Multi-Cost-Bounded Reachability Analysis of POMDPs

Alexander Bork[1] **Joost-Pieter Katoen**[1] **Tim Quatmann**[1] **Svenja Stein**[1]

[1]RWTH Aachen University, Aachen, Germany

Abstract

We consider multi-dimensional cost-bounded reachability probability objectives for partially observable Markov decision processes (POMDPs). The goal is to compute the maximal probability to reach a set of target states while simultaneously satisfying specified bounds on incurred costs. Such objectives generalise well-studied POMDP objectives by allowing multiple upper and lower bounds on different cost or reward measures, e.g. to naturally model scenarios where an agent acts under limited resources. We present a reduction of the multi-cost-bounded problem to unbounded reachability probabilities on an unfolding of the original POMDP. We employ a refined approach in case the agent is cost-aware—i.e., collected costs are fully observed—and also consider a setting where only partial information about the collected costs is known. Our approaches elegantly lift existing results from the fully observable MDP case to POMDPs. An empirical evaluation shows the potential of analysing POMDPs under multi-cost-bounded reachability objectives in practical settings.

1 INTRODUCTION

Partially observable Markov decision processes (POMDPs) are a powerful modelling formalism for sequential decision making in uncertain domains where non-determinism is present. They extend *Markov decision processes (MDPs)* [Puterman, 1994] for agents that in addition to uncertain transitions also only have incomplete information about the state of the environment [Smallwood and Sondik, 1973, Russell and Norvig, 2020]. POMDPs have applications in a plethora of domains, including robotics [Spaan and Vlassis, 2004], ecology [Chades et al., 2012], and cyber security [Miehling et al., 2018].

The classical planning problem in POMDPs is to compute a *policy*—a plan for resolving non-determinism in the system—that optimises a given objective. This problem is notoriously difficult for many kinds of objectives. Various approaches consider *finite horizons*, where the objective has to be satisfied in a finite amount of steps [Smallwood and Sondik, 1973], or *discounting*, where events in later stages become less relevant [Smith and Simmons, 2004, Kurniawati et al., 2008, Shani et al., 2013].

In recent years, work focusing on objectives *without discounting* over an *infinite* time horizon has emerged [Norman et al., 2017, Horák et al., 2018, Bork et al., 2022, Andriushchenko et al., 2022, 2023, Ho et al., 2024]. One commonly considered objective is to compute the maximal probability to reach a set of target states. In the field of probabilistic model checking, this *maximal reachability probability* objective is the basis for the analysis of models with respect to more involved logical specifications such as linear time temporal logic [Baier and Katoen, 2008].

Expected Costs vs. Cost-Bounded Reachability In many practical scenarios, the objective to reach a target is subject to hard constraints on resources. For example, an autonomous vehicle navigating towards a goal position has to consider its fuel level and emission levels of pollutants. A policy minimising the *expected* fuel and emission costs does not take the resource limits into account which may lead to the vehicle running out of fuel with unnecessarily high probability. In such scenarios, a policy that maximises the *probability* of a successful run—where the vehicle reaches the goal without running out of fuel and within pollution limits—is preferable. Thus, objectives that constrain the *expected* costs fail to capture scenarios where satisfying hard constraints on actually incurred costs is key for a run to be successful.

(Multi-)cost-bounded reachability probability objectives characterise such scenarios where a strict adherence to certain resource constraints is crucial. Numerical costs (or rewards) are assigned to transitions in the POMDP. The goal of the agent is to maximise its probability to reach the target

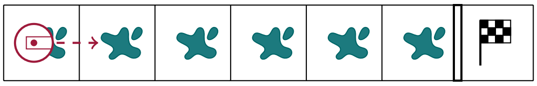


Figure 1: Hallway Cleaning Task

states while satisfying all bounds on the *actual* accumulated costs. In the most general setting, both upper and lower bounds on multiple cost measures can be considered simultaneously. Cost-bounded reachability objectives have mainly been studied for (fully observable) MDPs [Hahn and Hartmanns, 2016, Klein et al., 2018, Hartmanns et al., 2020].

Motivating Example Consider the scenario depicted in Figure 1. A robot is tasked with cleaning a hallway consisting of 6 tiles, all initially dirty. The robot starts in the left-most tile with an energy level of 60 units. In every step, the robot can either attempt to clean the current tile or move to the next tile. A cleaning attempt can fail with a probability of 0.2, leaving the tile dirty. The robot, however, is not able to observe if it has successfully cleaned a tile. Moving to the next tile always consumes one unit of energy. A cleaning attempt consumes either 2 or 4 units of energy, each with probability 0.5. The robot is successful if it cleans all 6 tiles and reaches the target position by moving in the right-most tile of the hallway without running out of energy.

Our goal is to find a policy for the robot that schedules in each step the best action based on the available information in order to maximise the probability that the task succeeds. We can model the scenario as a POMDP with two cost measures $\mathbf{C}_{\text{energy}}$ and $\mathbf{C}_{\text{clean}}$ (the latter assigning a cost of 1 when a tile is successfully cleaned and 0 otherwise), and the cost-bounded reachability query "*maximise the probability to reach the target while accumulating at most* 60 *cost units for* $\mathbf{C}_{\text{energy}}$ *and at least* 6 *cost units for* $\mathbf{C}_{\text{clean}}$."

Usual POMDP objectives fail to capture the objective we are interested in. In particular, minimising the expected energy use or maximising the expected number of cleaned tiles fails to consider the respective other constraint and does not accurately reflect the hard requirements on incurred costs. A multi-cost-bounded objective with cost bounds on both measures, however, adequately reflects our objective.

Contributions We formalise the problem of multi-cost-bounded reachability probability objectives on POMDPs and observe undecidability of its decision variant. We then consider three variations of this problem that all differ in the degree of observability of the accumulated cost so far. (As usual, we assume that the observations in the POMDP are visible to the decision-making agent.) For the extreme case in which the accumulated costs are completely invisible, we provide a transformation to an equivalent unbounded reachability problem on an often larger POMDP. The key concept is to encode cost collection in the state space. The resulting *cost-unfolding POMDP* can then be analysed using existing approximation methods for unbounded infinite-horizon reachability problems.

We then consider the other extreme case in which the accumulated cost is fully observable by the agent and can be used in its decision making. We show that the sequential approach of Hartmanns et al. [2020] for multi-cost-bounded reachability in MDPs, i.e., fully observable POMDPs, can be readily lifted to this setting of *cost-aware* POMDPs. Furthermore, we consider the novel setting where the agent cannot observe the exact costs gathered so far, but only certain *cost levels* (e.g. high, mid and low). We show that this setting can often be reduced to the analysis of a cost-aware POMDP.

Our algorithmic solutions are designed such that they leverage strengths of existing and (time- and space-)efficient techniques, and are thus able to directly profit from future advancements in unbounded reachability in POMDPs and cost-bounded reachability in MDPs.

We provide detailed proofs for main theoretical results and additional technical information in the appendix.

1.1 RELATED WORK

Multi-cost-bounded reachability objectives for (fully observable) MDPs are well-studied [Ohtsubo, 2004, Baier et al., 2014, Randour et al., 2017, Hahn and Hartmanns, 2016, Klein et al., 2018, Hartmanns et al., 2020]. We focus on related research dedicated to partially observable models.

Most closely related to our paper is the work on *risk-sensitive POMDPs* [Hou et al., 2016] where a special case of a cost-bounded reachability objective is considered. In particular, the authors present an unfolding of cost-bounds on the level of beliefs, similar to our unfolding on the level of the POMDP. They also consider the case of observable costs. However, our framework is more general by allowing for *multiple* bounds over different cost measures as well as mixtures of upper and lower bounds. In addition, our unfolding allows the use of non-belief-based solution methods for the problem.

Wu et al. [2019] consider an unfolding that encodes cost and step bounds in the belief space to solve a classification problem for *hidden model MDPs*, which are a special class of POMDPs. In that setup, the goal of an agent is to find out in which specific instance of structurally similar MDP environments it is located while not exceeding certain costs.

Another related formalism are constrained POMDPs (CPOMDPs) [Isom et al., 2008, Poupart et al., 2015, Santana et al., 2016]. In a CPOMDP, the objective is to maximise the expected sum of rewards subject to a bound on the *expected* cumulative costs. While classically, the setting considers discounted values with upper bounds on expected costs, there are extensions to an undiscounted setting [Kalagarla et al., 2025]. Furthermore, Undurti and How [2010] consider a setting where violation of bounds for expected and actual costs coincide. In contrast, our work considers mixtures of different bound types on the actual incurred costs and does not make assumptions on the POMDP or the cost bounds.

Chatterjee et al. [2016] consider the setting of minimising expected costs over all strategies in a POMDP that reach a target *almost surely*, i.e., with probability 1. This can be considered a related problem where a strict bound is placed on the reachability probability rather than the incurred costs.

2 PRELIMINARIES

We briefly outline the theoretical background for POMDPs. Further details can be found in Russell and Norvig [2020]. Baier and Katoen [2008, Chapter 10] gives an introduction to MDPs from a formal methods perspective.

Let $X \neq \emptyset$ be a countable set. A *(probability) distribution* over X is a function $\mu : X \to [0,1]$ with $\sum_{x \in X} \mu(x) = 1$. $\mathit{Dist}(X)$ is the set of distributions over X. We write $x \in \mu$ if $\mu(x) > 0$. The *support* of μ is $\mathit{supp}(\mu) := \{x \mid x \in \mu\}$. For $k \in \mathbb{N}$ and *vector* $\mathsf{x} = \langle x_1, \ldots, x_k \rangle \in X^k$, we write $\mathsf{x}[i] = x_i$ for the i-th element ($1 \leq i \leq k$).

MDP A *Markov decision process (MDP)* is a tuple $M = \langle S, \mathit{Act}, \mathbf{P}, s_{\mathit{init}} \rangle$ with a (finite or countably infinite) set S of states, a finite set Act of actions, a transition function $\mathbf{P}\colon S \times \mathit{Act} \to \mathit{Dist}(S)$, and an initial state $s_{\mathit{init}} \in S$. In every state s, an agent making decisions in the MDP chooses an action $a \in \mathit{Act}$ and the state is updated to state s' with probability $\mathbf{P}(s,a)(s')$. If $s' \in \mathbf{P}(s,a)$, we call (s,a,s') a transition and write $s \xrightarrow{a} s'$.

POMDP A *partially observable MDP (POMDP)* is a tuple $\mathcal{M} = \langle M, Z, \mathbf{O} \rangle$, where M is the underlying MDP with $|S| \in \mathbb{N}$, i.e., S is finite, Z is a finite set of observations, and $\mathbf{O}\colon S \times \mathit{Act} \times S \to \mathit{Dist}(Z)$ is an observation function.

In a POMDP, the agent does not have complete access to the current state of the system to base its decisions on. Instead, upon taking a transition $s \xrightarrow{a} s'$ the agent receives an *observation* z with probability $\mathbf{O}(s,a,s')(z)$. A *(finite, initial) path* in an MDP or POMDP is a sequence $\hat{\pi} = s_0 a_1 s_1 \ldots a_n s_n$ with $s_0 = s_{\mathit{init}}$ such that for all $0 < i \leq n$ we have $s_i \in \mathbf{P}(s_{i-1}, a_i)$. We denote by $|\hat{\pi}| := n$ the length, by $\hat{\pi}[i] := s_i$ the i-th state, and by $\mathit{last}(\hat{\pi}) := s_n$ the last state of $\hat{\pi}$. An *observation trace* of a POMDP is a sequence of observations $\tau = z_1 \ldots z_n \in Z^*$. Given a path $\hat{\pi} = s_0 a_1 s_1 \ldots a_n s_n$, the probability of observing trace $\tau = z_1 \ldots z_n$ in $\mathcal{M}$ is $P^{\mathcal{M}}(\tau|\hat{\pi}) = \prod_{i=0}^{n-1} \mathbf{O}(s_i, a_{i+1}, s_{i+1})(z_{i+1})$.

Policies resolve the non-determinism of POMDPs by determining the next action to play after observing an observation trace τ. Formally, a *policy* for $\mathcal{M}$ is a function $\sigma\colon Z^* \to \mathit{Act}$. We denote the set of policies for POMDP $\mathcal{M}$ by $\Sigma^{\mathcal{M}}$.

Belief MDP A *belief* is a tuple $b = \langle z, \mu_b \rangle$ of an observation $z \in Z \uplus \{z_{\mathit{init}}\}$, where $z_{\mathit{init}} \notin Z$ is a dedicated initial observation, and a probability distribution $\mu_b \in \mathit{Dist}(S)$ over POMDP states. [1] The distribution captures the evolution of an agent's information about its current state given histories of actions and observations, while the observation represents the last observation made in such a history.

An agent starts with the initial belief $b_{\mathit{init}} = \langle z_{\mathit{init}}, \mu_{b_{\mathit{init}}} \rangle$, with $\mu_{b_{\mathit{init}}}(s_{\mathit{init}}) = 1$. Beliefs are updated when an action is played and a new observation is received. The probability to observe $z \in Z$ after playing action a in belief $b = \langle \hat{z}, \mu_b \rangle$ is

$$P(z|b,a) = \sum_{s \in \mu_b} \mu_b(s) \cdot \sum_{s' \in S} \mathbf{P}(s,a)(s') \cdot \mathbf{O}(s,a,s')(z).$$

The successor belief of b after playing a and observing z is $\mathsf{succ}(b,a,z) = \langle z, \mu^{\mathsf{succ}}_{b,a,z} \rangle$ where $\mu^{\mathsf{succ}}_{b,a,z}$ is given by

$$\begin{aligned} \mu^{\mathsf{succ}}_{b,a,z}(s') &:= P(s'|b,a,z) \\ &= \frac{\sum_{s \in \mu_b} \mu_b(s) \cdot \mathbf{P}(s,a)(s') \cdot \mathbf{O}(s,a,s')(z)}{P(z|b,a)} \end{aligned}$$

if $P(z|b,a) > 0$ and *undefined* otherwise. Successive computation of successor beliefs yields an infinite-state fully observable MDP capturing the POMDP dynamics. This *belief MDP* [Åström, 1965] is the basis for many solution methods for analysis problems on POMDPs. Let $\mathcal{B}^n_{\mathcal{M}}$ be the set of beliefs *reachable in n steps*, given by $\mathcal{B}^0 := \{b_{\mathit{init}}\}$ and $\mathcal{B}^{n+1}_{\mathcal{M}} := \mathcal{B}^n_{\mathcal{M}} \cup \{\mathsf{succ}(b,a,z) \mid b \in \mathcal{B}^n_{\mathcal{M}}, a \in \mathit{Act}, z \in Z\}$. $\mathcal{B}_{\mathcal{M}} := \lim_{n \to \infty} \mathcal{B}^n_{\mathcal{M}}$ is the set of reachable beliefs.

The belief MDP of POMDP $\mathcal{M}$ is $\mathit{bel}(\mathcal{M}) = \langle \mathcal{B}_{\mathcal{M}}, \mathit{Act}, \mathbf{P}^B, b_{\mathit{init}} \rangle$ where $\mathbf{P}^B(b,a)(b') := P(z|b,a)$ if $b' = \mathsf{succ}(b,a,z)$ and $\mathbf{P}^B(b,a)(b') := 0$ otherwise.

Costs We annotate POMDPs with *costs* (also referred to as *rewards*). Let $k \in \mathbb{N}$. A *(k-dimensional) cost structure* for $\mathcal{M}$ is a function $\mathbf{C}\colon S \times \mathit{Act} \times S \to \mathbb{N}^k$. When taking the transition $s \xrightarrow{a} s'$, the cost values $\mathbf{C}(s,a,s') = \langle c_1, \ldots, c_k \rangle$ are collected. A cost structure allows the encoding of different, independent cost measures in its dimensions. For example, one dimension can model the expired time, while another models energy consumption. The *cumulative cost of a finite path* $\hat{\pi}$ in $\mathcal{M}$ with respect to $\mathbf{C}$ is $\mathsf{cost}_{\mathbf{C}}(\hat{\pi}) := \sum_{i=1}^{|\hat{\pi}|} \mathbf{C}(s_{i-1}, a_i, s_i)$. The set of all distinct cost vectors occurring in $\mathbf{C}$ is $\Gamma_{\mathbf{C}} := \{\mathbf{C}(s,a,s') \mid s,s' \in S, a \in \mathit{Act}\}$.

[1] In the literature, beliefs are typically considered to only be the distribution μ_b. We extend this definition by the explicit inclusion of the observation that yields the belief to simplify later definitions.

3 PROBLEM STATEMENT

The cost-bounded reachability (CBR) problem asks for the maximal probability to reach a set of states in the POMDP via paths that respect (multi-dimensional) bounds on the cumulative costs. Formally, for a k-dimensional cost structure $\mathbf{C}$, relations $\bowtie \in \{\leq, >\}^k$ and threshold $\mathsf{t} \in \mathbb{N}^k$, we call $(\mathbf{C} \bowtie \mathsf{t})$ a *(k-dimensional) cost bound* over $\mathbf{C}$. Cost bounds represent constraints on the cumulative cost of paths with respect to the corresponding cost structure. For a finite path $\hat{\pi}$, bound $(\mathbf{C} \bowtie \mathsf{t})$ is *active in dimension* i iff $\mathsf{cost}_{\mathbf{C}}(\hat{\pi})[i] \bowtie[i]\ \mathsf{t}[i]$. Moreover, $(\mathbf{C} \bowtie \mathsf{t})$ is *active* for $\hat{\pi}$ iff it is active in all dimensions $1 \leq i \leq k$, i.e., $\mathsf{cost}_{\mathbf{C}}(\hat{\pi}) \bowtie \mathsf{t}$, where the relations in $\bowtie$ are applied element-wise. As we consider only natural values for costs, relations $\geq$ and $<$ are supported by adapting the thresholds in t. Non-negative rational costs are supported by suitable scaling of $\mathbf{C}$ and t.

We fix a POMDP $\mathcal{M}$ and a k-dimensional cost bound $(\mathbf{C} \bowtie \mathsf{t})$. Given a policy σ, the *cost-bounded reachability probability* for a state set $T \subseteq S$ is

$$\mathsf{Pr}^{\mathcal{M}}_{\sigma}\left(\lozenge_{\mathbf{C}\bowtie\mathsf{t}}\, T\right) := \mathsf{Pr}^{\mathcal{M}}_{\sigma}\big\{\pi \in \mathit{Cyl}(\hat{\pi}) \bigm| \mathit{last}(\hat{\pi}) \in T \text{ and } (\mathbf{C} \bowtie \mathsf{t}) \text{ is active for } \hat{\pi}\big\},$$

where $\mathit{Cyl}(\hat{\pi})$ is the set of infinite extensions of finite path $\hat{\pi}$ and $\mathsf{Pr}^{\mathcal{M}}_{\sigma}$ denotes the standard probability measure for $\mathcal{M}$ under policy σ [Puterman, 1994]. We call $\mathsf{Pr}^{\mathcal{M}}_{\sigma}\left(\lozenge_{\mathbf{C}\bowtie\mathsf{t}}\, T\right)$ the *value* of policy σ. The *maximal* cost-bounded reachability probability is $\mathsf{Pr}^{\mathcal{M}}_{\max}\left(\lozenge_{\mathbf{C}\bowtie\mathsf{t}}\, T\right) := \sup_{\sigma\in\Sigma^{\mathcal{M}}} \mathsf{Pr}^{\mathcal{M}}_{\sigma}\left(\lozenge_{\mathbf{C}\bowtie\mathsf{t}}\, T\right)$. Similar to related problems in POMDPs, a policy realising the maximal cost-bounded reachability probability is not guaranteed to exist. We formulate our problem in terms of a two-sided ϵ-approximation.

Problem 1 ((Multi-)Cost-Bounded Reachability (CBR))**.** *For POMDP $\mathcal{M}$, $T \subseteq S$, cost bound $(\mathbf{C} \bowtie \mathsf{t})$, and $\varepsilon \in [0,1]$, compute $V^U \in [0,1]$ and a policy $\tilde{\sigma} \in \Sigma^{\mathcal{M}}$, such that $V^U - \varepsilon \leq \mathsf{Pr}^{\mathcal{M}}_{\tilde{\sigma}}\left(\lozenge_{\mathbf{C}\bowtie\mathsf{t}}\, T\right) \leq \mathsf{Pr}^{\mathcal{M}}_{\max}\left(\lozenge_{\mathbf{C}\bowtie\mathsf{t}}\, T\right) \leq V^U$.*

Theorem 1. *The decision variant of CBR is undecidable.*

Proof. By considering 0-dimensional cost structures, CBR subsumes unbounded, undiscounted indefinite-horizon reachability, which is undecidable [Madani et al., 2003]. □

We write $\mathsf{Pr}^{\mathcal{M}}_{\max}\left(\lozenge\, T\right)$ for the unbounded problem as a special case of CBR with 0-dimensional costs. Methods to tackle unbounded reachability using two-sided approximations have been described in the literature (see Section 1).

Decidability of several subclasses of CBR can be established. For example, *finite-horizon* reachability probabilities—which can be computed exactly [Smallwood and Sondik, 1973]—are a special instance of cost-bounded reachability probabilities, where *each* transition induces exactly a cost of 1 and the costs are bounded upwards by the horizon. Lifting this to arbitrary, but strictly positive costs yields the *risk sensitive* setup from Hou et al. [2016], which—as mentioned by the authors—is still decidable. Our setting is more general since we allow transitions with 0 costs in any dimension as well as queries with only lower bounds.

4 FROM COST-BOUNDED TO UNBOUNDED REACHABILITY

We present an extension of the unfolding approach [Andova et al., 2003, Ohtsubo, 2004] for the cost-bounded reachability problem in MDPs to the partially observable domain.

Definition 1 (Cost Epoch)**.** *A* (cost) epoch *of dimension k is a tuple $\mathsf{e} = \langle e_1, \dots, e_k\rangle \in (\mathbb{N} \cup \{\bot\})^k$. We denote the domain of all k-dimensional epochs by $\mathsf{E}_k := (\mathbb{N} \cup \{\bot\})^k$.*

Each entry $\mathsf{e}[i]$ of an epoch keeps track of the costs that can be accumulated until the bound $(\mathbf{C} \bowtie \mathsf{t})$ changes its status in dimension i (active to inactive or vice versa).

The initial epoch is defined by the threshold vector t. To evolve cost epochs, we subtract the costs collected in each dimension. For values below 0, we use the dedicated symbol $\bot$ to indicate that the bound changed its status from the initial one. Formally, this is captured by the *monus* operation.

Definition 2 (Monus for Epochs)**.** *The* monus *operator for cost epochs $\ominus : \mathsf{E}_k \times \mathbb{N}^k \to \mathsf{E}_k$ is given component-wise as*

$$(\mathsf{e} \ominus \mathsf{c})[i] := \begin{cases} \mathsf{e}[i] - \mathsf{c}[i] & \text{if } \bot \neq \mathsf{e}[i] \wedge \mathsf{e}[i] \geq \mathsf{c}[i], \\ \bot & \text{otherwise.} \end{cases}$$

We lift the notion of being active to epochs. The indicator function $\mathsf{actv}_{\mathbf{C}\bowtie\mathsf{t}} : \mathsf{E}_k \to \{0,1\}$ is 1 iff bound $(\mathbf{C} \bowtie \mathsf{t})$ is active in a given epoch, i.e., $\mathsf{actv}_{\mathbf{C}\bowtie\mathsf{t}}(\mathsf{e}) := 1$ if for all $1 \leq i \leq k$, $\bowtie[i] = \leq$ implies $\mathsf{e}[i] \neq \bot$ and $\bowtie[i] = >$ implies $\mathsf{e}[i] = \bot$, and $\mathsf{actv}_{\mathbf{C}\bowtie\mathsf{t}}(\mathsf{e}) := 0$ otherwise.

Using the idea of epochs, we construct a POMDP that enables us to reason about the activation of a bound on the level of states instead of the path level. We first recap the construction for an MDP as it is described in the literature, e.g. in Hartmanns et al. [2020]. Let the (finite) set of reachable epochs from an epoch e be given by

$$\mathsf{E}_k(\mathsf{e}) := \{\mathsf{e}' \in \mathsf{E}_k \mid \exists \mathsf{c} \in \mathbb{N}^k\colon\ \mathsf{e}' = \mathsf{e} \ominus \mathsf{c}\}.$$

Definition 3 (Bound Unfolding MDP)**.** *For MDP $M = \langle S, \mathit{Act}, \mathbf{P}, s_{\mathit{init}}\rangle$ and cost bound $(\mathbf{C} \bowtie \mathsf{t})$, the* bound unfolding MDP *is $\mathsf{un}_{\mathbf{C}\bowtie\mathsf{t}}(M) := \langle S \times \mathsf{E}_k(\mathsf{t}), \mathit{Act}, \mathbf{P}_{\mathsf{un}}, \langle s_{\mathit{init}}, \mathsf{t}\rangle\rangle$, and for $s_{\mathsf{e}} := \langle s, \mathsf{e}\rangle$, $s'_{\mathsf{e}'} := \langle s', \mathsf{e}'\rangle$, and $a \in \mathit{Act}$:*

$$\mathbf{P}_{\mathsf{un}}(s_{\mathsf{e}}, a)(s'_{\mathsf{e}'}) := \begin{cases} \mathbf{P}(s,a)(s') & \text{if } \mathsf{e}' = \mathsf{e} \ominus \mathbf{C}(s,a,s'), \\ 0 & \text{otherwise.} \end{cases}$$

Definition 4 (Bound Unfolding POMDP)**.** *Given a POMDP $\mathcal{M} = \langle M, Z, \mathbf{O}\rangle$ with underlying MDP $M = \langle S, Act, \mathbf{P}, s_{init}\rangle$ and cost bound* $(\mathbf{C} \bowtie \mathsf{t})$*, the* bound unfolding POMDP *is* $\mathsf{un}_{\mathbf{C}\bowtie\mathsf{t}}(\mathcal{M}) := \langle \mathsf{un}_{\mathbf{C}\bowtie\mathsf{t}}(M), Z, \mathbf{O}_{\mathsf{un}}\rangle$ *where for $s_{\mathsf{e}} := \langle s, \mathsf{e}\rangle$, $s'_{\mathsf{e}'} := \langle s', \mathsf{e}'\rangle$, and $a \in Act$: $\mathbf{O}_{\mathsf{un}}(s_{\mathsf{e}}, a, s'_{\mathsf{e}'}) := \mathbf{O}(s, a, s')$.*

Finally, we lift the notion of an active bound to states of the unfolding POMDP which enables us to reason about conformance to the bound on states instead of paths.

Definition 5 (Active States)**.** *A state of the unfolding $\langle s, \mathsf{e}\rangle \in S \times \mathsf{E}_k(\mathsf{t})$ is* active *iff* $\mathsf{actv}_{\mathbf{C}\bowtie\mathsf{t}}(\mathsf{e}) = 1$*. Given $T \subseteq S$, the set of* active T-states *is* $\mathsf{actv}_{\mathbf{C}\bowtie\mathsf{t}}(T) = \{\langle s_T, \mathsf{e}\rangle \in S \times \mathsf{E}_k(\mathsf{t}) \mid s_T \in T \wedge \mathsf{actv}_{\mathbf{C}\bowtie\mathsf{t}}(\mathsf{e}) = 1\}$.

As $\mathcal{M}$ and $\mathsf{un}_{\mathbf{C}\bowtie\mathsf{t}}(\mathcal{M})$ share the set Z of observations, their possible policies coincide, i.e., $\Sigma^{\mathcal{M}} = \Sigma^{\mathsf{un}_{\mathbf{C}\bowtie\mathsf{t}}(\mathcal{M})}$. We formalise the relationship between the original POMDP and its unfolding POMDP regarding cost-bounded reachability.

Theorem 2. *Given a POMDP $\mathcal{M}$, set $T \subseteq S$ and cost bound* $(\mathbf{C} \bowtie \mathsf{t})$*, it holds that for all policies $\sigma \in \Sigma^{\mathcal{M}}$:* $\mathsf{Pr}^{\mathcal{M}}_{\sigma}(\Diamond_{\mathbf{C}\bowtie\mathsf{t}} T) = \mathsf{Pr}^{\mathsf{un}_{\mathbf{C}\bowtie\mathsf{t}}(\mathcal{M})}_{\sigma}(\Diamond\ \mathsf{actv}_{\mathbf{C}\bowtie\mathsf{t}}(T))$.

Proof Sketch. Let f be the mapping from paths of $\mathsf{un}_{\mathbf{C}\bowtie\mathsf{t}}(\mathcal{M})$ to paths of $\mathcal{M}$ obtained by dropping the epochs from the states. f is bijective and $\mathsf{Pr}^{\mathcal{M}}_{\sigma}(\{f(\pi) \mid \pi \in \Pi\}) = \mathsf{Pr}^{\mathsf{un}_{\mathbf{C}\bowtie\mathsf{t}}(\mathcal{M})}_{\sigma}(\Pi)$ for any policy σ and set Π of paths in $\mathsf{un}_{\mathbf{C}\bowtie\mathsf{t}}(\mathcal{M})$. The claim follows by taking Π as the set of paths that reach $\mathsf{actv}_{\mathbf{C}\bowtie\mathsf{t}}(T)$. The corresponding paths in $\mathcal{M}$ reach T while the bound is active. □

We get the following result about maximal probabilities.

Corollary 1. *For POMDP $\mathcal{M}$, $T \subseteq S$ and* $(\mathbf{C} \bowtie \mathsf{t})$*:* $\mathsf{Pr}^{\mathcal{M}}_{\max}(\Diamond_{\mathbf{C}\bowtie\mathsf{t}} T) = \mathsf{Pr}^{\mathsf{un}_{\mathbf{C}\bowtie\mathsf{t}}(\mathcal{M})}_{\max}(\Diamond\ \mathsf{actv}_{\mathbf{C}\bowtie\mathsf{t}}(T))$.

Thus, to tackle CBR as in Problem 1, we can consider the unbounded indefinite-horizon reachability problem on the unfolding POMDP. Solution methods for this problem include smart exploration of the belief space [Norman et al., 2017, Bork et al., 2022, Ho et al., 2024] or the policy space [Andriushchenko et al., 2022].

5 COST-(LEVEL-)AWARE POMDPS

The general cost bound framework assumes that an agent's decisions are solely based on environmental observations the agent receives. However, costs might reflect quantities—such as the level of a battery—that the agent observes. We refine our cost-bounded analysis for the special case where the observation model captures the additional information provided by costs. This *cost-awareness* notion is related to the reward-based belief updates in Izadi and Precup [2005].

Definition 6 (Cost-Aware POMDP)**.** *A POMDP $\mathcal{M} = \langle M, Z, \mathbf{O}\rangle$ with $M = \langle S, Act, \mathbf{P}, s_{init}\rangle$ is* cost-aware *with respect to a k-dimensional cost structure $\mathbf{C}$ if for all $z \in Z$ there is $\mathsf{c}_z \in \mathbb{N}^k$ such that for any transition $s \xrightarrow{a} s'$ with $z \in supp(\mathbf{O}(s, a, s'))$ we have $\mathbf{C}(s, a, s') = \mathsf{c}_z$.*

In a cost-aware POMDP, all observations z can be assigned a cost vector c_z. *An observation z only occurs at transitions that yield costs equal to c_z, effectively guaranteeing that the collected costs are observable.* Cost-awareness implies that, given an observation trace $z_1 \ldots z_n \in Z^*$, an agent can derive the costs $\sum_{i=1}^{n} \mathsf{c}_{z_i}$ that have been accumulated so far. Thus, in the bound unfolding POMDP, an agent can always be certain about the cost epoch it is currently in as it has access to the history. This is reflected in the belief space: if the POMDP $\mathcal{M}$ is cost-aware, all reachable beliefs in the belief MDP of its unfolding $bel(\mathsf{un}_{\mathbf{C}\bowtie\mathsf{t}}(\mathcal{M}))$ only contain states that belong to the same epoch. Formally, if $\mathcal{M}$ is cost-aware, then for every reachable belief $b = \langle z, \mu\rangle$ of $bel(\mathsf{un}_{\mathbf{C}\bowtie\mathsf{t}}(\mathcal{M}))$ there is an epoch $\mathsf{e} \in \mathsf{E}_k$ such that $supp(\mu_b) \subseteq S \times \{\mathsf{e}\}$. This enables an optimised analysis of $bel(\mathsf{un}_{\mathbf{C}\bowtie\mathsf{t}}(\mathcal{M}))$ as discussed in the sequel.

5.1 SEQUENTIAL EPOCH ANALYSIS

Our approach so far is to analyse unbounded reachability for the unfolding POMDP $\mathsf{un}_{\mathbf{C}\bowtie\mathsf{t}}(\mathcal{M})$, e.g. via (abstractions of) its belief MDP $bel(\mathsf{un}_{\mathbf{C}\bowtie\mathsf{t}}(\mathcal{M}))$. However, such an analysis operates on the potentially large unfolding POMDP.

Recent works on fully observable MDPs avoid the construction of a large unfolding MDP by considering epochs one after another in a dynamic programming fashion [Hahn and Hartmanns, 2016, Klein et al., 2018, Hartmanns et al., 2020]. This *sequential epoch analysis* is based on the epoch dependency graph $\big\langle \mathsf{E}_k(\mathsf{t}), \big\{\langle \mathsf{e}, \mathsf{e} \ominus \mathsf{c}\rangle \bigm| \mathsf{c} \in \Gamma_{\mathbf{C}}\big\}\big\rangle$ which has an edge from epoch e to epoch e' iff a transition of the form $\langle s, \mathsf{e}\rangle \xrightarrow{a} \langle s', \mathsf{e}'\rangle$ exists in the unfolding POMDP. Since costs are non-negative, the epoch graph is acyclic (except for self-loops). The idea is to process epochs in a reversed topological order $\mathsf{e}_0, \mathsf{e}_1, \ldots, \mathsf{e}_n$ with $\mathsf{e}_0 = \bot^k$ and $\mathsf{e}_n = \mathsf{t}$. For each considered epoch e_i, an *epoch MDP*—which essentially is the restriction of the bound unfolding MDP to states with epoch e—is constructed and analysed while propagating results from previous epochs $\mathsf{e}_0, \ldots, \mathsf{e}_{i-1}$. Implementations can exploit similarities between different epoch MDPs. This way, the approach efficiently analyses properties of the large unfolding MDP without an explicit construction.

We lift sequential epoch analysis to POMDPs. In the general case, the POMDP dynamics do not allow a clear separation of epochs. In particular, when considering belief-based solution methods, beliefs may have states representing several different epochs in their support. We therefore focus on cost-aware POMDPs. Our approach is to perform sequential epoch analysis on the belief MDP—or a finite abstraction thereof. To this end, we lift cost bounds to the belief space. We fix a POMDP $\mathcal{M}$ and cost bound $(\mathbf{C} \bowtie \mathsf{t})$ such that $\mathcal{M}$ is cost-aware with respect to $\mathbf{C}$.

Definition 7 (Cost-Aware Belief Cost Bound)**.** *The* belief cost structures *for $\mathit{bel}(\mathcal{M})$ is $\mathbf{C}^B$ where for $s \in \mu_b$, $s' \in \mu_{b'}$, $\mathbf{C}^B((z, \mu_b), a, (z', \mu_{b'})) := \mathbf{C}(s, a, s')$. The belief cost bound is $\mathit{bel}(\mathbf{C} \bowtie \mathsf{t}) = (\mathbf{C}^B \bowtie \mathsf{t})$.*

$\mathbf{C}^B$ is well-defined as cost-awareness guarantees that $\mathbf{C}(s, a, s') = \mathbf{C}(q, a, q')$ for all $s, q \in \mu_b$ and $s', q' \in \mu_{b'}$. For cost-aware POMDPs, applying the cost unfolding and then constructing the belief MDP is equivalent to *first* constructing the belief MDP and *then* applying cost unfolding. For MDPs M_1 and M_2 we write $M_1 \cong M_2$ iff the reachable fragments are isomorphic, i.e., equal up to renaming.

Theorem 3. $\mathit{bel}(\mathsf{un}_{\mathbf{C}\bowtie\mathsf{t}}(\mathcal{M})) \cong \mathsf{un}_{\mathit{bel}(\mathbf{C}\bowtie\mathsf{t})}(\mathit{bel}(\mathcal{M}))$.

Proof Sketch. Let $\langle z, \mu_b\rangle$ be a state of $\mathit{bel}(\mathsf{un}_{\mathbf{C}\bowtie\mathsf{t}}(\mathcal{M}))$. Since $\mathcal{M}$ is cost-aware, there is an epoch e with $\mathit{supp}(\mu_b) \subseteq S \times \{\mathsf{e}\}$. $\langle z, \mu_b\rangle$ is isomorphic to state $\langle\langle z, \mu'\rangle, \mathsf{e}\rangle$ of $\mathsf{un}_{\mathit{bel}(\mathbf{C}\bowtie\mathsf{t})}(\mathit{bel}(\mathcal{M}))$, where $\mu'(s) = \mu_b(\langle s, \mathsf{e}\rangle)$. □

The sequential epoch analysis for MDPs outlined above can readily be applied to $\mathit{bel}(\mathcal{M})$ to show properties for its unfolding $\mathsf{un}_{\mathit{bel}(\mathbf{C}\bowtie\mathsf{t})}(\mathit{bel}(\mathcal{M}))$. Due to Theorem 3, analysis results immediately carry over to $\mathit{bel}(\mathsf{un}_{\mathbf{C}\bowtie\mathsf{t}}(\mathcal{M}))$. Approaches like the ones described in Norman et al. [2017], Bork et al. [2020, 2022] handle large or even infinite belief MDPs through abstraction, yielding a finite MDP $\mathit{abstr}(\mathit{bel}(\mathcal{M}))$ which over- or under-approximates the behaviour of $\mathit{bel}(\mathcal{M})$. If bound unfolding retains the abstraction—i.e., $\mathit{abstr}(\mathit{bel}(\mathsf{un}_{\mathbf{C}\bowtie\mathsf{t}}(\mathcal{M}))) \cong \mathsf{un}_{\mathit{bel}(\mathbf{C}\bowtie\mathsf{t})}(\mathit{abstr}(\mathit{bel}(\mathcal{M})))$—the sequential epoch analysis is compatible with such techniques. In our experiments, we use this observation for the cut-off abstraction of Bork et al. [2022] and the discretisation of Bork et al. [2020].

5.2 COST LEVEL AWARENESS

In many problem instances, neither full cost-awareness nor absolute unawareness are realistic. For example, a robot can have a rough estimate of its energy level (high, medium, low, empty) while not being aware of exactly how much energy it has spent yet. We capture this notion of *cost levels* using functions that assign in each dimension a *level* according to the collected cost. We assume uniform levels, i.e., we change the level in dimension i whenever an additional cost of some fixed $\mathsf{d}[i] > 0$ is collected. Moreover, 0 cost is its own level.

Definition 8 (Level Function)**.** *For $\mathsf{d} \in (\mathbb{N}\setminus\{0\})^k$, the* level function *$\mathit{lvl}_{\mathsf{d}} : \mathbb{N}^k \to \mathbb{N}^k$ is given by $\mathit{lvl}_{\mathsf{d}}(\mathsf{c})[i] := \left\lceil \frac{\mathsf{c}[i]}{\mathsf{d}[i]} \right\rceil$.*

We fix a POMDP $\mathcal{M}$, cost bound $(\mathbf{C} \bowtie \mathsf{t})$, target states T and a level function $\mathit{lvl}_{\mathsf{d}}$ and define a *level-aware* instance of CBR where in addition to the observations in $\mathcal{M}$, the agent can use information about the current cost level to make decisions. This level-aware instance is in general not equivalent to the CBR instance $\mathcal{M}$ with cost bounds $(\mathbf{C} \bowtie \mathsf{t})$. It introduces new observations that provide additional information which is not observable in the original model. However, the level-aware instance may more accurately capture the scenario we are interested in as an agent may have access to such information. Our definition of a level-aware instance decouples the modelling of observations arising from the environment and those arising from possible cost observation. This simplifies the modelling of such instances and allows, for example, the comparison of different degrees of cost levels. By choosing $\mathsf{d} = \langle 1, \ldots, 1\rangle$, we can define a variant of the original POMDP with *full* cost-awareness which we call the *cost-aware variant*. $\mathsf{d} = \mathsf{t}$ means that only activeness of bounds and the first collection of a non-zero cost can be observed for each dimension.

To incorporate cost level awareness into our framework, we present a transformation of $\mathcal{M}$, $\mathbf{C}$ and $\mathit{lvl}_{\mathsf{d}}$ into a new cost-aware POMDP (cf. Def. 6). The transformation encodes the cost that can still be collected until a new level is reached in the state space of the POMDP and introduces fresh observations to mark transitions in which one or more *level changes* occur. This way, an observation trace suffices to deduce the current level. Appendix B provides further details.

Definition 9 (Level Unfolding POMDP)**.** *The* level unfolding *with respect to $\mathit{lvl}_{\mathsf{d}}$ is the POMDP $\mathit{lvl}_{\mathsf{d}}(\mathcal{M}) = \langle M_{\mathit{lvl}_{\mathsf{d}}}, Z_{\mathit{lvl}_{\mathsf{d}}}, \mathbf{O}_{\mathit{lvl}_{\mathsf{d}}}\rangle$ and the cost structure $\mathbf{C}_{\mathit{lvl}_{\mathsf{d}}}$ with*

- $M_{\mathit{lvl}_{\mathsf{d}}} = \langle S_{\mathit{lvl}_{\mathsf{d}}}, \mathit{Act}, \mathbf{P}_{\mathit{lvl}_{\mathsf{d}}}, \langle s_{\mathit{init}}, \langle 0, \ldots, 0\rangle\rangle\rangle$,
- $S_{\mathit{lvl}_{\mathsf{d}}} = S \times \{\ell \in \mathbb{N}^k \mid \forall i : \ell[i] < \mathsf{d}[i]\}$,
- $Z_{\mathit{lvl}_{\mathsf{d}}} = Z \times \left\{\mathsf{c} \in \mathbb{N}^k \mid \forall i : \mathsf{c}[i] \leq \left\lceil c_i^{\max}/\mathsf{d}[i]\right\rceil\right\}$, *where* $c_i^{\max} = \max_{s,s' \in S, a \in \mathit{Act}} \mathbf{C}(s, a, s')[i]$,
- $\mathbf{P}_{\mathit{lvl}_{\mathsf{d}}}(\langle s, \ell\rangle, a, \langle s', \ell'\rangle) = \mathbf{P}(s, a, s')$ *if for all i:* $\ell'[i] = \ell[i] - \mathbf{C}(s, a, s')[i] \mod d$,
- $\mathbf{O}_{\mathit{lvl}_{\mathsf{d}}}(\langle s, \ell\rangle, a, \langle s', \ell'\rangle)(\langle z, \mathsf{c}\rangle) = \mathbf{O}(s, a, s')(z)$ *if for all i:* $\mathsf{c}[i] = \left\lceil \big(\mathbf{C}(s, a, s')[i] - \ell[i]\big)/\mathsf{d}[i]\right\rceil$,
- $\mathbf{C}_{\mathit{lvl}_{\mathsf{d}}}(\langle s, \ell\rangle, a, \langle s', \ell'\rangle) = \mathbf{C}(s, a, s')$,

and $\mathbf{P}_{\mathit{lvl}_{\mathsf{d}}}$ and $\mathbf{O}_{\mathit{lvl}_{\mathsf{d}}}$ are zero in all other cases.

The CBR instance with POMDP $\mathit{lvl}_{\mathsf{d}}(\mathcal{M})$, bounds $(\mathbf{C}_{\mathit{lvl}_{\mathsf{d}}} \bowtie \mathsf{t})$ and target states $T_{\mathsf{d}} = (T \times \mathbb{N}^k) \cap S_{\mathit{lvl}_{\mathsf{d}}}$ is the *level-aware variant* (w.r.t. $\mathit{lvl}_{\mathsf{d}}$) for $\mathcal{M}$, $(\mathbf{C} \bowtie \mathsf{t})$, and $T \subseteq S$.

When assuming that a level function captures when a bound becomes (in-)active—e.g. if an energy limit is exceeded—the level-aware instance can be reduced to an equivalent CBR instance on a *fully cost-aware* POMDP which can then be solved using the sequential approach. Mathematically, this is the case if for all i, $\mathsf{d}[i]$ divides $\mathsf{t}[i]$ (written $\mathsf{d}[i] \mid \mathsf{t}[i]$).

Theorem 4. *Let $lvl_{\mathsf{d}}(\mathcal{M})$ such that $\forall i : \mathsf{d}[i] \mid \mathsf{t}[i]$ and let $(\mathbf{L} \bowtie \mathsf{t}_{\mathsf{d}})$ be a cost bound for $lvl_{\mathsf{d}}(\mathcal{M})$ with $\forall i$: $\mathbf{L}(\langle s, \ell \rangle, a, \langle s', \ell' \rangle)[i] := \left\lceil \big(\mathbf{C}(s,a,s')[i] - \ell[i]\big)/\mathsf{d}[i] \right\rceil$ and $\mathsf{t}_{\mathsf{d}}[i] := \mathsf{t}[i]/\mathsf{d}[i]$. Then, $lvl_{\mathsf{d}}(\mathcal{M})$ is cost-aware w.r.t. $\mathbf{L}$ and*

$$\mathsf{Pr}_{\max}^{lvl_{\mathsf{d}}(\mathcal{M})}\left(\Diamond_{\mathbf{C}_{lvl_{\mathsf{d}}} \bowtie \mathsf{t}}\, T_{\mathsf{d}}\right) = \mathsf{Pr}_{\max}^{lvl_{\mathsf{d}}(\mathcal{M})}\left(\Diamond_{\mathbf{L} \bowtie \mathsf{t}_{\mathsf{d}}}\, T_{\mathsf{d}}\right).$$

Proof Sketch. $\mathbf{L}$ captures the number of level jumps that occur when a transition in $lvl_{\mathsf{d}}(\mathcal{M})$ is taken. Since $\mathsf{d}[i] \mid \mathsf{t}[i]$, the bound changes its activeness in dimension i after exactly $\mathsf{t}[i]/\mathsf{d}[i] + 1$ level jumps occurred. We can then show that for a path $\hat{\pi}$, $\mathsf{cost}_{\mathbf{C}_{lvl_{\mathsf{d}}}}(\hat{\pi}) \bowtie \mathsf{t}$ if and only if $\mathsf{cost}_{\mathbf{L}}(\hat{\pi}) \bowtie \mathsf{t}_{\mathsf{d}}$. □

6 EMPIRICAL EVALUATION

We evaluate the practicality of cost-bounded reachability analysis in POMDPs to answer the following questions:

(**Q**1) Are the presented approaches suitable for solving cost-bounded reachability problems in practice?

(**Q**2) How is this influenced by cost (level) awareness?

(**Q**3) Does applying the sequential approach on cost-aware belief MDPs improve performance compared to reducing the problem to unbounded reachability?

Implementation We extended the probabilistic model checking tool Storm [Hensel et al., 2022] to support CBR queries for POMDPs. Storm can analyse unbounded reachability in POMDPs via finite abstractions of the belief MDP [Bork et al., 2020, 2022] as well as cost-bounded reachability in (fully observable) MDPs via the sequential approach from Hartmanns et al. [2020]. On top of that, we implemented the construction of bound unfolding and level unfolding POMDPs, along with the exploration of cost-aware belief MDPs with belief cost bounds. Both constructions are integrated into the existing POMDP verification framework of Storm, enabling us to tackle CBR problems in three different configurations:

- UNFOLD: transform to an unbounded reachability problem on the unfolding POMDP (see Section 4), then verify (finite abstractions of) its belief MDP.
- CA-UNFOLD: Construct a cost-aware variant of the POMDP (see Section 5.2), then analyse it as in UNFOLD.
- CA-BEL-SEQ: Construct a cost-aware variant of the POMDP, then construct (a finite abstraction of) its cost-aware belief MDP and analyse CBR on this fully observable MDP using the sequential epoch approach see Section 5.1).

Table 1: Information on Benchmark Instances

Model	$\lvert S \rvert$	$\lvert Z \rvert$	$\bowtie$	$\lvert \mathsf{E} \rvert$
clean6	37	2	$\leq, >^{\dagger}$	413
clean12	73	2	$\leq, >^{\dagger}$	1508
incline	25	9	$\leq, \leq$	497
obstcl	25	10	$\leq, \leq$	83
resrc	721	155	$>, >, \leq$	$2107/4{\cdot}10^4$
rover	16	9	$>, \leq, \leq$	$7{\cdot}10^5/2{\cdot}10^7$
serv	$8{\cdot}10^4$	6016	$\leq$	$40/68$
walk40	84	44	$\leq$	82
walk120	244	124	$\leq$	82
water	34	5	$\leq, >$	$3{\cdot}10^4/3{\cdot}10^5$

CA-UNFOLD and CA-BEL-SEQ both first construct a cost-aware variant based on the input POMDP, extending it with additional observations which in general changes the optimal achievable value. We opt for this method of defining cost-aware instances to decouple the modelling of environmental observations and those observations stemming from costs. CA-UNFOLD and CA-BEL-SEQ differ in the underlying solution approach. Instead of full cost-awareness, they can also be used with cost level awareness by applying a level unfolding first. Moreover, cost-awareness can also be induced for only a subset of the dimensions $I \subseteq \{1, \ldots, k\}$. Storm uses a state-based observation model $\mathbf{O} : S \to Z$. We transform our observation model into such a state-based one by encoding observations in the state space [Chatterjee et al., 2016]. For CA-UNFOLD and CA-BEL-SEQ, this results in larger state spaces compared to the original POMDP due to the additional observations.

For each of the three configurations, we can either use the *cut-off* approach of Bork et al. [2022] or the *discretisation* approach of Bork et al. [2020] to obtain a finite abstraction of the belief MDP yielding sound lower or upper bounds for the optimal cost-bounded reachability probabilities. Both abstractions—noted below as CUT and DISCR—have a hyper-parameter that controls the size of the obtained belief MDP abstractions.

Our implementation is publicly available as part of the supplementary material of this paper [Bork et al., 2025].[2]

Benchmarks Since there is no established benchmark set for CBR problems, we use partially observable variants of some cost-bounded reachability problems from Hartmanns et al. [2020] (resrc, rover, serv). In addition, we consider three variants of grid world examples where reaching a goal is made difficult by either an incline (incline), obstacles (obstcl), water levels (water), or uncertain movements (walk40, walk120). Finally, we consider our motivating example (clean6) and a version with 12 tiles (clean12). The benchmarks are given in the guarded command language of Prism [Kwiatkowska et al., 2011]. Appendix D

[2] https://zenodo.org/records/15642233

Table 2: Overview of Obtained Value Bounds and Runtimes

Model	$\lvert\mathsf{E}\rvert$	$\lvert S_{\mathsf{un}}\rvert$	UNFOLD: CUT / DISCR		CA-UNFOLD: CUT / DISCR		CA-BEL-SEQ: CUT / DISCR	
clean6	413	3809	0.86 (<1s)	1 (713s)	0.929 (3.9s)*	0.971 (299s)	**0.929 (2.2s)**	**0.949 (<1s)**
clean12	1508	$3\cdot10^{4}$	0.77 (262s)	1 (61.9s)	0.708 (240s)	1 (128s)	**0.88 (381s)**	**0.957 (4.8s)**
incline	497	2094	0.989 (21.5s)	0.989 (<1s)	0.989 (1.1s)	**0.989 (<1s)**	**0.989 (<1s)**	**0.989 (<1s)**
obstcl	83	741	0.87 (<1s)*	0.87 (<1s)	**0.87 (<1s)**	**0.87 (<1s)**	**0.87 (<1s)**	**0.87 (<1s)**
resrc	2107	$2\cdot10^{5}$	$7\cdot10^{-15}$ (360s)	0.0312 (2.7s)	$7\cdot10^{-15}$ (345s)	0.0312 (3.6s)	**0.0312 (<1s)**	**0.0312 (<1s)**
resrc	$4\cdot10^{4}$	$6\cdot10^{6}$	$2\cdot10^{-73}$ (400s)	$3\cdot10^{-5}$ (81.1s)	$2\cdot10^{-73}$ (427s)	$3\cdot10^{-5}$ (101s)	**$3\cdot10^{-5}$ (16.3s)**	**$3\cdot10^{-5}$ (4.4s)**
rover	$7\cdot10^{5}$	$1\cdot10^{7}$	0.353 (337s)	0.853 (237s)	0.861 (462s)*	0.861 (406s)	**0.861 (12.2s)****	**0.861 (12.4s)**
rover	$2\cdot10^{7}$	?	-	-	-	-	**0.951 (466s)****	**0.951 (456s)**
serv	40	$10\cdot10^{4}$	0.0474 (111s)	0.378 (1.9s)	**0.0474 (107s)**	**0.378 (6.4s)**	-	0.378 (15.4s)
serv	68	$3\cdot10^{5}$	0.172 (282s)	0.636 (158s)	**0.169 (281s)**	**0.636 (528s)**	-	0.637 (139s)
walk40	82	6847	0.916 (97.3s)*	0.932 (1603s)	0.916 (174s)*	0.935 (1431s)	**0.916 (<1s)**	**0.93 (1295s)**
walk120	82	$2\cdot10^{4}$	0.867 (846s)	0.931 (542s)	0.869 (1681s)	0.931 (1309s)	**0.895 (11.0s)****	**0.926 (727s)**
water	$3\cdot10^{4}$	$6\cdot10^{5}$	$3\cdot10^{-123}$ (327s)	1 (144s)	0.166 (17.9s)*	0.166 (17.5s)	**0.166 (<1s)****	**0.166 (<1s)**
water	$3\cdot10^{5}$	$5\cdot10^{6}$	-	-	0.181 (268s)*	0.181 (269s)	**0.181 (4.2s)****	**0.181 (4.3s)**

provides further details.

Table 1 outlines the benchmark instances we consider, including the number of POMDP states $|S|$, the number of distinct observations $|Z|$, the relation of the cost bounds $\bowtie$ (also indicating the dimensionality k), and the number of reachable epochs $|\mathsf{E}|$ indicating the magnitude of the involved cost thresholds $\mathsf{t} = \langle t_1, \dots, t_k \rangle$. We consider two different thresholds for resrc, rover, serv, and water. The symbol † for the clean instances denotes that the second cost dimension remains unobservable, even for the cost-aware configurations. While some of the considered POMDPs have small state spaces, a high number of relevant epochs leads to intricate cost-bounded reachability queries. We emphasise that the complexity of the considered instances stems from the inclusion of cost bounds—the unbounded problem variants with the same target states we consider result in a maximal probability of 1 for all instances.

Setup We conducted experiments on Intel Xeon 8468 Sapphire systems (2.1 GHz) with memory limited to 64 GB. STORM runs on a single core. For each combination of benchmark instance, configuration, and abstraction type, we considered 25 different hyper-parameter assignments to capture different trade-offs between approximation accuracy and computational tractability. A time limit of 1800 seconds (walltime) was applied for the individual runs. A detailed setup description is provided in Appendix E.

Results Table 2 lists the best value bounds obtained within the time and memory limit as well as the time it took to obtain these bounds. CUT yields a *lower bound* on the optimal value, and DISCR yields an *upper bound* V^U (see Problem 1). Table entries are bold-faced when they depict the tightest lower or upper bound obtained within the fastest runtime. This does not include UNFOLD as it computes a different measure. A dash (-) indicates that no non-trivial bound was obtained within the time limit. Column $|S_{\mathsf{un}}|$ denotes the number of reachable states in the unfolding POMDP (if known). For CUT, an asterisk (*) indicates that the belief MDP was fully explored.

The plots at the top of Figure 2 show the obtained value bounds over analysis time for two selected instances. A data point $\langle x, y\rangle$ for configuration Z means that (lower or upper) value bound y was established within x seconds using configuration Z and appropriate hyper-parameters. Similarly, the plots at the bottom show obtained value bounds for our motivating example clean6, on the left when increasing the number of observable energy levels (intuitively providing more information for a policy) and on the right when increasing the energy budget. Tables containing additional details on the results and further plots are given in Appendix F.

Discussion Concerning (**Q**1), we see that all considered approaches produce non-trivial bounds on the optimal value for the majority of benchmark instances. Towards (**Q**2), we observe that the true optimal values in cost-(level-)aware variants of POMDPs is always at least as good as in the original POMDP. Our results show that cost-awareness does not affect the obtained upper value bounds for many of our benchmarks. However, the obtained lower bounds are often larger with cost-awareness enabled, resulting in tighter gaps between lower- and upper bounds. As we expect the maximal value to increase under cost-awareness, this indicates that the discretisation method works better for obtaining tight approximations when having more information about incurred costs. This is also indicated by the data shown in the bottom left of Figure 2. With an increasing number of observable cost levels within the cost threshold $\mathsf{t}[0]$, the obtained upper bounds decrease, while we expect the actual (unknown) maximal probability to increase when more in-

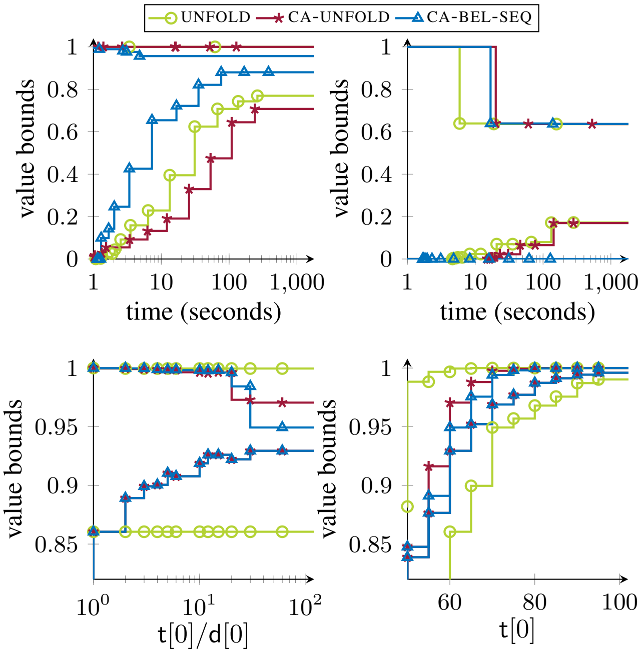


Figure 2: Value bounds obtained over time (top left: clean12, |E|=1508; top right: serv, |E|=40), for different observation levels (bottom left: clean6, |E|=413), and for increasing cost thresholds (bottom right: clean6, $168 \leq |E| \leq 658$).

formation is available for the policy, pointing towards tighter upper bounds. In addition, we see that with increasingly finer levels, the obtained values indeed increase, hinting at an increasingly higher true optimal value. An exception to our general observation is clean12 where CA-UNFOLD yields a smaller value than UNFOLD for CUT. This is a result of the state space increase for cost-aware variants caused by STORM's state-based observation model. The exploration of the cost-aware variant's belief space appears to be not as thorough as for the original POMDP within our given time limit. This results in a worse approximation (smaller lower bound) of a larger optimal value. Regarding (**Q**3), CA-BEL-SEQ often yields better approximations in less time compared to CA-UNFOLD. A notable exception is the serv case study. Here, the unfolding POMDP has similar size compared to the original POMDP ($|S_{\mathsf{un}}| \approx |S|$) as we have comparably few reachable epochs.

7 CONCLUSION

We proposed a general framework for the analysis of reachability probability objectives under multiple cost constraints on POMDPs. These objectives can be tackled by considering an unbounded objective on an unfolding of the POMDP. Observation of incurred costs enables an advanced technique based on a sequential analysis of cost epochs on the belief MDP. Awareness of cost levels provides a more realistic way to model certain scenarios and can often be reduced to the cost-aware setting. Our experiments using a prototype implementation in STORM indicate the suitability of cost-bounded reachability analysis for practical applications.

As future research, we propose the extension of our framework towards cost-bounded expected reward objectives and a more flexible notion of cost levels. Additionally, developing approximation methods tailored towards the cost-bounded setting is useful for increasing scalability.

Data Availability The supplementary material contains detailed proofs for our theoretical results and additional data from our evaluation. The artifact accompanying this paper [Bork et al., 2025] contains the source code of our implementation, model files used for the experimental evaluation and logfiles created during our experiments.

Acknowledgements

This work has been partially funded by the DFG RTG 2236/2 (UnRAVeL), the KI-Starter Project 'Verifying AI Systems under Partial Observability' of the Ministry of Culture and Science of the German State of North Rhine-Westphalia and the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No. 101008233 (MISSION). Experiments were performed with computing resources granted by RWTH Aachen University under project rwth1632.

References

Suzana Andova, Holger Hermanns, and Joost-Pieter Katoen. Discrete-time rewards model-checked. In *Formal Modeling and Analysis of Timed Systems: First International Workshop, FORMATS 2003*, volume 2791 of *Lecture Notes in Computer Science*, pages 88–104. Springer, 2003. doi: 10.1007/978-3-540-40903-8_8. URL `https://doi.org/10.1007/978-3-540-40903-8_8`.

Roman Andriushchenko, Milan Ceska, Sebastian Junges, and Joost-Pieter Katoen. Inductive synthesis of finite-state controllers for POMDPs. In *Thirty-Eighth Conference on Uncertainty in Artificial Intelligence, UAI 2022*, volume 180 of *Proceedings of Machine Learning Research*, pages 85–95. PMLR, 2022. URL `https://proceedings.mlr.press/v180/andriushchenko22a.html`.

Roman Andriushchenko, Alexander Bork, Milan Ceska, Sebastian Junges, Joost-Pieter Katoen, and Filip Macák. Search and explore: Symbiotic policy synthesis in POMDPs. In *Computer Aided Verification - 35th International Conference, CAV 2023*, volume 13966 of *Lecture Notes in Computer Science*, pages 113–135. Springer, 2023. doi:

10.1007/978-3-031-37709-9_6. URL https://doi.org/10.1007/978-3-031-37709-9_6.

Karl Johan Åström. Optimal control of Markov processes with incomplete state information. *Journal of Mathematical Analysis and Applications*, 10(1):174–205, 1965. doi: https://doi.org/10.1016/0022-247X(65)90154-X.

Christel Baier and Joost-Pieter Katoen. *Principles of Model Checking*. MIT Press, 2008. ISBN 978-0-262-02649-9.

Christel Baier, Marcus Daum, Clemens Dubslaff, Joachim Klein, and Sascha Klüppelholz. Energy-utility quantiles. In *NASA Formal Methods - 6th International Symposium, NFM 2014*, volume 8430 of *Lecture Notes in Computer Science*, pages 285–299. Springer, 2014. doi: 10.1007/978-3-319-06200-6_24. URL https://doi.org/10.1007/978-3-319-06200-6_24.

Leon Barrett and Srini Narayanan. Learning all optimal policies with multiple criteria. In *Machine Learning, Proceedings of the Twenty-Fifth International Conference (ICML 2008)*, volume 307 of *ACM International Conference Proceeding Series*, pages 41–47. ACM, 2008. doi: 10.1145/1390156.1390162. URL https://doi.org/10.1145/1390156.1390162.

Alexander Bork, Sebastian Junges, Joost-Pieter Katoen, and Tim Quatmann. Verification of indefinite-horizon POMDPs. In *Automated Technology for Verification and Analysis - 18th International Symposium, ATVA 2020*, volume 12302 of *Lecture Notes in Computer Science*, pages 288–304. Springer, 2020. doi: 10.1007/978-3-030-59152-6_16. URL https://doi.org/10.1007/978-3-030-59152-6_16.

Alexander Bork, Joost-Pieter Katoen, and Tim Quatmann. Under-approximating expected total rewards in POMDPs. In *Tools and Algorithms for the Construction and Analysis of Systems - 28th International Conference, TACAS 2022*, volume 13244 of *Lecture Notes in Computer Science*, pages 22–40. Springer, 2022. doi: 10.1007/978-3-030-99527-0_2. URL https://doi.org/10.1007/978-3-030-99527-0_2.

Alexander Bork, Joost-Pieter Katoen, Tim Quatmann, and Svenja Stein. Artifact for paper: Multi-cost-bounded reachability analysis of POMDPs, 2025. URL https://doi.org/10.5281/zenodo.15642232.

John L. Bresina, Ari K. Jónsson, Paul H. Morris, and Kanna Rajan. Activity planning for the mars exploration rovers. In *Proceedings of the Fifteenth International Conference on Automated Planning and Scheduling (ICAPS 2005)*, pages 40–49. AAAI, 2005. URL http://www.aaai.org/Library/ICAPS/2005/icaps05-005.php.

Iadine Chades, Josie Carwardine, Tara G. Martin, Samuel Nicol, Régis Sabbadin, and Olivier Buffet. MOMDPs: A solution for modelling adaptive management problems. In *Proceedings of the Twenty-Sixth AAAI Conference on Artificial Intelligence*, pages 267–273. AAAI Press, 2012. doi: 10.1609/AAAI.V26I1.8171. URL https://doi.org/10.1609/aaai.v26i1.8171.

Krishnendu Chatterjee, Martin Chmelik, Raghav Gupta, and Ayush Kanodia. Optimal cost almost-sure reachability in POMDPs. *Artif. Intell.*, 234:26–48, 2016. doi: 10.1016/J.ARTINT.2016.01.007. URL https://doi.org/10.1016/j.artint.2016.01.007.

Ernst Moritz Hahn and Arnd Hartmanns. A comparison of time- and reward-bounded probabilistic model checking techniques. In *Dependable Software Engineering: Theories, Tools, and Applications - Second International Symposium, SETTA 2016*, volume 9984 of *Lecture Notes in Computer Science*, pages 85–100, 2016. doi: 10.1007/978-3-319-47677-3_6. URL https://doi.org/10.1007/978-3-319-47677-3_6.

Arnd Hartmanns, Sebastian Junges, Joost-Pieter Katoen, and Tim Quatmann. Multi-cost bounded tradeoff analysis in MDP. *J. Autom. Reason.*, 64(7):1483–1522, 2020. doi: 10.1007/S10817-020-09574-9. URL https://doi.org/10.1007/s10817-020-09574-9.

Christian Hensel, Sebastian Junges, Joost-Pieter Katoen, Tim Quatmann, and Matthias Volk. The probabilistic model checker Storm. *Int. J. Softw. Tools Technol. Transf.*, 24(4):589–610, 2022. doi: 10.1007/S10009-021-00633-Z. URL https://doi.org/10.1007/s10009-021-00633-z.

Qi Heng Ho, Martin S. Feather, Federico Rossi, Zachary N Sunberg, and Morteza Lahijanian. Sound heuristic search value iteration for undiscounted POMDPs with reachability objectives. In *40th Conference on Uncertainty in Artificial Intelligence, UAI 2024*, 2024. URL https://openreview.net/forum?id=3zSiuXYtqf.

Karel Horák, Branislav Bosanský, and Krishnendu Chatterjee. Goal-HSVI: Heuristic search value iteration for goal POMDPs. In *Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence, IJCAI 2018*, pages 4764–4770. ijcai.org, 2018. doi: 10.24963/IJCAI.2018/662. URL https://doi.org/10.24963/ijcai.2018/662.

Ping Hou, William Yeoh, and Pradeep Varakantham. Solving risk-sensitive POMDPs with and without cost observations. In *Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence*, pages 3138–3144. AAAI Press, 2016. doi: 10.1609/AAAI.V30I1.10402. URL https://doi.org/10.1609/aaai.v30i1.10402.

Joshua D. Isom, Sean P. Meyn, and Richard D. Braatz. Piecewise linear dynamic programming for constrained pomdps. In *Proceedings of the Twenty-Third AAAI Conference on Artificial Intelligence, AAAI 2008*, pages 291–296. AAAI Press, 2008. URL http://www.aaai.org/Library/AAAI/2008/aaai08-046.php.

Masoumeh T. Izadi and Doina Precup. Using rewards for belief state updates in partially observable markov decision processes. In *Machine Learning: ECML 2005, 16th European Conference on Machine Learning*, volume 3720 of *Lecture Notes in Computer Science*, pages 593–600. Springer, 2005. doi: 10.1007/11564096_58. URL https://doi.org/10.1007/11564096_58.

Krishna Chaitanya Kalagarla, Dhruva Kartik, Dongming Shen, Rahul Jain, Ashutosh Nayyar, and Pierluigi Nuzzo. Optimal control of logically constrained partially observable and multiagent Markov decision processes. *IEEE Trans. Autom. Control.*, 70(1):263–277, 2025. doi: 10.1109/TAC.2024.3422213. URL https://doi.org/10.1109/TAC.2024.3422213.

Joachim Klein, Christel Baier, Philipp Chrszon, Marcus Daum, Clemens Dubslaff, Sascha Klüppelholz, Steffen Märcker, and David Müller. Advances in probabilistic model checking with PRISM: variable reordering, quantiles and weak deterministic Büchi automata. *Int. J. Softw. Tools Technol. Transf.*, 20(2):179–194, 2018. doi: 10.1007/S10009-017-0456-3. URL https://doi.org/10.1007/s10009-017-0456-3.

Hanna Kurniawati, David Hsu, and Wee Sun Lee. SARSOP: efficient point-based POMDP planning by approximating optimally reachable belief spaces. In Oliver Brock, Jeff Trinkle, and Fabio Ramos, editors, *Robotics: Science and Systems IV, Eidgenössische Technische Hochschule Zürich, Switzerland, June 25-28, 2008*. The MIT Press, 2008. doi: 10.15607/RSS.2008.IV.009. URL http://www.roboticsproceedings.org/rss04/p9.html.

Marta Z. Kwiatkowska, Gethin Norman, and David Parker. PRISM 4.0: Verification of probabilistic real-time systems. In *Computer Aided Verification - 23rd International Conference, CAV 2011*, volume 6806 of *Lecture Notes in Computer Science*, pages 585–591. Springer, 2011. doi: 10.1007/978-3-642-22110-1_47. URL https://doi.org/10.1007/978-3-642-22110-1_47.

Bruno Lacerda, David Parker, and Nick Hawes. Multi-objective policy generation for mobile robots under probabilistic time-bounded guarantees. In *Proceedings of the Twenty-Seventh International Conference on Automated Planning and Scheduling, ICAPS 2017*, pages 504–512. AAAI Press, 2017. URL https://aaai.org/ocs/index.php/ICAPS/ICAPS17/paper/view/15751.

Omid Madani, Steve Hanks, and Anne Condon. On the undecidability of probabilistic planning and related stochastic optimization problems. *Artif. Intell.*, 147(1-2):5–34, 2003. doi: 10.1016/S0004-3702(02)00378-8. URL https://doi.org/10.1016/S0004-3702(02)00378-8.

Erik Miehling, Mohammad Rasouli, and Demosthenis Teneketzis. A POMDP approach to the dynamic defense of large-scale cyber networks. *IEEE Trans. Inf. Forensics Secur.*, 13(10):2490–2505, 2018. doi: 10.1109/TIFS.2018.2819967. URL https://doi.org/10.1109/TIFS.2018.2819967.

Gethin Norman, David Parker, and Xueyi Zou. Verification and control of partially observable probabilistic systems. *Real Time Syst.*, 53(3):354–402, 2017. doi: 10.1007/S11241-017-9269-4. URL https://doi.org/10.1007/s11241-017-9269-4.

Yoshio Ohtsubo. Optimal threshold probability in undiscounted Markov decision processes with a target set. *Appl. Math. Comput.*, 149(2):519–532, 2004. doi: 10.1016/S0096-3003(03)00158-9. URL https://doi.org/10.1016/S0096-3003(03)00158-9.

Pascal Poupart, Aarti Malhotra, Pei Pei, Kee-Eung Kim, Bongseok Goh, and Michael Bowling. Approximate linear programming for constrained partially observable Markov decision processes. In *Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence*, pages 3342–3348. AAAI Press, 2015. doi: 10.1609/AAAI.V29I1.9655. URL https://doi.org/10.1609/aaai.v29i1.9655.

Martin L. Puterman. *Markov Decision Processes: Discrete Stochastic Dynamic Programming*. Wiley Series in Probability and Statistics. Wiley, 1994. doi: 10.1002/9780470316887.

Mickael Randour, Jean-François Raskin, and Ocan Sankur. Percentile queries in multi-dimensional Markov decision processes. *Formal Methods Syst. Des.*, 50(2-3):207–248, 2017. doi: 10.1007/S10703-016-0262-7. URL https://doi.org/10.1007/s10703-016-0262-7.

Stuart Russell and Peter Norvig. *Artificial Intelligence: A Modern Approach (4th Edition)*. Pearson, 2020. ISBN 9780134610993. URL http://aima.cs.berkeley.edu/.

Pedro Santana, Sylvie Thiébaux, and Brian Charles Williams. RAO*: An algorithm for chance-constrained POMDPs. In *Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence*, pages 3308–3314. AAAI Press, 2016. doi: 10.1609/AAAI.V30I1.10423. URL https://doi.org/10.1609/aaai.v30i1.10423.

Guy Shani, Joelle Pineau, and Robert Kaplow. A survey of point-based POMDP solvers. *Auton. Agents Multi Agent Syst.*, 27(1):1–51, 2013. doi: 10.1007/S10458-012-9200-2. URL `https://doi.org/10.1007/s10458-012-9200-2`.

Richard D. Smallwood and Edward J. Sondik. The optimal control of partially observable markov processes over a finite horizon. *Oper. Res.*, 21(5):1071–1088, 1973. doi: 10.1287/OPRE.21.5.1071. URL `https://doi.org/10.1287/opre.21.5.1071`.

Trey Smith and Reid G. Simmons. Heuristic search value iteration for POMDPs. In *20th Conference in Uncertainty in Artificial Intelligence, UAI 2004*, pages 520–527. AUAI Press, 2004.

Matthijs T. J. Spaan and Nikos Vlassis. A point-based POMDP algorithm for robot planning. In *Proceedings of the 2004 IEEE International Conference on Robotics and Automation, ICRA 2004*, pages 2399–2404. IEEE, 2004. doi: 10.1109/ROBOT.2004.1307420. URL `https://doi.org/10.1109/ROBOT.2004.1307420`.

Aditya Undurti and Jonathan P. How. An online algorithm for constrained pomdps. In *IEEE International Conference on Robotics and Automation, ICRA 2010*, pages 3966–3973. IEEE, 2010. URL `https://doi.org/10.1109/ROBOT.2010.5509743`.

Bo Wu, Mohamadreza Ahmadi, Suda Bharadwaj, and Ufuk Topcu. Cost-bounded active classification using partially observable markov decision processes. In *2019 American Control Conference, ACC 2019*, pages 1216–1223. IEEE, 2019. doi: 10.23919/ACC.2019.8814415. URL `https://doi.org/10.23919/ACC.2019.8814415`.

Multi-Cost-Bounded Reachability Analysis of POMDPs
(Supplementary Material)

Alexander Bork[1] **Joost-Pieter Katoen**[1] **Tim Quatmann**[1] **Svenja Stein**[1]

[1]RWTH Aachen University, Aachen, Germany

A EXAMPLE

We illustrate the unfolding of cost bounds for POMDPs to treat cost-bounded reachability probability problems by an example. Consider the POMDP $\mathcal{M}$ and 2-dimensional cost structure $\mathbf{C}$ depicted in Figure 3.

$\mathcal{M}$ contains 4 states and 2 actions α and β. The left-hand side of Figure 3 depicts the state-transition diagram of $\mathcal{M}$. For example, we have $\mathbf{P}(s_0, a)(s_0) = {}^1\!/_2$. Observations after a transition are deterministic, and the single possible observation after a transition is given next to the transition probability, so for example we have $\mathbf{O}(s_0, a, s_1)(z_0) = 1$. The way the observations are chosen means that we always observe z_0 if we enter s_0 or s_1, z_1 if we enter s_2 and z_2 if we enter t.

The cost vector of each transition is given on the right-hand side of Figure 3. In particular, we only have two transitions that do not have a cost of $\langle 0, 0\rangle$, namely $\mathbf{C}(s_1, \alpha, s_1) = \langle 1, 0\rangle$ and $\mathbf{C}(s_0, \alpha, s_0) = \langle 0, 1\rangle$.

Consider the cost bound $\mathsf{bnd} = (\mathbf{C}\ \langle \leq, > \rangle\ \langle 1, 0\rangle)$, i.e., in dimension 1, we want to collect at most 1 unit of cost, while in dimension 2, we want to collect more than 0 units. We are interested in the cost-bounded reachability probability $\mathsf{Pr}^{\mathcal{M}}_{\max}(\Diamond_{\mathsf{bnd}}\{t\})$. Satisfying the cost bounds requires to take transition $s_0 \xrightarrow{\alpha} s_0$ *at least* once while taking the self-loop $s_1 \xrightarrow{\alpha} s_1$ *at most* once.

The state-transition diagram of the fragment of the unfolding POMDP $\mathsf{un}_{\mathsf{bnd}}(\mathcal{M})$ reachable from the initial state is given in Figure 4. For illustration of the transitions in $\mathsf{un}_{\mathsf{bnd}}(\mathcal{M})$, we consider an example. Due to the bound thresholds, we get that the initial epoch t of the unfolding is $\langle 1, 0\rangle$. Thus, the initial unfolding state is $\langle s_0, \langle 1, 0\rangle\rangle$. Consider transition $s_0 \xrightarrow{\alpha} s_0$ in $\mathcal{M}$. We have $\mathbf{C}(s_0, \alpha, s_0) = \langle 0, 1\rangle$, thus the corresponding transition starting in $\langle s_0, \langle 1, 0\rangle\rangle$ changes the epoch to $\langle 1, 0\rangle \ominus \langle 0, 1\rangle = \langle 1, \bot\rangle$, resulting in the transition $\langle s_0, \langle 1, 0\rangle\rangle \xrightarrow{\alpha} \langle s_0, \langle 1, \bot\rangle\rangle$ with probability

$$\mathbf{P}_{\mathsf{un}}(\langle s_0, \langle 1, 0\rangle\rangle, \alpha)(\langle s_0, \langle 1, \bot\rangle\rangle) = \mathbf{P}(s_0, \alpha)(s_0) = {}^1\!/_2.$$

The observations resulting from $\langle s_0, \langle 1, 0\rangle\rangle \xrightarrow{\alpha} \langle s_0, \langle 1, \bot\rangle\rangle$ have the same probability as for $s_0 \xrightarrow{\alpha} s_0$ in the original POMDP, i.e., z_1 is observed with probability 1.

To analyse $\mathrm{Pr}^{\mathcal{M}}_{\max}(\Diamond_{\mathsf{bnd}}\{t\})$, we need to identify the active $\{t\}$-states. The active epochs in $\mathsf{un}_{\mathsf{bnd}}(\mathcal{M})$ are $\langle 1, \bot\rangle$ and $\langle 0, \bot\rangle$. Therefore, we have

$$\mathsf{actv}_{\mathsf{bnd}}(\{t\}) = \{\langle t, \langle 1, \bot\rangle\rangle, \langle t, \langle 0, \bot\rangle\rangle\}.$$

We can now consider the *unbounded* reachability probability

$$\mathsf{Pr}^{\mathsf{un}_{\mathsf{bnd}}(\mathcal{M})}_{\max}(\Diamond \mathsf{actv}_{\mathsf{bnd}}(\{t\})) = \mathsf{Pr}^{\mathsf{un}_{\mathsf{bnd}}(\mathcal{M})}_{\max}(\Diamond\{\langle t, \langle 1, \bot\rangle\rangle, \langle t, \langle 0, \bot\rangle\rangle\})$$

and solve the problem using methods known from the literature.

This results in a value of $\mathsf{Pr}^{\mathsf{un}_{\mathsf{bnd}}(\mathcal{M})}_{\max}(\Diamond \mathsf{actv}_{\mathsf{bnd}}(\{t\})) = {}^3\!/_8$, achieved by a policy that chooses α in the first 3 steps (in which only z_0 is observed) and then β.

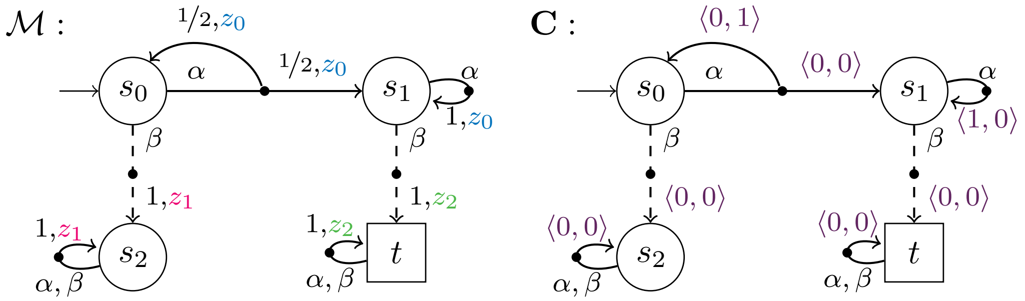


Figure 3: Example POMDP $\mathcal{M}$ and corresponding cost structure $\mathbf{C}$

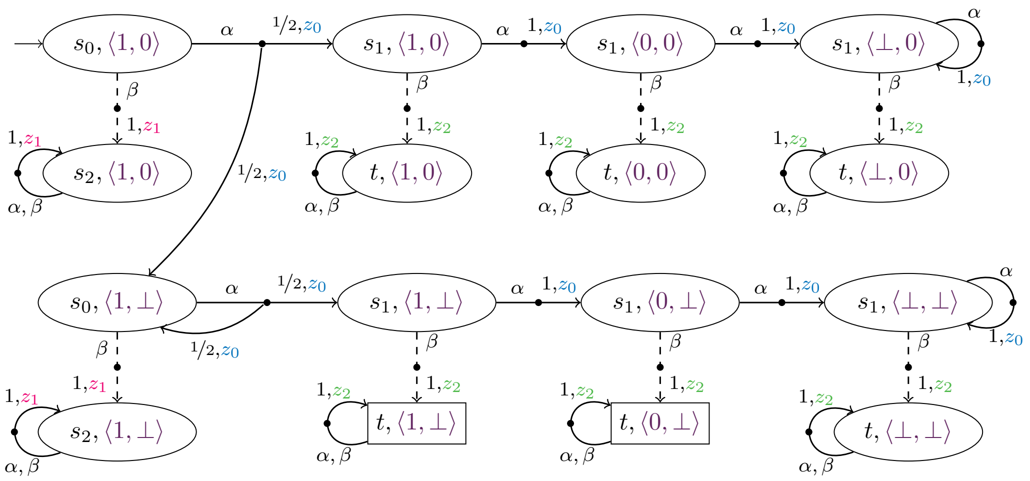

$s_0, \langle 1, 0\rangle$
$s_1, \langle 1, 0\rangle$
$s_1, \langle 0, 0\rangle$
$s_1, \langle \bot, 0\rangle$
$s_2, \langle 1, 0\rangle$
$t, \langle 1, 0\rangle$
$t, \langle 0, 0\rangle$
$t, \langle \bot, 0\rangle$
$s_0, \langle 1, \bot\rangle$
$s_1, \langle 1, \bot\rangle$
$s_1, \langle 0, \bot\rangle$
$s_1, \langle \bot, \bot\rangle$
$s_2, \langle 1, \bot\rangle$
$t, \langle 1, \bot\rangle$
$t, \langle 0, \bot\rangle$
$t, \langle \bot, \bot\rangle$
α
β
α, β
$1/2, z_0$
$1, z_0$
$1, z_1$
$1, z_2$

Figure 4: Reachable fragment of the cost-bounded unfolding $\mathsf{un}_{\mathsf{bnd}}(\mathcal{M})$ for example POMDP $\mathcal{M}$ with respect to cost bound $\mathsf{bnd} = (\mathbf{C}\ \langle \leq, > \rangle\ \langle 1, 0\rangle)$

By Theroem 2 we therefore get that

$$\mathsf{Pr}^{\mathcal{M}}_{\max}(\Diamond_{\mathsf{bnd}}\{t\}) = \mathsf{Pr}^{\mathsf{un}_{\mathsf{bnd}}(\mathcal{M})}_{\max}(\Diamond \mathsf{actv}_{\mathsf{bnd}}(\{t\})) = 3/8.$$

This example also showcases why a naïve unfolding approach that directly encodes the collected costs in the state space (and not the costs that remain until the bound changes its status) is inappropriate. In such a naïve unfolding, we have infinitely many copies of every state in the POMDP, and in particular there are infinitely many reachable target states for the unbounded reachability problem. In contrast, our unfolding results in a finite POMDP which can be treated with standard methods.

B DETAILED EXPLANATION OF LEVEL UNFOLDING POMDP

We explain the construction of the level unfolding POMDP (Def. 9). For the remainder of this section, fix a POMDP $\mathcal{M} = \langle M, Z, \mathbf{O}\rangle$ with $M = \langle S, Act, \mathbf{P}, s_{init}\rangle$, k-dimensional cost bound $(\mathbf{C} \bowtie \mathsf{t})$ and level function $lvl_{\mathsf{d}} : \mathbb{N}^k \to \mathbb{N}^k$ with $\mathsf{d} \in (\mathbb{N} \setminus \{0\})^k$.

The core idea of the level unfolding is that we keep track of jumps in the level using the observations of the unfolding POMDP. An observation also stores by how many levels we jump up when taking a transition. To keep track of when an incurred cost causes a level jump, we need to do bookkeeping in between jumps. We do this by storing the remaining cost until the next jump in the state space.

The state space $S_{lvl_{\mathsf{d}}}$ is defined such that for there is a copy of each state for each possible combination of costs with which we stay in the current level in any dimension. We keep track of this using a set of vectors $\ell \in \mathbb{N}$ where each entry $\ell[i]$ indicates the amount of cost which is allowed to be collected to stay in the current level of dimension i. We call this cost the *remainder*. Formally, the state space is defined as:

$$S_{lvl_{\mathsf{d}}} := S \times \{\ell \in \mathbb{N} \mid \forall 1 \leq i \leq k : 0 \leq \ell[i] < \mathsf{d}[i]\}$$

The transitions $\mathbf{P}_{lvl_{\mathsf{d}}}$ of the unfolding are defined such that for two states in the unfolding, the transition probability is the same as for the corresponding states in the original POMDP if their remainder vectors are compatible. In particular, that means that the transition models the correct transformation of the remainder. This is the case if for each dimension, the new remainder is the old remainder minus the cost of the transition, modulo $\mathsf{d}[i]$, effectively modelling that a level jump occurs every time a total of $\mathsf{d}[i]$ costs has been incurred and the remainder stores the progress in the current level. We get:

$$\mathbf{P}_{lvl_{\mathsf{d}}}(\langle s, \ell\rangle, a, \langle s', \ell'\rangle) := \begin{cases} \mathbf{P}(s, a, s') & \text{if } \forall 1 \leq i \leq k : \ell'[i] = \ell[i] - \mathbf{C}(s, a, s')[i] \mod \mathsf{d}[i], \\ 0 & \text{otherwise.} \end{cases}$$

In the unfolding, the observations are defined such that in addition to the observations of the original POMDP, we keep track of the level changes that occur when taking transitions. Thus, we consider copies of the original observations for each vector j of possible level changes in one step. In particular, we observe that in dimension i, in one step a transition $s \xrightarrow{a} s'$ can increase the level by at most $\left\lceil \frac{\mathbf{C}(s,a,s')[i]}{\mathsf{d}[i]} \right\rceil$. Thus it suffices to consider vectors j where the value in each dimension is at most $\left\lceil \frac{\max_{s,s' \in S, a \in Act} \mathbf{C}(s,a,s')[i]}{\mathsf{d}[i]} \right\rceil$, resulting in a finite set of observations. In particular, we get:

$$Z_{lvl_{\mathsf{d}}} := Z \times \left\{ \mathsf{j} \in \mathbb{N}^k \mid \forall 1 \leq i \leq k : 0 \leq \mathsf{j}[i] \leq \left\lceil \frac{\max_{s,s' \in S, a \in Act} \mathbf{C}(s, a, s')[i]}{\mathsf{d}[i]} \right\rceil \right\}$$

$\mathbf{O}_{lvl_{\mathsf{d}}}$, i.e., the observation function of the unfolding, is defined such that for every transition, the probability of observing a new observation $\langle z, \mathsf{j}\rangle$ after a transition is the same as the probability to observe z after taking the corresponding transition in the original POMDP exactly if j corresponds to the correct jumps in level in all dimensions. To ensure this, in addition to the costs of the transition, we need to consider the remainders ℓ in the origin state $\langle s, \ell\rangle$ as the smaller the the value $\ell[i]$, i.e., the deeper we already are in the current level, the fewer costs we need to collect to jump to the next level. This results in the following definition:

$$\mathbf{O}_{lvl_{\mathsf{d}}}(\langle s, \ell\rangle, a, \langle s', \ell'\rangle)(\langle z, \mathsf{j}\rangle) := \begin{cases} \mathbf{O}(s, a, s')(z) & \text{if } \forall 1 \leq i \leq k : \mathsf{j}[i] = \left\lceil \frac{\mathbf{C}(s,a,s')[i] - \ell[i]}{\mathsf{d}[i]} \right\rceil, \\ 0 & \text{otherwise.} \end{cases}$$

Finally, as the initial state we consider the copy of the original initial state s_{init} where all remainders are 0, i.e., $\langle s_{init}, \langle 0, \ldots, 0\rangle\rangle$. This captures the behaviour of the level function lvl_{d} where if we consider the input to be a vector of total incurred costs, as soon as *some* non-zero cost is collected in a dimension, the level jumps from 0 to a higher value.

Using the above components, we can define a level unfolding POMDP with respect to lvl_{d} as $lvl_{\mathsf{d}}(\mathcal{M}) = \langle M_{lvl_{\mathsf{d}}}, Z_{lvl_{\mathsf{d}}}, \mathbf{O}_{lvl_{\mathsf{d}}}\rangle$ with $M_{lvl_{\mathsf{d}}} = \langle S_{lvl_{\mathsf{d}}}, Act, \mathbf{P}_{lvl_{\mathsf{d}}}, \langle s_{init}, \langle 0, \ldots, 0\rangle\rangle\rangle$.

C PROOFS FOR MAIN RESULTS

C.1 DETAILS ON PATH PROBABILITY MEASURE

As a basis for further theoretical results, we recap the definition of the probability measure for paths.

Given POMDP $\mathcal{M} = \langle M, Z, \mathbf{O}\rangle$ with underlying MDP $M = \langle S, Act, \mathbf{P}, s_{init}\rangle$ and a policy $\sigma \in \Sigma^{\mathcal{M}}$, we define the functions

$$\sigma_a(\tau) := \begin{cases} 1 & \text{if } \sigma(\tau) = a, \\ 0 & \text{otherwise} \end{cases}$$

indicating if policy σ chooses action $a \in Act$ when trace τ is observed.

Towards the probability measure for paths, we observe that the probability of a path $\hat{\pi} = s_0 a_1 \dots s_n$ and an observation trace of compatible length $\tau = z_1 \dots z_n$ occurring under a policy σ is

$$P^{\mathcal{M}}_{\sigma}(\hat{\pi} \wedge \tau) = \prod_{i=1}^{n} \mathbf{P}(s_{i-1}, a_i, s_i) \cdot \sigma_{a_i}(\tau[..(i-1)]) \cdot \mathbf{O}(s_{i-1}, a_i, s_i)(z_i)$$

where

$$\tau[..k] := \begin{cases} z_1 \dots z_k & \text{if } k > 0, \\ \varepsilon & \text{otherwise.} \end{cases}$$

The overall probability of a a path is then the probability that it occurs with *any* observation trace. We get that overall for a finite path $\hat{\pi} = s_0 a_1 \dots s_n \in \mathit{Paths}^{\mathcal{M}}_{\mathrm{fin}}$, the probability of $\hat{\pi}$ under policy σ is

$$\mathsf{Pr}^{\mathcal{M}}_{\sigma}(\{\hat{\pi}\}) = \sum_{z_1 \dots z_n \in \mathit{ObsTraces}^{\mathcal{M}}} P^{\mathcal{M}}_{\sigma}(\hat{\pi} \wedge z_1 \dots z_n)$$

Using the probability for a finite path, the probability measure for infinite paths and by extension (measurable) sets of infinite paths is defined using the standard cylinder set construction. We refer to Baier and Katoen [2008] for details.

In the following, we identify LTL-like formulae for reachability with the sets of path they describe, i.e., for a set $T \subseteq S$ and $(\mathbf{C} \bowtie \mathsf{t})$, we define

$$\lozenge_{\mathbf{C}\bowtie\mathsf{t}} T := \big\{\pi \in \mathit{Cyl}(\hat{\pi}) \bigm| \mathit{last}(\hat{\pi}) \in T \text{ and } (\mathbf{C} \bowtie \mathsf{t}) \text{ is active for } \hat{\pi}\big\}$$

where $\mathit{Cyl}(\hat{\pi})$ is the set of infinite extensions of finite path $\hat{\pi}$.

C.2 PROOF OF THEOREM 2

Theorem 2. *Given a POMDP $\mathcal{M}$, set $T \subseteq S$ and cost bound $(\mathbf{C} \bowtie \mathsf{t})$, it holds that for all policies $\sigma \in \Sigma^{\mathcal{M}}$:*

$$\mathsf{Pr}^{\mathcal{M}}_{\sigma}\left(\lozenge_{\mathbf{C}\bowtie\mathsf{t}}\, T\right) = \mathsf{Pr}^{\mathsf{un}_{\mathbf{C}\bowtie\mathsf{t}}(\mathcal{M})}_{\sigma}\left(\lozenge\ \mathsf{actv}_{\mathbf{C}\bowtie\mathsf{t}}(T)\right)$$

Proof. Let $\mathcal{M} = \langle M, Z, \mathbf{O}\rangle$ with $M = \langle S, \mathit{Act}, \mathbf{P}, s_{init}\rangle$. Furthermore, $\mathsf{un}_{\mathbf{C}\bowtie\mathsf{t}}(\mathcal{M}) = \langle \mathsf{un}_{\mathbf{C}\bowtie\mathsf{t}}(M), Z, \mathbf{O}_{\mathsf{un}}\rangle$ with $\mathsf{un}_{\mathbf{C}\bowtie\mathsf{t}}(M) = \langle S \times \mathsf{E}_k(\mathsf{t}), \mathit{Act}, \mathbf{P}_{\mathsf{un}}, \langle s_{init}, \mathsf{t}\rangle\rangle$. We define a mapping

$$f : \mathit{Paths}^{\mathsf{un}_{\mathbf{C}\bowtie\mathsf{t}}(\mathcal{M})} \to \mathit{Paths}^{\mathcal{M}}$$

with

$$f(\langle s_0, \mathsf{e}_0\rangle a_1 \langle s_1, \mathsf{e}_1\rangle \dots \langle s_n, \mathsf{e}_n\rangle) := s_0 a_1 \dots s_n,$$

i.e., $f(\hat{\pi})$ is the path resulting from dropping the epoch component from $\hat{\pi}$. We show that f is bijective:

- *f is injective:* Consider two (finite, initial) paths

  $$\hat{\pi} = \langle s_0, \mathsf{e}_0\rangle a_1 \langle s_1, \mathsf{e}_1\rangle a_2 \dots \langle s_n, \mathsf{e}_n\rangle$$

  and

  $$\hat{\pi}' = \langle s'_0, \mathsf{e}'_0\rangle a'_1 \langle s'_1, \mathsf{e}'_1\rangle a'_2 \dots \langle s'_m, \mathsf{e}'_m\rangle$$

  of $\mathsf{un}_{\mathbf{C}\bowtie\mathsf{t}}(\mathcal{M})$ with $\hat{\pi} \neq \hat{\pi}'$. By definition, both paths start at the initial state $\langle s_{init}, \mathsf{t}\rangle$. We distinguish two cases:
  - If $\hat{\pi}$ and $\hat{\pi}'$ have different lengths ($n \neq m$), then $|f(\hat{\pi})| = |\hat{\pi}| \neq |\hat{\pi}'| = |f(\hat{\pi}')|$ and thus $f(\hat{\pi}) \neq f(\hat{\pi}')$.
  - Otherwise, let $0 < i \leq n = m$ be the first index where the paths disagree, i.e., $\langle s_{i-1}, \mathsf{e}_{i-1}\rangle = \langle s'_{i-1}, \mathsf{e}'_{i-1}\rangle$ and $a_i \neq a'_i$, $s_i \neq s'_i$, or $\mathsf{e}_i \neq \mathsf{e}'_i$. If $a_i = a'_i$ and $s_i = s'_i$, we get $\mathsf{e}_i = \mathsf{e}_{i-1} \ominus \mathbf{C}(s_{i-1}, a_i, s_i) = \mathsf{e}'_i$. Therefore, either $a_i \neq a'_i$ or $s_i \neq s'_i$ must hold and we immediately get $f(\hat{\pi}) \neq f(\hat{\pi}')$.
- *f is surjective:* we show that for all $\tilde{\pi} \in \mathit{Paths}^{\mathcal{M}}$, there exists $\hat{\pi} \in \mathit{Paths}^{\mathsf{un}_{\mathbf{C}\bowtie\mathsf{t}}(\mathcal{M})}$ such that $f(\hat{\pi}) = \tilde{\pi}$. Let $\tilde{\pi} = s_0 a_1 s_1 a_2 \dots s_n \in \mathit{Paths}^{\mathcal{M}}$. Consider

  $$\hat{\pi} = \langle s_0, \mathsf{t}\rangle a_1 \langle s_1, \mathsf{t} \ominus \mathbf{C}(s_0, a_1, s_1)\rangle \dots \langle s_n, (\dots(\mathsf{t} \ominus \mathbf{C}(s_0, a_1, s_1)) \ominus \dots) \ominus \mathbf{C}(s_{n-1}, a_n, s_n)\rangle$$

  where $\hat{\pi} \in \mathit{Paths}^{\mathsf{un}_{\mathbf{C}\bowtie\mathsf{t}}(\mathcal{M})}$ by definition of the unfolding MDP.
  We have $f(\hat{\pi}) = \tilde{\pi}$, thus f is surjective.

Therefore, f is bijective. Next, recall that $\mathcal{M}$ and $\mathsf{un}_{\mathbf{C}\bowtie\mathsf{t}}(\mathcal{M})$ share the same set of policies $\Sigma^{\mathcal{M}} = \Sigma^{\mathsf{un}_{\mathbf{C}\bowtie\mathsf{t}}(\mathcal{M})}$.

We show that f preserves the probability of measurable sets of (infinite) paths under a given policy $\sigma \in \Sigma^{\mathcal{M}}$. As those sets can be constructed from cylinder sets of finite paths, we can focus on finite paths.

We show that

$$\mathsf{Pr}_{\sigma}^{\mathsf{un}_{\mathbf{C}\bowtie\mathsf{t}}(\mathcal{M})}\left(\{\hat{\pi}\}\right) = \mathsf{Pr}_{\sigma}^{\mathcal{M}}\left(\{f(\hat{\pi})\}\right)$$

$$\begin{aligned}
\mathsf{Pr}_{\sigma}^{\mathsf{un}_{\mathbf{C}\bowtie\mathsf{t}}(\mathcal{M})}\left(\{\hat{\pi}\}\right) &= \sum_{z_1\ldots z_n \in \mathit{ObsTraces}^{\mathsf{un}_{\mathbf{C}\bowtie\mathsf{t}}(\mathcal{M})}} P_{\sigma}^{\mathsf{un}_{\mathbf{C}\bowtie\mathsf{t}}(\mathcal{M})}(\hat{\pi} \wedge z_1 \ldots z_n) \\
&= \sum_{z_1\ldots z_n \in \mathit{ObsTraces}^{\mathsf{un}_{\mathbf{C}\bowtie\mathsf{t}}(\mathcal{M})}} \prod_{i=1}^{n} \mathbf{P}_{\mathsf{un}}(\langle s_{i-1}, \mathsf{e}_{i-1}\rangle, a_i, \langle s_i, \mathsf{e}_i\rangle) \cdot \\
&\qquad \sigma_{a_i}(\tau[..(i-1)]]) \cdot \mathbf{O}(\langle s_{i-1}, \mathsf{e}_{i-1}\rangle, a_i, \langle s_i, \mathsf{e}_i\rangle)(z_i) \\
&= \sum_{z_1\ldots z_n \in \mathit{ObsTraces}^{\mathcal{M}}} \prod_{i=1}^{n} \mathbf{P}(s_{i-1}, a_i, s_i) \cdot \\
&\qquad \sigma_{a_i}(\tau[..(i-1)]]) \cdot \mathbf{O}(s_{i-1}, a_i, s_i)(z_i) \\
&= \sum_{z_1\ldots z_n \in \mathit{ObsTraces}^{\mathcal{M}}} P_{\sigma}^{\mathcal{M}}(f(\hat{\pi}) \wedge z_1 \ldots z_n) \\
&= \mathsf{Pr}_{\sigma}^{\mathcal{M}}\left(\{f(\hat{\pi})\}\right)
\end{aligned}$$

It remains to show that f correctly transforms the set of paths $\Diamond\ \mathsf{actv}_{\mathbf{C}\bowtie\mathsf{t}}(T)$, i.e., $\Diamond_{\mathbf{C}\bowtie\mathsf{t}}\, T = \{f(\pi) \mid \pi \in \Diamond\ \mathsf{actv}_{\mathbf{C}\bowtie\mathsf{t}}(T)\}$.

Let $\pi \in \Diamond\ \mathsf{actv}_{\mathbf{C}\bowtie\mathsf{t}}(T)$, i.e., $\pi \in \mathit{Cyl}(\hat{\pi})$ for a $\hat{\pi} = \langle s_0, \mathsf{e}_0\rangle a_1 \ldots \langle s_n, \mathsf{e}_n\rangle$ such that $\mathit{last}(\hat{\pi}) = \langle s_n, \mathsf{e}_n\rangle \in \mathsf{actv}_{\mathbf{C}\bowtie\mathsf{t}}(T) = \{\langle t, \mathsf{e}\rangle \in S \times \mathsf{E}_k(\mathsf{t}) \mid t \in T \wedge \mathsf{actv}_{\mathbf{C}\bowtie\mathsf{t}}(\mathsf{e}) = 1\}$. Consider $f(\hat{\pi}) = s_0 a_1 \ldots s_n$. By definition we have $s_n \in T$.

Consider further an arbitrary, but fixed dimension $1 \leq j \leq k$ of bound $(\mathbf{C} \bowtie \mathsf{t})$. We distinguish two cases:

$\underline{\bowtie[j] = (\leq)}$: as $\langle s_n, \mathsf{e}_n\rangle \in \mathsf{actv}_{\mathbf{C}\bowtie\mathsf{t}}(T)$, we have $\mathsf{actv}_{\mathbf{C}\bowtie\mathsf{t}}(\mathsf{e}_n) = 1$. Thus, $\mathsf{e}_n[j] \in \mathbb{N}$ and in particular $\mathsf{e}_n[j] \neq \bot$. By definition of $\mathbf{P}_{\mathsf{un}}$, we then know that for all transitions $\langle s_i, \mathsf{e}_i\rangle \xrightarrow{a_{i+1}} \langle s_{i+1}, \mathsf{e}_{i+1}\rangle$ along $\hat{\pi}$, $\mathsf{e}_i[j] - \mathbf{C}(s_i, a_{i+1}, s_{i+1})[j] \geq 0$ and therefore also that $\mathsf{t}[j] - \sum_{i=1}^{n} \mathbf{C}(s_{i-1}, a_i, s_i)[j] \geq 0$. Furthermore,

$$\mathsf{t}[j] - \sum_{i=1}^{n} \mathbf{C}(s_{i-1}, a_i, s_i)[j] = \mathsf{t}[j] - \mathsf{cost}_{\mathbf{C}}(f(\hat{\pi}))[j]$$

so $\mathsf{t}[j] - \mathsf{cost}_{\mathbf{C}}(f(\hat{\pi}))[j] \geq 0$ and $\mathsf{t}[j] \geq \mathsf{cost}_{\mathbf{C}}(f(\hat{\pi}))[j]$, so bound $(\mathbf{C} \bowtie \mathsf{t})$ is active in dimension j for $f(\hat{\pi})$.

$\underline{\bowtie[j] = (>)}$: as $\langle s_n, \mathsf{e}_n\rangle \in \mathsf{actv}_{\mathbf{C}\bowtie\mathsf{t}}(T)$, we have $\mathsf{actv}_{\mathbf{C}\bowtie\mathsf{t}}(\mathsf{e}_n) = 1$. Thus, $\mathsf{e}_n[j] = \bot$. By definition of $\mathbf{P}_{\mathsf{un}}$, we then know that there is a transition $\langle s_i, \mathsf{e}_i\rangle \xrightarrow{a_{i+1}} \langle s_{i+1}, \mathsf{e}_{i+1}\rangle$ along $\hat{\pi}$ such that $\mathsf{e}_i[j] - \mathbf{C}(s_i, a_{i+1}, s_{i+1})[j] < 0$.

Therefore also $\mathsf{t}[j] - \sum_{i=1}^{n} \mathbf{C}(s_{i-1}, a_i, s_i)[j] < 0$ holds. Furthermore, by the argument from before, we then get $\mathsf{t}[j] < \mathsf{cost}_{\mathbf{C}}(f(\hat{\pi}))[j]$, so bound $(\mathbf{C} \bowtie \mathsf{t})$ is active in dimension j for $f(\hat{\pi})$.

Thus, for $f(\hat{\pi})$, we have $\mathit{last}(f(\hat{\pi})) \in T$ and for all $1 \leq j \leq k$, $(\mathbf{C} \bowtie \mathsf{t})$ is active for $f(\hat{\pi})$ in dimension j, so for all $f(\pi) \in \mathit{Cyl}(f(\hat{\pi}))$, $f(\pi) \in \Diamond_{\mathbf{C}\bowtie\mathsf{t}}\, T$.

Now let $\pi_{\mathcal{M}} \in \Diamond_{\mathbf{C}\bowtie\mathsf{t}}\, T$. By bijectivity of f and with similar arguments as before, we get $f^{-1}(\pi_{\mathcal{M}}) \in \Diamond\ \mathsf{actv}_{\mathbf{C}\bowtie\mathsf{t}}(T)$.

Overall we get $\Diamond_{\mathbf{C}\bowtie\mathsf{t}}\, T = \{f(\pi) \mid \pi \in \Diamond\ \mathsf{actv}_{\mathbf{C}\bowtie\mathsf{t}}(T)\}$.

We conclude that

$$\mathsf{Pr}_{\sigma}^{\mathcal{M}}\left(\Diamond_{\mathbf{C}\bowtie\mathsf{t}}\, T\right) = \mathsf{Pr}_{\sigma}^{\mathsf{un}_{\mathbf{C}\bowtie\mathsf{t}}(\mathcal{M})}\left(\Diamond\ \mathsf{actv}_{\mathbf{C}\bowtie\mathsf{t}}(T)\right)$$

for all policies $\sigma \in \Sigma^{\mathcal{M}}$. □

C.3 PROOF OF THEOREM 3

Theorem 3. *For cost bound $(\mathbf{C} \bowtie \mathsf{t})$ and cost-aware POMDP $\mathcal{M}$ we have*

$$\mathit{bel}(\mathsf{un}_{\mathbf{C}\bowtie\mathsf{t}}(\mathcal{M})) \cong \mathsf{un}_{\mathit{bel}(\mathbf{C}\bowtie\mathsf{t})}(\mathit{bel}(\mathcal{M})).$$

Let $\mathcal{M} = \langle M, Z, \mathbf{O}\rangle$ with $M = \langle S, \mathit{Act}, \mathbf{P}, s_{\mathit{init}}\rangle$ be a cost-aware POMDP with respect to k-dimensional cost structure $\mathbf{C}$. Furthermore, let

$$\mathit{bel}(\mathsf{un}_{\mathbf{C}\bowtie\mathsf{t}}(\mathcal{M})) = \langle \mathcal{B}_{\mathsf{un}_{\mathbf{C}\bowtie\mathsf{t}}(\mathcal{M})}, \mathit{Act}, \mathbf{P}^{B}_{\mathsf{un}}, \langle z_{\mathit{init}}, \mu_{\mathit{init}}\rangle\rangle$$

with $\mu_{\mathit{init}}(\langle s_{\mathit{init}}, \mathsf{t}\rangle) = 1$ and

$$\mathsf{un}_{\mathit{bel}(\mathbf{C}\bowtie\mathsf{t})}(\mathit{bel}(\mathcal{M})) = \langle \mathcal{B}^{\mathcal{M}} \times \mathsf{E}_k(\mathsf{t}), \mathit{Act}, \mathbf{P}^{\mathit{bel}(\mathcal{M})}_{\mathsf{un}}, \langle b_{\mathit{init}}, \mathsf{t}\rangle\rangle.$$

To show that $\mathit{bel}(\mathsf{un}_{\mathbf{C}\bowtie\mathsf{t}}(\mathcal{M})) \cong \mathsf{un}_{\mathit{bel}(\mathbf{C}\bowtie\mathsf{t})}(\mathit{bel}(\mathcal{M}))$, we will show that we can identify each belief $\langle z, \mu\rangle$ over state-epoch tuples in $\mathit{bel}(\mathsf{un}_{\mathbf{C}\bowtie\mathsf{t}}(\mathcal{M}))$ with a belief-epoch tuple $\langle\langle z, b\rangle, \mathsf{e}\rangle$ in $\mathsf{un}_{\mathit{bel}(\mathbf{C}\bowtie\mathsf{t})}(\mathit{bel}(\mathcal{M}))$ by moving the probability distribution inside and the epoch out of the belief and vice versa.

We define a mapping $f : \mathcal{B}_{\mathsf{un}_{\mathbf{C}\bowtie\mathsf{t}}(\mathcal{M})} \to \mathcal{B}^{\mathcal{M}} \times \mathsf{E}_k(\mathsf{t})$ with

$$f(\langle z, \mu\rangle) := \langle\langle z, \mu_f\rangle, \mathsf{e}\rangle$$

where $\mu_f(s) := \mu(\langle s, \mathsf{e}\rangle)$ and e is the epoch of some $\langle s, \mathsf{e}\rangle \in \mu$.

For f to be well-defined, the epoch e must be unique for a fixed belief $\langle z, \mu\rangle$. As already stated in the main paper, this is indeed the case. We formalise the claim in the following lemma.

Lemma 1. *Given a belief $b = \langle z, \mu\rangle \in \mathcal{B}_{\mathsf{un}_{\mathbf{C}\bowtie\mathsf{t}}(\mathcal{M})}$, for all states $\langle s, \mathsf{e}\rangle, \langle s', \mathsf{e}'\rangle \in \mu$ it holds that $\mathsf{e} = \mathsf{e}'$, i.e., the epoch for all states in b is unique.*

Proof. We show the claim by induction over the structure of

$$\mathcal{B}_{\mathsf{un}_{\mathbf{C}\bowtie\mathsf{t}}(\mathcal{M})} = \lim_{n\to\infty} \mathcal{B}^{n}_{\mathsf{un}_{\mathbf{C}\bowtie\mathsf{t}}(\mathcal{M})}.$$

For $n = 0$, we have $\mathcal{B}^{0}_{\mathsf{un}_{\mathbf{C}\bowtie\mathsf{t}}(\mathcal{M})} = \{\langle z_{\mathit{init}}, \mu_{\mathit{init}}\rangle\}$ with $\mu_{\mathit{init}}(\langle s_{\mathit{init}}, \mathsf{t}\rangle) = 1$, so the claim holds.

Assume the claim holds for an arbitrary but fixed $n \in \mathbb{N}$. We show that then the claim also holds for $n + 1$. Consider $b = \langle z, \mu\rangle \in \mathcal{B}^{n+1}_{\mathsf{un}_{\mathbf{C}\bowtie\mathsf{t}}(\mathcal{M})}$. If $b \in \mathcal{B}^{n}_{\mathsf{un}_{\mathbf{C}\bowtie\mathsf{t}}(\mathcal{M})}$, the claim holds by assumption. Otherwise we have that $b = \langle z, \mu\rangle \in \mathcal{B}^{n+1}_{\mathsf{un}_{\mathbf{C}\bowtie\mathsf{t}}(\mathcal{M})} \setminus \mathcal{B}^{n}_{\mathsf{un}_{\mathbf{C}\bowtie\mathsf{t}}(\mathcal{M})}$. Thus, $b = \mathsf{succ}(b', a, z)$ for some $b' = \langle z', \mu'\rangle \in \mathcal{B}^{n}_{\mathsf{un}_{\mathbf{C}\bowtie\mathsf{t}}(\mathcal{M})}$ and $a \in \mathit{Act}$.

Towards a contradiction, assume that there are two states $\langle s, \mathsf{e}\rangle, \langle s', \mathsf{e}'\rangle \in \mu$ such that $\mathsf{e} \neq \mathsf{e}'$. Let $\langle q, \tilde{\mathsf{e}}\rangle, \langle q', \tilde{\mathsf{e}}\rangle \in \mu'$ where $\tilde{\mathsf{e}}$ denotes the unique epoch of b' and

$$\begin{aligned} &\mathbf{P}_{\mathsf{un}}(\langle q, \tilde{\mathsf{e}}\rangle, a, \langle s, \mathsf{e}\rangle) > 0, &&\mathbf{P}_{\mathsf{un}}(\langle q', \tilde{\mathsf{e}}\rangle, a, \langle s', \mathsf{e}'\rangle) > 0,\\ &\mathbf{O}_{\mathsf{un}}(\langle q, \tilde{\mathsf{e}}\rangle, a, \langle s, \mathsf{e}\rangle)(z) > 0, &&\mathbf{O}_{\mathsf{un}}(\langle q', \tilde{\mathsf{e}}\rangle, a, \langle s', \mathsf{e}'\rangle)(z) > 0, \end{aligned}$$

i.e., the states $\langle q, \tilde{\mathsf{e}}\rangle$ and $\langle q', \tilde{\mathsf{e}}\rangle$ respectively contribute to the probability of $\langle s, \mathsf{e}\rangle$ and $\langle s', \mathsf{e}'\rangle$ in μ.

By definition of the unfolding, we also get $\mathbf{O}(q, a, s)(z) > 0$ and $\mathbf{O}(q', a, s')(z) > 0$. As $\mathsf{e} \neq \mathsf{e}'$, we also have that $\tilde{\mathsf{e}} \ominus \mathbf{C}(q, a, s) \neq \tilde{\mathsf{e}} \ominus \mathbf{C}(q', a, s')$, implying that $\mathbf{C}(q, a, s) \neq \mathbf{C}(q', a, s')$. This, however, contradicts the cost-awareness of $\mathcal{M}$.

Therefore the claim also holds for $n + 1$.

We conclude the uniqueness of the epoch in belief b. □

Lemma 2. *The mapping $f : \mathcal{B}_{\mathsf{un}_{\mathbf{C}\bowtie\mathsf{t}}(\mathcal{M})} \to \mathcal{B}^{\mathcal{M}} \times \mathsf{E}_k(\mathsf{t})$ as defined above is well-defined.*

Proof. By Lemma 1, we have that the epoch e used in the mapping is non-ambiguous.

It remains to show that indeed $f(b) \in \mathcal{B}^{\mathcal{M}} \times \mathsf{E}_k(\mathsf{t})$ for all $b \in \mathcal{B}_{\mathsf{un}_{\mathbf{C}\bowtie\mathsf{t}}(\mathcal{M})}$.

We use induction over the structure of $\mathcal{B}_{\mathsf{un}_{\mathbf{C}\bowtie\mathsf{t}}(\mathcal{M})}$. In particular, we show that if for a belief $b \in \mathcal{B}_{\mathsf{un}_{\mathbf{C}\bowtie\mathsf{t}}(\mathcal{M})}$, $f(b) \in \mathcal{B}^{\mathcal{M}} \times \mathsf{E}_k(\mathsf{t})$ holds, then for all successors $b' \in \mathbf{P}^{B}_{\mathsf{un}}(b,a)$ for some $a \in \mathit{Act}$, $f(b') \in \mathbf{P}^{\mathit{bel}(\mathcal{M})}_{\mathsf{un}}(f(b),a)$.

As the base case, consider $\langle z_{init}, \mu_{init}\rangle \in \mathcal{B}_{\mathsf{un}_{\mathbf{C}\bowtie\mathsf{t}}(\mathcal{M})}$ with

$$\mu_{init}(\langle s_{init}, \mathsf{t}\rangle) = 1.$$

We get $f(\langle z_{init}, \mu_{init}\rangle) = \langle b_{init}, \mathsf{t}\rangle$.

$\langle b_{init}, \mathsf{t}\rangle \in \mathcal{B}^{\mathcal{M}} \times \mathsf{E}_k(\mathsf{t})$ by definition of $\mathcal{B}^{\mathcal{M}}$ and $\mathsf{E}_k(\mathsf{t})$, thus

$$f(\langle z_{init}, \mu_{init}\rangle) \in \mathcal{B}^{\mathcal{M}} \times \mathsf{E}_k(\mathsf{t}).$$

Now let $b = \langle z, \mu\rangle \in \mathcal{B}_{\mathsf{un}_{\mathbf{C}\bowtie\mathsf{t}}(\mathcal{M})}$ such that the claim holds, i.e.,

$$f(b) \in \mathcal{B}^{\mathcal{M}} \times \mathsf{E}_k(\mathsf{t}).$$

Let e denote the unique epoch of all $\langle s, \mathsf{e}\rangle \in \mu$ and let $f(b) = \langle\langle z, \mu_f\rangle, \mathsf{e}\rangle$ with $\mu_f(s) = \mu(\langle s, \mathsf{e}\rangle)$. Let $\beta = \langle z, \mu_f\rangle$.

We first show that $P(z'|b,a) = P(z'|\beta,a)$.

$$\begin{aligned}
P(z'|b,a) &= \sum_{\langle s,\mathsf{e}\rangle\in\mu} \mu(\langle s,\mathsf{e}\rangle)\cdot \sum_{\langle s',\mathsf{e}'\rangle\in S\times \mathsf{E}^k(\mathsf{t})} \mathbf{P}_{\mathsf{un}}(\langle s,\mathsf{e}\rangle,a)(\langle s',\mathsf{e}'\rangle)\cdot \mathbf{O}_{\mathsf{un}}(\langle s,\mathsf{e}\rangle,a,\langle s',\mathsf{e}'\rangle)(z')\\
&= \sum_{\langle s,\mathsf{e}\rangle\in\mu} \mu(\langle s,\mathsf{e}\rangle)\cdot \sum_{\langle s',\mathsf{e}'\rangle\in S\times \mathsf{E}^k(\mathsf{t})} \mathbf{P}_{\mathsf{un}}(\langle s,\mathsf{e}\rangle,a)(\langle s',\mathsf{e}'\rangle)\cdot \mathbf{O}(s,a,s')(z')\\
&= \sum_{\langle s,\mathsf{e}\rangle\in\mu} \mu(\langle s,\mathsf{e}\rangle)\cdot \sum_{s'\in S} \mathbf{P}(s,a)(s')\cdot \mathbf{O}(s,a,s')(z')\\
&= \sum_{s\in\mu_f} \mu_f(s)\cdot \sum_{s'\in S} \mathbf{P}(s,a)(s')\cdot \mathbf{O}(s,a,s')(z')\\
&= P(z'|\beta,a)
\end{aligned}$$

Let $b' = \langle z', \mu'\rangle \in \mathbf{P}^{B}_{\mathsf{un}}(b,a)$ for some $a \in \mathit{Act}$ with $z' \in Z$.

Then by definition of the belief MDP, $b' = \mathsf{succ}(b,a,z')$ and the unique epoch e' of all states $\langle s', \mathsf{e}'\rangle \in \mu'$ is $\mathsf{e}' = \mathsf{e} \ominus \mathbf{C}(s,a,s')$ for some $s \in \mu$ and $s' \in \mu'$. Cost vector $\mathbf{C}(s,a,s')$ is guaranteed to be unique as $\mathcal{M}$ is cost-aware.

As $b' = \mathsf{succ}(b,a,z')$, we have for $\langle s', \mathsf{e}'\rangle \in \mu'$

$$\begin{aligned}
\mu'(\langle s',\mathsf{e}'\rangle) &= \frac{\sum_{\langle s,\mathsf{e}\rangle\in\mu} \mu(\langle s,\mathsf{e}\rangle)\cdot \mathbf{P}_{\mathsf{un}}(\langle s,\mathsf{e}\rangle,a)(\langle s',\mathsf{e}'\rangle)\cdot \mathbf{O}_{\mathsf{un}}(\langle s,\mathsf{e}\rangle,a,\langle s',\mathsf{e}'\rangle)(z')}{P(z'|b,a)}\\
&= \frac{\sum_{\langle s,\mathsf{e}\rangle\in\mu} \mu_f(s)\cdot \mathbf{P}_{\mathsf{un}}(\langle s,\mathsf{e}\rangle,a)(\langle s',\mathsf{e}'\rangle)\cdot \mathbf{O}(s,a,s')(z')}{P(z'|b,a)}\\
&\overset{\mathsf{e}'=\mathsf{e}\ominus\mathbf{C}(s,a,s')}{=} \frac{\sum_{s\in\mu} \mu_f(s)\cdot \mathbf{P}(s,a)(s',\mathsf{e}')\cdot \mathbf{O}(s,a,s')(z')}{P(z'|b,a)}\\
&= \frac{\sum_{s\in\mu} \mu_f(s)\cdot \mathbf{P}(s,a)(s')\cdot \mathbf{O}(s,a,s')(z')}{P(z'|\beta,a)}
\end{aligned}$$

Let $\beta' = \mathsf{succ}(\beta,a,z') = \langle z', \mu'_f\rangle$ be the successor belief of β in $\mathit{bel}(\mathcal{M})$.

We have that $\mathbf{C}^{B}(\beta,a,\beta') = \mathbf{C}(s,a,s')$. So in $\mathsf{un}_{\mathit{bel}(\mathbf{C}\bowtie\mathsf{t})}(\mathit{bel}(\mathcal{M}))$, we have that

$$\mathbf{P}^{\mathit{bel}(\mathcal{M})}_{\mathsf{un}}(\langle\beta,\mathsf{e}\rangle,a)(\langle\beta',\tilde{\mathsf{e}}\rangle) = \mathbf{P}^{\mathit{bel}(\mathcal{M})}(\beta,a)(\beta') > 0$$

if and only if $\tilde{\mathsf{e}} = \mathsf{e} \ominus \mathbf{C}^B(\beta, a, \beta')$.

$\mathbf{C}^B(\beta, a, \beta') = \mathbf{C}(s, a, s')$ yields $\mathsf{e} \ominus \mathbf{C}^B(\beta, a, \beta') = \mathsf{e} \ominus \mathbf{C}(s, a, s') = \mathsf{e}'$ holds and we get that $\langle \beta', \mathsf{e}' \rangle \in \mathbf{P}_{\mathsf{un}}^{bel(\mathcal{M})}(\langle \beta, \mathsf{e} \rangle, a)$.

Additionally, we have that

$$\mu'_f(s') = \frac{\sum_{s \in \mu} \mu_f(s) \cdot \mathbf{P}(s,a)(s') \cdot \mathbf{O}(s,a,s')(z')}{P(z'|\beta, a)} = \mu'(\langle s', \mathsf{e}' \rangle)$$

for all $s' \in \mu'_f$ and we conclude that $f(b') = \langle \beta', \mathsf{e}' \rangle$.

Thus, we have shown that for an arbitrary successor b' of b, if $f(b) \in \mathcal{B}^{\mathcal{M}} \times \mathsf{E}_k(\mathsf{t})$, then also $f(b') \in \mathcal{B}^{\mathcal{M}} \times \mathsf{E}_k(\mathsf{t})$.

We conclude that $f(b) \in \mathcal{B}^{\mathcal{M}} \times \mathsf{E}_k(\mathsf{t})$ for all $b \in \mathcal{B}_{\mathsf{un}_{\mathbf{C} \bowtie \mathsf{t}}(\mathcal{M})}$.

Therefore, f is well-defined. □

Lemma 3. *The mapping $f : \mathcal{B}_{\mathsf{un}_{\mathbf{C} \bowtie \mathsf{t}}(\mathcal{M})} \to \mathcal{B}^{\mathcal{M}} \times \mathsf{E}_k(\mathsf{t})$ as defined above is a bijection.*

Proof. $\underline{f \text{ is injective}}$: We show that $\langle z, \mu \rangle \neq \langle z', \mu' \rangle$ implies

$$f(\langle z, \mu \rangle) = \langle \langle z, \mu_f \rangle, \mathsf{e} \rangle \neq f(\langle z', \mu' \rangle) = \langle \langle z', \mu'_f \rangle, \mathsf{e}' \rangle.$$

If $z \neq z'$, $\langle \langle z, \mu_f \rangle, \mathsf{e} \rangle \neq \langle \langle z', \mu'_f \rangle, \mathsf{e}' \rangle$ directly follows.

If $z = z'$, then $\mu \neq \mu'$ must hold for $\langle z, \mu \rangle \neq \langle z', \mu' \rangle$ to hold. Recall that e and e' are the unique epochs of states in μ' and μ' respectively. We distinguish two cases:

- $\mathsf{e} \neq \mathsf{e}'$. Then $\langle \langle z, \mu_f \rangle, \mathsf{e} \rangle \neq \langle \langle z', \mu'_f \rangle, \mathsf{e}' \rangle$ follows directly.
- $\mathsf{e} = \mathsf{e}'$. As $\mu \neq \mu'$, if $z = z'$ and $\mathsf{e} = \mathsf{e}'$, there exists an $\langle s, \mathsf{e} \rangle$ such that $\mu(\langle s, \mathsf{e} \rangle) \neq \mu'(\langle s, \mathsf{e} \rangle)$. Therefore, we have $\mu_f(s) \neq \mu'_f(s)$, establishing
$$\langle \langle z, \mu_f \rangle, \mathsf{e} \rangle \neq \langle \langle z', \mu'_f \rangle, \mathsf{e}' \rangle.$$

Thus f is injective.

$\underline{f \text{ is surjective}}$: we show that for all $\langle \langle z, \mu_f \rangle, \mathsf{e} \rangle \in \mathcal{B}^{\mathcal{M}} \times \mathsf{E}_k(\mathsf{t})$, there exists a $\langle z, \mu \rangle \in \mathcal{B}_{\mathsf{un}_{\mathbf{C} \bowtie \mathsf{t}}(\mathcal{M})}$ such that $\underline{f(\langle z, \mu \rangle) = \langle \langle z, \mu_f \rangle, \mathsf{e} \rangle}$.

Given $\langle \langle z, \mu_f \rangle, \mathsf{e} \rangle \in \mathcal{B}^{\mathcal{M}} \times \mathsf{E}_k(\mathsf{t})$, let $b = \langle z, \mu \rangle$ with

$$\mu(\langle s, \mathsf{e}' \rangle) := \begin{cases} \mu_f(s) & \text{if } \mathsf{e}' = \mathsf{e}, \\ 0 & \text{otherwise.} \end{cases}$$

$b \in \mathcal{B}_{\mathsf{un}_{\mathbf{C} \bowtie \mathsf{t}}(\mathcal{M})}$ can be established analogous to the proof for Lemma 2.

We have $f(b) = \langle \langle z, \mu_f \rangle, \mathsf{e} \rangle$.

As f is surjective and injective, f is a bijection. □

We can now proof the theorem.

Proof of Theorem 3. The bijective function f is the isomorphism establishing

$$bel(\mathsf{un}_{\mathbf{C} \bowtie \mathsf{t}}(\mathcal{M})) \cong \mathsf{un}_{bel(\mathbf{C} \bowtie \mathsf{t})}(bel(\mathcal{M})).$$

It remains to show that f indeed preserves transition probabilities, i.e.,

$$\mathbf{P}_{\mathsf{un}}^B(b, a)(b') = \mathbf{P}_{\mathsf{un}}^{bel(\mathcal{M})}(f(b), a)(f(b'))$$

Let $b = \langle z, \mu \rangle, b' = \langle z', \mu' \rangle \in \mathcal{B}_{\mathsf{un}_{\mathbf{C}\bowtie\mathsf{t}}(\mathcal{M})}$ and let $a \in Act$. Furthermore, let $f(b) = \langle \langle z, \mu_f \rangle, \mathsf{e} \rangle = \langle \beta, \mathsf{e} \rangle$ and $f(b') = \langle \langle z', \mu'_f \rangle, \mathsf{e}' \rangle = \langle \beta', \mathsf{e}' \rangle$.

We distinguish three cases:

- $b' \neq \mathsf{succ}(b, a, z')$ and $P(z'|b, a) = 0$. Then for all $\langle s, \mathsf{e} \rangle \in \mu$ and $\langle s', \mathsf{e}' \rangle \in P_{\mathsf{un}}(\langle s, \mathsf{e} \rangle, a)$, we have

  $$\mathbf{O}_{\mathsf{un}}(\langle s, \mathsf{e} \rangle, a, \langle s', \mathsf{e}' \rangle)(z') = 0$$

  and thus already

  $$\mathbf{O}(s, a, s')(z') = 0$$

  for all $s \in \mu_f$. Therefore, in belief MDP $bel(\mathcal{M})$, $\beta' = \langle z', \mu'_f \rangle$ is already not a successor of $\beta = \langle z, \mu \rangle$ and thus

  $$\mathbf{P}_{\mathsf{un}}^{bel(\mathcal{M})}(f(b), a)(f(b')) = \mathbf{P}^{bel(\mathcal{M})}(\beta, a)(\beta') = 0 = \mathbf{P}_{\mathsf{un}}^{B}(b, a)(b').$$

- $b' \neq \mathsf{succ}(b, a, z')$ and $P(z'|b, a) > 0$. Then there exists a $\langle s', \mathsf{e}' \rangle \in \mu'$ such that

  $$\begin{aligned} &\mu'(\langle s', \mathsf{e}' \rangle) \\ \neq &\frac{\sum_{\langle s, \mathsf{e} \rangle \in \mu} \mu(\langle s, \mathsf{e} \rangle) \cdot \mathbf{P}_{\mathsf{un}}(\langle s, \mathsf{e} \rangle, a)(\langle s', \mathsf{e}' \rangle) \cdot \mathbf{O}_{\mathsf{un}}(\langle s, \mathsf{e} \rangle, a, \langle s', \mathsf{e}' \rangle)(z')}{P(z'|b, a)} \\ = &\frac{\sum_{\langle s, \mathsf{e} \rangle \in \mu} \mu(\langle s, \mathsf{e} \rangle) \cdot \mathbf{P}_{\mathsf{un}}(\langle s, \mathsf{e} \rangle, a)(\langle s', \mathsf{e}' \rangle) \cdot \mathbf{O}(s, a, s')(z')}{P(z'|b, a)} \end{aligned}$$

  First consider the case if $\mathsf{e}' = \mathsf{e} \ominus \mathbf{C}(s, a, s')$. Then

  $$\begin{aligned} &\frac{\sum_{\langle s, \mathsf{e} \rangle \in \mu} \mu(\langle s, \mathsf{e} \rangle) \cdot \mathbf{P}_{\mathsf{un}}(\langle s, \mathsf{e} \rangle, a)(\langle s', \mathsf{e}' \rangle) \cdot \mathbf{O}(s, a, s')(z')}{P(z'|b, a)} \\ = &\frac{\sum_{\langle s, \mathsf{e} \rangle \in \mu} \mu(\langle s, \mathsf{e} \rangle) \cdot \mathbf{P}(s, a)(s') \cdot \mathbf{O}(s, a, s')(z')}{P(z'|b, a)} \\ = &\frac{\sum_{s \in \mu_f} \mu_f(s) \cdot \mathbf{P}(s, a)(s') \cdot \mathbf{O}(s, a, s')(z')}{P(z'|\beta, a)} \\ \neq &\mu'(\langle s', \mathsf{e}' \rangle) = \mu'_f(s') \end{aligned}$$

  If $\mathsf{e}' \neq \mathsf{e} \ominus \mathbf{C}(s, a, s')$, then

  $$\begin{aligned} &\frac{\sum_{\langle s, \mathsf{e} \rangle \in \mu} \mu(\langle s, \mathsf{e} \rangle) \cdot \mathbf{P}_{\mathsf{un}}(\langle s, \mathsf{e} \rangle, a)(\langle s', \mathsf{e}' \rangle) \cdot \mathbf{O}(s, a, s')(z')}{P(z'|b, a)} \\ = &\frac{\sum_{\langle s, \mathsf{e} \rangle \in \mu} \mu(\langle s, \mathsf{e} \rangle) \cdot 0 \cdot \mathbf{O}(s, a, s')(z')}{P(z'|b, a)} \\ = &0 \neq \mu'(\langle s', \mathsf{e}' \rangle) = \mu'_f(s') \end{aligned}$$

  Thus in both cases, β' is already not a successor belief of β, meaning that

  $$\mathbf{P}_{\mathsf{un}}^{bel(\mathcal{M})}(f(b), a)(f(b')) = 0 = \mathbf{P}_{\mathsf{un}}^{B}(b, a)(b').$$

- $b' = \mathsf{succ}(b, a, z')$: The proof of Lemma 2 already establishes that in this case $\mathbf{P}_{\mathsf{un}}^{bel(\mathcal{M})}(f(b), a)(f(b')) = \mathbf{P}_{\mathsf{un}}^{B}(b, a)(b')$ holds.

In all cases, f indeed preserves transition probabilities, i.e.,

$$\mathbf{P}_{\mathsf{un}}^{B}(b, a)(b') = \mathbf{P}_{\mathsf{un}}^{bel(\mathcal{M})}(f(b), a)(f(b'))$$

holds. We conclude that f establishes the isomorphism between $bel(\mathsf{un}_{\mathbf{C}\bowtie\mathsf{t}}(\mathcal{M}))$ and $\mathsf{un}_{bel(\mathbf{C}\bowtie\mathsf{t})}(bel(\mathcal{M}))$, i.e., we get that

$$bel(\mathsf{un}_{\mathbf{C}\bowtie\mathsf{t}}(\mathcal{M})) \cong \mathsf{un}_{bel(\mathbf{C}\bowtie\mathsf{t})}(bel(\mathcal{M})).$$

□

C.4 PROOF OF THEOREM 4

Fix a POMDP $\mathcal{M} = \langle M, Z, \mathbf{O}\rangle$ with $M = \langle S, Act, \mathbf{P}, s_{init}\rangle$, cost structure $\mathbf{C}$, and a level function lvl_{d}. Furthermore, let $lvl_{\mathsf{d}}(\mathcal{M})$ be the level unfolding POMDP according to Def. 9 and define the *level jump (cost) structure* $\mathbf{L} : S_{lvl_{\mathsf{d}}} \times Act \times S_{lvl_{\mathsf{d}}} \to \mathbb{N}^k$ as

$$\mathbf{L}(\langle s, \ell\rangle, a, \langle s', \ell'\rangle)[i] := \left\lceil \frac{\mathbf{C}(s, a, s')[i] - \ell[i]}{\mathsf{d}[i]} \right\rceil.$$

We will first show three lemmata that will help with the proof of Theorem 4.

Lemma 4. *Let* $\hat{\pi} = \langle s_0, \ell_0\rangle a_1 \langle s_1, \ell_1\rangle a_2 \dots a_n \langle s_n, \ell_n\rangle \in Paths_{\mathrm{fin}}^{lvl_{\mathsf{d}}(\mathcal{M})}$ *and* $\tilde{\pi} = \hat{\pi} a_{n+1} \langle s_{n+1}, \ell_{n+1}\rangle$ *with* $\mathbf{P}_{lvl_{\mathsf{d}}}(\langle s_n, \ell_n\rangle, a_{n+1}, \langle s_{n+1}, \ell_{n+1}\rangle) > 0$ *an extension of* $\hat{\pi}$*. Then it holds that*

$$\mathsf{d}[i] \cdot \mathsf{cost}_{\mathbf{L}}(\tilde{\pi})[i] - \ell_{n+1}[i] = \mathbf{C}(s_n, a_{n+1}, s_{n+1})[i] + \mathsf{d}[i] \cdot \mathsf{cost}_{\mathbf{L}}(\hat{\pi})[i] - \ell_n[i]$$

Proof. In the following, we use the facts that $(x \mod y) = x - y \cdot \left\lfloor \frac{x}{y} \right\rfloor$ and $\lfloor -x \rfloor = -\lceil x \rceil$. We get:

$$\begin{aligned}
&\mathsf{d}[i] \cdot \mathsf{cost}_{\mathbf{L}}(\tilde{\pi})[i] - \ell_{n+1}[i] \\
=&\mathsf{d}[i] \cdot \sum_{h=1}^{n+1} \left\lceil \frac{\mathbf{C}(s_{h-1}, a_h, s_h)[i] - \ell_{h-1}[i]}{\mathsf{d}[i]} \right\rceil - \ell_{n+1}[i] \\
=&\left(\mathsf{d}[i] \cdot \sum_{h=1}^{n+1} \left\lceil \frac{\mathbf{C}(s_{h-1}, a_h, s_h)[i] - \ell_{h-1}[i]}{\mathsf{d}[i]} \right\rceil \right) - \ell_{n+1}[i] \\
=&\left(\mathsf{d}[i] \cdot \sum_{h=1}^{n+1} \left\lceil \frac{\mathbf{C}(s_{h-1}, a_h, s_h)[i] - \ell_{h-1}[i]}{\mathsf{d}[i]} \right\rceil \right) - (\ell_n[i] - \mathbf{C}(s_n, a_{n+1}, s_{n+1})[i] \mod \mathsf{d}[i]) \\
=&\mathsf{d}[i] \cdot \left\lceil \frac{\mathbf{C}(s_n, a_{n+1}, s_{n+1})[i] - \ell_n[i]}{\mathsf{d}[i]} \right\rceil + \left(\mathsf{d}[i] \cdot \sum_{h=1}^{n} \left\lceil \frac{\mathbf{C}(s_{h-1}, a_h, s_h)[i] - \ell_{h-1}[i]}{\mathsf{d}[i]} \right\rceil \right) \\
&- (\ell_n[i] - \mathbf{C}(s_n, a_{n+1}, s_{n+1})[i] \mod \mathsf{d}[i]) \\
=&\mathsf{d}[i] \cdot \left\lceil \frac{\mathbf{C}(s_n, a_{n+1}, s_{n+1})[i] - \ell_n[i]}{\mathsf{d}[i]} \right\rceil + \mathsf{d}[i] \cdot \mathsf{cost}_{\mathbf{L}}(\hat{\pi}) - \\
&\left(\ell_n[i] - \mathbf{C}(s_n, a_{n+1}, s_{n+1})[i] - \mathsf{d}[i] \cdot \left\lfloor \frac{\ell_n[i] - \mathbf{C}(s_n, a_{n+1}, s_{n+1})[i]}{\mathsf{d}[i]} \right\rfloor \right) \\
=&\mathsf{d}[i] \cdot \left\lceil \frac{\mathbf{C}(s_n, a_{n+1}, s_{n+1})[i] - \ell_n[i]}{\mathsf{d}[i]} \right\rceil + \mathsf{d}[i] \cdot \mathsf{cost}_{\mathbf{L}}(\hat{\pi}) \\
&- \left(\ell_n[i] - \mathbf{C}(s_n, a_{n+1}, s_{n+1})[i] + \mathsf{d}[i] \cdot \left\lceil \frac{-\ell_n[i] + \mathbf{C}(s_n, a_{n+1}, s_{n+1})[i]}{\mathsf{d}[i]} \right\rceil \right) \\
=&\mathsf{d}[i] \cdot \left\lceil \frac{\mathbf{C}(s_n, a_{n+1}, s_{n+1})[i] - \ell_n[i]}{\mathsf{d}[i]} \right\rceil + \mathsf{d}[i] \cdot \mathsf{cost}_{\mathbf{L}}(\hat{\pi}) \\
&- \ell_n[i] + \mathbf{C}(s_n, a_{n+1}, s_{n+1})[i] - \mathsf{d}[i] \cdot \left\lceil \frac{\mathbf{C}(s_n, a_{n+1}, s_{n+1})[i] - \ell_n[i]}{\mathsf{d}[i]} \right\rceil \\
=&\mathbf{C}(s_n, a_{n+1}, s_{n+1})[i] + \mathsf{d}[i] \cdot \mathsf{cost}_{\mathbf{L}}(\hat{\pi}) - \ell_n[i]
\end{aligned}$$

□

The following lemma formalises the relationship between the cost structure on the level unfolding and the level jump structure.

Lemma 5. *Given* $\hat{\pi} = \langle s_0, \ell_0\rangle a_1 \langle s_1, \ell_1\rangle a_2 \dots a_n \langle s_n, \ell_n\rangle \in Paths_{\mathrm{fin}}^{lvl_{\mathsf{d}}(\mathcal{M})}$*, we have that*

$$\mathsf{cost}_{\mathbf{C}_{lvl_{\mathsf{d}}}}(\hat{\pi})[i] = \mathsf{d}[i] \cdot \mathsf{cost}_{\mathbf{L}}(\hat{\pi})[i] - \ell_n[i].$$

Proof. We proof the claim by induction over the length of $\hat{\pi}$. We denote by $\tilde{\pi}$ the prefix of $\hat{\pi}$ without the last transition, i.e., $\tilde{\pi} = \langle s_0, \ell_0\rangle a_1 \dots a_{|\hat{\pi}|-1}\langle s_{|\hat{\pi}|-1}, \ell_{|\hat{\pi}|-1}\rangle$

Let $|\hat{\pi}| = 0$. Then

$$\begin{aligned}
\mathsf{d}[i]\cdot\mathsf{cost}_{\mathbf{L}}(\hat{\pi})[i] - \ell_n[i] &= \mathsf{d}[i]\cdot\mathsf{cost}_{\mathbf{L}}(\hat{\pi})[i] - \ell_0[i]\\
&= \mathsf{d}[i]\cdot 0 - \langle 0,\dots,0\rangle[i]\\
&= 0\\
&= \mathsf{cost}_{\mathbf{C}_{\mathit{lvl}_\mathsf{d}}}(\hat{\pi})[i]
\end{aligned}$$

Assume the claim holds for an arbitrary, but fixed $n \in \mathbb{N}$.

Let $|\hat{\pi}| = n+1$

$$\begin{aligned}
\mathsf{cost}_{\mathbf{C}_{\mathit{lvl}_\mathsf{d}}}(\hat{\pi})[i] &= \sum_{h=1}^{n+1} \mathbf{C}_{\mathit{lvl}_\mathsf{d}}(\langle s_{h-1}, \ell_{h-1}\rangle a_h \langle s_h, \ell_h\rangle)[i]\\
&= \mathbf{C}_{\mathit{lvl}_\mathsf{d}}(\langle s_n, \ell_n\rangle a_{n+1}\langle s_{n+1}, \ell_{n+1}\rangle)[i] + \sum_{h=1}^{n} \mathbf{C}_{\mathit{lvl}_\mathsf{d}}(\langle s_{h-1}, \ell_{h-1}\rangle a_h \langle s_h, \ell_h\rangle)[i]\\
&= \mathbf{C}_{\mathit{lvl}_\mathsf{d}}(\langle s_n, \ell_n\rangle a_{n+1}\langle s_{n+1}, \ell_{n+1}\rangle)[i] + \mathsf{cost}_{\mathbf{C}_{\mathit{lvl}_\mathsf{d}}}(\tilde{\pi})[i]\\
&\overset{(IH)}{=} \mathbf{C}_{\mathit{lvl}_\mathsf{d}}(\langle s_n, \ell_n\rangle a_{n+1}\langle s_{n+1}, \ell_{n+1}\rangle)[i] + \mathsf{d}[i]\cdot\mathsf{cost}_{\mathbf{L}}(\tilde{\pi})[i] - \ell_n[i]\\
&= \mathbf{C}(s_n, a_{n+1}, s_{n+1})[i] + \mathsf{d}[i]\cdot\mathsf{cost}_{\mathbf{L}}(\tilde{\pi})[i] - \ell_n[i]\\
&\overset{\text{(Lem. 4)}}{=} \mathsf{d}[i]\cdot\mathsf{cost}_{\mathbf{L}}(\hat{\pi})[i] - \ell_{n+1}[i]
\end{aligned}$$

Thus the claim is shown for all lengths of $\hat{\pi}$. □

The following lemma shows that in the case that $\mathsf{d}[i] \mid \mathsf{t}[i]$ for all $1 \le i \le k$, we can define a bound on the jump structure such that for an arbitrary finite path, the new bound is active if and only if the original cost bound is active.

Lemma 6. *Let $\mathit{lvl}_\mathsf{d}(\mathcal{M})$, cost bound $(\mathbf{C}_{\mathit{lvl}_\mathsf{d}} \bowtie \mathsf{t})$ and T_d such that $\mathsf{d}[i] \mid \mathsf{t}[i]$ for $1 \le i \le k$. Furthermore, define t_d with $\mathsf{t}_\mathsf{d}[i] := \frac{\mathsf{t}[i]}{\mathsf{d}[i]}$ for all $1 \le i \le k$. Then for $\hat{\pi} \in \mathit{Paths}_{\mathrm{fin}}^{\mathit{lvl}_\mathsf{d}(\mathcal{M})}$, $(\mathbf{C}_{\mathit{lvl}_\mathsf{d}} \bowtie \mathsf{t})$ is active for $\hat{\pi}$ if and only if $(\mathbf{L} \bowtie \mathsf{t}_\mathsf{d})$ is active for $\hat{\pi}$.*

Proof. Fix a path $\hat{\pi} \in \mathit{Paths}_{\mathrm{fin}}^{\mathit{lvl}_\mathsf{d}(\mathcal{M})}$. To show that $(\mathbf{C}_{\mathit{lvl}_\mathsf{d}} \bowtie \mathsf{t})$ is active for $\hat{\pi}$ if and only if $(\mathbf{L} \bowtie \mathsf{t}_\mathsf{d})$ is active for $\hat{\pi}$, we need to show for all dimensions i that $(\mathbf{C}_{\mathit{lvl}_\mathsf{d}} \bowtie \mathsf{t})$ is active in dimension i for $\hat{\pi}$ if and only if $(\mathbf{L} \bowtie \mathsf{t}_\mathsf{d})$ is active in i for $\hat{\pi}$. Thus, we need to show that $\mathsf{cost}_{\mathit{lvl}_\mathsf{d}}(\hat{\pi})[i] {\bowtie}[i] \mathsf{t}[i]$ if and only if $\mathsf{cost}_{\mathbf{L}}(\hat{\pi})[i] {\bowtie}[i] \mathsf{t}_\mathsf{d}[i]$ for all $1 \le i \le k$.

Fix an arbitrary dimension $1 \le i \le k$. We get

$$\begin{aligned}
& & \mathsf{cost}_{\mathit{lvl}_\mathsf{d}}(\hat{\pi})[i] {\bowtie}[i] \mathsf{t}[i]\\
&\iff & \frac{\mathsf{cost}_{\mathit{lvl}_\mathsf{d}}(\hat{\pi})[i]}{\mathsf{d}[i]} {\bowtie}[i] \frac{\mathsf{t}[i]}{\mathsf{d}[i]}\\
&\overset{\text{Lem. 5}}{\iff} & \frac{\mathsf{d}[i]\cdot\mathsf{cost}_{\mathbf{L}}(\hat{\pi})[i] - \ell_n[i]}{\mathsf{d}[i]} {\bowtie}[i] \frac{\mathsf{t}[i]}{\mathsf{d}[i]}\\
&\iff & \frac{\mathsf{d}[i]\cdot\mathsf{cost}_{\mathbf{L}}(\hat{\pi})[i]}{\mathsf{d}[i]} - \frac{\ell_n[i]}{\mathsf{d}[i]} {\bowtie}[i] \frac{\mathsf{t}[i]}{\mathsf{d}[i]}\\
&\iff & \mathsf{cost}_{\mathbf{L}}(\hat{\pi})[i] - \frac{\ell_n[i]}{\mathsf{d}[i]} {\bowtie}[i] \frac{\mathsf{t}[i]}{\mathsf{d}[i]}\\
&\overset{(*)}{\iff} & \mathsf{cost}_{\mathbf{L}}(\hat{\pi})[i] {\bowtie}[i] \frac{\mathsf{t}[i]}{\mathsf{d}[i]}
\end{aligned}$$

It remains to show that equivalence $(*)$ holds. Observe that by definition of $S_{lvl_{\mathsf{d}}}$ it always holds that $0 \leq \frac{\ell_n[i]}{\mathsf{d}[i]} < 1$. Furthermore, as $\mathsf{d}[i] \mid \mathsf{t}[i]$, we get that $\frac{\mathsf{t}[i]}{\mathsf{d}[i]} \in \mathbb{N}$. Also observe that $\mathsf{cost}_{\mathbf{L}}(\hat{\pi})[i] \in \mathbb{N}$.

We distinguish the two cases for $\bowtie[i]$.

$\underline{\bowtie[i] = (>)}$: As $0 \leq \frac{\ell_n[i]}{\mathsf{d}[i]}$, it follows directly that

$$\mathsf{cost}_{\mathbf{L}}(\hat{\pi})[i] - \frac{\ell_n[i]}{\mathsf{d}[i]} > \frac{\mathsf{t}[i]}{\mathsf{d}[i]} \quad \Rightarrow \quad \mathsf{cost}_{\mathbf{L}}(\hat{\pi})[i] > \frac{\mathsf{t}[i]}{\mathsf{d}[i]}.$$

For

$$\mathsf{cost}_{\mathbf{L}}(\hat{\pi})[i] - \frac{\ell_n[i]}{\mathsf{d}[i]} > \frac{\mathsf{t}[i]}{\mathsf{d}[i]} \quad \Leftarrow \quad \mathsf{cost}_{\mathbf{L}}(\hat{\pi})[i] > \frac{\mathsf{t}[i]}{\mathsf{d}[i]},$$

observe that

$$\begin{aligned} & & \mathsf{cost}_{\mathbf{L}}(\hat{\pi})[i] &> \frac{\mathsf{t}[i]}{\mathsf{d}[i]} \\ \Rightarrow & & \mathsf{cost}_{\mathbf{L}}(\hat{\pi})[i] &\geq \frac{\mathsf{t}[i]}{\mathsf{d}[i]} + 1 \\ \Rightarrow & & \mathsf{cost}_{\mathbf{L}}(\hat{\pi})[i] - 1 &\geq \frac{\mathsf{t}[i]}{\mathsf{d}[i]} \\ \Rightarrow & & \mathsf{cost}_{\mathbf{L}}(\hat{\pi})[i] - \underbrace{\frac{\ell_n[i]}{\mathsf{d}[i]}}_{<1} &> \frac{\mathsf{t}[i]}{\mathsf{d}[i]}. \end{aligned}$$

Therefore, we conclude that

$$\mathsf{cost}_{\mathbf{L}}(\hat{\pi})[i] - \frac{\ell_n[i]}{\mathsf{d}[i]} > \frac{\mathsf{t}[i]}{\mathsf{d}[i]} \quad \Longleftrightarrow \quad \mathsf{cost}_{\mathbf{L}}(\hat{\pi})[i] > \frac{\mathsf{t}[i]}{\mathsf{d}[i]}$$

$\underline{\bowtie[i] = (\leq)}$: By contraposition of the previous case, it follows that

$$\mathsf{cost}_{\mathbf{L}}(\hat{\pi})[i] - \frac{\ell_n[i]}{\mathsf{d}[i]} \leq \frac{\mathsf{t}[i]}{\mathsf{d}[i]} \quad \Longleftrightarrow \quad \mathsf{cost}_{\mathbf{L}}(\hat{\pi})[i] \leq \frac{\mathsf{t}[i]}{\mathsf{d}[i]}$$

Thus, we have shown that for all options for $\bowtie[i]$, it holds that

$$\mathsf{cost}_{\mathbf{L}}(\hat{\pi})[i] - \frac{\ell_n[i]}{\mathsf{d}[i]} \bowtie[i] \frac{\mathsf{t}[i]}{\mathsf{d}[i]} \quad \Longleftrightarrow \quad \mathsf{cost}_{\mathbf{L}}(\hat{\pi})[i] \bowtie[i] \frac{\mathsf{t}[i]}{\mathsf{d}[i]}.$$

We conclude that $(\mathbf{C}_{lvl_{\mathsf{d}}} \bowtie \mathsf{t})$ is active in dimension i for $\hat{\pi}$ if and only if $(\mathbf{L} \bowtie \mathsf{t}_{\mathsf{d}})$ is active in i for $\hat{\pi}$ for all $1 \leq i \leq k$ and thus $(\mathbf{C}_{lvl_{\mathsf{d}}} \bowtie \mathsf{t})$ is active for $\hat{\pi}$ if and only if $(\mathbf{L} \bowtie \mathsf{t}_{\mathsf{d}})$ is active for $\hat{\pi}$. □

Finally, we prove Theorem 4.

Theorem 4. *Let $lvl_{\mathsf{d}}(\mathcal{M})$ such that for all $1 \leq i \leq k$: $\mathsf{d}[i] \bigm| \mathsf{t}[i]$ and let $(\mathbf{L} \bowtie \mathsf{t}_{\mathsf{d}})$ be a cost bound for $lvl_{\mathsf{d}}(\mathcal{M})$ with for all $1 \leq i \leq k$: $\mathbf{L}(\langle s, \ell\rangle, a, \langle s', \ell'\rangle)[i] := \left\lceil \big(\mathbf{C}(s, a, s')[i] - \ell[i]\big)/\mathsf{d}[i] \right\rceil$ and $\mathsf{t}_{\mathsf{d}}[i] := \mathsf{t}[i]/\mathsf{d}[i]$. Then, $lvl_{\mathsf{d}}(\mathcal{M})$ is cost-aware w.r.t. $\mathbf{L}$ and*

$$\mathsf{Pr}_{\max}^{lvl_{\mathsf{d}}(\mathcal{M})}\Big(\Diamond_{\mathbf{C}_{lvl_{\mathsf{d}}} \bowtie \mathsf{t}}\, T_{\mathsf{d}}\Big) = \mathsf{Pr}_{\max}^{lvl_{\mathsf{d}}(\mathcal{M})}\Big(\Diamond_{\mathbf{L} \bowtie \mathsf{t}_{\mathsf{d}}}\, T_{\mathsf{d}}\Big).$$

Proof. For the cost-awareness, we define for each observation $\langle z, \mathsf{j}\rangle$ the vector $\mathsf{c}_{\langle z, \mathsf{j}\rangle} = \mathsf{j}$. Fix an arbitrary transition $\langle s, \ell\rangle \xrightarrow{a} \langle s', \ell'\rangle$. We have by definition that $\langle z, \mathsf{j}\rangle \in supp(\mathbf{O}_{lvl_{\mathsf{d}}}(\langle s, \ell\rangle, a, \langle s', \ell'\rangle))$ if and only if $\mathsf{j}[i] = \left\lceil \frac{\mathbf{C}(s,a,s')[i] - \ell[i]}{\mathsf{d}[i]} \right\rceil$ (and $\mathbf{O}_{lvl_{\mathsf{d}}}(s, a, s'))(z) > 0$). Additionally, we have that $\mathbf{L}(\langle s, \ell\rangle, a, \langle s', \ell'\rangle)[i] = \left\lceil \frac{\mathbf{C}(s,a,s')[i] - \ell[i]}{\mathsf{d}[i]} \right\rceil = \mathsf{j}[i]$ for all $1 \leq i \leq k$ and thus $\mathbf{L}(\langle s, \ell\rangle, a, \langle s', \ell'\rangle) = \mathsf{j} = \mathsf{c}_{\langle z, \mathsf{j}\rangle}$. Therefore, $lvl_{\mathsf{d}}(\mathcal{M})$ is cost-aware w.r.t. $\mathbf{L}$.

Table 3: Overview of Benchmarking Instances

Model	$\lvert S \rvert$	Bounds	$\lvert \mathsf{E} \rvert$
clean6	37	$\mathbf{C}[1] \leq 60, \mathbf{C}[2] > 5$	413
clean12	73	$\mathbf{C}[1] \leq 120, \mathbf{C}[2] > 11$	1508
incline	25	$\mathbf{C}[1] \leq 75, \mathbf{C}[2] \leq 21$	497
obstcl	25	$\mathbf{C}[1] \leq 25, \mathbf{C}[2] \leq 7$	83
resrc	721	$\mathbf{C}[1] > 4, \mathbf{C}[2] > 4, \mathbf{C}[3] \leq 60$	2107
	721	$\mathbf{C}[1] > 14, \mathbf{C}[2] > 14, \mathbf{C}[3] \leq 180$	$4 \cdot 10^4$
rover	16	$\mathbf{C}[1] > 199, \mathbf{C}[2] \leq 360, \mathbf{C}[3] \leq 200$	$7 \cdot 10^5$
		$\mathbf{C}[1] > 599, \mathbf{C}[2] \leq 1080, \mathbf{C}[3] \leq 600$	$2 \cdot 10^7$
serv	$8 \cdot 10^4$	$\mathbf{C}[1] \leq 570$	40
		$\mathbf{C}[1] \leq 1000$	68
walk40	84	$\mathbf{C}[1] \leq 80$	82
walk120	244	$\mathbf{C}[1] \leq 80$	82
water	34	$\mathbf{C}[1] \leq 590, \mathbf{C}[2] > 49$	$3 \cdot 10^4$
		$\mathbf{C}[1] \leq 1790, \mathbf{C}[2] > 149$	$3 \cdot 10^5$

We now show that $\mathsf{Pr}^{lvl_{\mathsf{d}}(\mathcal{M})}_{\max}\left(\Diamond_{\mathbf{C}_{lvl_{\mathsf{d}}} \bowtie \mathsf{t}} T_{\mathsf{d}}\right) = \mathsf{Pr}^{lvl_{\mathsf{d}}(\mathcal{M})}_{\max}\left(\Diamond_{\mathbf{L} \bowtie \mathsf{t}_{\mathsf{d}}} T_{\mathsf{d}}\right)$ by showing that the two sets of paths are equal, i.e., that $\Diamond_{\mathbf{C}_{lvl_{\mathsf{d}}} \bowtie \mathsf{t}} T_{\mathsf{d}} = \Diamond_{\mathbf{L} \bowtie \mathsf{t}_{\mathsf{d}}} T_{\mathsf{d}}$.

$\underline{\Diamond_{\mathbf{C}_{lvl_{\mathsf{d}}} \bowtie \mathsf{t}} T_{\mathsf{d}} \subseteq \Diamond_{\mathbf{L} \bowtie \mathsf{t}_{\mathsf{d}}} T_{\mathsf{d}}}$

Let $\pi \in \Diamond_{\mathbf{C}_{lvl_{\mathsf{d}}} \bowtie \mathsf{t}} T_{\mathsf{d}}$. Then $\pi \in Cyl(\hat{\pi})$ for a $\hat{\pi} \in Paths^{lvl_{\mathsf{d}}(\mathcal{M})}_{\mathrm{fin}}$ where $last(\hat{\pi}) \in T_{\mathsf{d}}$ and $(\mathbf{C}_{lvl_{\mathsf{d}}} \bowtie \mathsf{t})$ is active for $\hat{\pi}$. Then, by Lemma 6, also $(\mathbf{L} \bowtie \mathsf{t}_{\mathsf{d}})$ is active for $\hat{\pi}$ and thus also $\pi \in \Diamond_{\mathbf{L} \bowtie \mathsf{t}_{\mathsf{d}}} T_{\mathsf{d}}$.

$\underline{\Diamond_{\mathbf{C}_{lvl_{\mathsf{d}}} \bowtie \mathsf{t}} T_{\mathsf{d}} \supseteq \Diamond_{\mathbf{L} \bowtie \mathsf{t}_{\mathsf{d}}} T_{\mathsf{d}}}$

Let $\pi \in \Diamond_{\mathbf{L} \bowtie \mathsf{t}_{\mathsf{d}}} T_{\mathsf{d}}$. Then $\pi \in Cyl(\hat{\pi})$ for a $\hat{\pi} \in Paths^{lvl_{\mathsf{d}}(\mathcal{M})}_{\mathrm{fin}}$ where $last(\hat{\pi}) \in T_{\mathsf{d}}$ and $(\mathbf{L} \bowtie \mathsf{t}_{\mathsf{d}})$ is active for $\hat{\pi}$. Then, by Lemma 6, also $(\mathbf{C}_{lvl_{\mathsf{d}}} \bowtie \mathsf{t})$ is active for $\hat{\pi}$ and thus also $\pi \in \Diamond_{\mathbf{C}_{lvl_{\mathsf{d}}} \bowtie \mathsf{t}} T_{\mathsf{d}}$.

We conclude that $\Diamond_{\mathbf{C}_{lvl_{\mathsf{d}}} \bowtie \mathsf{t}} T_{\mathsf{d}} = \Diamond_{\mathbf{L} \bowtie \mathsf{t}_{\mathsf{d}}} T_{\mathsf{d}}$ and therefore $\mathsf{Pr}^{lvl_{\mathsf{d}}(\mathcal{M})}_{\max}\left(\Diamond_{\mathbf{C}_{lvl_{\mathsf{d}}} \bowtie \mathsf{t}} T_{\mathsf{d}}\right) = \mathsf{Pr}^{lvl_{\mathsf{d}}(\mathcal{M})}_{\max}\left(\Diamond_{\mathbf{L} \bowtie \mathsf{t}_{\mathsf{d}}} T_{\mathsf{d}}\right)$ holds. □

D BENCHMARK PROBLEMS

We give a short overview over the models we use in our experimental evaluation. Table 3 contains more information about the specific instances we consider, in particular the bounds on the different cost structures and the resulting number of epochs $\lvert \mathsf{E} \rvert$. The bounds are chosen such that the resulting instances are challenging for the implementation while still resulting in non-trivial values for most configuration.

The files containing the models encoded in the PRISM language are part of the code & data appendix, located in the folder `models`. For an explanation of the format, we refer to the PRISM manual.[1]

clean This model is a generalised version of the cleaning robot scenario described in Section 1. A robot is placed in position 0 of a hallway consisting of N tiles, all initially dirty. In each step, it can decide between moving to the next tile, increasing its position by 1, or attempting to clean the tile at the current position. A cleaning attempt is successful with probability 0.8. The robot can repeatedly attempt to clean the same tile. Moving always consumes 1 unit of energy, while a cleaning attempt requires either 2 or 4 units, each with probability 0.5. If the robot moves to position N, i.e., out of the

[1] https://www.prismmodelchecker.org/manual/ThePRISMLanguage/Introduction

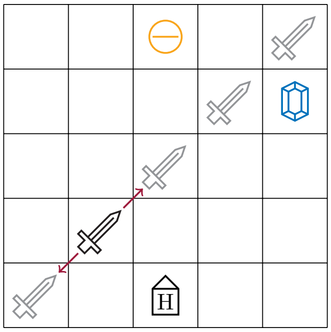


Figure 5: Resource Gathering Scenario

hallway, it has reached its target position. The robot can observe its position, but generally does not observe how many tiles it has already cleaned.

We use two cost dimensions to model the energy consumption of the robot and the amount of tiles it has successfully cleaned. The objective is to reach the target position with a specified bound on the energy while successfully cleaning at least a specified amount of tiles. We consider two sizes, $N = 6$ and $N = 12$.

incline The agent has to reach a target in a 5×5 grid of cells. The agent can move in any of the four cardinal directions. Each move is either *uphill*, *downhill* or neither (*flat*). With probability 0.5, an attempted move fails due to slipping. If the attempted move is uphill, the agent is staying in place if it slips. Similarly, slipping downhill causes the agent to overshoot and move a cell further in the chosen direction. With a flat move, slipping has no effect. In case a move would lead out of bounds, the agent moves as far as possible in the direction and then stays in place. The agent does not observe its current position, but knows its starting position in the south-west corner of the grid. We consider a map where the incline is such that all moves north and east are uphill and all moves south and west are downhill.

We consider a cost model where downhill steps consume 1 unit of energy, flat moves 2 units and uphill moves 3 units. The objective is to reach the target in the north-east corner of the grid. We are interesting in reaching the goal within the energy budget (modelled in the first dimension) and a maximum number of steps (modelled in the second dimension).

obstcl Similar to incline, the agent is supposed to reach a target in a 5×5 grid of cells. Only the outermost ring of cells can be traversed freely, the inner ring contains light forest and the center cell is dense forest. The agent cannot observe its position and it slips, i.e., stays in place when executing a move from light forest or dense forest with probability 0.25 or 0.5 respectively. Moves from free tiles, light forest tiles and dense forest tiles consume 1, 2 or 3 units of energy each. Like before, the objective is to reach the goal within the energy budget and a maximum number of steps (modelled by a bound on $\mathbf{C}[1]$ and $\mathbf{C}[2]$ respectively).

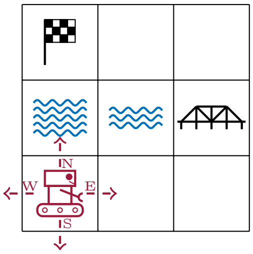
resrc This benchmark is a variant of the resource gathering model from Barrett and Narayanan [2008]. The scenario is similar to a problem arising in many strategy games. We provide a depiction of the setting in Figure 5. An agent, starting in a home base marked by H, is tasked with collecting two kinds of resources–*gold* (depicted by the coin) and *gems* (depicted by the gem)–in a grid environment. In any step,The agent can move in any of the four cardinal directions. When it reaches either the gold or the gem location, it picks up one unit of the respective resource. To collect it, the agent needs to return the resource to the base. The agent can hold at most one unit of each resource.

An *enemy* patrols the diagonal of the grid. It starts in the south-west corner and changes its position to one of the adjacent locations on the diagonal with each step of the agent. If the enemy is in one of the corners, it will certainly move to the adjacent location. When the agent and the enemy enter the same location at the same time, the agent loses all currently held resources and is teleported home without collecting anything. The objective is to maximise the probability to collect a minimum amount of each resource (lower bounds on $\mathbf{C}[1]$ for gold and $\mathbf{C}[2]$ for gems) within a given step bound (upper bound on $\mathbf{C}[3]$).

rover A partially observable version of the Mars rover task scheduling problem described in Hartmanns et al. [2020] based on Bresina et al. [2005]. The problem models the scheduling of a variety of experiments on Mars. Experiments have differing time and energy consumption and success probabilities. Upon success of an experiment, a certain scientific value is

Table 4: Values of the Tasks in the rover Problem

Task	Time	Energy	Sci. Value	Success Prob.
1	10	$\{3, 5\}$	10	1/2
2	5	$\{5, 7\}$	10	3/5
3	5	3	2	4/5
4	10	7	30	1/10



Tile	Energy Cost High	Energy Cost Normal
Deep Water	10	6
Shallow Water	6	2
Bridge	1	1
Land	1	1

Figure 6: Robot Navigation Task with Water Obstacles

collected. Energy consumption for some tasks is subject to uncertainty; the consumed energy has a high or low value with probability 0.5 each. The specific parameters for each task are given in Table 4.

The agent can schedule several experiments each day. It does not directly observe whether a task has been successful or not. The objective is to maximise the probability of achieving at least a certain cumulative scientific value without exceeding both time and energy limits. Scientific values is modelled in $\mathbf{C}[1]$, while time and energy are modelled in $\mathbf{C}[2]$ and $\mathbf{C}[3]$ respectively.

serv The *service* domain is a partially observable variant of the care home scenario described in Lacerda et al. [2017]. A robot is navigating a care home environment consisting of a central hallway with rooms adjacent to each side of the hallway. Overall, the map consists of 21 locations. The robot does not observe its location and each location change to an adjacent location can fail with probability 0.01.

The robot's routine consists of the main task of checking whether three occupants of specified rooms want water and deliver it if it is desired. The robot can collect bottles of water at a central kitchen area. At any point, it can only carry at most two bottles. For delivery, the robot first has to check whether the occupant wants the water. With a probability of 0.2, the occupant actually wants the water and the robot can deliver it. In addition, the robot has the secondary task of interacting with four designated occupants. Each action of the robot requires a certain amount of time. The central area, where the robot starts and which it has to cross to get to the kitchen, is crowded with a probability of 0.2 every time the robot enters it, causing an additional time cost.

We are interested in the probability that the robot delivers water to where it is required within the time limit (bound on $\mathbf{C}[1]$).

water The *water* problem considers an amphibious robot starting in the south-west corner of a regular grid consisting of nine cells. In every step, the robot moves in any of the four cardinal directions, where a move out of bounds makes the robot stay in place. The task is to visit the flag in the north-west corner and then return to the initial position multiple times. A river runs through the center row of the map, modelled as one tile of *deep water*, one tile of *shallow water*, and one *bridge* tile. All other tiles are *land* tiles. With a probability of 0.5, there are *high water* conditions, making crossing the river without the bridge more difficult. The robot is not able to observe the water conditions. The energy consumption for each move depends on the current tile and the water conditions as outlined on the right of Figure 6. A move out-of-bounds still consumes the respective energy.

The agent collects one unit of cost/reward (modelled in dimension $\mathbf{C}[2]$) every time it reaches the target position. The objective is to compute the probability that a minimum number of trips (bound on $\mathbf{C}[2]$) is completed within the given energy budget (bound in dimension 1).

walk An agent is starting at position 0 of a hallway consisting of $N{+}1$ positions. In each step, it can decide to either attempt to *move* to the adjacent position $i{+}1$, *observe* its current position i for a cost of 1 unit, or completely *stop* the process.

A move action fails with probability 0.5, resulting in the agent staying at position i. In position N, a move will always make the agent stay in place. The agent is only able to observe its current position if it executes an *observe* action. However, *observe* actions are unreliable and fail with probability 0.9.

The goal of the agent is to stop the process exactly in position $N{-}1$, leading to a target state. If it stops the process in any other position, it is trapped and cannot fulfil its goal any more.

The cost-bounded objective is to reach the target state, i.e., stopping in position $N{-}1$, while using at most a specified number of *observe* actions. We consider to sizes, $N{=}40$ and $N{=}120$.

E CONSIDERED HYPER-PARAMETERS AND DETAILED SETUP INFORMATION

For the experimental evaluation we consider 25 different hyper-parameter assignments each for CUT and DISCR. As described in the main paper, both methods are applied to yield finite abstractions of the belief MDP we analyse for the respective configuration UNFOLD, CA-UNFOLD or CA-BEL-SEQ.

We have chosen the hyper-parameter values such that we expect them to result in MDPs that STORM can handle in a reasonable amount of time, with larger values included optimistically.

The considered hyper-parameters are:

- CUT
  - **Considered hyper-parameter:** `size-threshold`
  - **Description:** The number of states up to which the belief MDP is explored. After the limit is reached, *cut-offs* are applied to approximate the dynamics of the belief MDP beyond that point. Larger values result in more accurate approximations, but exploration and analysis of the resulting MDP typically take longer. For more information, we refer to Bork et al. [2022].
  - **Considered values:** 2^{8}, 2^{9}, 2^{10}, 2^{11}, 2^{12}, 2^{13}, 2^{14}, 2^{15}, 2^{16}, 2^{17}, 2^{18}, 2^{19}, 2^{20}, 2^{21}, 2^{22}, 2^{23}, 2^{24}, 2^{25}, 2^{26}, 2^{27}, 2^{28}, 2^{29}, 2^{30}, 2^{31}, 2^{32}
- DISCR
  - **Considered hyper-parameter:** `resolution`
  - **Description:** The *resolution* of the belief grid used for discretisation. The resolution describes how coarse the belief space is approximated. With a resolution of `r`, the space is discretised to beliefs only containing probabilities that are multiples of $1/\mathtt{r}$. Typically, the higher the resolution, the better the approximation gets, at the cost of increased runtime. For more information, we refer to Bork et al. [2020].
  - **Considered values:** 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 12, 14, 15, 16, 18, 20, 21, 24, 25, 28, 30, 35, 36, 42, 49

We elaborate on the setup used for our experimental evaluation.

E.1 IMPLEMENTATION

- based on STORM 1.9.0
- built using *CMake* 3.26.3 and *GCC* 12.3.0
- relevant dependency: *Boost* 1.82
- for all relevant computations, in particular solving MDP queries, we use native implementations in STORM, i.e., no external libraries are used.

E.2 SYSTEM

We used several identical systems to conduct the benchmarking. We used the *Slurm* workload manager, version 22.05.4 for enforcing number of used cores and memory limits per instance.

- CPU: Intel Xeon 8468 Sapphire 2.1 GHz, limited to 4 cores per instance. The implementation runs single-threaded.
- RAM: limited to 64 GB per instance

Table 5: Further Details for Experiments (UNFOLD)

						UNFOLD: CUT			UNFOLD: DISCR			
Model	$\lvert S\rvert$	$\lvert Z\rvert$	k	$\lvert\mathsf{E}\rvert$	$\lvert S_{\mathsf{un}}\rvert$	$\lvert\mathcal{B}^{\mathsf{cut}}\rvert$	time	result	$\mathtt{r}$	$\lvert\mathcal{B}^{\mathsf{discr}}\rvert$	time	result
clean6	37	2	2	413	3809	$2{\cdot}10^4$	<1s	≥ 0.86	8	$9{\cdot}10^7$	713s	≤ 1
clean12	73	2	2	1508	$3{\cdot}10^4$	$4{\cdot}10^6$	262s	≥ 0.77	4	$2{\cdot}10^7$	61.9s	≤ 1
incline	25	9	2	497	2094	$4{\cdot}10^6$	21.5s	≥ 0.989	1	2005	<1s	≤ 0.989
obstcl	25	10	2	83	741	$1{\cdot}10^4$	<1s	$= 0.87$	1	725	<1s	≤ 0.87
resrc	721	155	3	2107	$2{\cdot}10^5$	$7{\cdot}10^7$	360s	$\geq 7{\cdot}10^{-15}$	2	$2{\cdot}10^5$	2.7s	≤ 0.0312
resrc	721	155	3	$4{\cdot}10^4$	$6{\cdot}10^6$	$7{\cdot}10^7$	400s	$\geq 2{\cdot}10^{-73}$	2	$6{\cdot}10^6$	81.1s	$\leq 3{\cdot}10^{-5}$
rover	16	9	3	$7{\cdot}10^5$	$1{\cdot}10^7$	$8{\cdot}10^6$	337s	≥ 0.353	2	$3{\cdot}10^7$	237s	≤ 0.853
rover	16	9	3	$2{\cdot}10^7$	?	-	-	-	-	-	-	-
serv	$8{\cdot}10^4$	6016	1	40	$1{\cdot}10^5$	$2{\cdot}10^7$	111s	≥ 0.0474	1	$2{\cdot}10^4$	1.9s	≤ 0.378
serv	$8{\cdot}10^4$	6016	1	68	$3{\cdot}10^5$	$3{\cdot}10^7$	282s	≥ 0.172	3	$1{\cdot}10^7$	158s	≤ 0.636
walk40	84	44	1	82	6847	$4{\cdot}10^5$	97.3s	$= 0.916$	36	$3{\cdot}10^7$	1603s	≤ 0.932
walk120	244	124	1	82	$2{\cdot}10^4$	$1{\cdot}10^6$	846s	≥ 0.867	24	$2{\cdot}10^7$	542s	≤ 0.931
water	34	5	2	$3{\cdot}10^4$	$6{\cdot}10^5$	$3{\cdot}10^7$	327s	$\geq 3{\cdot}10^{-123}$	2	$3{\cdot}10^7$	144s	≤ 1
water	34	5	2	$3{\cdot}10^5$	$5{\cdot}10^6$	8198	114s	≥ 0	1	$5{\cdot}10^6$	162s	≤ 1

- OS: Rocky Linux 8.10
- no GPUs are used for our experiments

F ADDITIONAL RESULTS

Detailed Result Tables Tables 5, 6, and 7 extend our experimental data provided in the main paper. For each table, the first five columns are as in Table 1, except that column 'k' depicts the dimension of the cost bound. The columns 'time' and 'result' in the CUT and DISCR sections of the individual tables repeat the information from Table 2, where each of the three tables consider a different configuration—UNFOLD in Table 5, CA-UNFOLD in Table 6, and CA-BEL-SEQ in Table 7. The remaining columns provide additional information concerning

- the number of states of the transformed POMDP after incorporation of cost-awareness and/or unfolding (Columns '$\lvert S_{\mathsf{un}}\rvert$', '$\lvert S^{\mathsf{ca}}_{\mathsf{un}}\rvert$', and '$\lvert S^{\mathsf{ca}}\rvert$'),
- the number of states of the considered belief MDP abstraction (Columns '$\lvert\mathcal{B}^{\mathsf{cut}}\rvert$' and '$\lvert\mathcal{B}^{\mathsf{discr}}\rvert$'), and
- used hyper parameter for discretization (Column '$\mathtt{r}$', see also Section E).

Comparing unfolding and sequential epoch analysis on cost-aware POMDPs (CA-UNFOLD VS. CA-BEL-SEQ), we observe that the latter handles significantly smaller state spaces while usually achieving similar or tighter approximations in less time.

Additional Plots Figure 2 in the main paper shows the evolution of value bounds obtained over time for two benchmarks. We provide similar plots for the remaining benchmarks in Figures 7 to 13.

Log Files As part of the supplementary material, we provide all raw log files generated for our experimental evaluation in the folder

```
code_data_appendix/logs/raw.
```

They document runs of our implementation for all settings UNFOLD, CA-UNFOLD and CA-BEL-SEQ for both CUT and DISCR with the hyper-parameters described in Section E. In addition to the raw log files (.log) there are also JSON files containing the relevant information extracted from the logs.

The files are named according to the scheme:

$$\text{storm.}\underbrace{\text{CONFIG}}\,\underbrace{\text{C/D}}\,\underbrace{\text{PARAM}}\,.\,\underbrace{\text{EXPERIMENT}}\_\underbrace{\text{INSTANCE}}$$

where CONFIG is one of

Table 6: Further Details for Experiments (CA-UNFOLD)

						CA-UNFOLD: CUT			CA-UNFOLD: DISCR			
Model	$\lvert S\rvert$	$\lvert Z\rvert$	k	$\lvert \mathsf{E}\rvert$	$\lvert S^{\mathsf{ca}}_{\mathsf{un}}\rvert$	$\lvert\mathcal{B}^{\mathsf{cut}}\rvert$	time	result	$\mathtt{r}$	$\lvert\mathcal{B}^{\mathsf{discr}}\rvert$	time	result
clean6	37	2	2	413	5907	$9\cdot 10^{5}$	3.9s	$= 0.929$	30	$3\cdot 10^{7}$	299s	≤ 0.971
clean12	73	2	2	1508	$4\cdot 10^{4}$	$3\cdot 10^{7}$	240s	≥ 0.708	7	$2\cdot 10^{7}$	128s	≤ 1
incline	25	9	2	497	5092	$3\cdot 10^{5}$	1.1s	≥ 0.989	1	4966	<1s	≤ 0.989
obstcl	25	10	2	83	1081	4087	<1s	≥ 0.87	2	1560	<1s	≤ 0.87
resrc	721	155	3	2107	$2\cdot 10^{5}$	$7\cdot 10^{7}$	345s	$\geq 7\cdot 10^{-15}$	2	$2\cdot 10^{5}$	3.6s	≤ 0.0312
resrc	721	155	3	$4\cdot 10^{4}$	$6\cdot 10^{6}$	$7\cdot 10^{7}$	427s	$\geq 2\cdot 10^{-73}$	2	$6\cdot 10^{6}$	101s	$\leq 3\cdot 10^{-5}$
rover	16	9	3	$7\cdot 10^{5}$	$2\cdot 10^{7}$	$1\cdot 10^{7}$	462s	$= 0.861$	30	$1\cdot 10^{7}$	406s	≤ 0.861
rover	16	9	3	$2\cdot 10^{7}$	?	-	-	-	-	-	-	-
serv	$8\cdot 10^{4}$	6016	1	40	$4\cdot 10^{5}$	$2\cdot 10^{7}$	107s	≥ 0.0474	1	$6\cdot 10^{4}$	6.4s	≤ 0.378
serv	$8\cdot 10^{4}$	6016	1	68	$1\cdot 10^{6}$	$3\cdot 10^{7}$	281s	≥ 0.169	3	$4\cdot 10^{7}$	528s	≤ 0.636
walk40	84	44	1	82	$1\cdot 10^{4}$	$8\cdot 10^{5}$	174s	$= 0.916$	30	$3\cdot 10^{7}$	1431s	≤ 0.935
walk120	244	124	1	82	$3\cdot 10^{4}$	$2\cdot 10^{6}$	1681s	≥ 0.869	24	$4\cdot 10^{7}$	1309s	≤ 0.931
water	34	5	2	$3\cdot 10^{4}$	$1\cdot 10^{6}$	$1\cdot 10^{6}$	17.9s	$= 0.166$	42	$1\cdot 10^{6}$	17.5s	≤ 0.166
water	34	5	2	$3\cdot 10^{5}$	$1\cdot 10^{7}$	$1\cdot 10^{7}$	268s	$= 0.181$	42	$1\cdot 10^{7}$	269s	≤ 0.181

Table 7: Further Details for Experiments (CA-BEL-SEQ)

						CA-BEL-SEQ: CUT			CA-BEL-SEQ: DISCR			
Model	$\lvert S\rvert$	$\lvert Z\rvert$	k	$\lvert \mathsf{E}\rvert$	$\lvert S^{\mathsf{ca}}\rvert$	$\lvert\mathcal{B}^{\mathsf{cut}}\rvert$	time	result	$\mathtt{r}$	$\lvert\mathcal{B}^{\mathsf{discr}}\rvert$	time	result
clean6	37	2	2	413	62	$1\cdot 10^{5}$	2.2s	≥ 0.929	42	$2\cdot 10^{4}$	<1s	≤ 0.949
clean12	73	2	2	1508	122	$3\cdot 10^{7}$	381s	≥ 0.88	36	$3\cdot 10^{5}$	4.8s	≤ 0.957
incline	25	9	2	497	68	516	<1s	≥ 0.989	1	69	<1s	≤ 0.989
obstcl	25	10	2	83	43	8194	<1s	≥ 0.87	3	204	<1s	≤ 0.87
resrc	721	155	3	2107	746	4098	<1s	≥ 0.0312	2	660	<1s	≤ 0.0312
resrc	721	155	3	$4\cdot 10^{4}$	746	4098	16.3s	$\geq 3\cdot 10^{-5}$	2	660	4.4s	$\leq 3\cdot 10^{-5}$
rover	16	9	3	$7\cdot 10^{5}$	29	23	12.2s	$= 0.861$	21	30	12.4s	≤ 0.861
rover	16	9	3	$2\cdot 10^{7}$	29	23	466s	$= 0.951$	28	28	456s	≤ 0.951
serv	$8\cdot 10^{4}$	6016	1	40	$4\cdot 10^{5}$	262	1.7s	≥ 0	1	$4\cdot 10^{5}$	15.4s	≤ 0.378
serv	$8\cdot 10^{4}$	6016	1	68	$4\cdot 10^{5}$	262	1.7s	≥ 0	2	$3\cdot 10^{6}$	139s	≤ 0.637
walk40	84	44	1	82	126	8195	<1s	≥ 0.916	42	$2\cdot 10^{6}$	1295s	≤ 0.93
walk120	244	124	1	82	366	$6\cdot 10^{4}$	11.0s	$= 0.895$	30	$1\cdot 10^{6}$	727s	≤ 0.926
water	34	5	2	$3\cdot 10^{4}$	67	84	<1s	$= 0.166$	8	84	<1s	≤ 0.166
water	34	5	2	$3\cdot 10^{5}$	67	84	4.2s	$= 0.181$	10	84	4.3s	≤ 0.181

- unf : UNFOLD,
- caunf : CA-UNFOLD,
- belseq : CA-BEL-SEQ,

and C/D indicates whether the run uses CUT (c) or DISCR (d), followed by the considered hyper-parameter PARAM where for (c) the parameter is the exponent, i.e., we apply a size threshold of 2^{PARAM}. EXPERIMENT $\in \{\text{main}, \text{lvls}, \text{bnds}\}$ denotes the kind of experiment the logfile is referring to, where the latter two refer to the runs for the plots at the bottom of Figure 2. All other results belong to 'main'. Instances are named by the model identifier, followed by an (internally used) identifier of the bounded reachability query and the considered bound values. Note that the implementation uses non-strict inequalities for $>$.

Interactive Table As a more convenient way to view all results obtained during our experiments, we provide interactive tables.

The tables are given in the HTML files

- `code_data_appendix/logs/table/table.html` (for the main experiments)
- `code_data_appendix/logs/lvlstable/table.html` (for the data used in the plot at the bottom left of Figure 2)
- `code_data_appendix/logs/bndstable/table.html` (for the data used in the plot at the bottom right of Figure 2)

and can be viewed in a web browser. The columns of the tables are named similar to the naming scheme for the log files. In addition, we provide columns indicating the best result obtained within 10, 100, 1000 and 1800 seconds. The latter coincides with the values considered in the main paper. Columns can be hidden for a more clearly arranged view. A result cell contains the computed approximation values (also indicating if it is an under- or over-approximation) and the runtime (walltime) in seconds. Clicking on a result shows an overview as well as the raw log used to obtain the result.

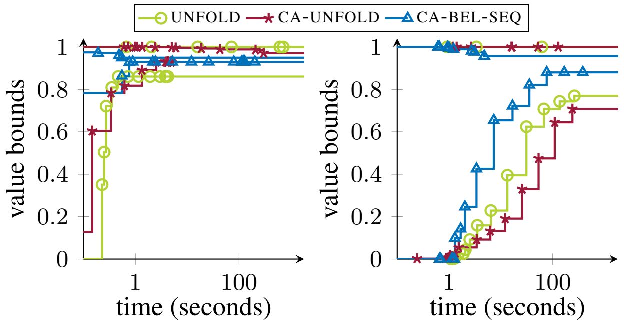


Figure 7: Value bounds obtained for clean6, |E|=497 (left) and clean12, |E|=1508 (right).

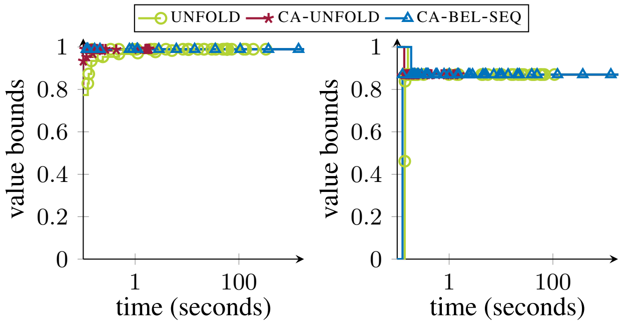


Figure 8: Value bounds obtained for incline, |E|=497 (left) and obstcl, |E|=83 (right).

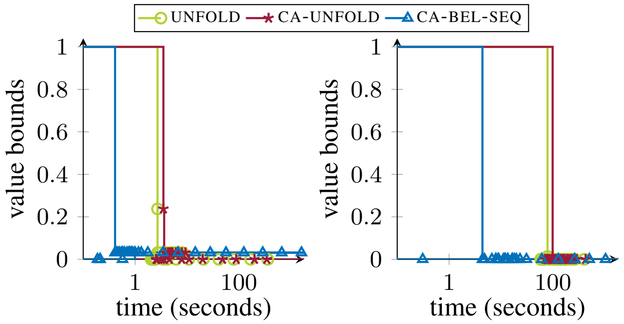


Figure 9: Value bounds obtained for resrc, |E|=2·2107 (left) and |E|=$4{\cdot}10^4$ (right).

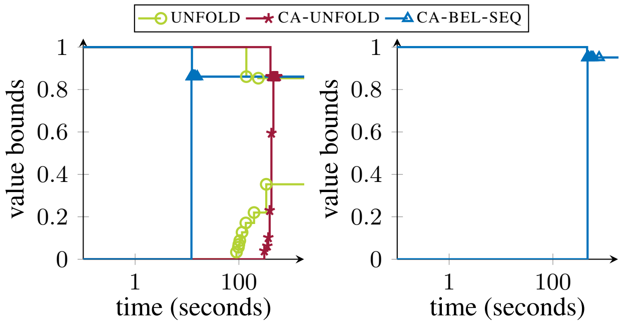


Figure 10: Value bounds obtained for rover, |E|=$7{\cdot}10^5$ (left) and |E|=$2{\cdot}10^7$ (right).

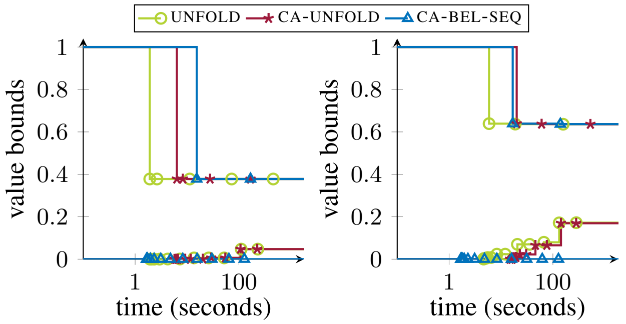


Figure 11: Value bounds obtained for serv, |E|=40 (left) and |E|=68 (right).

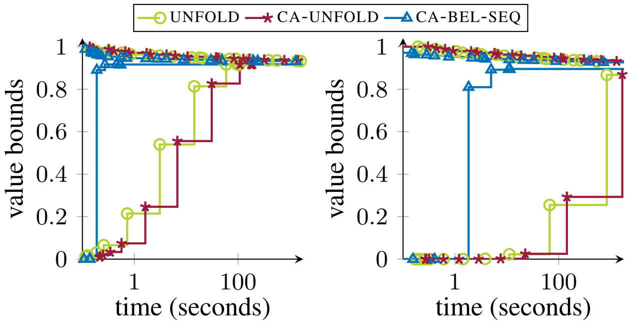


Figure 12: Value bounds obtained for walk40, |E|=82 (left) and walk120, |E|=82 (right).

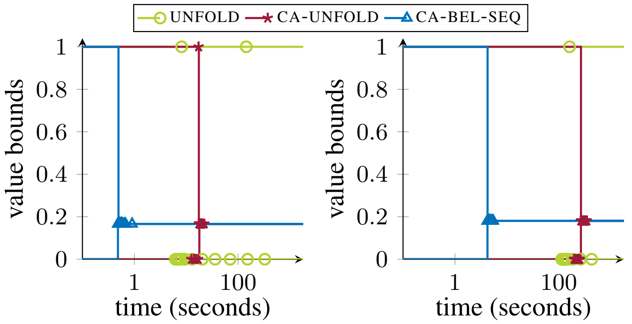


Figure 13: Value bounds obtained for water, |E|=$3{\cdot}10^4$ (left) and |E|=$3{\cdot}10^5$ (right).