Positivity-hardness results on Markov decision processes

Received Mar 6, 2023
Revised Oct 11, 2023
Accepted Feb 23, 2024
Published Apr 7, 2024

Key words and phrases
Positivity problem for linear recurrence sequence, Skolem problem, Markov decision processes

Jakob Piribauer [a]
Christel Baier [a]

[a] Technische Universität Dresden

arXiv:2302.13675v2 [cs.LO] 5 Apr 2024

ABSTRACT. This paper investigates a series of optimization problems for one-counter Markov decision processes (MDPs) and integer-weighted MDPs with finite state space. Specifically, it considers problems addressing termination probabilities and expected termination times for one-counter MDPs, as well as satisfaction probabilities of energy objectives, conditional and partial expectations, satisfaction probabilities of constraints on the total accumulated weight, the computation of quantiles for the accumulated weight, and the conditional value-at-risk for accumulated weights for integer-weighted MDPs. Although algorithmic results are available for some special instances, the decidability status of the decision versions of these problems is unknown in general.

The paper demonstrates that these optimization problems are inherently mathematically difficult by providing polynomial-time reductions from the Positivity problem for linear recurrence sequences. This problem is a well-known number-theoretic problem whose decidability status has been open for decades and it is known that decidability of the Positivity problem would have far-reaching consequences in analytic number theory. So, the reductions presented in the paper show that an algorithmic solution to any of the investigated problems is not possible without a major breakthrough in analytic number theory. The reductions rely on the construction of MDP-gadgets that encode the initial values and linear recurrence relations of linear recurrence sequences. These gadgets can flexibly be adjusted to prove the various Positivity-hardness results.

This paper is an extension of work published at ICALP 2020 [**58**]. It extends the conference version by the results for one-counter MDPs, energy objectives, quantiles, and cost problems. These additional results are also presented in the PhD thesis [**57**]. Furthermore, full proofs omitted in the conference version and a detailed and improved description of all constructions are provided.

Funding This work was partly funded by the DFG Grant 389792660 as part of TRR 248 (Foundations of Perspicuous Software Systems).

Cite as Jakob Piribauer, Christel Baier. Positivity-hardness results on Markov decision processes. TheoretiCS, Volume 3 (2024), Article 9, 1-47.

https://theoretics.episciences.org
DOI 10.46298/theoretics.24.9

1. Introduction

When modelling and analyzing computer systems and their interactions with their environment, two qualitatively different kinds of uncertainty about the evolution of the system execution play a central role: non-determinism and probabilism. If a system is, for example, employed in an unknown environment or depends on user inputs or concurrent processes, modelling the system as non-deterministic accounts for all possible external influences, sequences of user inputs, or possible orders in which concurrent events take place. If transition probabilities between the states of a system, such as the failure probability of components or the probabilities in a probabilistic choice employed in a randomized algorithm, are known or can be estimated, it is appropriate to model this behavior as probabilistic. A pure worst- or best-case analysis is not very informative in such cases and the additional probabilistic information available should be put to use. *Markov decision processes (MDPs)* are a standard operational model combining non-deterministic and probabilistic behavior and are widely used in operations research, artificial intelligence, and verification among others.

In each state of an MDP, there is a non-deterministic choice from a set of actions. Each action specifies a probability distribution over the possible successor states according to which a transition is chosen randomly. Typical optimization problems on MDPs require resolving the non-deterministic choices by specifying a *scheduler* such that a quantitative objective function is optimized. For example, the standard model-checking problem asks for the minimal or maximal probability that an execution satisfies a given linear-time property. Here, minimum and maximum range over all resolutions of the non-deterministic choices, i.e., over all schedulers. This model-checking problem is known to be 2EXPTIME-complete if the property is given in linear temporal logic (LTL) [29] and solvable in polynomial time if the property is given by a deterministic automaton [30, 10]. Many quantitative aspects of a system can be modeled by equipping an MDP with weights that are collected in each step. These weights might represent time, energy consumption, utilities, or generally speaking any sort of costs or rewards incurred. Classical optimization problems in this context that are known to be solvable in polynomial time include the optimization of the expected value of the total accumulated weight before a target state is reached, the so-called *stochastic shortest path problem* (SSPP) [16, 30, 5], the expected value of the reward earned on average per step, the so-called expected *mean payoff* or *long-run average*, or the expected *discounted accumulated weight* where after each step a discount factor is applied to all future weights (for the latter two, see, e.g., [40, 60]).

Of course, there is a vast landscape of further optimization problems on finite-state MDPs that have been analyzed. We are, nevertheless, 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. Undecidability results have been established for more expressive models.

This applies, e.g., to recursive MDPs [32], MDPs with two or more weight functions [12, 61], or partially observable MDPs [44, 9].

In this paper, we will investigate a series of optimization problems that have been studied in the literature, but are open in general. We will show that these problems possess an inherent mathematical difficulty that makes algorithmic solutions impossible without a major breakthrough in analytic number theory. Formally, this result is obtained by reductions from the *Positivity problem* for *linear recurrence sequences*, a number theoretic problem whose decidability status has been open for many decades (see, e.g., [39, 51, 52]).

1.1 Positivity problem

DEFINITION 1.1 (Positivity problem). The Positivity problem for linear recurrence sequences asks whether such a sequence stays non-negative. More formally, 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 Positivity problem asks to decide whether $u_n \geq 0$ for all n.[1] The number k is called the order of the linear recurrence sequence.

The Positivity problem is closely related to the famous *Skolem problem*. The Skolem problem asks whether there is an n such that $u_n = 0$ for a given linear recurrence sequence $(u_n)_{n\geq 0}$. It is well-known that the Skolem problem is polynomial-time reducible to the Positivity problem (see, e.g., [33]). The Positivity problem and the Skolem problem are outstanding problems in the fields of number theory and theoretical computer science (see, e.g., [39, 51, 52]) and their decidability has been open for many decades. Deep results establish decidability for both problems for linear recurrence sequences of low order or for restricted classes of sequences [62, 66, 53, 54, 55]. A proof of decidability or undecidability of the Positivity problem for arbitrary sequences, however, withstands all known number-theoretic techniques. In [54], it is shown that decidability of the Positivity problem (already for linear recurrence sequences of order 6) would entail a major breakthrough in the field of Diophantine approximation of transcendental numbers, an area of analytic number theory.

We call a problem to which the Positivity problem is reducible *Positivity-hard*. From a complexity theoretic point of view, the Positivity problem is known to be at least as hard as the decision problem for the universal fragment of the theory of the reals with addition, multiplication, and order [55], a problem known to be coNP-hard and to lie in PSPACE [22]. As most of the problems we will address are PSPACE-hard, the reductions in this paper do not

1 We do not distinguish between the Positivity problem and its complement in the sequel. So, we also refer to the problem whether there is an n such that $u_n < 0$ as the Positivity problem.

provide new lower bounds on the computational complexity. The hardness results in this paper hence refer to the far-reaching consequences on major open problems that a decidability result would imply. Furthermore, of course, the undecidability of the Positivity problem would entail the undecidability of any Positivity-hard problem.

1.2 Problems under investigation and related work on these problems

In the sequel, we briefly describe the problems studied in this paper and describe related work on these problems. In general, the decidability status of all of these problems is open and we will prove them to be Positivity-hard.

Energy objectives, one-counter MDPs, and quantiles. If weights model a resource like energy that can be consumed and gained during a system execution, a natural problem is to determine the worst- or best-case probability that the system never runs out of the resource. This is known as the *energy objective*. There has been work on combinations of the energy objective with further objectives such as parity objectives [23, 46] and expected mean payoffs [20]. Previous work on this objective focused on the possibility to satisfy the objective (or the combination of objectives) almost surely. The quantitative problem whether it is possible to satisfy an energy objective with probability greater than some threshold p is open.

The complement of the energy objective can be found in the context of *one-counter MDPs* (see [19, 18, 21]): Equipping an MDP with a counter that can be increased and decreased can be used to model a simple form of recursion and can be seen as a special case of pushdown MDPs. The process is said to terminate as soon as the counter value drops below 0 and the standard task is to compute maximal or minimal termination probabilities. In one-counter MDPs that terminate almost surely, one furthermore can ask for the extremal expected termination times, i.e. the expected number of steps until termination. On the positive side, for one-counter MDPs, it is decidable in polynomial time whether there is a scheduler that ensures termination with probability 1 [19]. Furthermore, *selective termination*, which requires termination to occur inside a specified set of states can be decided in exponential time [19]. On the other hand, the computation of the optimal value and the quantitative decision problem whether the optimal value exceeds a threshold p are left open in the literature. For selective termination, even the question whether the supremum of termination probabilities over all schedulers is 1 is open. Furthermore, also the problem to compute the minimal or maximal expected termination time of a one-counter MDP that terminates almost surely under any scheduler is open. There are, however, approximation algorithms for the optimal termination probability [18] and for the expected termination time of almost surely terminating one-counter MDPs [21]. One-counter MDPs can be seen as a special case of recursive MDPs [32]. For general recursive MDPs, the qualitative decision problem whether the maximal termination probability is 1 is undecidable while for restricted forms, so-called 1-exit recursive MDPs, the qualitative and also

the quantitative problem is decidable in polynomial space [32]. One-counter MDPs can be seen as a special case of 1-box recursive MDPs in the terminology of [32], a restriction orthogonal to 1-exit recursive MDPs.

The termination probability of one-counter MDPs and the satisfaction probability of the energy objective are closely related to the computation of *quantiles* (see [64, 7, 61]). Given a probability value p, here the task is to compute the best bound b such that the maximal or minimal probability that the accumulated weight exceeds the bound is at most or at least p. The decision version whether the maximal or minimal probability that the accumulated weight before reaching a target state exceeds b is at least or at most p is also known as the *cost problem* (see [37, 38, 5]). The computation of quantiles and the cost problem have been addressed for MDPs with non-negative weights and are solvable in exponential time in this setting [64, 37]. The decision version of the cost problem with non-negative weights is furthermore PSPACE-hard for a single inequality on the accumulated weight and EXPTIME-complete if a Boolean combination of inequality constraints on the accumulated weight is considered [37]. For the setting with arbitrary weights, [5] provides solutions to the qualitative question whether a constraint on the accumulated weight is satisfied with probability 1 (or > 0). Further, it is known that the quantitative problem is undecidable if multiple objectives with multiple weight functions have to be satisfied simultaneously [61].

Non-classical stochastic shortest path problems (SSPPs). The classical SSPP described above requires that a goal state is reached almost surely. In many situations, however, there might be no schedulers reaching the target with probability 1 or schedulers that miss the target with positive probability are of interest, too. Two non-classical variants that drop this requirement are the conditional SSPP (see [11, 59]) and the partial SSPP (see [25, 26]). In the conditional SSPP, the goal is to optimize the conditional expected accumulated weight before reaching the target under the condition that the target is reached. In other words, the average weight of all paths reaching the target has to be optimized. In the partial SSPP, paths not reaching the target are not ignored, but assigned weight 0. Possible applications for these non-classical SSPPs include the analysis of probabilistic programs where no guarantees on almost sure termination can be given (see, e.g., [36, 42, 13, 24, 50]), the analysis of fault-tolerant systems where error scenarios might occur with small, but positive probability, or the trade-off analysis with conjunctions of utility and cost constraints that are achievable with positive probability, but not almost surely (see, e.g., [8]). In [25] and [11], partial and conditional expectations, respectively, have been addressed in the setting of non-negative weights. In both-cases, the optimal value can be computed in exponential time [25, 11] while the threshold problem is PSPACE-hard [59, 11]. In MDPs with positive and negative weights, it is known that the optimal values might be irrational and that optimal schedulers might require infinite memory [59].

Conditional expectations also play an important role for some risk measures. The *conditional value-at-risk (CVaR)* is an established risk measure (see, e.g., [65, 1]) defined as the conditional expected outcome under the condition that the outcome belongs to the p worst outcomes for a given probability value p. In the context of optimization problems on weighted MDPs, the CVaR has been studied for mean-payoffs and weighted reachability where only one terminal weight is collected per run (see [43]), and for the accumulated weight before reaching a target state in MDPs with non-negative weights (see [3]). The CVaR for accumulated weights can be optimized in MDPs with non-negative weights in exponential time [58, 48].

1.3 Contribution

We develop a technique to provide reductions from the Positivity problem to threshold problems on MDPs, asking whether the optimal value of a quantity *strictly* exceeds a given rational threshold. The resulting reductions are based on the construction of MDP-gadgets that allow to encode the linear recurrence relation of a linear recurrence sequence and the initial values, respectively. The approach turns out to be quite flexible. By adjusting the gadgets encoding initial values, we can provide reductions of the same overall structure for several of the optimization problems we discussed. Through further chains of reductions depicted in Figure 1, we establish Positivity-hardness for the full series of optimization problems under investigation. The main result of this paper consequently is the following:

Main result. The Positivity problem is polynomial-time reducible to the threshold problems for the optimal values of the following quantities:

- termination probabilities of one-counter MDPs,
- expected termination times of almost surely terminating one-counter MDPs,
- the satisfaction probabilities of energy objectives in MDPs with weights in $\mathbb{Z}$,
- the probability to satisfy an inequality on the accumulated weight (cost problem) in MDPs with weights in $\mathbb{Z}$,
- conditional expectations (conditional SSPP) in MDPs with weights in $\mathbb{Z}$,
- partial expectations (partial SSPP) in MDPs with weights in $\mathbb{Z}$,
- conditional values-at-risk for accumulated weights (before reaching a goal) in MDPs with weights in $\mathbb{Z}$, and
- a two-sided version of partial expectations in MDPs with two non-negative weight functions with values in $\mathbb{N}$.

Furthermore, an algorithm for

- the computation of quantiles for accumulated weights in MDPs with weights in $\mathbb{Z}$

would imply the decidability of the Positivity-problem.

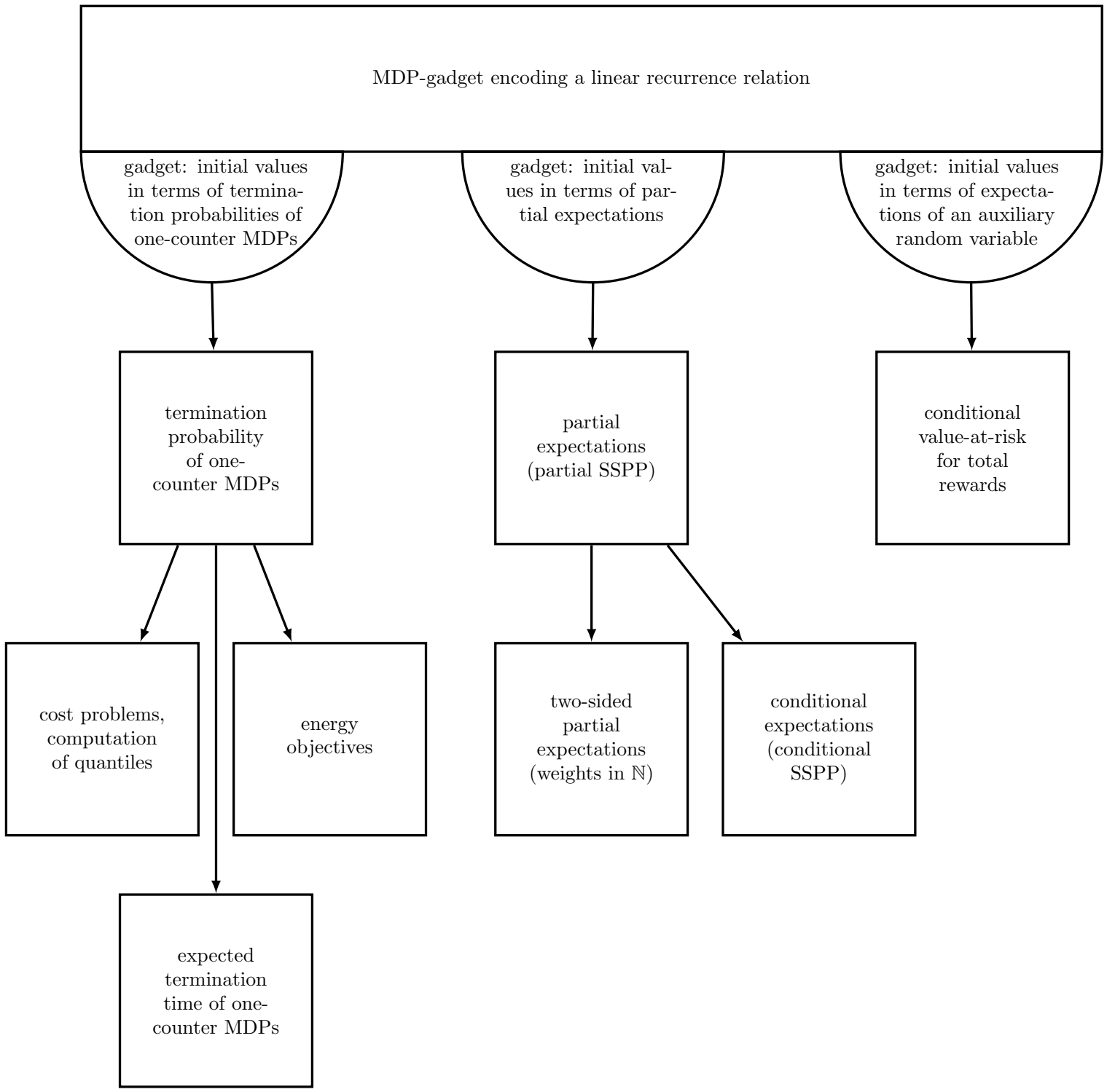

Figure 1. Overview of the dependencies between the Positivity-hardness results. The squares refer to the threshold problems for the respective quantities.

1.4 Related work on Skolem- and Positivity-hardness in verification

In [4], the Positivity-hardness of decision problems for Markov chains has been established. The problems studied in [4] are (1) 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 is at least p, and (2) 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. A connection between similar problems and the Skolem problem and Positivity problem has also been conjectured in [14, 2]. These decision problems are of quite different nature than the problems studied here. In particular, the problems are shown to be Positivity-hard in Markov chains. In contrast, e.g., partial and conditional expectations in Markov chains can be computed in polynomial time [59] and the threshold problem for the termination probability of recursive Markov chains, which subsume one-counter Markov chains, can be solved in polynomial space [31]. So, the Positivity-hardness of the corresponding problems on MDPs is not inherited from Positivity-hardness on Markov chains. Instead, our reductions show how the non-determinism in MDPs allows encoding linear recurrence sequences in terms of optimal values of various quantitative objectives by forcing an optimal scheduler to take certain decisions. Consequently, the reductions are of a different nature than the reductions in [4]. There, the behavior of a Markov chains in n steps can directly be expressed by P^n where P is the transition probability matrix, which resembles the matrix formulation of the Positivity problem, which asks for a matrix M and an initial vector v whether there is an n such that $M^n v$ lies within a half-space H.

In this context also the results of [27] and [45] 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. In other areas of formal verification, the Skolem problem and the Positivity problem play an important role in the context of the termination of linear programs [15, 63, 17, 52].

The Positivity-hardness results leave the possibility open that the problems under consideration are undecidable. Remarkable undecidability results in this context are presented in [41]: The hardness of deciding almost sure termination and almost sure termination with finite expected termination time for purely probabilistic programs formulated in the probabilistic fragment of probabilistic guarded command language (pGCL) [47] is pinpointed to levels of the arithmetical hierarchy (for details on the arithmetical hierarchy, see, e.g., [49]). The results reach up to Π^0_3-completeness for deciding universal almost sure termination with finite expected termination time (Π^0_1-complete problems are already undecidable while still co-recursively enumerable). Undecidability is not surprising as the programs subsume ordinary programs. But the universal halting problem for ordinary programs is only Π^0_2-complete showing that deciding universal termination with finite expected termination time of probabilistic programs is strictly harder. Similarly deciding termination from a given initial configuration is Σ^0_1-complete for or-

dinary programs (halting problem) while deciding almost sure termination with finite expected termination time for probabilistic programs from a given initial configuration is Σ_2^0-complete. Operational semantics of pGCL-programs can be given as infinite-state MDPs [36]. Applied to the purely probabilistic fragment, this leads to infinite-state Markov chains.

1.5 Outline

In the following Section 2, we provide necessary definitions and present our notation. In Section 3, we outline the general structure of the gadget-based reductions from the Positivity-problem and construct an MDP-gadget in which a linear recurrence relation can be encoded in terms of the optimal values for a variety of optimization problems (Section 3.2). Afterwards, we construct gadgets encoding also the initial values of a linear recurrence sequence and provide the reductions from the Positivity problems and all subsequent reductions as depicted in Figure 1 (Section 4). We conclude with final remarks and an outlook on future work (Section 5).

2. Preliminaries

We assume some familiarity with Markov decision processes and briefly introduce our notation in the sequel. More details can be found in text books such as [60].

Markov decision process. A *Markov decision process* (MDP) is a tuple $\mathcal{M} = (S, Act, P, s_{init})$ 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 the transition probability function for which we require that $\sum_{t \in S} P(s,\alpha,t) \in \{0,1\}$ for all $(s,\alpha) \in S \times Act$, and $s_{init} \in S$ is the initial state. Depending on the context, we enrich MDPs with a weight function $wgt\colon S \times Act \to \mathbb{Z}$, a finite set of atomic propositions AP and a labeling function $L\colon S \to 2^{\mathsf{AP}}$, or a designated set of goal states *Goal*. The *size* of an MDP $\mathcal{M}$, denoted by $size(\mathcal{M})$, is the sum of the number of states plus the total sum of the lengths of the encodings of the non-zero probability values $P(s,\alpha,s')$ as fractions of co-prime integers in binary and, if present, the lengths of the encodings of the weight values $wgt(s,\alpha)$ in binary.

We write $Act(s)$ for the set of actions that are enabled in a state s, i.e., $\alpha \in Act(s)$ if and only if $\sum_{t \in S} P(s,\alpha,t) = 1$. Whenever the process is in a state s, a non-deterministic choice between the enabled actions $Act(s)$ has to be made. We call a state *absorbing* if the only enabled actions lead to the state itself with probability 1 and weight 0. If there are no enabled actions, we call a state *terminal* or a *trap state*. The paths of $\mathcal{M}$ are finite or infinite sequences $s_0\,\alpha_0\,s_1\,\alpha_1\,s_2\,\alpha_2\ldots$ where states and actions alternate such that $P(s_i,\alpha_i,s_{i+1}) > 0$ for all $i \geq 0$. Throughout this section, we assume that all states are reachable from the initial state in any MDP, i.e., that there is a finite path from s_{init} to each state s. We extend the weight function to finite paths. For a

finite path $\pi = s_0\,\alpha_0\,s_1\,\alpha_1\,\ldots\alpha_{k-1}\,s_k$, we denote its accumulated weight by

$$wgt(\pi) = wgt(s_0, \alpha_0) + \ldots + wgt(s_{k-1}, \alpha_{k-1}).$$

Similarly, we extend the transition probability function to finite paths and write

$$P(\pi) = P(s_0, \alpha_0, s_1) \cdot \ldots \cdot P(s_{k-1}, \alpha_{k-1}, s_k).$$

A *one-counter MDP* is an MDP equipped with a counter. Each state-action pair increases or decreases the counter or leaves the counter unchanged. A one-counter MDP is said to terminate if the counter value drops below zero. We view one-counter MDPs as MDPs with a weight-function $wgt\colon S \times Act \to \{-1, 0, +1\}$. In this formulation a one-counter MDP terminates when a prefix π of a path satisfies $wgt(\pi) < 0$.

A *Markov chain* is an MDP in which the set of actions is a singleton. There are no non-deterministic choices in a Markov chain and hence we drop the set of actions. Consequently, a Markov chain is a tuple $\mathcal{M} = (S, P, s_{init})$, possibly extended with a weight function, a labeling, or a designated set of goal states. The transition probability function P is a function from $S \times S$ to $[0, 1] \cap \mathbb{Q}$ such that $\sum_{t \in S} P(s, t) \in \{0, 1\}$ for all $s \in S$.

Scheduler. A *scheduler* for an MDP $\mathcal{M} = (S, Act, P, s_{init})$ is a function $\mathfrak{S}$ that assigns to each finite path π not ending in trap state a probability distribution over $Act(last(\pi))$ where $last(\pi)$ denotes the last state of π. This probability distribution indicates which of the enabled actions is chosen with which probability under $\mathfrak{S}$ after the process has followed the finite path π.

We allow schedulers to be *randomized* and *history-dependent*. By restricting the possibility to randomize over actions or by restricting the amount of information from the history of a run that can affect the choice of a scheduler, we obtain the following types of schedulers: A scheduler $\mathfrak{S}$ is called *deterministic* if it does not make use of the possibility to randomize over actions, i.e., if $\mathfrak{S}(\pi)$ is a Dirac distribution for each path π. Such a scheduler $\mathfrak{S}$ can be viewed as a function that assigns an action to each finite path π. A scheduler $\mathfrak{S}$ is called *memoryless* if $\mathfrak{S}(\pi) = \mathfrak{S}(\pi')$ for all finite paths π, π' with $last(\pi) = last(\pi')$. In this case, $\mathfrak{S}$ can be viewed as a function that assigns to each state s a distribution over $Act(s)$. A memoryless deterministic scheduler hence can be seen as a function from states to actions. In an MDP with a weight function, a scheduler $\mathfrak{S}$ is said to be *weight-based* if $\mathfrak{S}(\pi) = \mathfrak{S}(\pi')$ for all finite paths π, π' with $wgt(\pi) = wgt(\pi')$ and $last(\pi) = last(\pi')$. Such a scheduler assigns distributions over actions to state-weight pairs from $S \times \mathbb{Z}$.

Probability measure. Given an MDP $\mathcal{M} = (S, Act, P, s_{init})$ and a scheduler $\mathfrak{S}$, we obtain a probability measure $\Pr^{\mathfrak{S}}_{\mathcal{M},s}$ on the set of maximal paths of $\mathcal{M}$ that start in s: For each finite path $\pi = s_0\,\alpha_0\,s_1\,\alpha_1\,\ldots\alpha_{k-1}\,s_k$ with $s_0 = s$, we denote the cylinder set of all its maximal extensions

by $Cyl(\pi)$. The probability mass of this cylinder set is then given by

$$\Pr^{\mathfrak{S}}_{\mathcal{M},s}(Cyl(\pi)) = P(\pi) \cdot \Pi_{i=0}^{k-1} \mathfrak{S}(s_0 \ldots s_i)(\alpha_i).$$

Recall that $\mathfrak{S}(s_0 \ldots s_i)$ is a probability distribution over actions and that $\mathfrak{S}(s_0 \ldots s_i)(\alpha_i)$ denotes the probability that the scheduler $\mathfrak{S}$ chooses action α after the prefix $s_0 \ldots s_i$ of π. The set of cylinder sets forms the basis of the standard tree topology on the set of maximal paths. By Carathéodory's extension theorem, we can extend the pre-measure $\Pr^{\mathfrak{S}}_{\mathcal{M},s}(Cyl(\pi))$ defined on the cylinder sets to a probability measure on the Borel σ-algebra of the space of maximal paths with the standard tree topology. We sometimes drop the subscript s if s is the initial state s_{init} of $\mathcal{M}$. In a Markov chain $\mathcal{N}$, we drop the reference to a scheduler and write $\Pr_{\mathcal{N},s}$.

Let X be a random variable on the set of maximal paths of $\mathcal{M}$ starting in s, i.e., X is a function assigning values from $\mathbb{R} \cup \{-\infty, +\infty\}$ to maximal paths. We denote the expected value of X under the probability measure $\Pr^{\mathfrak{S}}_{\mathcal{M},s}$ by $\mathbb{E}^{\mathfrak{S}}_{\mathcal{M},s}(X)$.

The values we are typically interested in are the worst- or best-case probabilities of an event or the worst- or best-case expected values of a random variable. Worst or best case refers to the possible ways to resolve the non-deterministic choices. Hence, these values are formally expressed by taking the supremum or infimum over all schedulers. Given an MDP $\mathcal{M}$, a state s, and an event, i.e., a set of maximal paths, E, or a random variable X on the maximal paths of $\mathcal{M}$, we define

$$\Pr^{\max}_{\mathcal{M},s}(E) = \sup_{\mathfrak{S}} \Pr^{\mathfrak{S}}_{\mathcal{M},s}(E), \qquad \Pr^{\min}_{\mathcal{M},s}(E) = \inf_{\mathfrak{S}} \Pr^{\mathfrak{S}}_{\mathcal{M},s}(E),$$

$$\mathbb{E}^{\max}_{\mathcal{M},s}(X) = \sup_{\mathfrak{S}} \mathbb{E}^{\mathfrak{S}}_{\mathcal{M},s}(X), \text{ and} \qquad \mathbb{E}^{\min}_{\mathcal{M},s}(X) = \inf_{\mathfrak{S}} \mathbb{E}^{\mathfrak{S}}_{\mathcal{M},s}(X),$$

where inf and sup range over all schedulers $\mathfrak{S}$ for $\mathcal{M}$.

We use LTL-like notation such as "$\Diamond$(accumulated weight < 0)" to denote the event that a prefix of a path has a negative accumulated weight. Note that this event expresses the termination of a one-counter MDP in our view of one-counter MDPs as MDPs with a weight-function taking only values in $\{-1, 0, +1\}$.

Classical stochastic shortest path problem. Let $\mathcal{M}$ be an MDP with a weight function $wgt\colon S \times Act \to \mathbb{Z}$ and a designated set of terminal goal states $Goal$. We define the following random variable $\diamondplus Goal$ on maximal paths ζ of $\mathcal{M}$ as follows:

$$\diamondplus Goal(\zeta) = \begin{cases} wgt(\zeta) & \text{if } \zeta \vDash \Diamond Goal, \\ undefined & \text{otherwise.} \end{cases}$$

The expected accumulated weight before reaching $Goal$ under a scheduler $\mathfrak{S}$ is given by the expected value $\mathbb{E}^{\mathfrak{S}}_{\mathcal{M},s_{init}}(\diamondplus Goal)$. It is evident that this expected value is only defined if

$\Pr^{\mathfrak{S}}_{\mathcal{M},s_{init}}(\diamondsuit Goal) = 1$. The *classical stochastic shortest path problem* asks for the optimal value

$$\mathbb{E}^{\max}_{\mathcal{M},s_{init}}(\diamondsuit Goal) = \sup_{\mathfrak{S}} \mathbb{E}^{\mathfrak{S}}_{\mathcal{M},s_{init}}(\diamondsuit Goal)$$

where the supremum ranges over all schedulers $\mathfrak{S}$ with $\Pr^{\mathfrak{S}}_{\mathcal{M},s_{init}}(\diamondsuit Goal) = 1$. The classical stochastic shortest path problem can be solved in polynomial time [16, 30, 5].

3. Outline of the Positivity-hardness proofs

The Positivity-hardness results in this paper are obtained by sequences of reductions depicted in Figure 1. The key steps for these sequences are the three direct reductions from the Positivity-problem to the threshold problems for the maximal termination probability of one-counter MDPs, the maximal partial expectation, and the maximal conditional value-at-risk, respectively.

3.1 Structure of the MDP constructed for the direct reductions from the Positivity problem

The three direct reductions from the Positivity problem (at the top of Figure 1) follow a modular approach: The MDPs constructed for the reductions are obtained by putting together three gadgets as sketched in Figure 2. One gadget encodes a linear recurrence relation exploiting the dependency of optimal values from different starting states after different amounts of weight have been accumulated in the history of a run onto each other. A second gadget encodes the initial values of a linear recurrence sequence. Together, these two gadget allow us to encode linear recurrence sequences. Finally, an initial gadget is added in which each positive amount of weight w is accumulated with positive probability. Afterwards, the gadget is left and a scheduler has to decide how to leave the initial gadget. The optimal decision if weight w has been accumulated directly corresponds to whether the wth member of the given linear recurrence sequence is non-negative.

More precisely, let a rational linear recurrence sequence be given in terms of the initial values $u_0, \ldots, u_{k-1}$ and the coefficients $\alpha_1, \ldots, \alpha_k$ of the linear recurrence relation. The three gadgets are connected via two states s and t as depicted in Figure 2. In state t and s, actions $\gamma_0, \ldots, \gamma_{k-1}$ and $\delta_0, \ldots, \delta_{k-1}$, respectively, leading to the gadget encoding the initial values and action γ and δ, respectively leading to the gadget encoding the linear recurrence relation are enabled. The gadgets will be constructed such that an optimal scheduler has to choose action γ_i or δ_i if the accumulated weight in state t or s is a value i with $0 \leq i < k$ and that it has to choose action γ if the accumulated weight is at least k. After γ or δ is chosen, the accumulated weight is decreased within the gadget encoding the linear recurrence relation before the MDP moves back to the states s and t with positive probability.

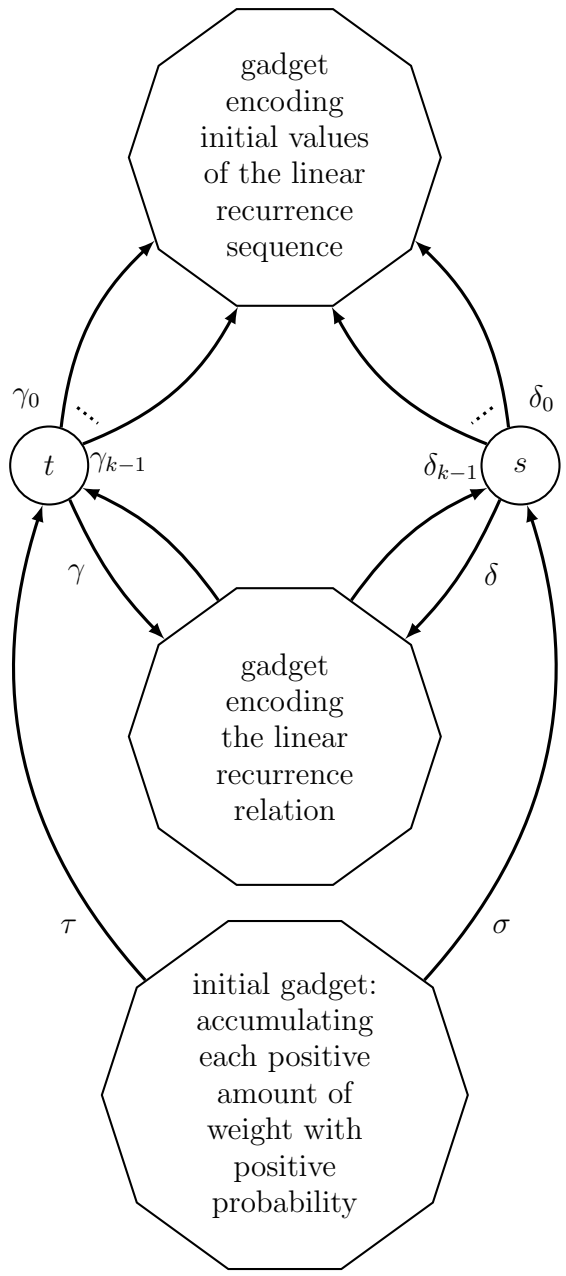

Figure 2. Interplay between the MDP-gadgets.

Let us now denote the maximal possible value for the quantity of interest when starting in one of the states t and s with accumulated weight w by $V(t,w)$ and $V(s,w)$. The linear recurrence relation will be found in the difference $d(w) \stackrel{\text{def}}{=} V(t,w) - V(s,w)$. If the accumulated weight is $0 \leq i < k$, the gadget encoding the initial values will make sure that $d(i) = V(t,i) - V(s,i) = u_i$. For each of the three direct reductions from the Positivity problem, we construct one such gadget tailored to the three respective quantities.

For accumulated weights w of at least k, the gadget encoding the recurrence will exploit the dependency of the optimal values $V(t,w)$ and $V(s,w)$ on the optimal values when starting with lower accumulated weight. This gadget can be used in all reductions and will be described in the next subsection.

Put together, these two gadgets ensure that $d(w) = u_w$ for all $w \geq 0$. To complete the reductions, we add an initial gadget $\mathcal{I}$ depicted in Figure 3 in which each positive amount

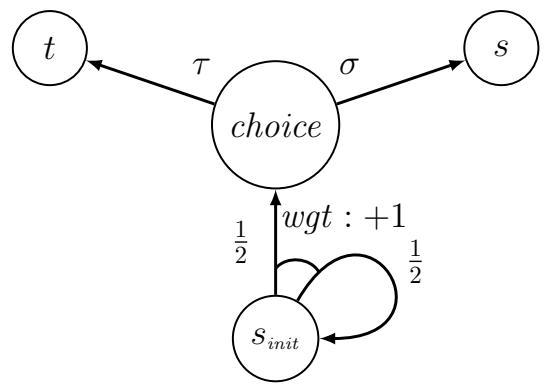

Figure 3. The initial gadget $\mathcal{I}$.

of weight w is accumulated with positive probability. Afterwards, a scheduler has to choose whether to move to state t or state s via the actions τ and σ, respectively. It is optimal to move to t if and only if $u_w \geq 0$. Let now $\mathfrak{S}$ be the scheduler always choosing τ in the initial gadget and afterwards behaving optimally when choosing from $\gamma_0, \ldots, \gamma_{k-1}$ and γ or $\delta_0, \ldots, \delta_{k-1}$ and δ as described above. This scheduler is optimal if and only if the given linear recurrence sequence is non-negative. The final step to complete the reduction is to compute the value $V^{\mathfrak{S}}(s_{init}, 0)$ that is achieved by $\mathfrak{S}$ starting from the initial state. In all three reductions, we can compute this rational value via converging matrix series. The optimal value $V^{\max}(s_{init}, 0)$ that can be achieved from the initial state now satisfies

$$V^{\max}(s_{init}, 0) \leq V^{\mathfrak{S}}(s_{init}, 0)$$

if and only if the given linear recurrence sequence is non-negative.

3.2 MDP-gadget for linear recurrence relations

In this section, we demonstrate how to construct the gadget ensuring that the difference of optimal values $V(t, w) - V(s, w)$ follows a given linear recurrence *relation* with respect to different weight levels w. In the next section, the initial values of a linear recurrence sequence will be encoded in MDP-gadgets tailored to the different quantities we address.

Optimality equations. Let us start by the following observations on the well-known relation between the optimal values at different states in the classical stochastic shortest path problem, i.e., the maximal expected accumulated weights before reaching a goal state (defined in Section 2). Let $\mathcal{M} = (S, Act, P, s_{init}, wgt, Goal)$ be an MDP. The solution to the classical stochastic shortest path problem satisfies the so-called *Bellman equation*. If $V(s)$ denotes the value when starting in state s, i.e., the maximal expected accumulated weight before reaching *Goal* from

state s, then

$$V(s) = \max_{\alpha \in Act(s)} wgt(s,\alpha) + \sum_{t \in S} P(s,\alpha,t) \cdot V(t)$$

for $s \notin Goal$ and $V(s) = 0$ for $s \in Goal$. This simple form of optimality equation implies the existence of optimal memoryless deterministic schedulers for the classical stochastic shortest path problem (in case optimal schedulers exist, i.e., if the optimal values are finite).

For problems like the optimization of the termination probability of one-counter MDPs, it is, however, clearly not sufficient to consider the optimal values only in dependency of the starting state. The counter-value, i.e. the weight that has been accumulated so far, is essential. So, let $V(s,w)$ denote the maximal termination probability of a one-counter MDP when starting in state s with counter value w. Letting $V(s,w) = 1$ if $w < 0$, we obtain the following equation for all states s and all values $w \geq 0$:

$$V(s,w) = \max_{\alpha \in Act(s)} \sum_{t \in S} P(s,\alpha,t) \cdot V(t, w + wgt(s,\alpha)). \qquad (*)$$

Already in this equation, the value $V(s,w)$ hence possibly depends on values of the form $V(s,w-i)$ for some i. We want to exploit this interrelation to encode linear recurrence relations

$$u_{n+k} = \alpha_1 u_{n+k-1} + \cdots + \alpha_k u_n$$

into the optimal values $V(s,w)$. Of course, the values $P(s,\alpha,t)$ are all non-negative. So, we cannot directly encode a linear recurrence into the optimal values for different weight levels at one state as the coefficients might be negative. To overcome this problem, we instead consider the difference $V(t,w) - V(s,w)$ for two different states s and t.

Scaling down coefficients of a linear recurrence sequence. Given the coefficients $\alpha_1, \ldots, \alpha_k$, and initial values $u_0 = \beta_0, \ldots, u_{k-1} = \beta_{k-1}$ of a linear recurrence sequence, we have to assume that these are all sufficiently small for the following constructions. So, let us clarify why we can assume this without loss of generality and let us provide precise bounds. Let $(u_n)_{n\geq 0}$ be a linear recurrence sequence specified 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 $n \geq 0$. For any $\mu > 0$ and $\lambda > 0$, the sequence $(v_n)_{n\geq 0}$ defined by $v_n = \mu \cdot \lambda^n \cdot u_n$ for all n is non-negative if and only if $(u_n)_{n\geq 0}$ is non-negative. Furthermore, it satisfies $v_i = \mu \cdot \lambda^i \cdot \beta_i$ for $i < k$ and

$$v_{n+k} = \lambda \cdot \alpha_1 \cdot v_{n+k-1} + \lambda^2 \cdot \alpha_2 \cdot v_{n+k-2} + \cdots + \lambda^k \cdot \alpha_k \cdot v_n.$$

By choosing λ and μ appropriately, we can scale down the initial values and coefficients of the recurrence relation for any given input.

To obtain precise bounds that will be used throughout the following sections, let $\alpha \stackrel{\text{def}}{=} \sum_{i=1}^{k} |\alpha_i|$. and let $\lambda \stackrel{\text{def}}{=} \min\left(\frac{1}{\alpha\cdot(5k+5)}, \frac{1}{(5k+5)}\right)$. So, if $\alpha > 1$, then $\lambda = \frac{1}{\alpha\cdot(5k+5)}$ and else $\lambda = \frac{1}{(5k+5)}$. The value λ can be computed in polynomial time. As the numerical value of k is linear in the size of

the given original input, the coefficients $\alpha'_1 \stackrel{\text{def}}{=} \lambda \cdot \alpha_1, \alpha'_2 \stackrel{\text{def}}{=} \lambda^2 \cdot \alpha_2, \ldots, \alpha'_k \stackrel{\text{def}}{=} \lambda^k \cdot \alpha_k$ of the linear recurrence of the sequence $(v_n)_{n\geq 0}$ can be computed in polynomial time as well. The choice of λ ensures that $\sum_{i=1}^{k} |\alpha'_i| < \frac{1}{5k+5}$.

Let now $\alpha' \stackrel{\text{def}}{=} \sum_{i=1}^{k} |\alpha'_i|$ and $\beta \stackrel{\text{def}}{=} \max_{0\leq j<k} |\beta_j|$. We can choose $\mu \stackrel{\text{def}}{=} \frac{\min(\alpha',1)}{4k^{2k+2}\cdot\beta}$. Again, since the value k is linear in the size of the original input, μ can be computed in polynomial time. The initial values of the new sequence $(v_n)_{n\geq 0}$ are now $\beta'_i \stackrel{\text{def}}{=} v_i = \mu \cdot \lambda^i \cdot \beta_i$ for $i < k$, computable in polynomial time. The choice of μ guarantees that $\max_{0\leq j<k} \beta'_j < \min(\frac{1}{4k^{2k+2}}, \frac{\alpha'}{4})$.

Since this transformation can be carried out in polynomial time, we can w.l.o.g. from now on work under the following assumption:

ASSUMPTION 3.1. Given the coefficients $\alpha_1, \ldots, \alpha_k$, and initial values $u_0 = \beta_0, \ldots, u_{k-1} = \beta_{k-1}$ of a linear recurrence sequence, we assume that

$$\alpha \stackrel{\text{def}}{=} \sum_{i=1}^{k} |\alpha_i| < \frac{1}{5k+5} \text{ and that } \max_{0\leq j<k} \beta_j < \min(\frac{1}{4k^{2k+2}}, \frac{\alpha}{4}).$$

MDP-gadget for linear recurrence relations. Given the coefficients $\alpha_1, \ldots, \alpha_k$ of a linear recurrence relation satisfying Assumption 3.1, we construct the MDP-gadget depicted in Figure 4. The gadget contains states s, t, and *trap* 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 *trap*. 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 analogously. If α_i is negative, action δ reaches state t_i with probability $|\alpha_i|$. Otherwise it reaches s_i with probability α_i. The state *trap* is absorbing. As the gadget depends on the inputs $\bar{\alpha} = (\alpha_1, \ldots, \alpha_k)$, we call it $\mathsf{G}_{\bar{\alpha}}$.

This gadget $\mathsf{G}_{\bar{\alpha}}$ will be integrated into MDPs without further outgoing edges from states $s_1, \ldots, s_k, t_1, \ldots, t_k$. For any optimization problem for which the optimal values V depend on the state and the weight accumulated so far and satisfy equation $(*)$, we can encode a linear recurrence in an MDP containing this gadget (and possibly further actions for state t and s): If we know that an optimal scheduler chooses action γ in state t and action δ in state s if the accumulated weight is w, then

$$\begin{aligned} V(t,w) - V(s,w) = &\left(1 - \sum_{i=1}^{k} |\alpha_i|\right)(V(goal,w) - V(goal,w)) \quad + \\ &\sum_{1\leq i\leq k,\, \alpha_i\geq 0} \alpha_i V(t,w-i) - \alpha_i V(s,w-i) \quad + \\ &\sum_{1\leq i\leq k,\, \alpha_i<0} (-\alpha_i)V(s,w-i) + (-\alpha_i)V(t,w-i) \\ = &\sum_{i=1}^{k} \alpha_i \cdot (V(t,w-i) - V(s,w-i)). \end{aligned}$$

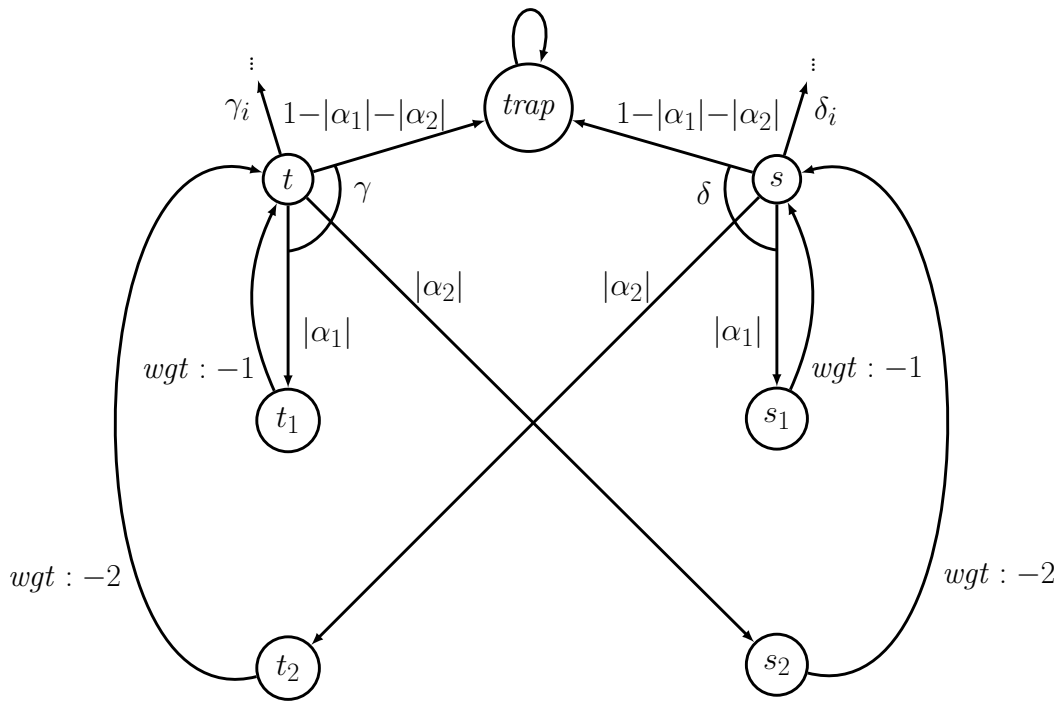

Figure 4. The gadget $G_{\bar{\alpha}}$ to encode linear recurrence relations. The example here is depicted for a linear recurrence of depth 2 with $\alpha_1 \geq 0$ and $\alpha_2 < 0$. The outgoing actions γ_i and δ_i lead to the gadget encoding initial values as depicted in Figure 2.

Note that this linear recurrence relation also holds for the optimal values in the classical stochastic shortest path problem for example. So, the gadget alone is not yet enough for a hardness proof. The missing ingredient is the encoding of the initial values of a linear recurrence sequence. In order to include the encoding of the initial values in our approach, it is necessary that optimal schedulers cannot be chosen to be memoryless. The optimal decisions have to depend on the weight that has been accumulated in the history of a run. If this is the case, we aim to encode the initial values by adding further outgoing actions to the states t and s. By fine-tuning the weights and probabilities of these actions, we can achieve that for small weights w some of the new actions are optimal while for large weights the actions γ and δ of the gadget are optimal. If we manage to design the other actions such that the differences $V(t, w+i) - V(s, w+i)$ are equal to given starting values β_i for a sequence of weights $w, w+1, \ldots, w+k-1$ while actions γ and δ are optimal for weights of at least $w+k$, we can encode arbitrary linear recurrence sequences. This is the goal of the subsequent section.

4. Reductions from the Positivity problem

To encode initial values of a linear recurrence sequence, we construct further MDP gadgets. For the termination probability and expected termination time of one-counter MDPs and for partial expectations, we can construct these gadgets directly. For the conditional value-at-risk, we use an intermediate auxiliary random variable. Putting together these gadgets with the

gadget $G_{\bar{\alpha}}$ from the previous section, we obtain the basis for the Positivity-hardness results of the respective threshold problems. The Positivity-hardness of the remaining problems is obtained as a consequence of these results via further reductions. An overview of the chains of reductions used is presented in Figure 1.

4.1 One-counter MDPs, energy objectives, cost problems, and quantiles

The first problem we will show to be Positivity-hard is the threshold problem for the optimal termination probability of one-counter MDPs. From this result, Positivity-hardness results for energy objectives, cost problems, and the computation of quantiles follow easily. Afterwards, we adjust the reduction to show Positivity-hardness of the threshold problem for the optimal expected termination time of almost-surely terminating one-counter MDPs.

Termination probability of one-counter MDPs. We formulated the termination of a one-counter MDP in terms of weighted MDPs $\mathcal{M}$. Recall that a one-counter MDP terminates if the counter value drops below zero. If we consider the weight that has been accumulated instead of the counter value, the quantities we are interested are $\Pr^{\text{opt}}_{\mathcal{M}}(\Diamond \text{ accumulated weight} < 0)$ for opt = max and opt = min. The main result we prove in this section is the following:

THEOREM 4.1. *The Positivity problem is reducible in polynomial time to the following problems: Given an MDP $\mathcal{M}$ and a rational $\vartheta \in (0,1)$,*

1. *decide whether* $\Pr^{\max}_{\mathcal{M},s_{init}}(\Diamond(\textit{accumulated weight} < 0)) > \vartheta$.
2. *decide whether* $\Pr^{\min}_{\mathcal{M},s_{init}}(\Diamond(\textit{accumulated weight} < 0)) < \vartheta$.

Note that if weights are encoded in unary, we can transform a weighted MDP to a one-counter MDP that can only increase or decrease the counter value by 1 in each step in polynomial time. The MDPs that are constructed from a linear recurrence sequence of depth k in the proof of Theorem 4.1 will contain only weights with an absolute value of at most k. So, they can be transformed to one-counter MDPs in time linear in the size of the original input and we conclude that the following two threshold problems for the optimal termination probability of one-counter MDPs are Positivity-hard:

COROLLARY 4.2. *The Positivity problem is reducible in polynomial time to the following problems: Given a one-counter MDP $\mathcal{M}$ viewed as an MDP with weights in $\{-1,0,+1\}$ and a rational $\vartheta \in (0,1)$,*

1. *decide whether* $\Pr^{\max}_{\mathcal{M},s_{init}}(\Diamond(\textit{accumulated weight} < 0)) > \vartheta$.
2. *decide whether* $\Pr^{\min}_{\mathcal{M},s_{init}}(\Diamond(\textit{accumulated weight} < 0)) < \vartheta$.

Among the direct reductions from the Positivity problem we present, the construction of the gadget encoding the initial values of a linear recurrence sequence is arguably the simplest for these optimal termination probabilities. In the formulation with weighted MDPs,

the termination of a one-counter MDP is moreover the complement of the energy objective "□ accumulated weight ≥ 0". We will first prove Positivity-hardness for the threshold problem for maximal termination probabilities and outline the necessary adjustments to show Positivity-hardness also for the threshold problem for minimal termination probabilities afterwards.

We split the proof of Theorem 4.1 into four parts. First, we provide the construction of an MDP from a linear recurrence sequence. Then, we show that the linear recurrence sequence is correctly encoded in this MDP in terms of the maximal termination probabilities. To complete the proof of item 1, we then show how to compute the threshold ϑ for the threshold problem and how this establishes the correctness of the reduction. Finally, we show how to adapt the construction to prove hardness of the threshold problem for minimal termination probabilities.

Proof of Theorem 4.1(1): construction of the MDP. Given a linear recurrence sequence in terms of the rational coefficients $\alpha_1, \ldots, \alpha_k$ of the linear recurrence relation as well as the rational initial values $\beta_0, \ldots, \beta_{k-1}$ for $k \geq 2$, our first goal is to construct an MDP $\mathcal{M}$ and a rational $\vartheta \in (0,1)$ such that

$$\Pr^{\max}_{\mathcal{M},s_{init}}(\Diamond(\text{ accumulated weight} < 0)) > \vartheta \qquad \text{if and only if} \qquad u_n < 0 \text{ for some } n \geq 0.$$

By Assumption 3.1, we can assume that the input values are sufficiently small. More precisely, we assume that $\sum_{i=1}^{k} |\alpha_i| < 1/(k+1)$ and that $0 \leq \beta_j < 1/(k+1)$ for all $0 \leq j \leq k-1$, which is ensured by the bounds in Assumption 3.1, and because the Positivity problem becomes trivial if one of the values β_j with $0 \leq j \leq k-1$ is negative.

We denote the supremum of possible termination probabilities in terms of the current state s and counter value (accumulated weight) w by $p(s,w)$. More precisely, in an MDP $\mathcal{M}$ for $w \geq 0$, we define

$$p(s,w) \stackrel{\text{def}}{=} \Pr^{\max}_{\mathcal{M},s}(\Diamond \text{ accumulated weight } < -w).$$

The values $p(s,w)$ in an MDP with state space S now satisfy the optimality equation $(*)$ from Section 3.2 (where $p(s,w)$ takes the role of $V(s,w)$ in $(*)$), which we restate here for convenience. We have $p(s,w) = 1$ for all states s and all $w < 0$ and

$$p(s,w) = \max_{\alpha \in Act(s)} \sum_{t \in S} P(s,\alpha,t) \cdot p(t, w + wgt(s,\alpha)) \qquad \text{for all } s \in S \text{ and } w \geq 0.$$

So, to capture the linear recurrence relation, we will be able to make use of the gadget $\mathsf{G}_{\bar{\alpha}}$ from Section 3.2. The missing ingredient is a gadget to encode the initial values of a linear recurrence sequence.

The new gadget $O_{\bar{\beta}}$ encoding the initial values $\bar{\beta}$ is depicted in Figure 5 and works as follows: For $0 \leq j \leq k-1$, the action γ_j enabled in t leads to state x_j with probability $\frac{k-j}{k+1} + \beta_j$. By assumption on β_j, this probability is less than $\frac{k-j+1}{k+1}$. The remaining probability leads to *trap*. In state s, the action δ_j leads to y_j with probability $\frac{k-j}{k+1}$ and to *trap* with the remaining probability.

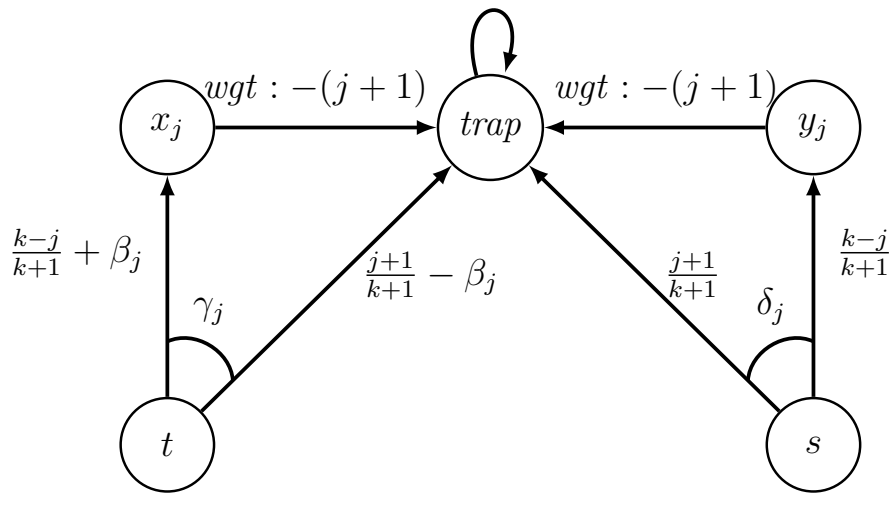

Figure 5. Gadget $\mathcal{O}_{\bar{\beta}}$ encoding initial values of a linear recurrence sequence in terms of maximal termination probabilities of one-counter MDPs.

For $0 \leq j \leq k-1$, one reaches *trap* from x_j and y_j with probability 1 and a counter change of $-(j+1)$.

Now, we glue together the initial gadget $\mathcal{I}$ defined in Section 3.1, the gadget encoding the linear recurrence relation $\mathsf{G}_{\bar{\alpha}}$ from Section 3.2, and the new gadget $\mathcal{O}_{\bar{\beta}}$ at states t, s, and *trap*. The resulting MDP $\mathcal{M}$ is depicted in Figure 6 – for better readability, it is depicted for $k = 2$ and assuming that $\alpha_1 \geq 0$ while $\alpha_2 < 0$.

Proof of Theorem 4.1(1): correctness of the encoding of the linear recurrence sequence. In this paragraph, we show that the initial linear recurrence sequence is indeed encoded in the maximal termination probabilities when starting from states t and s with different counter values, i.e., values of accumulated weight as described in Section 3.1. More precisely, let $(u_n)_{n\geq 0}$ be the linear recurrence sequence given by the initial values $\beta_0, \ldots \beta_{k-1}$ and the coefficients $\alpha_1, \ldots, \alpha_k$ of the linear recurrence relation. We prove the following:

LEMMA 4.3. *For each $w \geq 0$, we have*

$$p(t, w) - p(s, w) = u_w$$

where $p(r, w)$ denotes the maximal termination probability from state $r \in \{s, t\}$ when starting with accumulated weight w as defined above.

PROOF. For the correct interplay of the gadgets $\mathsf{G}_{\bar{\alpha}}$ and $\mathcal{O}_{\bar{\beta}}$, the optimal decisions in states t and s for different values of accumulated weights, i.e., different counter-values, are crucial. In order to terminate, the accumulated weight has to drop below 0 before reaching *trap*. As soon as the trap state is reached with non-negative accumulated weight, the process cannot terminate anymore. The optimal decision in order to maximize the termination probability

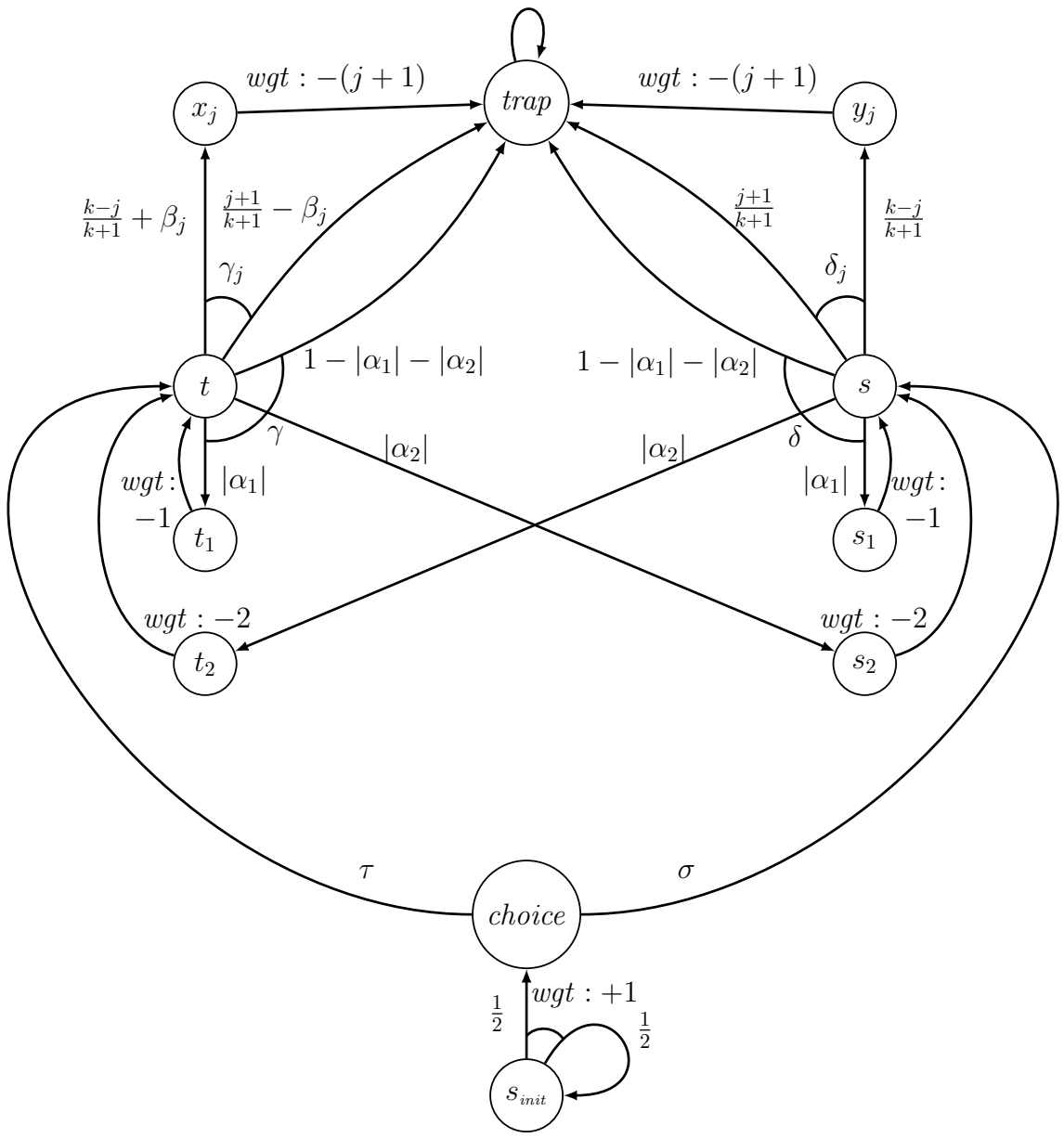

Figure 6. Full MDP for the reduction to the threshold problem for termination probabilities of one-counter MDPs. The MDP contains the upper part for all $0 \leq j \leq k-1$. The middle part is depicted for $k = 2$, $\alpha_1 \geq 0$, and $\alpha_2 < 0$.

in state t is now easy to determine. Let ℓ be the current weight. If $0 \leq \ell \leq k-1$, choosing action γ leads to termination with probability less than $1/(k+1)$ as *trap* is reached immediately with probability at least $k/(k+1)$ due to our assumption that $\sum_{i \leq k} |\alpha_i| < 1/(k+1)$. Choosing action γ_j makes it impossible to terminate if $\ell > j$. If $\ell \leq j$, then choosing γ_j lets the process terminate if x_j is reached. This happens with probability $\frac{k-j}{k+1} + \beta_j$. As $\beta_j < 1/(k+1)$ for all j, the maximal termination probability is reached when choosing γ_ℓ. If $\ell \geq k$, then γ_j leads to termination with probability 0 for all j. Hence, action γ is optimal. Analogously, we see that the optimal choice in state s with weight ℓ is δ_ℓ if $\ell \leq k-1$ and δ otherwise.

The linear recurrence sequence $(u_n)_{n \geq 0}$ now can be found in terms of the difference

$$d(w) \stackrel{\text{def}}{=} p(t, w) - p(s, w).$$

For counter value $w \leq k-1$, we have seen that γ_w and δ_w, respectively, are the optimal actions. Hence, $d(w) = u_w$ in this case as we have just seen that the optimal termination probability when starting with weight $w \leq k-1$ is $\frac{k-w}{k+1} + \beta_w$ in t and $\frac{k-w}{k+1}$ in s. Furthermore, for $w > k-1$, actions γ and δ are optimal. So by the construction of gadget $\mathsf{G}_{\bar{\alpha}}$,

$$\begin{aligned} p(t,w) - p(s,w) = &\left(1 - \sum_{i=1}^{k} |\alpha_i|\right)(p(trap,w) - p(trap,w)) \quad + \\ &\sum_{1 \leq i \leq k,\, \alpha_i \geq 0} \alpha_i p(t, w{-}i) - \alpha_i p(s, w{-}i) \quad + \\ &\sum_{1 \leq i \leq k,\, \alpha_i < 0} (-\alpha_i) V(s, w{-}i) + (-\alpha_i) p(t, w{-}i) \\ = &\sum_{i=1}^{k} \alpha_i \cdot (p(t, w{-}i) - p(s, w{-}i)). \end{aligned}$$

So, the sequence of differences satisfies the linear recurrence relation given by $\alpha_1, \ldots, \alpha_k$. Therefore, $d(w) = u_w$ for all $w \geq 0$. ■

Proof of Theorem 4.1(1): computation of the threshold ϑ. The state *choice* is reached with any positive accumulated weight with positive probability. For the optimal choices in the state *choice* with accumulated weight w, we observe that choosing τ is optimal if and only if $d(w) \geq 0$. By Lemma 4.3, this holds if and only if $u_w \geq 0$.

Consider now the scheduler $\mathfrak{S}$ which always chooses τ in state *choice* and afterwards behaves according to the optimal choices as described in the proof of Lemma 4.3. This scheduler $\mathfrak{S}$ is optimal if and only if the sequence $(u_n)_{n \geq 0}$ is non-negative. To complete the reduction, we will compute the value

$$\vartheta \stackrel{\text{def}}{=} \Pr^{\mathfrak{S}}_{\mathcal{M}, s_{init}}(\Diamond(\text{accumulated weight } < 0)).$$

We will see that ϑ is a rational computable in polynomial time and we know that

$$\Pr^{\max}_{\mathcal{M}, s_{init}}(\Diamond(\text{accumulated weight } < 0)) \leq \vartheta$$

if and only if the scheduler $\mathfrak{S}$ is optimal which is the case if and only if $(u_n)_{n \geq 0}$ is non-negative.

LEMMA 4.4. *In the constructed MDP $\mathcal{M}$, the value $\vartheta = \Pr^{\mathfrak{S}}_{\mathcal{M}, s_{init}}(\Diamond(\text{accumulated weight } < 0))$ can be computed in polynomial time.*

PROOF. In order to compute the value ϑ, we first provide a recursive expression of the maximal termination probabilities $p(t,w)$ and $p(s,w)$. By the definition of $\mathfrak{S}$, these are precisely the termination probabilities under $\mathfrak{S}$ when starting from t or s with some positive accumulated weight $w \in \mathbb{N}$ because $\mathfrak{S}$ behaves optimally as soon as state t or s has been reached.

For this recursive expression, we consider the following Markov chain C for $n \in \mathbb{N}$ that is also depicted in Figure 7 – for better readability, it is depicted for the case $k = 2$ there:

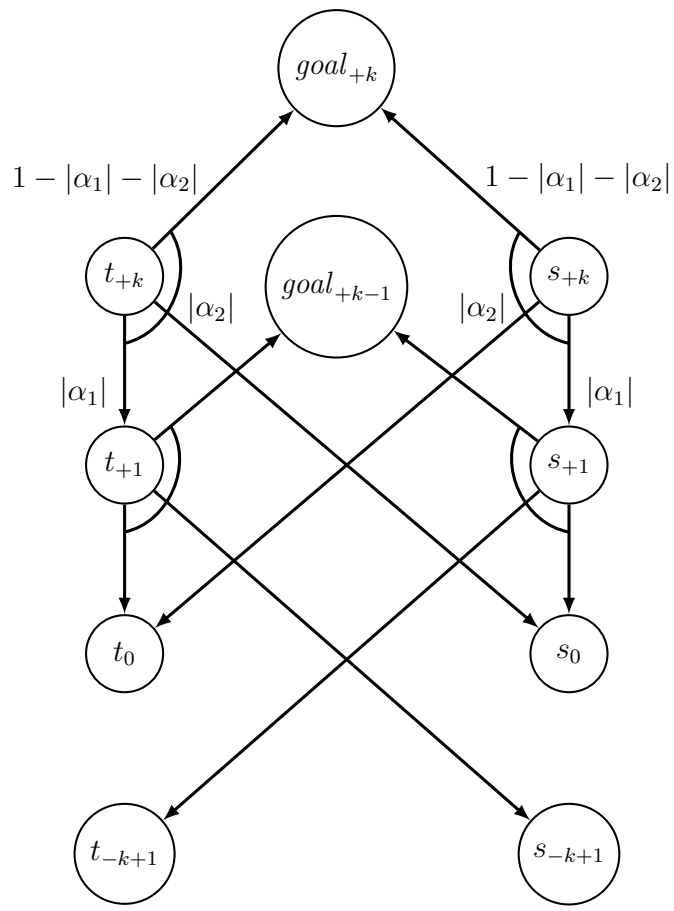

Figure 7. The Markov chain C depicted for $k = 2$ with $\alpha_1 \geq 0$ and $\alpha_2 < 0$.

The Markov chain C has $5k$ states named $t_{-k+1}, \ldots, t_{+k}$, $s_{-k+1}, \ldots, s_{+k}$, and $goal_{+1}, \ldots, goal_{+k}$. States $t_{-k+1}, \ldots, t_0$, $s_{-k+1}, \ldots, s_0$, and $goal_{+1}, \ldots, goal_{+k}$ are terminal. For $0 < i, j \leq k$, there are transitions from t_{+i} to t_{+i-j} with probability α_j if $\alpha_j > 0$, to s_{+i-j} with probability $|\alpha_j|$ if $\alpha_j < 0$, and to $goal_{+i}$ with probability $1 - |\alpha_1| - \ldots - |\alpha_k|$. Transitions from s_{+i} are defined analogously.

The idea behind this Markov chain is that the reachability probabilities describe how, for arbitrary $n \in \mathbb{N}$ and $1 \leq i \leq k$, the values $p(t, nk + i)$ and $p(s, nk + i)$ depend on the values $p(t, (n-1)k + j)$ and $p(s, (n-1)k + j)$ for $1 \leq j \leq k$. The transitions in C behave as γ and δ in $\mathcal{M}$, but the decrease in the accumulated weight is explicitly encoded into the state space. Namely, for $n \in \mathbb{N}$ and $0 < i \leq k$, we have

$$p(t, nk + i) = \sum_{j=1}^{k} \big(\Pr_{C,t_{+i}}(\Diamond t_{-k+j}) \cdot p(t, (n-1)k + j) + \Pr_{C,t_{+i}}(\Diamond s_{-k+j}) \cdot p(s, (n-1)k + j) \big) \qquad (*)$$

and analogously for $p(s, nk + i)$. We now group the optimal values together in the following column vectors

$$v_n = (p(t, nk + k - 1), p(t, nk + k - 2), \ldots, p(t, nk), p(s, nk + k - 1), \ldots, p(s, nk))^\top$$

for $n \in \mathbb{N}$. In other words, this vector contains the optimal values for the partial expectation when starting in t or s with an accumulated weight from $\{nk, \ldots, nk + k - 1\}$. The vector v_0 is

the column vector

$$(p(t,k-1),\dots,p(t,0),\dots p(s,k-1),\dots,p(s,0))^\top$$

and these values occur as transition probabilities in $\mathcal{M}$ under the actions $\gamma_{k-1},\dots,\gamma_0$ and $\delta_{k-1},\dots,\delta_0$.

As the reachability probabilities in C are rational and computable in polynomial time, we conclude from equation $(*)$ that there is a matrix $A \in \mathbb{Q}^{2k\times 2k}$ computable in polynomial time such that $v_{n+1} = Av_n$ for all $n \in \mathbb{N}$. So, $v_n = A^n v_0$ for all $n \in \mathbb{N}$.

As state *choice* is reached with weight w with probability $(1/2)^w$ for all $w \geq 1$, the value $\vartheta = \sum_{w=1}^{\infty}(1/2)^w p(t,w)$. Let $c = (\frac{1}{2^k}, \frac{1}{2^{k-1}},\dots,\frac{1}{2^1},0,\dots,0)$. Observe that for all $n \in \mathbb{N}$,

$$\left(\frac{1}{2^k}\right)^n \cdot c \cdot v_n = \sum_{i=1}^{k} \frac{1}{2^{nk+i}} p(t,nk+i).$$

Hence, we can write

$$\begin{aligned}\vartheta &= \sum_{n=0}^{\infty}\left(\frac{1}{2^k}\right)^n \cdot c \cdot v_n - p(t,0) = c \cdot \sum_{n=0}^{\infty}\left(\frac{1}{2^k}\right)^n \cdot v_n - p(t,0)\\ &= c \cdot \sum_{n=0}^{\infty}\left(\frac{1}{2^k}\right)^n \cdot A^n \cdot v_0 - p(t,0) = c \cdot \left(\sum_{n=0}^{\infty}\left(\frac{1}{2^k}\cdot A\right)^n\right) \cdot v_0 - p(t,0).\end{aligned}$$

We have to subtract $p(t,0)$ as the state *choice* cannot be reached with weight 0, but the summand $1 \cdot p(t,0)$ occurs in the sum. As $p(t,0) = \frac{k}{k+1} + \beta_0$, this does not cause a problem.

We claim that the matrix series involved converges to a rational matrix. We observe that the maximal row sum in A is at most $|\alpha_1|+\dots+|\alpha_k| < 1$ because the rows of the matrix contain exactly the probabilities to reach $t_0, \dots t_{-k+1}$, s_0, …, and s_{-k+1} from a state t_{+i} or s_{+i} in C for $1 \leq i \leq k$. But the probability to reach $goal_{+i}$ from these states is already $1-|\alpha_1|-\dots-|\alpha_k|$. Hence, $\|A\|_\infty$, the operator norm induced by the maximum norm $\|\cdot\|_\infty$, which equals $\max_i \sum_{j=1}^{2k}|A_{ij}|$, is less than 1. So, in particular, also $\|\frac{1}{2^k}A\|_\infty < 1$ and hence the Neumann series $\sum_{n=0}^{\infty}\left(\frac{1}{2^k}A\right)^n$ converges to $\left(I_{2k} - \frac{1}{2^k}A\right)^{-1}$ where I_{2k} is the identity matrix of size $2k\times 2k$. So,

$$\vartheta = c \cdot \left(I_{2k} - \frac{1}{2^k}A\right)^{-1} \cdot v_0 - p(t,0)$$

is computable in polynomial time. ■

All in all, this finishes the proof of item (1) of Theorem 4.1: We have seen that the MDP $\mathcal{M}$ and the threshold ϑ can be constructed in time polynomial in the size of the representations of $\alpha_1,\dots,\alpha_k$ and $\beta_0,\dots,\beta_{k-1}$. As $\vartheta = \Pr^{\mathfrak{S}}_{\mathcal{M},s_{init}}(\Diamond(\text{accumulated weight } < 0))$, we furthermore know that

$$\Pr^{\max}_{\mathcal{M},s_{init}}(\Diamond(\text{accumulated weight } < 0)) > \vartheta$$

if and only if the scheduler $\mathfrak{S}$ is not optimal. By Lemma 4.3, this is the case if and only if the given linear recurrence sequence $(u_n)_{n\geq 0}$ has a negative member: If $u_w < 0$ for some $w \in \mathbb{N}$, then the following scheduler $\mathfrak{T}$ achieves a value greater than ϑ. The scheduler $\mathfrak{T}$ behaves like $\mathfrak{S}$ except when in state *choice* with accumulated weight w. In this case, $\mathfrak{T}$ chooses σ instead of τ. As *choice* is reached with accumulated weight w with positive probability and $p(s,w) > p(t,w)$ this scheduler outperforms $\mathfrak{S}$ as it behaves optimally when reaching state t with accumulated weight w as shown in the proof of Lemma 4.3. If on the other hand $u_w \geq 0$ for all $w \in \mathbb{N}$, then $p(t,w) > p(s,w)$ for all w and hence action τ is always optimal in state *choice*. As $\mathfrak{S}$ behaves optimally once t or s is reached and always chooses τ, scheduler $\mathfrak{S}$ is indeed optimal in this case.

Finally, we want to emphasize again that the absolute values of the weights in the constructed MDP are at most k. Hence, if we want to view $\mathcal{M}$ as a one-counter MDP in which the counter value can only be increased or decreased by 1 in each step, the constructed MDP becomes only polynomially larger after we replace the transitions with a weight $+w$ or $-w$ for a $1 \leq w \leq k$ by a sequence of w states decreasing or increasing the counter value, which allowed us to conclude Corollary 4.2.

Proof of Theorem 4.1(2). The construction we provided so far shows that the threshold problem for the *maximal* termination probability of one-counter MDPs is Positivity-hard. Using exactly the same ideas, we can show that the threshold problem for the *minimal* termination probability is Positivity-hard as well. Let us describe the necessary changes in the construction that are also depicted in Figure 8. We rename the state *trap* to *trap′* and add a transition with weight $-k$ to a new absorbing state *trap*. For all $0 \leq j \leq k-1$, now state *trap* is reached directly with probability 1 and weight $-j$ from the states x_j and y_j. Furthermore, the probability to reach x_j when choosing γ_j in t is changed to $\frac{j+1}{k+1} + \beta_j$ and the probability to reach *trap′* is adjusted accordingly. The analogous change is performed for δ_j. Now, it is easy to check that the optimal choice to minimize the termination probability in state t is to choose γ if the accumulated weight is $\geq k$. In this case the probability of termination is less than $\frac{1}{k+1}$. If the accumulated weight is $0 \leq \ell < k$, the optimal choice is γ_ℓ. The analogous result holds in state s. From then on the proof is analogous to the proof for the maximal termination probability with the change that we have to consider the scheduler $\mathfrak{S}$ always choosing σ in the state *choice* this time. This scheduler is optimal to minimize the termination probability if and only if the given linear recurrence sequence is non-negative. With these adjustments, we conclude:

COROLLARY 4.5. *The Positivity problem is reducible in polynomial time to the following problem: Given an MDP $\mathcal{M}$ and a rational $\vartheta \in (0,1)$, decide whether*

$$\Pr^{\min}_{\mathcal{M},s_{init}}(\Diamond(\textit{accumulated weight} < 0)) < \vartheta.$$

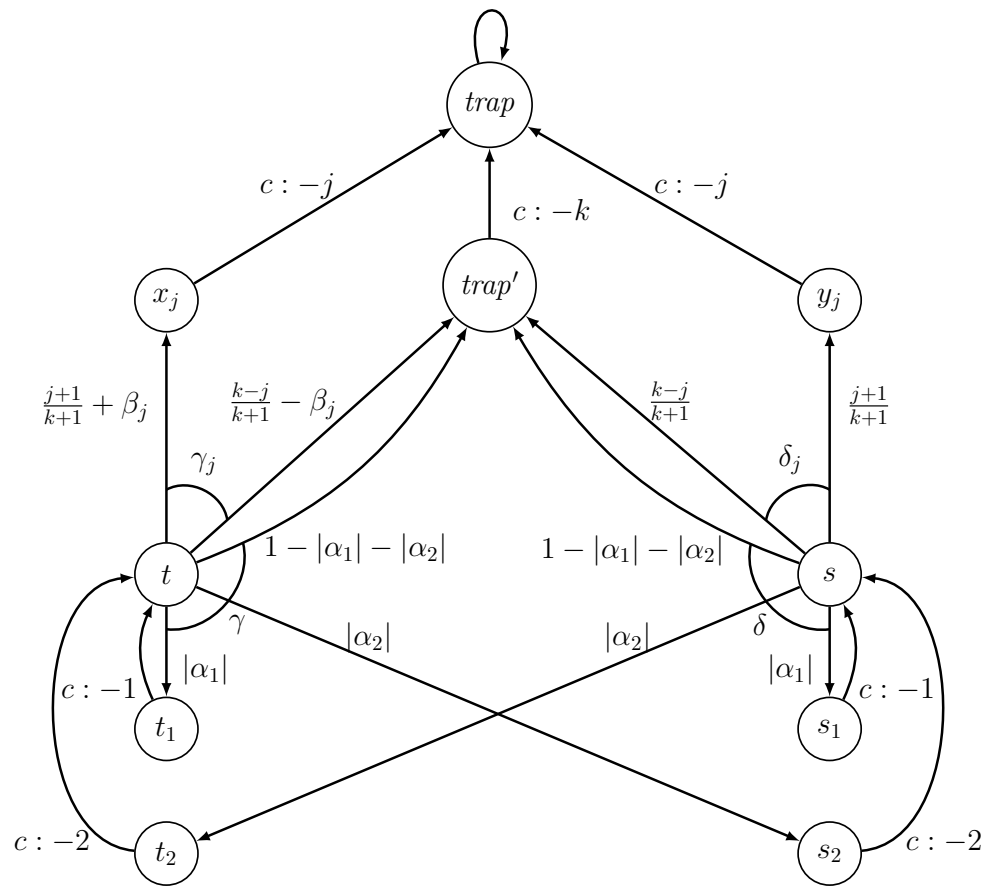

Figure 8. Necessary changes to the construction for the result for minimal termination probabilities. The initial component of the MDP is omitted here and stays unchanged.

REMARK 4.6. There is no obvious way to adjust the construction such that the Positivity-hardness of the question whether $\Pr^{\max}_{\mathcal{M},s_{init}}(\Diamond(\text{accumulated weight} < 0)) \geq \vartheta$ would follow. One attempt would be to provide an ε such that $\Pr^{\max}_{\mathcal{M},s_{init}}(\Diamond(\text{accumulated weight} < 0)) > \vartheta$ if and only if $\Pr^{\min}_{\mathcal{M},s_{init}}(\Diamond(\text{accumulated weight} < 0)) \geq \vartheta + \varepsilon$. 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 problem. The analogous observation applies to the question whether $\Pr^{\min}_{\mathcal{M},s_{init}}(\Diamond(\text{accumulated weight} < 0)) \leq \vartheta$ and all Positivity-hardness results in the sequel.

Energy objectives. As the energy objective $\Box(\text{accumulated weight} \geq 0)$ is satisfied if and only if $\Diamond(\text{accumulated weight} < 0)$ does not hold, the Positivity-hardness of the threshold problem of the optimal satisfaction probability of an energy objective follows easily. As

$$\Pr^{\max}_{\mathcal{M},s_{init}}(\Box(\text{accumulated weight} \geq 0)) = 1 - \Pr^{\min}_{\mathcal{M},s_{init}}(\Diamond(\text{accumulated weight} < 0)),$$

we conclude:

COROLLARY 4.7. *The Positivity problem is reducible in polynomial time to the following problems: Given an MDP $\mathcal{M}$ and a rational $\vartheta \in (0,1)$,*

1. *decide whether* $\Pr^{\max}_{\mathcal{M},s_{init}}(\Box(\textit{accumulated weight} \geq 0)) > \vartheta$.
2. *decide whether* $\Pr^{\min}_{\mathcal{M},s_{init}}(\Box(\textit{accumulated weight} \geq 0)) < \vartheta$.

Cost problems and quantiles. The proof of the Positivity-hardness of the threshold problem for the termination probability of one-counter MDPs in fact also serves as a proof that cost problems and the computation of quantiles of the accumulated weight before reaching a goal state are Positivity-hard. Observe that in the MDP constructed for Theorem 4.1 and Corollary 4.5, almost all paths ζ under any scheduler satisfy $\Diamond(\text{accumulated weight} < 0)$ if and only if they satisfy $\diamondplus \textit{trap}(\zeta) < 0$ if and only if their total accumulated weight is less than 0. Thus, we obtain the following corollary:

COROLLARY 4.8. *The Positivity problem is reducible in polynomial time to the following problems: Given an MDP $\mathcal{M}$ with a designated set of trap states Goal and a rational $\vartheta \in (0,1)$,*

1. *decide whether* $\Pr^{\max}_{\mathcal{M},s_{init}}(\diamondplus \textit{Goal} < 0) > \vartheta$.
2. *decide whether* $\Pr^{\min}_{\mathcal{M},s_{init}}(\diamondplus \textit{Goal} < 0) < \vartheta$.

The analogous result also holds for the total accumulated weight.

Termination times of one-counter MDPs. To conclude the section, we show that not only the threshold problems for optimal termination probabilities, but also for the optimal expected termination times in one-counter MDPs that terminate almost surely is Positivity-hard. We again work with weighted MDPs. Let T be the random variable that assigns to each path in a weighted MDP $\mathcal{M}$ the length of the shortest prefix π such that $wgt(\pi) < 0$. To reflect precisely the behavior of a one-counter MDP, we now will work with MDPs where the weight is reduced or increased by at most 1 in each step. We make a small change to the MDP constructed for the proof of Corollary 4.5 that is depicted in Figure 8. The initial component (that is not depicted) stays unchanged. For the remaining transitions, all transition reduce the weight or leave it unchanged. The transitions with weight 0 do not occur directly after each other except for the loop at the state *trap* that we adjust in a moment. Hence, we can add additional auxiliary states such that along each path starting from s or t not reaching the state *trap*, the weight is left unchanged and reduced by 1 in an alternating fashion. So, if a path starts in state s or t with accumulated weight w and terminates (i.e., reaches accumulated weight -1) before reaching the state *trap* this takes $2(w+1)$ steps. Now, we replace the loop at the state *trap* by the gadget depicted in Figure 9 and let us call the resulting MDP $\mathcal{N}$. So, when reaching *trap* the accumulated weight is increased by 1 before it is reduced in every other step until termination. That means that if a path starting in state s or t with weight w does not terminate before reaching *trap*, the termination time is $2(w+1)+3$ steps.

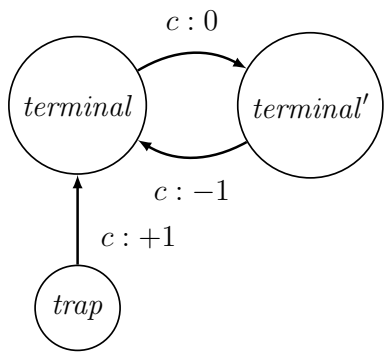

Figure 9. Necessary changes to the construction for the result for for maximal expected termination times.

Now, let $\mathfrak{S}$ be a scheduler and denote the probability not to terminate before reaching *trap* under $\mathfrak{S}$ by $p^{\mathfrak{S}}$. For the expected termination time T in $\mathcal{N}$, we now have

$$\mathbb{E}^{\mathfrak{S}}_{\mathcal{N},s_{init}} = \left(\sum_{i=1}^{\infty}(1/2)^i(i+2(i+1))\right) + 3 \cdot p^{\mathfrak{S}} = 8 + 3 \cdot p^{\mathfrak{S}}.$$

The summands $(1/2)^i(i+2(i+1))$ correspond to the probability to accumulated weight i in the initial component which takes i steps and the $2(i+1)$ steps needed to terminate by alternatingly leaving the weight unchanged and reducing it by 1. The three additional steps after *trap* occur precisely with probability $p^{\mathfrak{S}}$.

Not terminating before *trap* corresponds exactly to not terminating at all in the MDP constructed for Corollary 4.5. The termination probability there is hence $1 - p^{\mathfrak{S}}$ for any scheduler $\mathfrak{S}$. It is hence possible to terminate with a probability less than ϑ in that MDP if and only if it is possible to reach an expected termination time of more than $11 - 3\vartheta$ in $\mathcal{N}$. By Corollary 4.5 and the fact that termination is reached almost surely in $\mathcal{N}$ under any scheduler, we hence conclude:

COROLLARY 4.9. *Let $\mathcal{M}$ be a one-counter MDP with initial state s_{init} that terminates almost surely under any scheduler, let ϑ be a rational, and let T be the random variable assigning the termination time to runs. The Positivity problem is polynomial-time reducible to the problem whether*

$$\mathbb{E}^{\max}_{\mathcal{M},s_{init}}(T) > \vartheta.$$

The analogous argument with similar changes to the MDP used in the proof of Theorem 4.1 can be used to show the analogous result for the problem whether $\mathbb{E}^{\min}_{\mathcal{M},s_{init}}(T) < \vartheta$.

4.2 Partial and conditional stochastic shortest path problems

Our next goal is to prove that the partial and conditional SSPPs are Positivity-hard. Note that this stands in strong contrast to the classical SSPP, which is solvable in polynomial time [16, 30, 5]. We start by providing a formal definition of the decision versions of these two problems.

Let $\mathcal{M}$ be an MDP with a designated set of terminal states *Goal*. We define the random variable $\oplus Goal$ on maximal paths ζ of $\mathcal{M}$:

$$\oplus Goal(\zeta) = \begin{cases} wgt(\zeta) & \text{if } \zeta \vDash \Diamond Goal, \\ 0 & \text{otherwise.} \end{cases}$$

The objective in the *partial SSPP* is to maximize the expected value of $\oplus Goal$ which we call the *partial expected accumulated weight*, or *partial expectation* for short, i.e., to compute the value

$$\mathbb{PE}_{\mathcal{M}}^{\max} \stackrel{\text{def}}{=} \mathbb{E}_{\mathcal{M},s_{init}}^{\max}(\oplus Goal) = \sup_{\mathfrak{S}} \mathbb{E}_{\mathcal{M},s_{init}}^{\mathfrak{S}}(\oplus Goal)$$

where the supremum ranges over all schedulers $\mathfrak{S}$. The threshold problem asks, given a rational ϑ, whether

$$\mathbb{PE}_{\mathcal{M}}^{\max} > \vartheta.$$

Note that the minimization of the partial expectation can be reduced to the maximization by multiplying all weights in $\mathcal{M}$ with -1.

The *conditional expectation* under a scheduler $\mathfrak{S}$ that reaches *Goal* with positive probability is the value

$$\mathbb{CE}_{\mathcal{M}}^{\mathfrak{S}} \stackrel{\text{def}}{=} \mathbb{E}_{\mathcal{M}}^{\mathfrak{S}}(\oplus Goal \mid \Diamond Goal).$$

Again, we are interested in the maximal value

$$\mathbb{CE}_{\mathcal{M}}^{\max} \stackrel{\text{def}}{=} \sup_{\mathfrak{S}} \mathbb{CE}^{\mathfrak{S}}$$

where the supremum ranges over all schedulers $\mathfrak{S}$ with $\Pr_{\mathcal{M}}^{\mathfrak{S}}(\Diamond Goal) > 0$. Consequently, the threshold problem asks for a given rational ϑ whether

$$\mathbb{CE}^{\max} > \vartheta.$$

Again, multiplying all weights with -1 reduces the minimization of the conditional expectation to the maximization. Furthermore, given a further set of states F, the problem to maximize $\mathbb{E}_{\mathcal{M}}^{\mathfrak{S}}(\diamondplus Goal \mid \Diamond F)$ among all schedulers $\mathfrak{S}$ that reach F with positive probability can be reduced to the conditional SSPP in our formulation as shown in [11][2].

Partial SSPP. In the sequel, we will provide a direct reduction from the Positivity problem to the partial SSPP using our modular approach via MDP-gadgets to prove the following result:

THEOREM 4.10. *The Positivity problem is polynomial-time reducible to the decision version of the partial SSPP, i.e., the question whether*

$$\mathbb{PE}_{\mathcal{M}}^{\max} > \vartheta$$

2 In [11], only MDPs with non-negative weights are considered. The reduction of [11], however, does not require the restriction to non-negative weights.

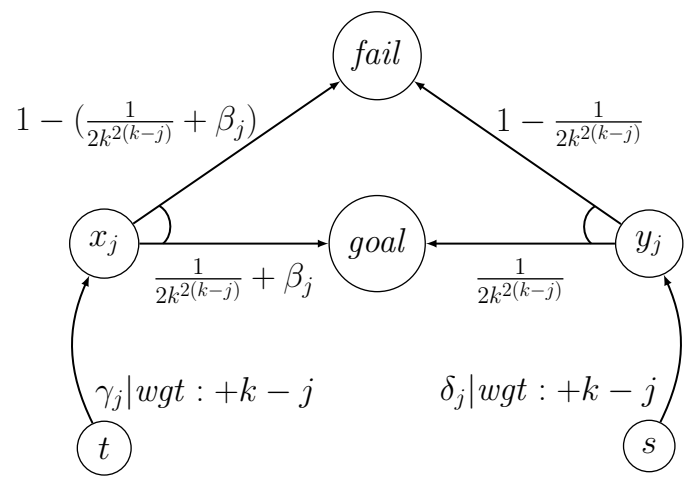

Figure 10. The gadget $\mathcal{P}_{\bar{\beta}}$ encoding the initial values in the reduction to the threshold problem for partial expectations.

for a given MDP $\mathcal{M}$ and a given rational ϑ.

Again, we split up the proof of the theorem into the construction of the MDP with the proof of the correctness of the encoding of the linear recurrence sequence and the computation of the threshold ϑ.

Proof of Theorem 4.10: construction of the MDP and correctness of the encoding of a linear recurrence sequence. 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 $0 \leq j \leq k-1$ via $u_0 = \beta_0, \ldots, u_{k-1} = \beta_{k-1}$ and $u_{n+k} = \alpha_1 u_{n+k-1} + \cdots + \alpha_k u_n$ for all $n \geq 0$. By Assumption 3.1, we can assume w.l.o.g. that $\sum_i |\alpha_i| < \frac{1}{4}$ and that $0 \leq \beta_j < \frac{1}{4k^{2k+2}}$ for all j.

We begin by constructing a gadget $\mathcal{P}_{\bar{\beta}}$ that encodes the initial values $\beta_0, \ldots, \beta_{k-1}$. The gadget is depicted in Figure 10 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 furthermore reuse the initial gadget $\mathcal{I}$ and the gadget encoding the linear recurrence relation $\mathsf{G}_{\bar{\alpha}}$ from the previous section. In the gadget $\mathsf{G}_{\bar{\alpha}}$, we rename the absorbing state *trap* to the terminal state *goal* which is the target state for the partial SSPP. As before, we glue together the three gadgets $\mathcal{I}$, $\mathsf{G}_{\bar{\alpha}}$ and $\mathcal{P}_{\bar{\beta}}$ at states s, t, and *goal*. Let us call the full MDP that we obtain in this way $\mathcal{M}$ which is depicted in Figure 11. We denote the state space by S.

The somewhat complicated choices of probability values lead to the following lemma showing the correct interplay between the gadgets constructed via straight-forward computations.

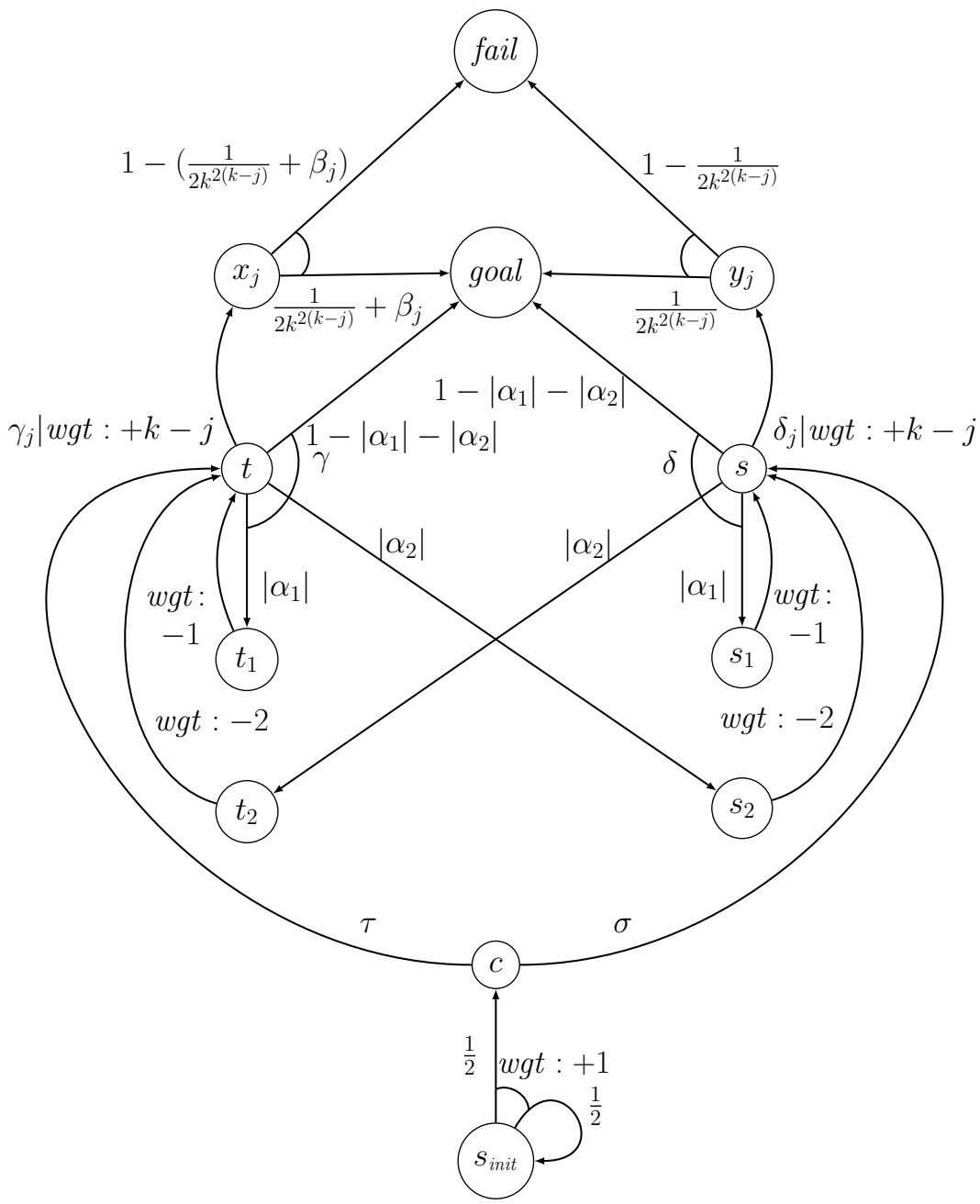

Figure 11. The full MDP for the Positivity-hardness proof for partial expectations. The MDP contains the upper part for all $0 \leq j \leq k-1$. The middle part is depicted for $k = 2$, $\alpha_1 \geq 0$, and $\alpha_2 < 0$.

LEMMA 4.11. *Consider the full MDP $\mathcal{M}$. 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.*

PROOF. Suppose action γ_i is chosen in state t when starting with weight $-(k-1)+j$. So, state x_i is reached with weight $-(k-1)+j+(k-i) = 1+j-i$. Then the partial expectation achieved from this situation is

$$(1+j-i)(\frac{1}{2k^{2(k-i)}}+\beta_i).$$

For $i > j$ this value is ≤ 0 and hence γ_i is certainly not optimal. For $i = j$, we obtain a partial expectation of

$$\frac{1}{2k^{2(k-j)}} + \beta_j.$$

For $i < j$, state x_i is reached with weight $1 + j - i \leq k$. Further, $\beta_i \leq \frac{1}{4k^{2k+2}}$ and $\frac{1}{2k^{2(k-i)}} \leq \frac{1}{2k^{2(k-j)} \cdot k^2}$. So, the partial expectation obtained via γ_i is at most

$$\frac{k}{2k^{2(k-j)} \cdot k^2} + \frac{k}{4k^{2k+2}} < \frac{1}{2k^{2(k-j)}}.$$

So, indeed action γ_j maximizes the partial expectation among the actions γ_i with $0 \leq i \leq k-1$ when the accumulated weight in state t is $-(k-1) + j$. The argument for state s is the same with $\beta_i = 0$ for all i. It is easy to see that for accumulated weight $-(k-1) + j$ with $0 \leq j \leq k-1$ actions γ or δ are not optimal in state t or s: If *goal* is reached immediately, the weight is not positive and otherwise states t or s are reached with lower accumulated weight again. The values β_j are chosen small enough such that also a switch from state t to s while accumulating negative weight does not lead to a higher partial expectation.

For positive accumulated weight w, the optimal partial expectation when choosing γ first is at least $\frac{3}{4}w$ by construction and the fact that a positive value can be achieved from any possible successor state via one of the actions γ_j and δ_j with $0 \leq j \leq k-1$. Choosing γ_j on the other hands results in a partial expectation of at most $(k+w) \cdot (\frac{1}{4k^{2k+2}} + \frac{1}{2k^2})$ which is easily seen to be less than $\frac{3}{4}w$ as $k \geq 2$. ■

For each weight w, denote by $e(t, w)$ and $e(s, w)$ the optimal partial expectation when starting in state t or s with accumulated weight w in $\mathcal{M}$ as if the respective state was reached from the initial state with weight w and probability 1. For each weight $w \geq -k+1$, denote by $d(w)$ the difference $e(t, w) - e(s, w)$ between these optimal partial expectation when starting in state t and s with weight w. Comparing action γ_j and δ_j for starting weight $-(k-1)+j$, we conclude from the previous lemma that the difference between optimal values $d(-(k-1)+j)$ is equal to β_j, for $0 \leq j \leq k-1$.

The important fact we use next is that for partial expectations, the optimal values $e(r, w)$ for states $r \in S \setminus \{goal\}$ and starting weights $w \in \mathbb{Z}$ satisfies the optimality equation $(*)$ from Section 3.2 when setting $e(goal, w) = w$ as already shown in [25]:

$$e(r, w) = \max_{\alpha \in Act(r)} \sum_{r' \in S} P(r, \alpha, r') \cdot e(r', w + wgt(s, \alpha)).$$

By the fact that $\mathsf{G}_{\bar{\alpha}}$ encodes the given linear recurrence relation as soon as γ and δ are the optimal actions as shown in Section 3.2, we conclude the following lemma:

LEMMA 4.12. *Consider the linear recurrence sequence $(u_n)_{n \geq 0}$ given above by $\alpha_1, \ldots, \alpha_k$ and $\beta_0, \ldots, \beta_{k-1}$ and the MDP $\mathcal{M}$ constructed from this sequence. We have*

$$d(-(k-1) + n) = u_n$$

for all n with the values $d(w)$ just defined.

Proof of Theorem 4.10: computation of the threshold ϑ. Let us now consider a run of the MDP $\mathcal{M}$. For any $w > 0$, state c is reached with accumulated weight w with positive probability. As before, an optimal scheduler has to decide whether the partial expectation when starting with weight w is better in state s or t: Action τ is optimal in c for accumulated weight w if and only if $d(w) \geq 0$. Once t or s is reached, the optimal actions are given by Lemma 4.11. Let $\mathfrak{S}$ be the scheduler that always chooses τ in c and actions $\gamma, \gamma_0, \ldots, \gamma_{k-1}, \delta, \ldots$ as described in Lemma 4.11. We conclude that $\mathfrak{S}$ is optimal if and only if the given linear recurrence sequence is non-negative. The remaining step is hence in our reduction is hence to prove that the partial expectation under $\mathfrak{S}$ is rational and can be computed in polynomial time:

LEMMA 4.13. *Let $\mathfrak{S}$ be the scheduler for the constructed MDP $\mathcal{M}$ always choosing τ in c and actions $\gamma, \gamma_0, \ldots, \gamma_{k-1}, \delta, \ldots$ as described in Lemma 4.11. The value $\mathbb{PE}^{\mathfrak{S}}_{\mathcal{M}}$ is rational and computable in polynomial time.*

PROOF. Recall that the scheduler $\mathfrak{S}$ chooses γ and δ, respectively, as long as the accumulated weight is positive. For an accumulated weight of $-(k-1)+j$ for $0 \leq j \leq k-1$, it chooses actions γ_j and δ_j, respectively.

Analogously to the proof of Lemma 4.4, we want to recursively express the partial expectations under $\mathfrak{S}$ starting from t or s with some positive accumulated weight $n \in \mathbb{N}$ which we again denote by $e(t,n)$ and $e(s,n)$, respectively. In order to do so, we reuse the following Markov chain C from Lemma 4.4 also depicted in Figure 7 which we briefly recall here: The Markov chain C has $5k$ states named $t_{-k+1}, \ldots, t_{+k}$, $s_{-k+1}, \ldots, s_{+k}$, and $goal_{+1}, \ldots, goal_{+k}$. States $t_{-k+1}, \ldots, t_0$, $s_{-k+1}, \ldots, s_0$, and $goal_{+1}, \ldots, goal_{+k}$ are absorbing. For $0 < i, j \leq k$, there are transitions from t_{+i} to t_{+i-j} with probability α_j if $\alpha_j > 0$, to s_{+i-j} with probability $|\alpha_j|$ if $\alpha_j < 0$, and to $goal_{+i}$ with probability $1 - |\alpha_1| - \ldots - |\alpha_k|$. Transitions from s_{+i} are defined analogously.

The idea behind this Markov chain is that the reachability probabilities describe how, for arbitrary $n \in \mathbb{N}$ and $1 \leq i \leq k$, the values $e(t, nk+i)$ and $e(s, nk+i)$ depend on n and the values $e(t,(n-1)k+j)$ and $e(s,(n-1)k+j)$ for $1 \leq j \leq k$. The transitions in C behave as γ and δ in $\mathcal{M}$, but the decrease in the accumulated weight is explicitly encoded into the state space. Namely, for $n \in \mathbb{N}$ and $0 < i \leq k$, we have

$$\begin{aligned} e(t, nk+i) = &\sum_{j=1}^{k} \big(\Pr_{C,t_{+i}}(\Diamond t_{-k+j}) \cdot e(t,(n-1)k+j) + \Pr_{C,t_{+i}}(\Diamond s_{-k+j}) \cdot e(s,(n-1)k+j)\big) \\ &+ \sum_{j=1}^{k} \Pr_{C,t_{+i}}(\Diamond goal_{+j}) \cdot (nk+j) \end{aligned} \tag{$**$}$$

and analogously for $e(s, nk+i)$. We now group the optimal values together in the following column vectors

$$v_n = (e(t, nk+k), e(t, nk+k-1), \dots, e(t, nk+1), e(s, nk+k), \dots, e(s, nk+1))^\top$$

for $n \in \mathbb{N}$. In other words, this vector contains the optimal values for the partial expectation when starting in t or s with an accumulated weight from $\{nk+1, \dots, nk+k\}$. Further, we define the vector containing the optimal values for weights in $\{-k+1, \dots, 0\}$ which are the least values of accumulated weight reachable under scheduler $\mathfrak{S}$.

$$v_{-1} = (e(t, 0), e(t, -1), \dots, e(t, -k+1), e(s, 0), e(s, -1), \dots, e(s, -k+1))^\top.$$

As we have seen, these values are given as follows:

$$e(t, -k+1+j) = \frac{1}{2k^{2(k-j)}} + \beta_j \text{ and } e(s, -k+1+j) = \frac{1}{2k^{2(k-j)}}$$

for $0 \leq j \leq k-1$.

As the reachability probabilities in C are rational and computable in polynomial time, we conclude from $(**)$ that there are a matrix $A \in \mathbb{Q}^{2k \times 2k}$, and vectors a and b in $\mathbb{Q}^{2k}$ computable in polynomial time such that

$$v_n = Av_{n-1} + na + b,$$

for all $n \in \mathbb{N}$. We claim that the following explicit representation for $n \geq -1$ satisfies this recursion:

$$v_n = A^{n+1}v_{-1} + \sum_{i=0}^{n}(n-i)A^i a + \sum_{i=0}^{n} A^i b.$$

We show this by induction: Clearly, this representation yields the correct value for v_{-1}. So, assume $v_n = A^{n+1}v_{-1} + \sum_{i=0}^{n}(n-i)A^i a + \sum_{i=0}^{n} A^i b$. Then,

$$\begin{aligned} v_{n+1} &= A(A^{n+1}v_{-1} + \sum_{i=0}^{n}(n-i)A^i a + \sum_{i=0}^{n} A^i b) + (n+1)a + b \\ &= A^{n+2}v_{-1} + \left(\sum_{i=0}^{n}(n-i)A^{i+1}a\right) + (n+1)A^0 a + \left(\sum_{i=1}^{n+1} A^i b\right) + A^0 b \\ &= A^{n+2}v_{-1} + \sum_{i=0}^{n+1}(n+1-i)A^i a + \sum_{i=0}^{n+1} A^i b. \end{aligned}$$

So, we have an explicit representation for v_n. The value we are interested in is

$$\mathbb{PE}_{\mathcal{M}}^{\mathfrak{S}} = \sum_{\ell=1}^{\infty}(1/2)^\ell e(t, \ell).$$

Let $c = (\frac{1}{2^k}, \frac{1}{2^{k-1}}, \ldots, \frac{1}{2^1}, 0, \ldots, 0)$. Then,

$$(\frac{1}{2^k})^n c \cdot v_n = \sum_{i=1}^{k} \frac{1}{2^{nk+i}} e(t, nk+i).$$

Hence, we can write

$$\begin{aligned}
\mathbb{PE}_{\mathcal{M}}^{\mathfrak{S}} &= \sum_{n=0}^{\infty} (\frac{1}{2^k})^n c \cdot v_n = c \cdot \sum_{n=0}^{\infty} (\frac{1}{2^k})^n v_n \\
&= c \cdot \sum_{n=0}^{\infty} (\frac{1}{2^k})^n (A^{n+1} v_{-1} + \sum_{i=0}^{n} (n-i) A^i a + \sum_{i=0}^{n} A^i b) \\
&= c \cdot \big((\sum_{n=0}^{\infty} (\frac{1}{2^k})^n A^{n+1}) v_{-1} + (\sum_{n=0}^{\infty} (\frac{1}{2^k})^n \sum_{i=0}^{n} (n-i) A^i) a + (\sum_{n=0}^{\infty} (\frac{1}{2^k})^n \sum_{i=0}^{n} A^i) b \big).
\end{aligned}$$

We claim that all of the matrix series involved converge to rational matrices. As in the proof of Lemma 4.4, we observe that the maximal row sum in A is at most $|\alpha_1| + \ldots + |\alpha_k| < 1$ because the rows of the matrix contain exactly the probabilities to reach $t_0, \ldots t_{-k+1}$, $s_0, \ldots$, and s_{-k+1} from a state t_{+i} or s_{+i} in C for $1 \leq i \leq k$. But the probability to reach $goal_{+i}$ from these states is already $1 - |\alpha_1| - \ldots - |\alpha_k|$. Hence, $\|A\|_\infty$, the operator norm induced by the maximum norm $\|\cdot\|_\infty$, which equals $\max_i \sum_{j=1}^{2k} |A_{ij}|$, is less than 1. So, of course also $\|\frac{1}{2^k} A\|_\infty < 1$ and hence the Neumann series $\sum_{n=0}^{\infty} (\frac{1}{2^k} A)^n$ converges to $(I_{2k} - \frac{1}{2^k} A)^{-1}$ where I_{2k} is the identity matrix of size $2k \times 2k$. So,

$$\sum_{n=0}^{\infty} (\frac{1}{2^k})^n A^{n+1} = A \sum_{n=0}^{\infty} (\frac{1}{2^k} A)^n = A (I_{2k} - \frac{1}{2^k} A)^{-1}.$$

Note that $\|A\|_\infty < 1$ also implies that $I_{2k} - A$ is invertible. We observe that for all n,

$$\sum_{i=0}^{n} A^i = (I_{2k} - A)^{-1} (I_{2k} - A^{n+1})$$

which is shown by straight-forward induction. Therefore,

$$\begin{aligned}
\sum_{n=0}^{\infty} (\frac{1}{2^k})^n \sum_{i=0}^{n} A^i &= (I_{2k} - A)^{-1} \left(\sum_{n=0}^{\infty} (\frac{1}{2^k})^n I_{2k} - A \sum_{n=0}^{\infty} (\frac{1}{2^k} A)^n \right) \\
&= (I_{2k} - A)^{-1} \left(\frac{2^k}{2^k - 1} I_{2k} - A (I_{2k} - \frac{1}{2^k} A)^{-1} \right).
\end{aligned}$$

Finally, we show by induction that

$$\sum_{i=0}^{n} (n-i) A^i = (I_{2k} - A)^{-2} (A^{n+1} - A + n(I_{2k} - A)).$$

This is equivalent to

$$(I_{2k} - A)^2 \sum_{i=0}^{n} (n-i)A^i = A^{n+1} - A + n(I_{2k} - A).$$

For $n = 0$, both sides evaluate to 0. So, we assume the claim holds for n.

$$\begin{aligned}
(I_{2k} - A)^2 \sum_{i=0}^{n+1} (n+1-i)A^i &= (I_{2k} - A)^2 \sum_{i=0}^{n} (n-i)A^i + (I_{2k} - A)^2 \sum_{i=0}^{n} A^i \\
&\overset{\text{IH}}{=} A^{n+1} - A + n(I_{2k} - A) + (I_{2k} - A)^2 \sum_{i=0}^{n} A^i \\
&= A^{n+1} - A + n(I_{2k} - A) + (I_{2k} - A)^2 (I_{2k} - A)^{-1} (I_{2k} - A^{n+1}) \\
&= A^{n+1} - A + n(I_{2k} - A) + I_{2k} - A - A^{n+1} + A^{n+2} \\
&= A^{n+2} - A + (n+1)(I_{2k} - A).
\end{aligned}$$

The remaining series is the following:

$$\begin{aligned}
\sum_{n=0}^{\infty} (\frac{1}{2^k})^n \sum_{i=0}^{n} (n-i)A^i &= \sum_{n=0}^{\infty} (\frac{1}{2^k})^n (I_{2k} - A)^{-2} (A^{n+1} - A + n(I_{2k} - A)) \\
&= (I_{2k} - A)^{-2} \left(\sum_{n=0}^{\infty} (\frac{1}{2^k})^n A^{n+1} - \sum_{n=0}^{\infty} (\frac{1}{2^k})^n A + \sum_{n=0}^{\infty} (\frac{1}{2^k})^n n(I_{2k} - A) \right) \\
&= (I_{2k} - A)^{-2} \left(A(I_{2k} - \frac{1}{2^k}A)^{-1} - \frac{2^k}{2^k - 1} A + \frac{2^k}{(2^k-1)^2} (I_{2k} - A) \right).
\end{aligned}$$

We conclude that all expressions in the representation of $\mathbb{PE}_{\mathcal{M}}^{\mathfrak{S}}$ above are rational and computable in polynomial time. ■

As we have also seen before, the originally given linear recurrence sequence contains a negative member if and only if the scheduler $\mathfrak{S}$ is not optimal: If the sequence is non-negative, the scheduler $\mathfrak{S}$ is optimal as it behaves optimally once t or s has been reached and as it always moves to state t instead of s from *choice* for any value w of accumulated weight, which is optimal by Lemma 4.12. If the sequence is negative at position n, then a scheduler that behaves like $\mathfrak{S}$, except for choosing σ for accumulated weight $w = n - (k-1)$ in state *choice* is better than $\mathfrak{S}$ as $e(s, w) > e(t, w)$ in this case. Note that for positions $n = 0, \ldots, k-1$, we may assume that u_n is non-negative as otherwise the sequence trivially becomes negative. So, the originally given linear recurrence sequence contains a negative member if and only if $\mathbb{PE}_{\mathcal{M}}^{\max} > \mathbb{PE}_{\mathcal{M}}^{\mathfrak{S}}$ for the MDP $\mathcal{M}$ constructed from the linear recurrence sequence in polynomial time above. This finishes the proof of Theorem 4.10.

Conditional SSPP. For the Positivity-hardness of the threshold problem for conditional expectations, we provide a reduction from the threshold problem for partial expectations in the

following lemma. Note that a reduction in the other direction is provided in [59] rendering the two problems polynomial-time inter-reducible.

LEMMA 4.14. *The threshold problem for the partial SSPP is polynomial-time reducible to the threshold problem of the conditional SSPP.*

PROOF. Let $\mathcal{M}$ be an MDP with a designated terminal target state *goal* and let ϑ be a rational number. We construct an MDP $\mathcal{N}$ such that $\mathbb{PE}_{\mathcal{M}}^{\max} > \vartheta$ if and only if $\mathbb{CE}_{\mathcal{N}}^{\max} > \vartheta$. We obtain $\mathcal{N}$ by adding a new initial state s'_{init}, renaming the state *goal* to *goal'*, and adding a new state *goal* to $\mathcal{M}$. In s'_{init}, one action with weight 0 is enabled leading to the old initial state s_{init} and to *goal* with probability 1/2 each. From *goal'* there is one new action leading to *goal* with probability 1 and weight $+\vartheta$.

Each scheduler $\mathfrak{S}$ for $\mathcal{M}$ can be seen as a scheduler for $\mathcal{N}$ and vice versa. Now, we observe that for any scheduler $\mathfrak{S}$,

$$\mathbb{CE}_{\mathcal{N}}^{\mathfrak{S}} = \frac{1/2(\mathbb{PE}_{\mathcal{M}}^{\mathfrak{S}} + \Pr_{\mathcal{M}}^{\mathfrak{S}}(\Diamond goal)\vartheta)}{1/2 + 1/2\Pr_{\mathcal{M}}^{\mathfrak{S}}(\Diamond goal)} = \frac{\mathbb{PE}_{\mathcal{M}}^{\mathfrak{S}} + \Pr_{\mathcal{M}}^{\mathfrak{S}}(\Diamond goal)\vartheta}{1 + \Pr_{\mathcal{M}}^{\mathfrak{S}}(\Diamond goal)}.$$

Hence, $\mathbb{PE}_{\mathcal{M}}^{\max} > \vartheta$ if and only if $\mathbb{CE}_{\mathcal{N}}^{\max} > \vartheta$. ■

Together with the Positivity-hardness of the threshold problem for partial expectations (Theorem 4.10), we conclude:

THEOREM 4.15. *The Positivity problem is reducible in polynomial time to the following problem: Given an MDP $\mathcal{M}$ and a rational ϑ, decide whether $\mathbb{CE}_{\mathcal{M}}^{\max} > \vartheta$.*

Two-sided partial SSPP. To conclude this section, we prove the Positivity-hardness of a two-sided version of the partial SSPP with two non-negative weight functions. The key idea is that, instead of using arbitrary integer weights, we can simulate the non-monotonic behavior of the accumulated weight along a path in the partial SSPP with arbitrary weights with two non-negative weight functions. In the definition of the random variable $\oplus Goal$, we can replace the choice that paths not reaching *Goal* are assigned weight 0 by a second weight function. Let $\mathcal{M} = (S, Act, \Pr, s_{init}, wgt_{goal}, wgt_{fail}, goal, fail)$ be an MDP with two designated terminal states *goal* and *fail* and two non-negative weight functions $wgt_{goal}\colon S \times Act \to \mathbb{N}$ and $wgt_{fail}\colon S \times Act \to \mathbb{N}$. Assume that the probability $\Pr_{\mathcal{M},s_{init}}^{\min}(\Diamond\{goal, fail\}) = 1$. Define the following random variable X on maximal paths ζ:

$$X(\zeta) = \begin{cases} wgt_{goal}(\zeta) & \text{if } \zeta \vDash \Diamond goal, \\ wgt_{fail}(\zeta) & \text{if } \zeta \vDash \Diamond fail. \end{cases}$$

Due to the assumption that *goal* or *fail* is reached almost surely under any scheduler, the expected value $\mathbb{E}_{\mathcal{M},s_{init}}^{\mathfrak{S}}(X)$ is well-defined for all schedulers $\mathfrak{S}$ for $\mathcal{M}$. We call the value $\mathbb{E}_{\mathcal{M},s_{init}}^{\max}(X) = \sup_{\mathfrak{S}} \mathbb{E}_{\mathcal{M},s_{init}}^{\mathfrak{S}}(X)$ the optimal *two-sided partial expectation*. We can show that the threshold

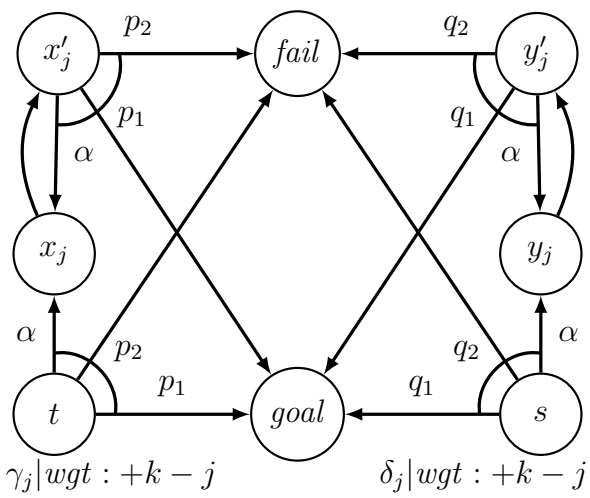

Figure 12. The gadget $\mathcal{T}_{\bar{\beta}}$ encoding initial values in terms of two-sided partial expectations.

problem for the two-sided partial expectation is Positivity-hard as well by a small adjustment of the construction above.

THEOREM 4.16. *The Positivity problem is polynomial-time reducible to the following problem: Given an MDP $\mathcal{M} = (S, Act, \Pr, s_{init}, wgt_{goal}, wgt_{fail}, goal, fail)$ as above and a rational ϑ, decide whether $\mathbb{E}^{\max}_{\mathcal{M},s_{init}}(X) > \vartheta$.*

PROOF. Given the parameters $\alpha_1, \ldots, \alpha_k$ and $\beta_0, \ldots, \beta_{k-1}$ of a rational linear recurrence sequence, we can construct an MDP $\mathcal{M}' = (S, Act, \Pr, s_{init}, wgt, goal, fail)$ with one weight function $wgt\colon S \times Act \to \mathbb{Z}$ similar to the MDP $\mathcal{M}$ depicted in Figure 11. W.l.o.g., we again assume that $\sum_i |\alpha_i| < \frac{1}{4}$ and that $0 \leq \beta_j < \frac{1}{4k^{2k+2}}$ for all j. The non-negativity of the values β_j for all j can be assumed as the Positivity problem is trivial otherwise. The initial gadget and the gadget $\mathsf{G}_{\bar{\alpha}}$ are as before. The gadget $\mathcal{P}_{\bar{\beta}}$, however, is slightly modified and replaced by the gadget $\mathrm{x}\mathcal{T}_{\bar{\beta}}$ depicted in Figure 12. For this gadget, we define $\alpha = \sum_{i=1}^{k} |\alpha_i|$, $p_1 = (1-\alpha)(\frac{1}{2k^{2(k-j)}} + \beta_j)$, $p_2 = (1-\alpha)(1 - (\frac{1}{2k^{2(k-j)}} + \beta_j))$, $q_1 = (1-\alpha)\frac{1}{2k^{2(k-j)}}$, and $q_2 = (1-\alpha)(1 - \frac{1}{2k^{2(k-j)}})$. With the transitions as in the figure, the probability to reach *goal* or *fail* and the weight accumulated does not change when choosing action γ_j or δ_j compared to the gadget $\mathcal{P}_{\bar{\beta}}$. The only difference is that the expected time to reach *goal* or *fail* changes. The steps alternate between probability $1-\alpha$ and probability 0 to reach *goal* or *fail* – just as in the gadget $\mathsf{G}_{\bar{\alpha}}$. In this way, it makes no difference for the expected time before reaching *goal* or *fail* when a scheduler stops choosing γ and δ. We can, in fact, compute the expected time T to reach *goal* or *fail* from s_{init} under any scheduler quite easily: Reaching t or s takes 3 steps in expectation. Afterwards, the number of steps taken is $1 + 2\ell$ with probability $\alpha^\ell \cdot (1-\alpha)$. In expectation, this yields

$$\sum_{\ell=0}^{\infty}(1+2\ell)\alpha^\ell \cdot (1-\alpha) = \left(\sum_{\ell=0}^{\infty} 2(\ell+1)\alpha^\ell \cdot (1-\alpha)\right) - 1 = \frac{2}{1-\alpha} - 1.$$

additional steps. So,

$$T = 3 + \frac{2}{1-\alpha} - 1 = 2 + \frac{2}{1-\alpha}.$$

The optimal scheduler $\mathfrak{S}$ for the partial expectation in $\mathcal{M}'$ is the same as in the MDP $\mathcal{M}$ above. Also, the value ϑ of this scheduler can be computed as in Lemma 4.13. So, $\mathbb{PE}^{\max}_{\mathcal{M}',s_{init}} > \vartheta$ if and only if the given linear recurrence sequence is eventually negative.

Note that all weights in $\mathcal{M}'$ are $\geq -k$. We define two new weight functions to obtain an MDP $\mathcal{N}$ from $\mathcal{M}'$: We let $wgt_{goal}(s,\alpha) = wgt(s,\alpha)+k$ and $wgt_{fail}(s,\alpha) = +k$ for all $(s,\alpha) \in S \times Act$. Both weight functions take only non-negative integer values.

Any scheduler $\mathfrak{S}$ for $\mathcal{M}'$ can be viewed as a scheduler for $\mathcal{N}$, and vice versa, as the two MDPs only differ in the weight functions. Further, we observe that for each maximal path ζ ending in *goal* or *fail* in $\mathcal{M}'$ and at the same time in $\mathcal{N}$, we have $X(\zeta) = \oplus goal(\zeta) + k \cdot length(\zeta)$. (Recall that $\oplus goal(\zeta)$ equals $wgt(\zeta)$ if ζ reaches *goal* and 0 if ζ reaches *fail*.) As the expected time before *goal* or *fail* is reached is constant, namely T under any scheduler, it follows that for all schedulers $\mathfrak{T}$ we have

$$\mathbb{E}^{\mathfrak{T}}_{\mathcal{N},s_{init}}(X) = \mathbb{PE}^{\mathfrak{T}}_{\mathcal{M}',s_{init}} + k \cdot T.$$

Therefore, $\mathbb{E}^{\max}_{\mathcal{N},s_{init}}(X) > \vartheta + k \cdot T$ if and only if the given linear recurrence sequence eventually becomes negative. ■

4.3 Conditional value-at-risk for accumulated weights

Lastly, we aim to prove the Positivity-hardness of the threshold problem for the CVaR in this section. To this end, we provide a further direct reduction from the Positivity-problem to the threshold problem for the expected value of an auxiliary random variable closely related to the CVaR using our MDP-gadgets.

Conditional Value-at-Risk. Given an MDP $\mathcal{M} = (S, Act, P, s_{init}, wgt, Goal)$ 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 $VaR^{\mathfrak{S}}_p(X) = \sup\{r \in \mathbb{R} | \Pr^{\mathfrak{S}}_{\mathcal{M}}(X \leq r) \leq p\}$. So, the value-at-risk is the point at which the cumulative distribution function of X reaches or exceeds p. The conditional value-at-risk is now the expectation of X under the condition that the outcome belongs to the p worst outcomes – in this case, the p lowest outcomes. Denote $VaR^{\mathfrak{S}}_p(X)$ by v. Following the treatment of random variables that are not continuous in general in [43], we define the conditional value-at-risk as follows:

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

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 $p - \Pr^{\mathfrak{S}}_{\mathcal{M}}(X < v)$ of the outcomes which are exactly v into account as well. To provide worst-case guarantees or to find risk-averse policies, we are

interested in the maximal and minimal conditional value-at-risk

$$CVaR_p^{\max}(X) = \sup_{\mathfrak{S}} CVaR_p^{\mathfrak{S}}(X) \text{ and } CVaR_p^{\min}(X) = \inf_{\mathfrak{S}} CVaR_p^{\mathfrak{S}}(X).$$

In our formulation here, low outcomes are considered to be bad. Completely analogously, one can define the conditional value-at-risk for the highest p outcomes.

The main result of the section is the following:

THEOREM 4.17. *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*

$$CVaR_p^{\max}(\diamondminus goal) > \vartheta.$$

We will use an auxiliary optimization problem to prove this result. We begin with 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 $VaR_{1/2}^{\mathfrak{S}}(\diamondminus goal) = 0$ 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 $\diamondminus goal$ on paths ζ by

$$\diamondminus goal(\zeta) = \min(\diamondminus goal(\zeta), 0).$$

Now, the conditional value-at-risk for the probability value $1/2$ under a scheduler $\mathfrak{S}$ in $\mathcal{N}$ is given by $CVaR_{1/2}^{\mathfrak{S}}(\diamondminus goal) = 2 \cdot \mathbb{E}_{\mathcal{N},s_{init}}^{\mathfrak{S}}(\diamondminus goal) = \frac{2}{3} \cdot \mathbb{E}_{\mathcal{M},s_{init}}^{\mathfrak{S}}(\diamondminus goal)$. So, the result follows from the following lemma:

LEMMA 4.18. *The Positivity problem is polynomial-time reducible to the following problem: Given an MDP $\mathcal{M}$ and a rational ϑ, decide whether $\mathbb{E}_{\mathcal{M},s_{init}}^{\max}(\diamondminus goal) > \vartheta$.*

PROOF. The first important observation is that the optimal expectation $e(q,w)$ of $\diamondminus goal$ for different starting states q and starting weights w satisfies equation $(*)$ from Section 3.2, i.e., $e(q,w) = \sum_{r \in S} P(q,\alpha,r) \cdot e(r, w{+}wgt(q,\alpha))$ if an optimal scheduler chooses actions α in state $q \neq goal$ when the accumulated weight is w. The value $e(goal,w)$ is w if $w \leq 0$ and 0 otherwise. This allows us to reuse the gadget $\mathsf{G}_{\bar{\alpha}}$ to encode a linear recurrence relation.

We again adjust the gadget encoding the initial values of a linear recurrence sequence. 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 using Assumption 3.1, 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 new gadget that encodes the initial values of a linear recurrence sequence is depicted in Figure 13. In states t and s, there is a choice between actions γ_j and δ_j, respectively, for

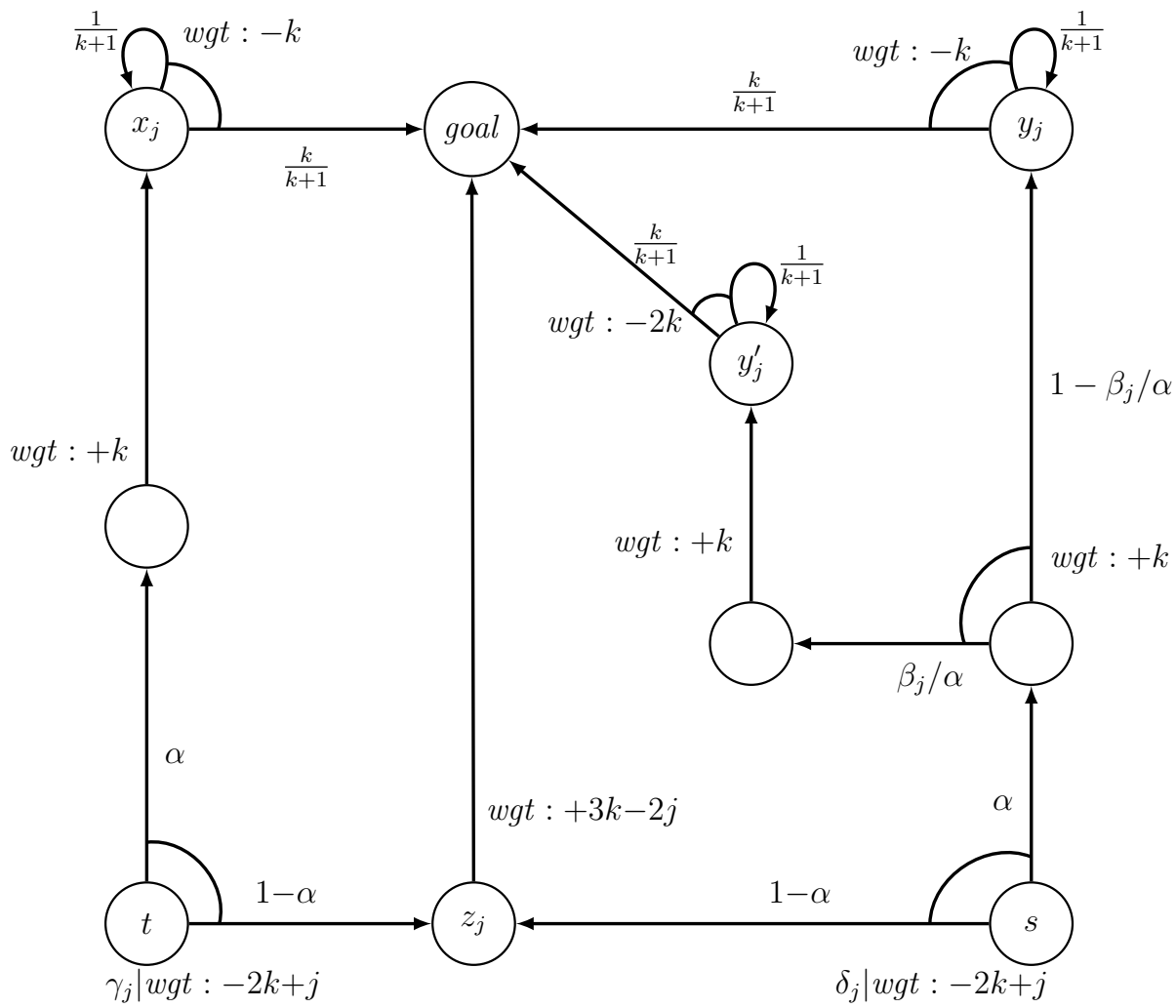

Figure 13. The gadget encoding initial values for the reduction to the threshold problem for the conditional value-at-risk. The gadget contains the depicted states and actions for each $0 \leq j \leq k-1$. The probability α is $\sum_{1 \leq i \leq k} |\alpha_i|$.

$0 \leq j \leq k-1$. After gluing together this gadget with the gadget $\mathsf{G}_{\bar{\alpha}}$ at states t, s, and *goal*, we prove that the interplay between the gadgets is correct: Let $0 \leq j \leq k-1$. Starting with accumulated weight $-k+j$ in state t, the action γ_j maximizes the partial expectation among the actions $\gamma_0, \ldots, \gamma_{k-1}$. Likewise, δ_j is optimal when starting in s with weight $-k+j$. If the accumulated weight is non-negative in state s or t, then γ or δ are optimal. The idea is that for positive starting weights, the tail loss of γ_i and δ_i is relatively high while for weights just below 0, the chance to reach *goal* with positive weight again outweighs this tail loss.

First, we estimate the expectation of $\diamondsuit goal$ when choosing δ_i and δ while the accumulated weight is $-k+j$ in s. If $i > j$, then δ_i and δ lead to *goal* directly with probability $1-\alpha$ and weight ≤ -1. So, the expectation is less than $-(1-\alpha) \leq -1+\frac{1}{5(k+1)}$.

If $i \leq j$, then with probability $1-\alpha$ *goal* is reached with positive weight, hence $\diamondsuit goal$ is 0 on these paths.

With probability β_i, goal is reached via y'_j. In this case all runs reach *goal* with negative weight. On the way to y'_j weight $2k$ is added, but afterwards subtracted again at least once. In expectation weight $2k$ is subtracted $\frac{k+1}{k}$ many times. Furthermore, $-2k+i$ is added to the

starting weight of $-k+j$. So, these paths contribute $\beta_i \cdot (2k - 2k\frac{k+1}{k} - 3k+j+i) = (-3k+j+i-2) \cdot \beta_i$ to the expectation of $\diamondminus goal$.

The remaining paths reach *goal* via y_j and all reach *goal* with negative weight as well. The probability to reach y_j is $\alpha - \beta_j$. On the way to y_j, the initial weight of $-k+i$ is changed to $-2k+j+i$. Afterwards, weight $-k$ is accumulated $\frac{k+1}{k}$-many times in expectation. So, these remaining paths contribute $(-3k+j+i-1) \cdot (\alpha - \beta_i)$. So, all in all the expectation of $\diamondminus goal$ in this situation is $\alpha \cdot (-3k+j+i-1) - \beta_i$. Now, as $\alpha \leq \frac{1}{5(k+1)}$ and $\beta_i \leq \frac{\alpha}{3}$ for all i, we see that $\alpha \cdot (-3k+j+i-1) - \beta_i \geq -(3k+2)\alpha \geq -1 + \frac{1}{5(k+1)}$. The optimum with $i \leq j$ is obtained for $i = j$ as $\beta_i \leq \alpha/3$ for all i. Hence indeed δ_j is the optimal action. For γ_j the same proof with $\beta_i = 0$ for all i leads to the same result.

Now assume that the accumulated weight in t or s is $\ell \geq 0$. Then, all actions lead to *goal* with a positive weight with probability $1 - \alpha$. In this case $\diamondminus goal$ is 0. However, a scheduler $\mathfrak{S}$ which always chooses γ and δ is better than a scheduler choosing γ_j or δ_j for any $j \leq k-1$. Under scheduler $\mathfrak{S}$ starting from s or t a run returns to $\{s, t\}$ with probability α while accumulating weight $\geq -k$ and the process is repeated. After choosing γ_j or δ_j the run moves to x_j, y_j or y'_j while accumulating a negative weight. From then on, in each step it will stay in that state with probability greater than α and accumulate weight $\leq -k$. Hence, the expectation of $\diamondminus goal$ is lower under γ_j or δ_j than under $\mathfrak{S}$. Therefore indeed γ and δ are the best actions for non-negative accumulated weight in states s and t.

Let now $e(t, w)$ and $e(s, w)$ denote the optimal expectations of $\diamondminus goal$ when starting in t or s with weight w. Further, let $d(w) = e(t, w) - e(s, w)$. From the argument above, we also learn that the difference $d(-k+j)$ is equal to β_j, for $0 \leq j \leq k-1$. Put together with the linear recurrence encoded in $\mathsf{G}_{\bar{\alpha}}$ this shows that $d(-k+w) = u_w$ for all w where $(u_n)_{n \in \mathbb{N}}$ is the linear recurrence sequence specified by the α_i, β_j, $1 \leq i \leq k$, and $0 \leq j \leq k-1$.

Finally, we add the same initial component as in the previous section to obtain an MDP $\mathcal{M}$. Let $\mathfrak{S}$ be the scheduler always choosing τ in state c and afterwards following the optimal actions as described above is optimal if and only if the linear recurrence sequence stays non-negative. The remaining argument goes completely analogously to the proof of Theorem 4.1. Grouping together the optimal values in vectors v_n with $2k$ entries as done there, we can use the same Markov chain as in that proof to obtain a matrix A such that $v_{n+1} = Av_n$. This allows us to compute the rational value $\vartheta = \mathbb{E}^{\mathfrak{S}}_{\mathcal{M},s_{init}}(\diamondminus goal)$ via a matrix series in polynomial time and $\mathbb{E}^{\max}_{\mathcal{M},s_{init}}(\diamondminus goal) > \vartheta$ if and only if the given linear recurrence sequence is eventually negative. ■

By the discussion above, this lemma directly implies Theorem 4.17. With adaptions similar to the previous section, it is possible to obtain the analogous result for the minimal expectation of $\diamondminus goal$. This implies that also the threshold problem whether the minimal conditional value-at-risk is less than a threshold ϑ, $CVaR^{\min}_p(\diamondminus goal) < \vartheta$, is Positivity-hard.

5. Conclusion

The Positivity-hardness results established in this paper show that a series of problems on finite-state MDPs that have been studied and left open in the literature exhibit an inherent mathematical difficulty. A decidability result for any of these problems would imply a major break-through in analytic number theory. At the heart of our Positivity-hardness proofs lies the construction of modular MDPs consisting of three gadgets. This construction provides a versatile proof strategy to establish Positivity-hardness results: It allowed us to provide three direct reductions from the Positivity problem by constructing structurally identical MDPs that only differ in the gadget encoding the initial values. The further chains of reductions depicted in Figure 1 established Positivity-hardness for a landscape of different problems on one-counter MDPs and integer-weighted MDPs.

The proof technique might be applicable to further threshold problems associated to optimization problems on MDPs. A main requirement for the direct applicability of the technique is that the optimal values $V(s,w)$ in terms of the current state s and the weight w accumulated so far, or a similar quantity that can be increased and decreased, satisfy an optimality equation of the form

$$V(s,w) = \max_{\alpha \in Act(s)} \sum_{t \in S} P(s,\alpha,t) \cdot V(t, w + wgt(s,\alpha)).$$

In addition, the optimum must not be achievable with memoryless schedulers, but the optimal decisions have to depend on the accumulated weight to make it possible to encode initial values of a linear recurrence sequence. This combination of conditions is quite common as we have seen.

Furthermore, our Positivity-hardness results can be used to establish Positivity-hardness of further decision problems on MDPs, which are on first sight of a rather different nature: In [58, 57], it is shown how our proof of the Positivity-hardness of the two-sided partial SSPP can be modified to prove the Positivity-hardness of two problems concerning the long-run satisfaction of path properties, namely the threshold problem for long-run probabilities and the model-checking problem of frequency-LTL. Both of these problems address the degree to which a property is satisfied by the sequence of suffixes of a run in order to analyze the long-run behavior of systems. The long-run probability of a property φ in an MDP $\mathcal{M}$ under a scheduler $\mathfrak{S}$ is the expected long-run average of the probability that a suffix generated by $\mathfrak{S}$ in $\mathcal{M}$ satisfies φ. Similarly, frequency-LTL extends LTL by an operator that requires a certain percentage of the suffixes of a run to satisfy a property. Long-run probabilities and frequency-LTL in MDPs have been investigated in [6] and [34, 35], respectively, where decidable special cases of the mentioned decision problems have been identified. In general, however, the decidability status of these problems is open. The reductions in [58, 57] show how the two-sided partial SSPP can be encoded into the long-run probability as well as the long-run frequency of the satisfaction of

a simple regular co-safety property, i.e., the negation of a safety property, yielding Positivity-hardness for the threshold problem for long-run probabilities and the model-checking problem of frequency-LTL in MDPs.

It is worth mentioning that in the special case of Markov chains, several of the problems investigated here are decidable: In Markov chains, partial and conditional expectations can be computed in polynomial time [59]. Furthermore, one-counter Markov chains constitute a special case of recursive Markov chains, for which the threshold problem for the termination probability can be decided in polynomial space [31]. Remarkably however, the threshold problem for the probability that the accumulated cost satisfies a Boolean combination of inequality constraints in finite-state Markov chains is open [38].

Finally, the Positivity-hardness results leave the possibility open that some or all of the problems we studied are in fact harder than the Positivity problem. In particular, it could be the case that the problems are undecidable and that a proof of the undecidability would yield no implications for the Positivity problem. For this reason, investigating whether some or all of the threshold problems are reducible to the Positivity problem constitutes a very interesting – and challenging – direction for future work. Such an inter-reducibility result would show that studying any of the discussed optimization problems on MDPs could be a worthwhile direction of research to settle the decidability status of the Positivity-problem. Some hope for an inter-reducibility result can be drawn from the fact that the optimal values are approximable for several of the problems – for termination probabilities and expected termination times of one-counter MDPs, this was shown in [18, 21] and for partial and conditional expectations in [59]. This indicates that there is at least a major difference to undecidable problems in a similar context such as the emptiness problem for probabilistic finite automata where the optimal value cannot be approximated [56, 28].

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 (6)

[2] **Manindra Agrawal, S. Akshay, Blaise Genest, and P. S. Thiagarajan**. Approximate verification of the symbolic dynamics of Markov chains. *Journal of the ACM*, 62(1):2:1–2:34, 2015. DOI (8)

[3] **Mohamadreza Ahmadi, Anushri Dixit, Joel W. Burdick, and Aaron D. Ames**. Risk-averse stochastic shortest path planning. *60th IEEE Conference on Decision and Control (CDC)*, pages 5199–5204. IEEE, 2021. DOI (6)

[4] **S. Akshay, Timos Antonopoulos, Joël Ouaknine, and James Worrell**. Reachability problems for Markov chains. *Information Processing Letters*, 115(2):155–158, 2015. DOI (8)

[5] **Christel Baier, Nathalie Bertrand, Clemens Dubslaff, Daniel Gburek, and Ocan Sankur**. Stochastic shortest paths and weight-bounded properties in Markov decision processes. *33rd Annual ACM/IEEE Symposium on Logic in Computer Science (LICS)*, pages 86–94. ACM, 2018. DOI (2, 5, 12, 28)

[6] **Christel Baier, Nathalie Bertrand, Jakob Piribauer, and Ocan Sankur**. Long-run satisfaction of path properties. *34th Annual ACM/IEEE Symposium on Logic in Computer Science (LICS)*, pages 1–14. IEEE, 2019. DOI (43)

[7] **Christel Baier, Marcus Daum, Clemens Dubslaff, Joachim Klein, and Sascha Klüppelholz**. Energy-utility quantiles. *6th International Symposium - NASA Formal Methods (NFM)*, volume 8430 of *Lecture Notes in Computer Science*, pages 285–299. Springer, 2014. DOI (5)

[8] **Christel Baier, Clemens Dubslaff, Joachim Klein, Sascha Klüppelholz, and Sascha Wunderlich.** Probabilistic model checking for energy-utility analysis. *Horizons of the Mind. A Tribute to Prakash Panangaden - Essays Dedicated to Prakash Panangaden on the Occasion of His 60th Birthday*, volume 8464 of *Lecture Notes in Computer Science*, pages 96–123. Springer, 2014. DOI (5)

[9] **Christel Baier, Marcus Größer, and Nathalie Bertrand.** Probabilistic ω-automata. *Journal of the ACM*, 59(1):1:1–1:52, 2012. DOI (3)

[10] **Christel Baier and Joost-Pieter Katoen.** Principles of Model Checking. MIT Press, 2008. (2)

[11] **Christel Baier, Joachim Klein, Sascha Klüppelholz, and Sascha Wunderlich.** Maximizing the conditional expected reward for reaching the goal. *23rd International Conference on Tools and Algorithms for the Construction and Analysis of Systems (TACAS)*, volume 10206 of *Lecture Notes in Computer Science*, pages 269–285, 2017. DOI (5, 29)

[12] **Christel Baier, Joachim Klein, Sascha Klüppelholz, and Sascha Wunderlich.** Weight monitoring with linear temporal logic: complexity and decidability. *Joint meeting of the 23rd EACSL Annual Conference on Computer Science Logic (CSL) and the 29th Annual ACM/IEEE Symposium on Logic in Computer Science (LICS)*, 11:1–11:10. ACM, 2014. DOI (3)

[13] **Gilles Barthe, Thomas Espitau, Luis María Ferrer Fioriti, and Justin Hsu.** Synthesizing probabilistic invariants via Doob's decomposition. *28th International Conference on Computer Aided Verification (CAV), Part I*, volume 9779 of *Lecture Notes in Computer Science*, pages 43–61. Springer, 2016. DOI (5)

[14] **Danièle Beauquier, Alexander Moshe Rabinovich, and Anatol Slissenko.** A logic of probability with decidable model checking. *Journal of Logic and Computation*, 16(4):461–487, 2006. DOI (8)

[15] **Amir M. Ben-Amram, Samir Genaim, and Abu Naser Masud.** On the termination of integer loops. *ACM Transactions on Programming Languages Systems*, 34(4):16:1–16:24, 2012. DOI (8)

[16] **Dimitri P. Bertsekas and John N. Tsitsiklis.** An analysis of stochastic shortest path problems. *Mathematics of Operations Research*, 16(3):580–595, 1991. DOI (2, 12, 28)

[17] **Mark Braverman.** Termination of integer linear programs. *18th International Conference on Computer Aided Verification (CAV)*, volume 4144 of *Lecture Notes in Computer Science*, pages 372–385. Springer, 2006. DOI (8)

[18] **Tomás Brázdil, Václav Brozek, Kousha Etessami, and Antonín Kucera.** Approximating the termination value of one-counter MDPs and stochastic games. *Information and Computation*, 222:121–138, 2013. DOI (4, 44)

[19] **Tomás Brázdil, Václav Brozek, Kousha Etessami, Antonín Kucera, and Dominik Wojtczak.** One-counter Markov decision processes. *21st Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 863–874. SIAM, 2010. DOI (4)

[20] **Tomás Brázdil, Antonín Kucera, and Petr Novotný.** Optimizing the expected mean payoff in energy markov decision processes. *14th International Symposium on Automated Technology for Verification and Analysis (ATVA)*, volume 9938 of *Lecture Notes in Computer Science*, pages 32–49, 2016. DOI (4)

[21] **Tomás Brázdil, Antonín Kucera, Petr Novotný, and Dominik Wojtczak.** Minimizing expected termination time in one-counter Markov decision processes. *39th International Colloquium on Automata, Languages, and Programming (ICALP)*, volume 7392 of *Lecture Notes in Computer Science*, pages 141–152. Springer, 2012. DOI (4, 44)

[22] **John F. Canny.** Some algebraic and geometric computations in PSPACE. *20th Annual ACM Symposium on Theory of Computing (STOC)*, pages 460–467. ACM, 1988. DOI (3)

[23] **Krishnendu Chatterjee and Laurent Doyen.** Energy and mean-payoff parity Markov decision processes. *36th International Symposium on Mathematical Foundations of Computer Science (MFCS)*, volume 6907 of *Lecture Notes in Computer Science*, pages 206–218. Springer, 2011. DOI (4)

[24] **Krishnendu Chatterjee, Hongfei Fu, and Amir Kafshdar Goharshady.** Termination analysis of probabilistic programs through Positivstellensatz's. *28th International Conference on Computer Aided Verification (CAV)*, volume 9779 of *Lecture Notes in Computer Science*, pages 3–22. Springer, 2016. DOI (5)

[25] **Taolue Chen, Vojtech Forejt, Marta Z. Kwiatkowska, David Parker, and Aistis Simaitis.** Automatic verification of competitive stochastic systems. *Formal Methods in System Design*, 43(1):61–92, 2013. DOI (5, 32)

[26] **Taolue Chen, Vojtech Forejt, Marta Z. Kwiatkowska, David Parker, and Aistis Simaitis.** PRISM-games: A model checker for stochastic multi-player games. *19th International Conference on Tools and Algorithms for the Construction and Analysis of Systems (TACAS)*, volume 7795 of *Lecture Notes in Computer Science*, pages 185–191. Springer, 2013. DOI (5)

[27] **Ventsislav Chonev, Joël Ouaknine, and James Worrell.** On the Skolem problem for continuous linear dynamical systems. *43rd International Colloquium on Automata, Languages, and Programming (ICALP)*, volume 55 of *LIPIcs*, 100:1–100:13. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2016. DOI (8)

[28] **Anne Condon and Richard J. Lipton.** On the complexity of space bounded interactive proofs (extended abstract). *30th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 462–467. IEEE Computer Society, 1989. DOI (44)

[29] **Costas Courcoubetis and Mihalis Yannakakis**. The complexity of probabilistic verification. *Journal of the ACM*, 42(4):857–907, 1995. DOI (2)

[30] **Luca de Alfaro**. Computing minimum and maximum reachability times in probabilistic systems. *10th International Conference on Concurrency Theory (CONCUR)*, volume 1664 of *Lecture Notes in Computer Science*, pages 66–81. Springer, 1999. DOI (2, 12, 28)

[31] **Kousha Etessami and Mihalis Yannakakis**. Recursive Markov chains, stochastic grammars, and monotone systems of nonlinear equations. *Journal of the ACM*, 56(1):1:1–1:66, 2009. DOI (8, 44)

[32] **Kousha Etessami and Mihalis Yannakakis**. Recursive Markov decision processes and recursive stochastic games. *Journal of the ACM*, 62(2):11:1–11:69, 2015. DOI (3–5)

[33] **Graham Everest, Alfred J. van der Poorten, Igor E. Shparlinski, and Thomas Ward**. Recurrence Sequences, volume 104 of *Mathematical surveys and monographs*. American Mathematical Society, 2003. URL (3)

[34] **Vojtech Forejt and Jan Krcál**. On frequency LTL in probabilistic systems. *26th International Conference on Concurrency Theory (CONCUR)*, volume 42 of *LIPIcs*, pages 184–197. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2015. DOI (43)

[35] **Vojtech Forejt, Jan Krcál, and Jan Kretínský**. Controller synthesis for MDPs and frequency $\mathrm{LTL}_{\setminus \mathrm{GU}}$. *20th International Conference on Logic for Programming, Artificial Intelligence, and Reasoning (LPAR)*, volume 9450 of *Lecture Notes in Computer Science*, pages 162–177. Springer, 2015. DOI (43)

[36] **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. DOI (5, 9)

[37] **Christoph Haase and Stefan Kiefer**. The odds of staying on budget. *42nd International Colloquium on Automata, Languages, and Programming (ICALP)*, volume 9135 of *Lecture Notes in Computer Science*, pages 234–246. Springer, 2015. DOI (5)

[38] **Christoph Haase, Stefan Kiefer, and Markus Lohrey**. Computing quantiles in markov chains with multi-dimensional costs. *32nd Annual ACM/IEEE Symposium on Logic in Computer Science (LICS)*, pages 1–12. IEEE Computer Society, 2017. DOI (5, 44)

[39] **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. (3)

[40] **Arie Hordijk and Lodewijk Kallenberg**. Linear programming and Markov decision chains. *Management Science*, 25(4):352–362, 1979. DOI (2)

[41] **Benjamin Lucien Kaminski and Joost-Pieter Katoen**. On the hardness of almost-sure termination. *40th International Symposium on Mathematical Foundations of Computer Science (MFCS)*, volume 9234 of *Lecture Notes in Computer Science*, pages 307–318. Springer, 2015. DOI (8)

[42] **Joost-Pieter Katoen, Friedrich Gretz, Nils Jansen, Benjamin Lucien Kaminski, and Federico Olmedo**. Understanding probabilistic programs. *Correct System Design - Symposium in Honor of Ernst-Rüdiger Olderog on the Occasion of His 60th Birthday*, volume 9360 of *Lecture Notes in Computer Science*, pages 15–32. Springer, 2015. DOI (5)

[43] **Jan Kretínský and Tobias Meggendorfer**. Conditional value-at-risk for reachability and mean payoff in Markov decision processes. *33rd Annual ACM/IEEE Symposium on Logic in Computer Science (LICS)*, pages 609–618. ACM, 2018. DOI (6, 39)

[44] **Omid Madani, Steve Hanks, and Anne Condon**. On the undecidability of probabilistic planning and related stochastic optimization problems, volume 147 of number 1-2, pages 5–34. Elsevier, 2003. DOI (3)

[45] **Rupak Majumdar, Mahmoud Salamati, and Sadegh Soudjani**. On decidability of time-bounded reachability in CTMDPs. *47th International Colloquium on Automata, Languages, and Programming (ICALP)*, volume 168 of *LIPIcs*, 133:1–133:19. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2020. DOI (8)

[46] **Richard Mayr, Sven Schewe, Patrick Totzke, and Dominik Wojtczak**. MDPs with energy-parity objectives. *32nd Annual ACM/IEEE Symposium on Logic in Computer Science (LICS)*, pages 1–12. IEEE Computer Society, 2017. DOI (4)

[47] **Annabelle McIver and Carroll Morgan**. Abstraction, Refinement and Proof for Probabilistic Systems. Monographs in Computer Science. Springer, 2005. DOI (8)

[48] **Tobias Meggendorfer**. Risk-aware stochastic shortest path. *36th AAAI Conference on Artificial Intelligence (AAAI)*, pages 9858–9867. AAAI Press, 2022. DOI (6)

[49] **Piergiorgio Odifreddi**. Classical recursion theory: The theory of functions and sets of natural numbers. Elsevier, 1992. (8)

[50] **Federico Olmedo, Friedrich Gretz, Nils Jansen, Benjamin Lucien Kaminski, Joost-Pieter Katoen, and Annabelle McIver**. Conditioning in probabilistic programming. *ACM Transactions of Programming Languages and Systems*, 40(1):4:1–4:50, 2018. DOI (5)

[51] **Joël Ouaknine and James Worrell**. Decision problems for linear recurrence sequences. *6th International Workshop on Reachability Problems (RP)*, volume 7550 of *Lecture Notes in Computer Science*, pages 21–28. Springer, 2012. DOI (3)

[52] **Joël Ouaknine and James Worrell**. On linear recurrence sequences and loop termination. *ACM SIGLOG News*, 2(2):4–13, 2015. DOI (3, 8)

[53] **Joël Ouaknine and James Worrell**. On the positivity problem for simple linear recurrence sequences, *41st International Colloquium on Automata, Languages, and Programming (ICALP)*, volume 8573 of *Lecture Notes in Computer Science*, pages 318–329. Springer, 2014. DOI (3)

[54] **Joël Ouaknine and James Worrell**. Positivity problems for low-order linear recurrence sequences. *25TH Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 366–379. SIAM, 2014. DOI (3)

[55] **Joël Ouaknine and James Worrell**. Ultimate positivity is decidable for simple linear recurrence sequences. *41st International Colloquium on Automata, Languages, and Programming (ICALP)*, volume 8573 of *Lecture Notes in Computer Science*, pages 330–341. Springer, 2014. DOI (3)

[56] **Azaria Paz**. Introduction to Probabilistic Automata. Academic Press, 1971. DOI (44)

[57] **Jakob Piribauer**. On non-classical stochastic shortest path problems. PhD thesis, Technische Universität Dresden, Germany, 2021. URL (1, 43)

[58] **Jakob Piribauer and Christel Baier**. On Skolem-hardness and saturation points in Markov decision processes. *47th International Colloquium on Automata, Languages, and Programming (ICALP)*, volume 168 of *LIPIcs*, 138:1–138:17. Schloss Dagstuhl–Leibniz-Zentrum für Informatik, 2020. DOI (1, 6, 43)

[59] **Jakob Piribauer and Christel Baier**. Partial and conditional expectations in markov decision processes with integer weights. *22nd International Conference on Foundations of Software Science and Computation Structures (FoSSaCS)*, volume 11425 of *Lecture Notes in Computer Science*, pages 436–452. Springer, 2019. DOI (5, 8, 37, 44)

[60] **Martin L. Puterman**. Markov Decision Processes: Discrete Stochastic Dynamic Programming. John Wiley & Sons, Inc., 1994. DOI (2, 9)

[61] **Mickael Randour, Jean-François Raskin, and Ocan Sankur**. Percentile queries in multi-dimensional Markov decision processes. *Formal Methods in Systems Design*, 50(2-3):207–248, 2017. DOI (3, 5)

[62] **Tarlok N. Shorey, Robert Tijdeman, and Max Mignotte**. The distance between terms of an algebraic recurrence sequence. *Journal für die reine und angewandte Mathematik*, 1984(349):63–76, 1984. DOI (3)

[63] **Ashish Tiwari**. Termination of linear programs. *16th International Conference on Computer Aided Verification (CAV)*, volume 3114 of *Lecture Notes in Computer Science*, pages 70–82. Springer, 2004. DOI (8)

[64] **Michael Ummels and Christel Baier**. Computing quantiles in Markov reward models. *16th International Conference on Foundations of Software Science and Computation Structures (FoSSaCS)*, volume 7794 of *Lecture Notes in Computer Science*, pages 353–368. Springer, 2013. DOI (5)

[65] **Stanislav Uryasev**. Conditional value-at-risk: optimization algorithms and applications. *IEEE/IAFE/INFORMS Conference on Computational Intelligence for Financial Engineering (CIFEr)*, pages 49–57. IEEE, 2000. DOI (6)

[66] **N. K. Vereshchagin**. The problem of appearance of a zero in a linear recurrence sequence. *Mathematical Notes of the Academy of Sciences of the USSR*, 38:609–615, 1985. DOI (3)

This work is licensed under the Creative Commons Attribution 4.0 International License.
http://creativecommons.org/licenses/by/4.0/
© Jakob Piribauer, Christel Baier.