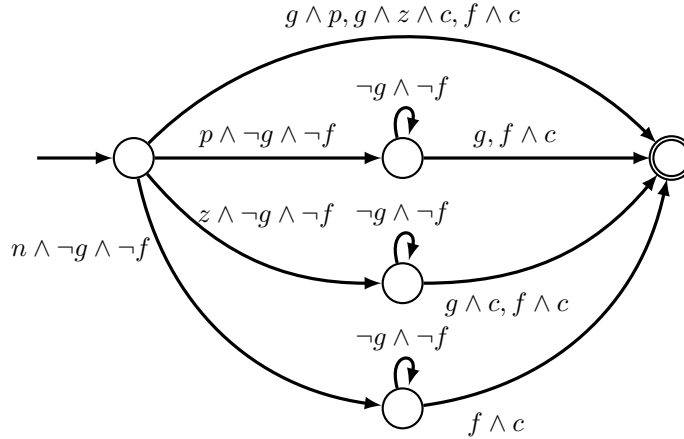
Figure 3 The NFA $\mathcal{A}$ expressing a property of the form $d \vee \bigvee_{i=1}^{3}(c_i \wedge (a \cup b_i))$.

8 above are, however, at most k. As the MDP has more than $2k$ states, transforming $\mathcal{K}$ to weights -1, 0, and $+1$ only leads to a polynomial blow-up. As this MDP has no non-trivial end-components, $\{goal, fail\}$ is visited infinitely often with probability 1 under any scheduler. Let $\mathsf{AP} = \{n, z, p, c, g, f\}$ be a set of atomic propositions representing *negative* (-1), *zero* (0), and *positive* $(+1)$ weight, *coin flip*, *goal*, and *fail*, respectively. We construct an MDP $\mathcal{L}$: The states *goal* and *fail* are duplicated while one copy of each is labeled with c and whenever *goal* or *fail* are entered in the MDP $\mathcal{K}$, both of the two copies in $\mathcal{L}$ are equally likely. For a formal definition see [39]. In Figure 3, we depict the NFA $\mathcal{A}$ used for the encoding. The NFA $\mathcal{A}$ is constructed such that in $\mathcal{L}$ any run starting in a state labeled *zero* or reaching *fail* before *goal* is accepted with probability $1/2$ due to the *coin flips*. A run starting in a state labeled *positive* and reaching *goal* before *fail* is accepted while such a path starting in a state labeled *negative* is not. This leads to the following lemma that proves Theorem 9.

▶ **Lemma 10.** *For the MDPs $\mathcal{K}$ and $\mathcal{L}$ constructed above, we have $\mathit{WLF}_{\mathcal{K}}^{\max} = \frac{1}{2} + \frac{1}{2}\mathbb{LP}_{\mathcal{L}}^{\max}(\mathcal{A})$.*

Proof sketch. It is quite easy to see that the claim holds for finite-memory schedulers as we can rely on steady state probabilities in the resulting Markov chain. That the supremum over all schedulers agrees with the supremum over finite-memory schedulers on both sides follows from Fatou's lemma. Details can be found in [39].   ◀

A consequence of this result is that model checking of frequency-LTL in MDPs is at least as hard as the Skolem problem. The decidability of the model-checking problem for the full logic frequency-LTL has been left open, but set as a goal in [23, 24]. Obtaining this goal by proving the decidability of the model-checking problem hence would settle the decidability of the Skolem problem. The frequency-globally modality $G_{\inf}^{>\vartheta}(\varphi)$ is defined to hold on a path π iff $\liminf_{n\to\infty} \frac{1}{n+1} \sum_{i=0}^{n} \mathbb{1}_{\pi[i...]\models\varphi} > \vartheta$, i.e. iff the long-run average number of positions at which a suffix satisfying φ starts exceeds ϑ.

▶ **Theorem 11.** *There is a polynomial-time reduction from the Positivity problem to the following qualitative model checking problem for frequency LTL for a fixed LTL-formula φ: Given an MDP $\mathcal{M}$ and a rational ϑ, is $\Pr_{\mathcal{M}}^{\max}(G_{\inf}^{>\vartheta}(\varphi)) = 1$?*

Proof sketch. The proof uses the reduction to the threshold problem for the long-run probability of the co-safety property expressed by $\mathcal{A}$. This property is captured by a simple LTL-formula φ (see Figure 3). For finite-memory schedulers $\mathfrak{S}$ inducing a single bottom

strongly connected component, we see that $G_{\inf}^{>\vartheta}(\varphi)$ holds with probability 1 iff the expected long-run probability of φ is greater than ϑ. That it is enough to consider such schedulers follows from the argument using Fatou's lemma again. For more details see [39].   ◀

4   Saturation points

Despite the inherent mathematical difficulty shown by the Skolem-hardness results so far, all of the problems studied here are solvable in exponential time under a natural restriction. For the problems on weighted MDPs, this restriction only allows non-negative weights while for the long-run notions the restriction to constrained reachability properties ($a \, U \, b$) leads to solvability. For the partial and the conditional SSPP [8, 16] and for long-run probabilities [4], the computability of optimal values under these restrictions has been shown. The algorithms exploit the existence of *saturation points*, a bound on the accumulated weight or the consecutive visits to certain states before optimal schedulers can behave memorylessly. We will extend this picture by providing a simple saturation point for the computation of the optimal conditional value-at-risk for the classical SSPP in MDPs with non-negative weights. Afterwards, we transfer the saturation-point algorithm from [4] to weighted long-run frequencies in the setting of non-negative weights. As a consequence, we obtain an exponential-time algorithm for the qualitative model-checking problem of a frequency-LTL formula for which no solutions were known. To conclude the section, we provide accompanying PSPACE lower bounds for the partial SSPP and weighted long-run frequencies with non-negative weights.

4.1   Conditional value-at-risk for the classical SSPP

Let $\mathcal{M}$ be an MDP with non-negative weights. In the classical SSPP, it is decidable in polynomial time whether the optimal expected accumulated weight before reaching the goal is bounded. If this is the case, the usual preprocessing step removes end components [20, 3] and transforms the MDP such that exactly the schedulers reaching the goal with probability 1 can be mimicked in the transformed MDP. So in the sequel, we assume that the absorbing state *goal* forms the only end component. Given a rational probability value $p \in (0, 1)$, we are interested in the value $CVaR_p^{\max}(\lozenge\!\!\!\!\!\lozenge goal)$. Note that in our formulation the worst outcomes are the paths with the lowest accumulated weight before reaching the goal. Below we will sketch how to treat the case where high outcomes are considered bad.

▶ **Theorem 12.** *Given an MDP $\mathcal{M} = (S, s_{init}, Act, P, wgt, goal)$ with non-negative weights and no end-components except for one absorbing state goal as well as a rational probability value $p \in (0, 1)$, the value $CVaR_p^{\max}(\lozenge\!\!\!\!\!\lozenge goal)$ is computable in pseudo-polynomial time.*

Proof sketch. As there are no end components, we can provide a *saturation point* $K \in \mathbb{N}$ such that paths accumulate a weight of more than K with probability less than $1 - p$. Then, paths reaching an accumulated weight of K do not belong to the worst p outcomes. We construct an MDP with the state space $S \times \{0, \ldots, K\}$ that encodes the accumulated weight of a path up to K. Letting states of the form $(goal, i)$ be terminal with weight i and of the form (s, K) be terminal with weight K, we can then rely on the algorithm computing the conditional value-at-risk for weighted reachability in [31]. As K can be chosen of pseudo-polynomial size and this algorithm runs in time polynomial in the size of the constructed MDP, this leads to a pseudo-polynomial time algorithm. For details see [39].   ◀

Note that the behavior of a scheduler on paths with accumulated weight above K does not matter at all for the conditional value-at-risk. If we want to consider the case where long paths are considered as bad, we can multiply all weights by -1 and use the definitions as

before. The idea here now is to compute a saturation point $-K$ such that the probability for a path to accumulate weight less than $-K$ is smaller than p. So, we know that a path with weight less than $-K$ belongs to the p worst paths. On these paths, the best thing to do in order to maximize the conditional value-at-risk is to maximize the expected accumulated weight before reaching the goal. This can be done by a memoryless deterministic scheduler simultaneously for all states and the values are computable in polynomial time [20]. Then we construct the MDP $\mathcal{N}$ as above but change the terminal weights as follows: states of the form $(goal, i)$ get weight $-i$ and states of the form (s, K) get weight $-K + \mathbb{E}_{\mathcal{M},s}^{\max}(\lozenge goal)$ where $\mathcal{M}$ is the MDP in which all weights are already multiplied by -1. Afterwards the problem can be solved by the techniques for weighted reachability from [31] again.

4.2 Weighted long-run frequencies and frequency-LTL

The existence of a saturation point for long-run probabilities of constrained reachability properties was shown in [4]. This result can easily be adapted to weighted long-run frequencies following the same arguments. First, it is shown by an application of Fatou's lemma that optimal weighted long-run frequency can be approximated by finite-memory schedulers. Afterwards, it is shown that the memory needed for the optimization can be restricted further: A saturation point $K \in \mathbb{N}$ is provided such that only scheduler keeping track of the accumulated weight up to K have to be considered. The adaptions necessary to the proof in [4] are worked out in the extended version [39] and lead to the following result:

▶ **Theorem 13.** *The maximal value $WLF_{\mathcal{M}}^{\max}$ in an MDP $\mathcal{M}$ with non-negative weights is computable in pseudo-polynomial time.*

▶ **Corollary 14.** *Given an MDP $\mathcal{M}$ and a rational ϑ, it can be checked in pseudo-polynomial time whether $\Pr_{\mathcal{M}}^{\max}(G_{\inf}^{>\vartheta}(a \operatorname{U} b)) = 1$.*

Proof. The semantics of $G_{\inf}^{>\vartheta}(\neg Fail \operatorname{U} Goal)$ on a path π agree with the semantics of $wlf(\pi) > \vartheta$ if all weights are set to $+1$. Now, we can check for each end component $\mathcal{E}$ of $\mathcal{M}$ whether $WLF_{\mathcal{E}}^{\max} > \vartheta$. If that is the case, there is a finite memory scheduler for $\mathcal{E}$ inducing only one BSCC achieving a weighted long-run frequency greater than ϑ. Under this scheduler almost all paths π satisfy $wlf(\pi) > \vartheta$. Afterwards, it remains to check whether end components with such a scheduler can be reached with probability 1 in $\mathcal{M}$. ◀

In [24], the fragment of frequency-LTL in which no until-operators occur in the scope of a globally operator has been studied. The formula in the corollary is hence of the simplest form of frequency-LTL formulas for which no solution to the qualitative model-checking problem has been known. Remarkably, the formula used in the Skolem-hardness proof (Theorem 11) is only slightly more complicated as it contains a Boolean combination of constrained reachability properties and atomic propositions under the frequency-globally operator.

4.3 PSPACE lower bounds

For the conditional SSPP with non-negative weights [8] and the long-run probability of constrained reachability properties [4], PSPACE and NP lower bounds, respectively, are known indicating that the pseudo-polynomial time algorithms for the computation can probably not be significantly improved. The threshold problem of the conditional SSPP is already PSPACE-hard in acyclic MDPs with non-negative weights as shown in [8]. In [38], it has been shown that the threshold problem of the conditional SSPP is polynomial-time

reducible to the threshold problem for the partial SSPP. This reduction generates an MDP with negative weights, even when all weights in the original MDP are non-negative. Here, we provide a new polynomial reduction for acyclic MDPs from the threshold problem for the conditional SSPP to the threshold problem of the partial SSPP that preserves the non-negativity of weights (see [39]).

▶ **Proposition 15.** *The threshold problem of the partial SSPP is PSPACE-hard in acyclic MDPs with non-negative weights. It is contained in PSPACE for acyclic MDPs with arbitrary integer weights.*

In an acyclic MDP, we can add intermediate states on transitions such that all paths have the same length ℓ. If we additionally add transitions form *goal* and *fail* back to the initial state, the maximal weighted long-run frequency is just the maximal partial expectation divided by ℓ. This allows us to conclude:

▶ **Proposition 16.** *The threshold problem for weighted long-run frequencies, "Does $WLF_{\mathcal{M}}^{\max} \bowtie \vartheta$ hold?", in MDPs with non-negative weights is PSPACE-hard.*

5   Conclusion

We identified a variety of optimization problems – some of which seemed rather unrelated on first sight – with a Skolem-hard threshold problem on MDPs. The results show that an algorithm for the exact solution to these optimization problems would imply a major breakthrough. For the partial and conditional SSPP, however, approximation algorithms were provided in [38]. Investigating the possibility to approximate optimal values might lead to algorithms useful in practice for the other objectives studied here. Further, the problems have a pseudo-polynomial solution under natural restrictions. The key result, the existence of a saturation point, has been established in the setting of stochastic multiplayer games for partial expectations [16]. This raises the question to which extend the saturation point results for the other problems can be transferred to stochastic multiplayer games.

To the best of our knowledge, the conditional value-at-risk for accumulated weights has not been addressed before. While we showed Skolem-hardness in the general setting, the computation of the optimal value is possible in exponential time in the setting of non-negative weights. Studying lower bounds for the complexity of the threshold problem and the combination of constraints on the expected accumulated weight before reaching the goal, the value-at-risk, and the conditional value-at-risk in this setting are left as future work.

References

1   Carlo Acerbi and Dirk Tasche. Expected shortfall: A natural coherent alternative to value at risk. *Economic Notes*, 31(2):379–388, 2002. `doi:10.1111/1468-0300.00091`.

2   S Akshay, Timos Antonopoulos, Joël Ouaknine, and James Worrell. Reachability problems for Markov chains. *Information Processing Letters*, 115(2):155–158, 2015.

3   Christel Baier, Nathalie Bertrand, Clemens Dubslaff, Daniel Gburek, and Ocan Sankur. Stochastic shortest paths and weight-bounded properties in Markov decision processes. In *Proceedings of the 33rd Annual ACM/IEEE Symposium on Logic in Computer Science (LICS'18)*, pages 86–94. ACM, 2018. `doi:10.1145/3209108.3209184`.

4   Christel Baier, Nathalie Bertrand, Jakob Piribauer, and Ocan Sankur. Long-run satisfaction of path properties. In *Proceedings of the Thirty-Fourth Annual ACM/IEEE Symposium on Logic in Computer Science (LICS'19)*, pages 1–14. IEEE, 2019.

5   Christel Baier, Clemens Dubslaff, Joachim Klein, Sascha Klüppelholz, and Sascha Wunderlich. Probabilistic model checking for energy-utility analysis. In Franck van Breugel, Elham Kashefi, Catuscia Palamidessi, and Jan Rutten, editors, *Horizons of the Mind. A Tribute to Prakash Panangaden*, volume 8464 of *LNCS*, pages 96–123. Springer, 2014.

6   Christel Baier, Marcus Größer, and Nathalie Bertrand. Probabilistic ω-automata. *Journal of the ACM*, 59(1):1:1–1:52, 2012. `doi:10.1145/2108242.2108243`.

7   Christel Baier, Joachim Klein, Sascha Klüppelholz, and Sascha Wunderlich. Weight monitoring with linear temporal logic: Complexity and decidability. In *Proceedings of the 23rd Conference on Computer Science Logic and the 29th Symposium on Logic In Computer Science (CSL-LICS'14)*, pages 11:1–11:10. ACM, 2014. `doi:10.1145/2603088.2603162`.

8   Christel Baier, Joachim Klein, Sascha Klüppelholz, and Sascha Wunderlich. Maximizing the conditional expected reward for reaching the goal. In Axel Legay and Tiziana Margaria, editors, *Tools and Algorithms for the Construction and Analysis of Systems (TACAS'17)*, volume 10206 of *Lecture Notes in Computer Science*, pages 269–285. Springer, 2017.

9   Gilles Barthe, Thomas Espitau, Luis María Ferrer Fioriti, and Justin Hsu. Synthesizing probabilistic invariants via Doob's decomposition. In Swarat Chaudhuri and Azadeh Farzan, editors, *Proceedings of the 28th International Conference on Computer Aided Verification (CAV'16), Part I*, volume 9779 of *LNCS*, pages 43–61. Springer, 2016.

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

11   Tomás Brázdil, Václav Brožek, Kousha Etessami, Antonín Kučera, and Dominik Wojtczak. One-counter Markov decision processes. In *Proceedings of the twenty-first annual ACM-SIAM symposium on Discrete Algorithms (SODA'10)*, pages 863–874. SIAM, 2010.

12   Tomás Brázdil, Antonín Kucera, and Petr Novotný. Optimizing the expected mean payoff in energy Markov decision processes. In *14th International Symposium on Automated Technology for Verification and Analysis (ATVA'16)*, volume 9938 of *Lecture Notes in Computer Science*, pages 32–49, 2016.

13   Pavol Cerný, Thomas A. Henzinger, and Arjun Radhakrishna. Simulation distances. *Theoretical Computer Science*, 413(1):21–35, 2012.

14   Krishnendu Chatterjee and Laurent Doyen. Energy and mean-payoff parity Markov decision processes. In *Proceedings of the 36th International Symposium on Mathematical Foundations of Computer Science (MFCS'11)*, volume 6907 of *Lecture Notes in Computer Science*, pages 206–218. Springer, 2011.

15   Krishnendu Chatterjee, Hongfei Fu, and Amir Kafshdar Goharshady. Termination analysis of probabilistic programs through Positivstellensatz's. In Swarat Chaudhuri and Azadeh Farzan, editors, *Proceedings of the 28th International Conference on Computer Aided Verification (CAV'16), Part I*, volume 9779 of *LNCS*, pages 3–22. Springer, 2016.

16   Taolue Chen, Vojtěch Forejt, Marta Kwiatkowska, David Parker, and Aistis Simaitis. Automatic verification of competitive stochastic systems. *Formal Methods in System Design*, 43(1):61–92, 2013.

17   Hana Chockler, Orna Kupferman, and Moshe Y. Vardi. Coverage metrics for temporal logic model checking. *Formal Methods in System Design*, 28(3):189–212, 2006.

18   Ventsislav Chonev, Joël Ouaknine, and James Worrell. On the Skolem problem for continuous linear dynamical systems. In *43rd International Colloquium on Automata, Languages, and Programming, (ICALP'16)*, volume 55 of *LIPIcs*, pages 100:1–100:13. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2016.

19   Costas Courcoubetis and Mihalis Yannakakis. The complexity of probabilistic verification. *Journal of the ACM*, 42(4):857–907, 1995.

20   Luca de Alfaro. Computing minimum and maximum reachability times in probabilistic systems. In Jos C. M. Baeten and Sjouke Mauw, editors, *Proceedings of the 10th International Conference on Concurrency Theory (CONCUR'99)*, volume 1664 of *Lecture Notes in Computer Science*, pages 66–81. Springer, 1999.

21   Kousha Etessami and Mihalis Yannakakis. Recursive Markov decision processes and recursive stochastic games. In Luís Caires, Giuseppe F. Italiano, Luís Monteiro, Catuscia Palamidessi, and Moti Yung, editors, *Proceedings of the 32nd International Colloquium on Automata, Languages and Programming (ICALP'05)*, volume 3580 of *Lecture Notes in Computer Science*, pages 891–903. Springer, 2005.

22   Graham Everest, Alfred Jacobus Van Der Poorten, Igor Shparlinski, Thomas Ward, et al. *Recurrence sequences*, volume 104. American Mathematical Society, 2003.

23   Vojtech Forejt and Jan Krcál. On frequency LTL in probabilistic systems. In Luca Aceto and David de Frutos-Escrig, editors, *26th International Conference on Concurrency Theory (CONCUR'15)*, volume 42 of *LIPIcs*, pages 184–197. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2015. `doi:10.4230/LIPIcs.CONCUR.2015.184`.

24   Vojtech Forejt, Jan Krcál, and Jan Kretínský. Controller synthesis for MDPs and frequency LTL$_{\backslash GU}$. In Martin Davis, Ansgar Fehnker, Annabelle McIver, and Andrei Voronkov, editors, *20th International Conference on Logic for Programming, Artificial Intelligence, and Reasoning (LPAR'15)*, volume 9450 of *Lecture Notes in Computer Science*, pages 162–177. Springer, 2015. `doi:10.1007/978-3-662-48899-7_12`.

25   Friedrich Gretz, Joost-Pieter Katoen, and Annabelle McIver. Operational versus weakest pre-expectation semantics for the probabilistic guarded command language. *Performance Evaluation*, 73:110–132, 2014.

26   Christoph Haase and Stefan Kiefer. The odds of staying on budget. In Magnús M. Halldórsson, Kazuo Iwama, Naoki Kobayashi, and Bettina Speckmann, editors, *Proceedings of the 42nd International Colloquium Automata, Languages, and Programming (ICALP'15)*, volume 9134, pages 234–246. Springer, 2015.

27   Vesa Halava, Tero Harju, Mika Hirvensalo, and Juhani Karhumäki. Skolem's problem–on the border between decidability and undecidability. Technical report, Technical Report 683, Turku Centre for Computer Science, 2005.

28   Thomas A. Henzinger. Quantitative reactive modeling and verification. *Computer Science - Research and Development*, 28(4):331–344, November 2013. `doi:10.1007/s00450-013-0251-7`.

29   Thomas A. Henzinger and Jan Otop. From model checking to model measuring. In *24th International Conference on Concurrency Theory (CONCUR'13)*, volume 8052 of *Lecture Notes in Computer Science*, pages 273–287. Springer, 2013.

30   Joost-Pieter Katoen, Friedrich Gretz, Nils Jansen, Benjamin Lucien Kaminski, and Federico Olmedo. Understanding probabilistic programs. In Roland Meyer, André Platzer, and Heike Wehrheim, editors, *Correct System Design - Proceedings of the Symposium in Honor of Ernst-Rüdiger Olderog on the Occasion of His 60th Birthday*, volume 9360 of *LNCS*, pages 15–32. Springer, 2015.

31   Jan Kretínský and Tobias Meggendorfer. Conditional value-at-risk for reachability and mean payoff in Markov decision processes. In *Proceedings of the 33rd Annual ACM/IEEE Symposium on Logic in Computer Science (LICS'18)*, pages 609–618. ACM, 2018. `doi:10.1145/3209108.3209176`.

32   Orna Kupferman and Moshe Y. Vardi. Robust satisfaction. In *10th International Conference on Concurrency Theory (CONCUR'99)*, volume 1664 of *Lecture Notes in Computer Science*, pages 383–398. Springer, 1999.

33   Omid Madani, Steve Hanks, and Anne Condon. On the undecidability of probabilistic planning and infinite-horizon partially observable Markov decision problems. In *Proceedings of the Sixteenth National Conference on Artificial Intelligence (AAAI'99)*, pages 541–548. MIT Press, 1999.

34   Rupak Majumdar, Mahmoud Salamati, and Sadegh Soudjani. On decidability of time-bounded reachability in CTMDPs. In *47th International Colloquium on Automata, Languages, and Programming, (ICALP'20)*, to appear.

35   Richard Mayr, Sven Schewe, Patrick Totzke, and Dominik Wojtczak. MDPs with energy-parity objectives. In *32nd Annual ACM/IEEE Symposium on Logic in Computer Science (LICS'17)*, IEEE Computer Society, pages 1–12, 2017.

36   Federico Olmedo, Friedrich Gretz, Nils Jansen, Benjamin Lucien Kaminski, Joost-Pieter Katoen, and Annabelle Mciver. Conditioning in probabilistic programming. *ACM Transactions on Programming Languages and Systems*, 40(1):4:1–4:50, 2018.

37   Joël Ouaknine and James Worrell. Positivity problems for low-order linear recurrence sequences. In *Proceedings of the twenty-fifth annual ACM-SIAM symposium on Discrete algorithms (SODA'14)*, pages 366–379. SIAM, 2014.

38   Jakob Piribauer and Christel Baier. Partial and conditional expectations in Markov decision processes with integer weights. In Mikolaj Bojanczyk and Alex Simpson, editors, *Proceedings of the 22nd International Conference on Foundations of Software Science and Computation Structures (FoSSaCS'19)*, volume 11425 of *Lecture Notes in Computer Science*, pages 436–452. Springer, 2019.

39   Jakob Piribauer and Christel Baier. On Skolem-hardness and saturation points in Markov decision processes, 2020. available at `arXiv:2004.11441 [cs.LO]`.

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

41   Mickael Randour, Jean-François Raskin, and Ocan Sankur. Variations on the stochastic shortest path problem. In *International Workshop on Verification, Model Checking, and Abstract Interpretation*, pages 1–18. Springer, 2015.

42   Mickael Randour, Jean-François Raskin, and Ocan Sankur. Percentile queries in multi-dimensional Markov decision processes. *Formal methods in system design*, 50(2-3):207–248, 2017.

43   Paulo Tabuada and Daniel Neider. Robust linear temporal logic. In *25th EACSL Annual Conference on Computer Science Logic (CSL)*, volume 62 of *LIPIcs*, pages 10:1–10:21. Schloss Dagstuhl - Leibniz-Zentrum fuer Informatik, 2016.

44   Michael Ummels and Christel Baier. Computing quantiles in Markov reward models. In Frank Pfenning, editor, *Proc. of the 16th International Conference on Foundations of Software Science and Computation Structures (FoSSaCS'13)*, volume 7794 of *Lecture Notes in Computer Science*, pages 353–368. Springer, 2013. `doi:10.1007/978-3-642-37075-5_23`.

45   Stanislav Uryasev. Conditional value-at-risk: optimization algorithms and applications. In *Proc. Computational Intelligence and Financial Engineering (CIFEr'00)*, pages 49–57. IEEE, 2000.