POWVER

Technical Report 2017-06

Title: **Multi-objective Robust Strategy Synthesis for Interval Markov Decision Processes**

Author: Ernst Moritz Hahn, Vahid Hashemi, Holger Hermanns, Morteza Lahijanian, Andrea Turrini

Report Number: 2017-06

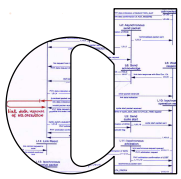
ERC Project: Power to the People. Verified.

ERC Project ID: 695614

Funded Under: H2020-EU.1.1. – EXCELLENT SCIENCE

Host Institution: Universität des Saarlandes, Dependable Systems and Software

Published In: QEST 2017



POWER TO THE PEOPLE.
VERIFIED.
erc

Multi-objective Robust Strategy Synthesis for Interval Markov Decision Processes*

Ernst Moritz Hahn[1,2], Vahid Hashemi[1], Holger Hermanns[1], Morteza Lahijanian[3], and Andrea Turrini[2]

[1] Saarland University, Saarland Informatics Campus, Saarbrücken, Germany
[2] State Key Laboratory of Computer Science, Institute of Software Chinese Academy of Sciences, Beijing, China
[3] Department of Computer Science, University of Oxford, Oxford, UK

Abstract. Interval Markov decision processes (*IMDP*s) generalise classical MDPs by having interval-valued transition probabilities. They provide a powerful modelling tool for probabilistic systems with an additional variation or uncertainty that prevents the knowledge of the exact transition probabilities. In this paper, we consider the problem of multi-objective robust strategy synthesis for interval MDPs, where the aim is to find a robust strategy that guarantees the satisfaction of multiple properties at the same time in face of the transition probability uncertainty. We first show that this problem is **PSPACE**-hard. Then, we provide a value iteration-based decision algorithm to approximate the Pareto set of achievable points. We finally demonstrate the practical effectiveness of our proposals by applying them on several real-world case studies.

1 Introduction

Interval Markov Decision Processes (*IMDP*s) extend the classical *Markov Decision Processes* (*MDP*s) by including uncertainty over the transition probabilities. Instead of a single value for the probability of taking a transition, *IMDP*s allow ranges of probabilities given as closed intervals. *IMDP*s are thus a powerful modelling tool for probabilistic systems with an additional variation or uncertainty concerning the knowledge of exact transition probabilities. They are well suited to represent realistic stochastic systems that, for instance, evolve in unknown environments with bounded behaviour or do not preserve the Markov property.

Since their introduction (under the name of bounded-parameter *MDP*s) [15], *IMDP*s have been receiving a lot of attention in the formal verification community. They are particularly viewed as the appropriate abstraction model for uncertain

* This work is supported by the ERC Advanced Investigators Grant 695614 (POWVER), by the CAS/SAFEA International Partnership Program for Creative Research Teams, by the National Natural Science Foundation of China (Grants No. 61550110506 and 61650410658), by the Chinese Academy of Sciences Fellowship for International Young Scientists, by the CDZ project CAP (GZ 1023), and by EPSRC Mobile Autonomy Program Grant EP/M019918/1.

systems with large state spaces, including continuous dynamical systems, for the purpose of analysis, verification, and control synthesis. Several model checking and control synthesis techniques have been developed [31,32,34] causing a boost in the applications of *IMDP*s, ranging from verification of continuous stochastic systems (e.g., [22]) to robust strategy synthesis for robotic systems (e.g., [24–26,34]).

In recent years, there has been an increasing interest in multi-objective strategy synthesis for probabilistic systems [5,10,13,14,21,27,29,30,33]. The goal is first to provide a complete trade-off analysis of several, possibly conflicting, quantitative properties and then to synthesise a strategy that guarantees the desired behaviour. Such properties, for instance, ask to "find a robot strategy that maximises p_{safe}, the probability of successfully completing a track by safely maneuvering between obstacles, while minimising t_{travel}, the total expected travel time". This example has competing objectives: maximising p_{safe}, which requires the robot to be conservative, and minimising t_{travel}, which causes the robot to be reckless. In such contexts, the interest is in the *Pareto curve* of the possible solution points: the set of all pairs of $(p_{\text{safe}}, t_{\text{travel}})$ for which an increase in the value of p_{safe} must induce an increase in the value of t_{travel}, and vice versa. Given a point on the curve, the computation of the corresponding strategy is asked.

Existing multi-objective synthesis frameworks are limited to *MDP* models. The algorithms use iterative methods (similar to value iteration) for the computation of the Pareto curve and rely on reductions to linear programming for strategy synthesis. As discussed above, *MDP*s, however, are constrained to single-valued transition probabilities, posing severe limitations for many real-world systems.

In this paper, we present a novel technique for multi-objective strategy synthesis for *IMDP*s. Our aim is to synthesise a robust strategy that guarantees the satisfaction of the multi-objective property, despite the additional uncertainty over the transition probabilities. Our approach views the uncertainty as making adversarial choices among the available transition probability distributions induced by the intervals, as the system evolves along state transitions. We refer to this as the *controller synthesis* semantics. We first analyse the problem complexity, proving that it is **PSPACE**-hard and then develop a value iteration-based decision algorithm to approximate the Pareto curve. We present promising results on a variety of case studies, obtained by prototypical implementations of all algorithms, to show the effectiveness of our approach.

Related work. Related work can be grouped into two main categories: uncertain Markov model formalisms and model checking/synthesis algorithms.

Firstly, regarding the modelling frameworks, various probabilistic modelling formalisms with uncertain transitions are studied in the literature. Interval Markov Chains (*IMC*s) [19,20] or abstract Markov chains [12] extend standard discrete-time Markov Chains (*MC*s) with interval uncertainties. They do not feature the non-deterministic choices of transitions. Uncertain *MDP*s [32] allow more general sets of distributions to be associated with each transition, not only those described by intervals. They usually are restricted to *rectangular uncertainty sets* requiring that the uncertainty is linear and independent for any two transitions of any two states. Parametric *MDP*s [16], to the contrary, allow

such dependencies as every probability is described as a rational function of a finite set of global parameters. *IMDP*s extend *IMC*s by inclusion of nondeterminism and are a subset of uncertain *MDP*s and parametric *MDP*s.

Secondly, regarding the algorithms, several verification methods for uncertain Markov models have been proposed. The problems of computing reachability probabilities and expected total reward for *IMC*s and *IMDP*s were first investigated in [8,35]. Then, several of their PCTL and LTL model checking algorithms were introduced in [2,6,8] and [22,32,34], respectively. As regards to strategy synthesis algorithms, the work in [16,28] considered synthesis for parametric *MDP*s and *MDP*s with ellipsoidal uncertainty in the verification community. In the control community, such synthesis problems were mostly studied for uncertain Markov models in [15,28,35] with the aim to maximise expected finite-horizon (un)discounted rewards. All these works, however, consider solely single objective properties, and their extension to multi-objective synthesis is not trivial.

Multi-objective model checking of probabilistic models with respect to various quantitative objectives has been recently investigated in a few works. The works in [11,13,14,21] focused on multi-objective verification of ordinary *MDP*s. In [7], these algorithms were extended to the more general models of 2-player stochastic games. These models, however, cannot capture the continuous uncertainty in the transition probabilities as *IMDP*s do. For the purposes of synthesis though, it is possible to transform an *IMDP* into a 2-player stochastic game; nevertheless, such a transformation raises an extra exponential factor to the complexity of the decision problem. This exponential blowup has been avoided in our setting.

2 Preliminaries

For a set X, denote by $\mathrm{Disc}(X)$ the sets of discrete probability distributions over X. A discrete probability distribution ρ is a function $\rho\colon X \to \mathbb{R}_{\geq 0}$ such that $\sum_{x\in X} \rho(x) = 1$; for $X' \subseteq X$, we write $\rho(X')$ for $\sum_{x\in X'} \rho(x)$. Given $\rho \in \mathrm{Disc}(X)$, we denote by $\mathrm{Supp}(\rho)$ the set $\{\, x \in X \mid \rho(x) > 0 \,\}$, and by δ_x, where $x \in X$, the *Dirac* distribution such that $\delta_x(y) = 1$ for $y = x$, 0 otherwise. For a distribution ρ, we also write $\rho = \{\, (x, p_x) \mid x \in X \,\}$ where p_x is the probability of x.

For a vector $\mathbf{x} \in \mathbb{R}^n$ we denote by x_i, its i-th component, and we call $\mathbf{x}$ a *weight vector* if $x_i \geq 0$ for all i and $\sum_{i=1}^{n} x_i = 1$. The Euclidean inner product $\mathbf{x} \cdot \mathbf{y}$ of two vectors $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ is defined as $\sum_{i=1}^{n} x_i \cdot y_i$. For a set of vectors $S = \{\mathbf{s}_1, \ldots, \mathbf{s}_t\} \subseteq \mathbb{R}^n$, we say that $\mathbf{s} \in \mathbb{R}^n$ is a *convex combination* of elements of S, if $\mathbf{s} = \sum_{i=1}^{t} w_i \cdot \mathbf{s}_i$ for some weight vector $\mathbf{w} \in \mathbb{R}^t_{\geq 0}$. Furthermore, we denote by $S{\downarrow}$ the *downward closure* of the convex hull of S which is defined as $S{\downarrow} = \{\, \mathbf{y} \in \mathbb{R}^n \mid \mathbf{y} \leq \mathbf{z} \text{ for some convex combination } \mathbf{z} \text{ of } S \,\}$. For a given convex set X, we say that a point $\mathbf{x} \in X$ is on the boundary of X, denoted by $\mathbf{x} \in \partial X$, if for every $\varepsilon > 0$ there is a point $\mathbf{y} \notin X$ such that the Euclidean distance between $\mathbf{x}$ and $\mathbf{y}$ is at most ε. Given a downward closed set $X \in \mathbb{R}^n$, for any $\mathbf{z} \in \mathbb{R}^n$ such that $\mathbf{z} \in \partial X$ or $\mathbf{z} \notin X$, there is a weight vector $\mathbf{w} \in \mathbb{R}^n$ such that $\mathbf{w} \cdot \mathbf{z} \geq \mathbf{w} \cdot \mathbf{x}$ for all $\mathbf{x} \in X$ [3]. We say that $\mathbf{w}$ separates $\mathbf{z}$ from $X{\downarrow}$. Given a set $Y \subseteq \mathbb{R}^k$, we call a vector $\mathbf{y} \in Y$ *Pareto optimal* in Y if there does not exist a vector $\mathbf{z} \in Y$ such that

$\mathbf{y} \leq \mathbf{z}$ and $\mathbf{y} \neq \mathbf{z}$. We define the *Pareto set* or *Pareto curve* of Y to be the set of all Pareto optimal vectors in Y, i.e., Pareto set $\mathcal{Y} = \{\, \mathbf{y} \in Y \mid \mathbf{y} \text{ is Pareto optimal} \,\}$.

2.1 Interval Markov Decision Processes

We now define *Interval Markov Decision Processes* (*IMDP*s) as an extension of *MDP*s, which allows for the inclusion of transition probability uncertainties as *intervals*. *IMDP*s belong to the family of uncertain *MDP*s and allow to describe a set of *MDP*s with identical (graph) structures that differ in distributions associated with transitions. Formally,

Definition 1 (*IMDP*s). *An* Interval Markov Decision Process *(IMDP)* $\mathcal{M}$ *is a tuple* $(S, \bar{s}, \mathcal{A}, I)$*, where* S *is a finite set of* states*,* $\bar{s} \in S$ *is the initial state,* $\mathcal{A}$ *is a finite set of* actions*, and* $I \colon S \times \mathcal{A} \times S \to \mathbb{I} \cup \{[0,0]\}$ *is a total* interval transition probability function where $\mathbb{I} = \{\, [a,b] \mid 0 < a \leq b \leq 1 \,\}$.

Given $s \in S$ and $a \in \mathcal{A}$, we call $\mathfrak{h}_s^a \in \mathrm{Disc}(S)$ a *feasible distribution* reachable from s by a, denoted by $s \xrightarrow{a} \mathfrak{h}_s^a$, if, for each state $s' \in S$, we have $\mathfrak{h}_s^a(s') \in I(s,a,s')$. We denote the set of feasible distributions for state s and action a by $\mathcal{H}_s^a$, i.e., $\mathcal{H}_s^a = \{\, \mathfrak{h}_s^a \in \mathrm{Disc}(S) \mid s \xrightarrow{a} \mathfrak{h}_s^a \,\}$ and we denote the set of available actions at state $s \in S$ by $\mathcal{A}(s)$, i.e., $\mathcal{A}(s) = \{\, a \in \mathcal{A} \mid \mathcal{H}_s^a \neq \emptyset \,\}$. We assume that $\mathcal{A}(s) \neq \emptyset$ for all $s \in S$. We define the *size* of $\mathcal{M}$, written $|\mathcal{M}|$, as the number of non-zero entries of I, i.e., $|\mathcal{M}| = |\{\, (s,a,s',\iota) \in S \times \mathcal{A} \times S \times \mathbb{I} \mid I(s,a,s') = \iota \,\}| \in \mathcal{O}(|S|^2 \cdot |\mathcal{A}|)$.

A *path* ξ in $\mathcal{M}$ is a finite or infinite sequence of alternating states and actions $\xi = s_0 a_0 s_1 \ldots$, ending with a state if finite, such that for each $i \geq 0$, $I(s_i, a_i, s_{i+1}) \in \mathbb{I}$. The i-th state (action) along the path ξ is denoted by $\xi[i]$ ($\xi(i)$) and, if the path is finite, we denote by $\mathit{last}(\xi)$ its last state. The sets of all finite and infinite paths in $\mathcal{M}$ are denoted by *FPaths* and *IPaths*, respectively.

The nondeterministic choices between available actions and feasible distributions present in an *IMDP* are resolved by strategies and natures, respectively.

Definition 2 (Strategy and Nature in *IMDP*s). *Given an IMDP* $\mathcal{M}$*, a* strategy *is a function* $\sigma \colon \mathit{FPaths} \to \mathrm{Disc}(\mathcal{A})$ *such that for each* $\xi \in \mathit{FPaths}$*,* $\sigma(\xi) \in \mathrm{Disc}(\mathcal{A}(\mathit{last}(\xi)))$*. A* nature *is a function* $\pi \colon \mathit{FPaths} \times \mathcal{A} \to \mathrm{Disc}(S)$ *such that for each* $\xi \in \mathit{FPaths}$ *and* $a \in \mathcal{A}(s)$*,* $\pi(\xi, a) \in \mathcal{H}_s^a$ *where* $s = \mathit{last}(\xi)$*. The sets of all strategies and all natures are denoted by* Σ *and* Π*, respectively.*

Given a finite path ξ of an *IMDP*, a strategy σ, and a nature π, the system evolution proceeds as follows: let $s = \mathit{last}(\xi)$. First, an action $a \in \mathcal{A}(s)$ is chosen probabilistically by σ. Then, π resolves the uncertainties and chooses one feasible distribution $\mathfrak{h}_s^a \in \mathcal{H}_s^a$. Finally, the next state s' is chosen according to the distribution $\mathfrak{h}_s^a$, and the path ξ is extended by s'.

A strategy σ and a nature π induce a probability measure over paths as follows. The basic measurable events are the cylinder sets of finite paths, where the *cylinder set* of a finite path ξ is the set $\mathit{Cyl}_\xi = \{\, \xi' \in \mathit{IPaths} \mid \xi \text{ is a prefix of } \xi' \,\}$. The probability $\Pr_{\mathcal{M}}^{\sigma,\pi}$ of a state s' is defined to be $\Pr_{\mathcal{M}}^{\sigma,\pi}[\mathit{Cyl}_{s'}] = \delta_{\bar{s}}(s')$ and

the probability $\Pr^{\sigma,\pi}_{\mathcal{M}}[Cyl_{\xi as'}]$ of traversing a finite path $\xi as'$ is defined to be $\Pr^{\sigma,\pi}_{\mathcal{M}}[Cyl_{\xi as'}] = \Pr^{\sigma,\pi}_{\mathcal{M}}[Cyl_{\xi}] \cdot \sigma(\xi)(a) \cdot \pi(\xi, a)(s')$. Then, $\Pr^{\sigma,\pi}_{\mathcal{M}}$ extends uniquely to the σ-field generated by cylinder sets.

In order to model additional quantitative measures of an *IMDP*, we associate rewards to the enabled actions. This is done by means of *reward structures*.

Definition 3 (Reward Structure). *A* reward structure *for an IMDP is a function* $\mathtt{r}\colon S \times \mathcal{A} \to \mathbb{R}$ *that assigns to each state-action pair* (s, a)*, where* $s \in S$ *and* $a \in \mathcal{A}(s)$*, a reward* $\mathtt{r}(s, a) \in \mathbb{R}$*. Given a path* ξ *and* $k \in \mathbb{N} \cup \{\infty\}$*, the total accumulated reward in* k *steps for* ξ *over* $\mathtt{r}$ *is* $\mathtt{r}[k](\xi) = \sum_{i=0}^{k-1} \mathtt{r}(\xi[i], \xi(i))$.

Note that we allow negative rewards in this definition, but that due to later assumptions their use is restricted.

As an example of *IMDP* with a reward structure, consider the *IMDP* $\mathcal{M}$ depicted in Fig. 1. The set of states is $S = \{s, t, u\}$ with s being the initial one. The set of actions is $\mathcal{A} = \{a, b\}$, and the non-zero transition probability intervals are $I(s, a, t) = [\frac{1}{3}, \frac{2}{3}]$, $I(s, a, u) = [\frac{1}{10}, 1]$, $I(s, b, t) = [\frac{2}{5}, \frac{3}{5}]$, $I(s, b, u) = [\frac{1}{4}, \frac{2}{3}]$, $I(t, a, t) = I(u, b, u) = [1, 1]$, and $I(t, b, t) = I(u, a, u) = [0, 0]$. The underlined numbers indicate the reward structure $\mathtt{r}$ such that $\mathtt{r}(s, a) = 3$, $\mathtt{r}(s, b) = 1$, and $\mathtt{r}(t, a) = \mathtt{r}(u, b) = 0$. Note that since $\mathcal{H}^b_t = \mathcal{H}^a_u = \emptyset$, then $\mathtt{r}(t, b)$ and $\mathtt{r}(u, a)$ are undefined.

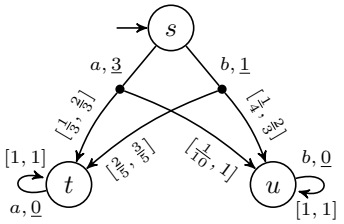


Fig. 1: An example of *IMDP*.

3 Multi-objective Robust Strategy Synthesis for IMDPs

In this paper, we consider two main classes of properties for *IMDP*s; the *probability of reaching a target* and the *expected total reward*. The reason that we focus on these properties is that their algorithms usually serve as the basis for more complex properties. For instance, they can be easily extended to answer queries with linear temporal logic properties as shown in [11]. To this aim, we lift the satisfaction definitions of these two classes of properties from *MDP*s in [13, 14] to *IMDP*s by encoding the notion of robustness for strategies.

Note that all proofs are contained in the extended version of the paper [17].

Definition 4 (Reachability Predicate & its Robust Satisfaction). *A* reachability predicate $[T]^{\leq k}_{\sim p}$ *consists of a set of target states* $T \subseteq S$*, a relational operator* $\sim \in \{\leq, \geq\}$*, a rational probability bound* $p \in [0, 1] \cap \mathbb{Q}$ *and a time bound* $k \in \mathbb{N} \cup \{\infty\}$*. It indicates that the probability of reaching* T *within* k *time steps satisfies* $\sim p$.

Robust satisfaction of $[T]^{\leq k}_{\sim p}$ *by IMDP* $\mathcal{M}$ *under strategy* $\sigma \in \Sigma$ *is denoted by* $\mathcal{M}\!\downarrow_{\sigma} \models_{\Pi} [T]^{\leq k}_{\sim p}$ *and indicates that the probability of the set of all paths that reach* T *under* σ *satisfies the bound* $\sim p$ *for every choice of nature* $\pi \in \Pi$*. Formally,* $\mathcal{M}\!\downarrow_{\sigma} \models_{\Pi} [T]^{\leq k}_{\sim p}$ *iff* $\Pr^{\sigma}_{\mathcal{M}}(\Diamond^{\leq k}\, T) \sim p$ *where* $\Pr^{\sigma}_{\mathcal{M}}(\Diamond^{\leq k}\, T) = \mathrm{opt}_{\pi \in \Pi} \Pr^{\sigma,\pi}_{\mathcal{M}}\{\, \xi \in \mathit{IPaths} \mid \exists i \leq k : \xi[i] \in T \,\}$ *and* $\mathrm{opt} = \min$ *if* $\sim = \geq$ *and* $\mathrm{opt} = \max$ *if* $\sim = \leq$*. Furthermore,* σ *is referred to as a robust strategy.*

Definition 5 (Reward Predicate & its Robust Satisfaction). *A* reward predicate $[\mathtt{r}]^{\leq k}_{\sim r}$ *consists of a reward structure* $\mathtt{r}$*, a time bound* $k \in \mathbb{N} \cup \{\infty\}$*, a relational operator* $\sim \in \{\leq, \geq\}$ *and a reward bound* $r \in \mathbb{Q}$*. It indicates that the expected total accumulated reward within* k *steps satisfies* $\sim r$*.*

Robust satisfaction of $[\mathtt{r}]^{\leq k}_{\sim r}$ *by IMDP* $\mathcal{M}$ *under strategy* $\sigma \in \Sigma$ *is denoted by* $\mathcal{M}\restriction_\sigma \models_\Pi [\mathtt{r}]^{\leq k}_{\sim r}$ *and indicates that the expected total reward over the set of all paths under* σ *satisfies the bound* $\sim r$ *for every choice of nature* $\pi \in \Pi$*. Formally,* $\mathcal{M}\restriction_\sigma \models_\Pi [\mathtt{r}]^{\leq k}_{\sim r}$ *iff* $ExpTot^{\sigma,k}_{\mathcal{M}}[\mathtt{r}] \sim r$ *where* $ExpTot^{\sigma,k}_{\mathcal{M}}[\mathtt{r}] = \mathrm{opt}_{\pi \in \Pi} \int_\xi \mathtt{r}[k](\xi) \,\mathrm{dPr}^{\sigma,\pi}_{\mathcal{M}}$ *and* $\mathrm{opt} = \min$ *if* $\sim = \geq$ *and* $\mathrm{opt} = \max$ *if* $\sim = \leq$*. Furthermore,* σ *is referred to as the robust strategy.*

For the purpose of algorithm design, we also consider weighted sum of rewards.

Definition 6 (Weighted Reward Sum). *Given a weight vector* $\mathbf{w} \in \mathbb{R}^n$*, vector of time bounds* $\mathbf{k} = (k_1, \ldots, k_n) \in (\mathbb{N} \cup \{\infty\})^n$ *and reward structures* $\mathtt{r} = (\mathtt{r}_1, \ldots, \mathtt{r}_n)$ *for IMDP* $\mathcal{M}$*, the* weighted reward sum $\mathbf{w} \cdot \mathtt{r}[\mathbf{k}]$ *over a path* ξ *is defined as* $\mathbf{w} \cdot \mathtt{r}[\mathbf{k}](\xi) = \sum_{i=1}^n w_i \cdot \mathtt{r}_i[k](\xi)$*. The* expected total weighted sum *is defined as* $ExpTot^{\sigma,\mathbf{k}}_{\mathcal{M}}[\mathbf{w} \cdot \mathtt{r}] = \max_{\pi \in \Pi} \int_\xi \mathbf{w} \cdot \mathtt{r}[\mathbf{k}](\xi) \,\mathrm{dPr}^{\sigma,\pi}_{\mathcal{M}}$ *for bounds* $\leq$ *and accordingly minimises over natures for* $\geq$*; for a given strategy* σ*, we have:* $ExpTot^{\sigma,\mathbf{k}}_{\mathcal{M}}[\mathbf{w} \cdot \mathtt{r}] = \sum_{i=1}^n w_i \cdot ExpTot^{\sigma,k_i}_{\mathcal{M}}[\mathtt{r}_i]$*.*

3.1 Multi-objective Queries

Multi-objective properties for *IMDP*s essentially require multiple predicates to be satisfied at the same time under the same strategy for every choice of the nature. We now explain how to formalise multi-objective queries for *IMDP*s.

Definition 7 (Multi-objective Predicate). *A* multi-objective predicate *is a vector* $\varphi = (\varphi_1, \ldots, \varphi_n)$ *of reachability or reward predicates. We say that* φ *is satisfied by IMDP* $\mathcal{M}$ *under strategy* σ *for every choice of nature* $\pi \in \Pi$*, denoted by* $\mathcal{M}\restriction_\sigma \models_\Pi \varphi$ *if, for each* $1 \leq i \leq n$*, it is* $\mathcal{M}\restriction_\sigma \models_\Pi \varphi_i$*. We refer to* σ *as a robust strategy. Furthermore, we call* φ *a basic multi-objective predicate if it is of the form* $([\mathtt{r}_1]^{\leq k_1}_{\geq r_1}, \ldots, [\mathtt{r}_n]^{\leq k_n}_{\geq r_n})$*, i.e., it includes only lower-bounded reward predicates.*

We formulate multi-objective queries for *IMDP*s in three ways, namely *synthesis queries*, *quantitative queries* and *Pareto queries*. Due to lack of space, we only focus on the synthesis queries and discuss the other types of queries in [17, Appendix C].

Definition 8 (Synthesis Query). *Given an IMDP* $\mathcal{M}$ *and a multi-objective predicate* φ*, the* synthesis query *asks if there exists a robust strategy* $\sigma \in \Sigma$ *such that* $\mathcal{M}\restriction_\sigma \models_\Pi \varphi$*.*

Note that the synthesis queries check for the existence of a robust strategy that satisfies a multi-objective predicate φ for every resolution of nature. In order to avoid unusual behaviours in strategy synthesis such as infinite total expected reward, we restrict the usage of rewards by assuming reward-finiteness for the strategies that satisfy the reachability predicates in φ.

Assumption 1 (Reward-finiteness) *Suppose that an IMDP $\mathcal{M}$ and a synthesis query φ are given. Let $\varphi = ([T_1]^{\leq k_1}_{\sim p_1}, \ldots, [T_n]^{\leq k_n}_{\sim p_n}, [\mathtt{r}_{n+1}]^{\leq k_{n+1}}_{\sim r_{n+1}}, \ldots, [\mathtt{r}_m]^{\leq k_m}_{\sim r_m})$. We say that φ is reward-finite if for each $n+1 \leq i \leq m$ such that $k_i = \infty$, $\sup\{\, ExpTot^{\sigma,k_i}_{\mathcal{M}}[\mathtt{r}_i] \mid \mathcal{M}\restriction_\sigma \models_{\Pi} ([T_1]^{\leq k_1}_{\sim p_1}, \ldots, [T_n]^{\leq k_n}_{\sim p_n}) \,\} < \infty$.*

Due to lack of space, we provide in [17, Appendix B] a method to check for this assumption, a preprocessing procedure that removes actions with non-zero rewards from the end components of the *IMDP*, and a proof for its correctness. Therefore, in the rest of the paper, we assume that all queries are reward-finite, and the *IMDP* does not include actions with non-zero rewards in its end components. Furthermore, for the soundness of our analysis we also require that for any *IMDP* $\mathcal{M}$ and φ given as in Assumption 1: *(i)* each reward structure $\mathtt{r}_i$ assigns only non-negative values; *(ii)* φ is reward-finite; and *(iii)* for indices $n+1 \leq i \leq m$ such that $k_i = \infty$, either all $\sim_i$s are $\leq$ or all are $\geq$.

3.2 Robust Strategy Synthesis

We first study the computational complexity of multi-objective robust strategy synthesis problem for *IMDP*s. Formally,

Theorem 9. *Given an IMDP $\mathcal{M}$ and a multi-objective predicate φ, the problem of synthesising a strategy $\sigma \in \Sigma$ such that $\mathcal{M}\restriction_\sigma \models_{\Pi} \varphi$ is **PSPACE**-hard.*

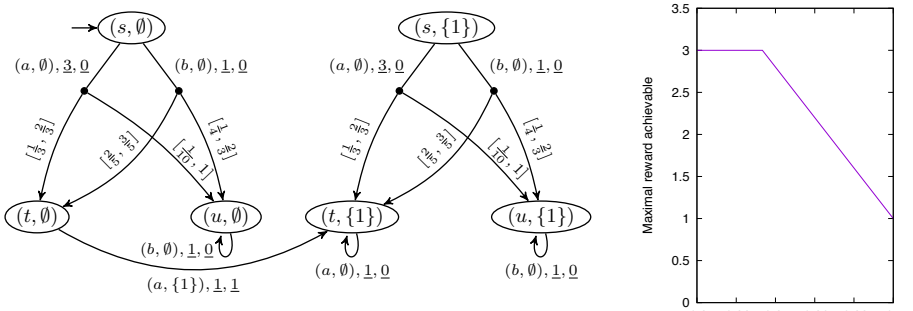
As the first step towards derivation of a solution approach for the robust strategy synthesis problem, we need to convert all reachability predicates to reward predicates and therefore, to transform an arbitrarily given query to a query over a basic predicate on a modified *IMDP*. This can be simply done by adding, once for all, a reward of one at the time of reaching the target set and also negating the objective of predicates with upper-bounded relational operators. We correct and extend the procedure in [14] to reduce a general multi-objective predicate on an *IMDP* model to a basic form.

Proposition 10. *Given an IMDP $\mathcal{M} = (S, \bar{s}, \mathcal{A}, I)$ and a multi-objective predicate $\varphi = ([T_1]^{\leq k_1}_{\sim_1 p_1}, \ldots, [T_n]^{\leq k_n}_{\sim_n p_n}, [\mathtt{r}_{n+1}]^{\leq k_{n+1}}_{\sim_{n+1} r_{n+1}}, \ldots, [\mathtt{r}_m]^{\leq k_m}_{\sim_m r_m})$, let $\mathcal{M}' = (S', \bar{s}', \mathcal{A}', I')$ be the IMDP whose components are defined as follows: $S' = S \times 2^{\{1,\ldots,n\}}$; $\bar{s}' = (\bar{s}, \emptyset)$; $\mathcal{A}' = \mathcal{A} \times 2^{\{1,\ldots,n\}}$; and for all $s, s' \in S$, $a \in \mathcal{A}$, and $v, v', v'' \subseteq \{1, \ldots, n\}$,*

$$I'((s,v),(a,v'),(s',v'')) = \begin{cases} I(s,a,s') & \text{if } v' = \{\, i \mid s \in T_i \,\} \setminus v \text{ and } v'' = v \cup v', \\ [0,0] & \text{otherwise.} \end{cases}$$

Now, let $\varphi' = ([\mathtt{r}_{T_1}]^{\leq k_1+1}_{\geq p'_1}, \ldots, [\mathtt{r}_{T_n}]^{\leq k_n+1}_{\geq p'_n}, [\bar{\mathtt{r}}_{n+1}]^{\leq k_{n+1}}_{\geq r'_{n+1}}, \ldots, [\bar{\mathtt{r}}_m]^{\leq k_m}_{\geq r'_m})$ where, for each $i \in \{1, \ldots, n\}$,

$$p'_i = \begin{cases} p_i & \text{if } \sim_i = \geq, \\ -p_i & \text{if } \sim_i = \leq; \end{cases} \quad \text{and} \quad \mathtt{r}_{T_i}((s,v),(a,v')) = \begin{cases} 1 & \text{if } i \in v' \text{ and } \sim_i = \geq, \\ -1 & \text{if } i \in v' \text{ and } \sim_i = \leq, \\ 0 & \text{otherwise;} \end{cases}$$



(a) The transformed *IMDP* $\mathcal{M}'$ (b) Pareto curve.

Fig. 2: Example of *IMDP* transformation. (a) The *IMDP* $\mathcal{M}'$ generated from $\mathcal{M}$ shown in Fig. 1. (b) Pareto curve for the property $([\mathtt{r}_T]^{\leq 2}_{\max}, [\mathtt{r}]^{\leq 1}_{\max})$.

and, for each $j \in \{n+1, \ldots, m\}$,

$$r'_j = \begin{cases} r_j & \text{if } \sim_j = \geq, \\ -r_j & \text{if } \sim_j = \leq; \end{cases} \quad \text{and} \quad \bar{\mathtt{r}}_j((s,v),(a,v')) = \begin{cases} \mathtt{r}_j(s,a) & \text{if } \sim_j = \geq, \\ -\mathtt{r}_j(s,a) & \text{if } \sim_j = \leq. \end{cases}$$

Then φ *is satisfiable in* $\mathcal{M}$ *if and only if* φ' *is satisfiable in* $\mathcal{M}'$.

We therefore need to only consider the basic multi-objective predicates of the form $([\mathtt{r}_1]^{\leq k_1}_{\geq r_1}, \ldots, [\mathtt{r}_n]^{\leq k_n}_{\geq r_n})$ for the purpose of robust strategy synthesis. For a basic multi-objective predicate, we define its Pareto curve as follows.

Definition 11 (Pareto Curve of a Multi-objective Predicate). *Given an IMDP* $\mathcal{M}$ *and a basic multi-objective predicate* $\varphi = ([\mathtt{r}_1]^{\leq k_1}_{\geq r_1}, \ldots, [\mathtt{r}_n]^{\leq k_n}_{\geq r_n})$*, we define the set of achievable values with respect to* φ *as* $A_{\mathcal{M},\varphi} = \{\, (r_1, \ldots, r_n) \in \mathbb{R}^n \mid ([\mathtt{r}_1]^{\leq k_1}_{\geq r_1}, \ldots, [\mathtt{r}_n]^{\leq k_n}_{\geq r_n}) \text{ is satisfiable} \,\}$*. We define the Pareto curve of* φ *to be the Pareto curve of* $A_{\mathcal{M},\varphi}$ *and denote it by* $\mathcal{P}_{\mathcal{M},\varphi}$.

To illustrate the transformation presented in Proposition 10, consider again the *IMDP* depicted in Fig. 1. Assume that the target set is $T = \{t\}$ and consider the property $\varphi = ([T]^{\leq 1}_{\geq \frac{1}{3}}, [\mathtt{r}]^{\leq 1}_{\geq \frac{1}{4}})$. The reduction converts φ to the property $\varphi' = ([\mathtt{r}_T]^{\leq 2}_{\geq \frac{1}{3}}, [\mathtt{r}]^{\leq 1}_{\geq \frac{1}{4}})$ on the modified $\mathcal{M}'$ depicted in Fig. 2a. We show two different reward structures $\bar{\mathtt{r}}$ and $\mathtt{r}_T$ besides each action, respectively. In Fig. 2b we show the Pareto curve for this property. As we see, until required probability $\frac{1}{3}$ to reach T, the maximal reward value is 3. Afterwards, the reward obtainable linearly decreases, until at required probability $\frac{2}{5}$ it is just 1. For higher required probabilities, the problem becomes infeasible. The reason for this behaviour is that, up to minimal probability $\frac{1}{3}$, action a can be chosen in state s, because the lower interval bound to reach t is $\frac{1}{3}$, which in turn leads to a reward of 3 being obtained. For higher reachability probabilities required, choosing action b with a

Algorithm 1: Algorithm for solving robust synthesis queries

Input: An *IMDP* $\mathcal{M}$, multi-objective predicate $\varphi = ([\mathtt{r}_1]^{\leq k_1}_{\geq r_1}, \ldots, [\mathtt{r}_n]^{\leq k_n}_{\geq r_n})$
Output: true if there exists a strategy $\sigma \in \Sigma$ such that $\mathcal{M}\restriction_\sigma \models_\Pi \varphi$, false if not.

1 **begin**
2 $X := \emptyset$; $\mathbf{r} := (\mathtt{r}_1, \ldots, \mathtt{r}_n)$;
3 $\mathbf{k} := (k_1, \ldots, k_n)$; $\mathbf{r} := (r_1, \ldots, r_n)$;
4 **while** $\mathbf{r} \notin X{\downarrow}$ **do**
5 Find $\mathbf{w}$ separating $\mathbf{r}$ from $X{\downarrow}$;
6 Find strategy σ maximising $ExpTot^{\sigma,\mathbf{k}}_{\mathcal{M}}[\mathbf{w} \cdot \mathbf{r}]$;
7 $\mathbf{g} := (ExpTot^{\sigma,k_i}_{\mathcal{M}}[\mathtt{r}_i])_{1 \leq i \leq n}$;
8 **if** $\mathbf{w} \cdot \mathbf{g} < \mathbf{w} \cdot \mathbf{r}$ **then**
9 **return** false;
10 $X := X \cup \{\mathbf{g}\}$;
11 **return** true;

certain probability is required, which however provides a lower reward. There is no strategy with which t is reached with a probability larger than $\frac{2}{5}$.

It is not difficult to see that the Pareto curve is in general an infinite set, and therefore, it is usually not possible to derive an exact representation of it in polynomial time. However, it can be shown that an ε-approximation of it can be computed efficiently [11]. In the rest of this section, we describe an algorithm to solve the synthesis query. We follow the well-known *normalisation* approach in order to solve the multi-objective predicate which is essentially based on normalising multiple objectives into one single objective. It is known that the optimal solution of the normalised (single-objective) predicate, if it exists, is the Pareto optimal solution of the multi-objective predicate [9].

The robust synthesis procedure is detailed in Algorithm 1. It basically aims to construct a sequential approximation to the Pareto curve $\mathcal{P}_{\mathcal{M},\varphi}$ while the quality of approximations gets better and more precise along the iterations. In other words, along the course of Algorithm 1 a sequence of weight vectors $\mathbf{w}$ are generated and corresponding to each of them, a $\mathbf{w}$-weighted sum of n objectives is optimised through lines 6-7. The optimal strategy σ is then used to generate a point $\mathbf{g}$ on the Pareto curve $\mathcal{P}_{\mathcal{M},\varphi}$. We collect all these points in the set X. The multi-objective predicate φ is satisfiable once we realise that $\mathbf{r}$ belongs to $X{\downarrow}$.

The optimal strategies for the multi-objective robust synthesis queries are constructed following the approach of [14] and as a result of termination of Algorithm 1. In particular, when Algorithm 1 terminates, a sequence of points $\mathbf{g}^1, \ldots, \mathbf{g}^t$ on the Pareto curve $\mathcal{P}_{\mathcal{M},\varphi}$ are generated each of which corresponds to a deterministic strategy $\sigma_{\mathbf{g}^j}$ for the current point $\mathbf{g}^j$. The resulting optimal strategy σ_{opt} is subsequently constructed from these using a randomised weight vector $\alpha \in \mathbb{R}^t$ satisfying $r_i \leq \sum_{j=1}^{t} \alpha_i \cdot {g_i}^j$ [17, Appendix E].

Remark 12. It is worthwhile to mention that the synthesis query for *IMDP*s cannot be solved on the *MDP*s generated from *IMDP*s by computing all feasible

extreme transition probabilities and then applying the algorithm in [14]. The latter is a valid approach provided the cooperative semantics is applied for resolving the two sources of nondeterminism in *IMDP*s. With respect to the competitive semantics needed here, one can instead transform *IMDP*s to $2\frac{1}{2}$-player games [1] and then along the lines of the previous approach apply the algorithm in [7]. Unfortunately, the transformation to (*MDP*s or) $2\frac{1}{2}$-player games induces an exponential blowup, adding an exponential factor to the worst case time complexity of the decision problem. Our algorithm avoids this by solving the robust synthesis problem directly on the *IMDP* so that the core part, i.e., lines 6- 7 of Algorithm 1 can be solved with time complexity polynomial in $|\mathcal{M}|$.

Algorithm 2 represents a value iteration-based algorithm which extends the value iteration-based algorithm in [14] and adjusts it for *IMDP* models by encoding the notion of robustness. The core difference is indicated in lines 6 and 16 where the optimal strategy is computed so as to be robust against any choice of nature.

Theorem 13. *Algorithm 1 is sound, complete and has runtime exponential in* $|\mathcal{M}|$, $\mathbf{k}$, *and* n.

Remark 14. It is worthwhile to mention that our robust strategy synthesis approach can also be applied to *MDP*s with richer formalisms for uncertainties such as likelihood or ellipsoidal uncertainties while preserving the computational complexity. In particular, in every inner optimisation problem in Algorithm 1, the optimality of a Markovian deterministic strategy and nature is guaranteed as long as the uncertainty set is convex, the set of actions is finite and the inner optimisation problem which minimises/maximises the objective function over the choices of nature achieves its optimum (cf. [31, Proposition 4.1]). Furthermore, due to the convexity of the generated optimisation problems, the computational complexity of our approach remains intact.

4 Case Studies

We implemented the proposed multi-objective robust strategy synthesis algorithm and applied them to two case studies: (1) motion planning for a robot with noisy continuous dynamics and (2) autonomous nondeterministic tour guides drawn from [4,18]. All experiments completed in few seconds on a standard laptop PC.

4.1 Robot Motion Planning under Uncertainty

In robot motion planning, designers often seek a plan that simultaneously satisfies multiple objectives [23], e.g., *maximising the chances of reaching the target while minimising the energy consumption*. These objectives are usually in conflict with each other; hence, presenting the Pareto curve, i.e., the set of achievable points with optimal trade-off between the objectives, is helpful to the designers. They can then choose a point on the curve according to their desired guarantees and obtain the corresponding plan (strategy) for the robot. In this case study, we considered

Algorithm 2: Value iteration algorithm to solve lines 6-7 of Algorithm 1

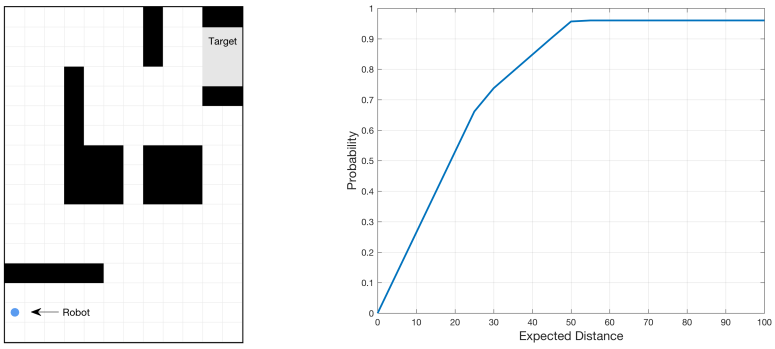
Input: An *IMDP* $\mathcal{M}$, weight vector w, reward structures $\mathbf{r} = (\mathbf{r}_1, \ldots, \mathbf{r}_n)$, time-bound vector $\mathbf{k} \in (\mathbb{N} \cup \{\infty\})^n$, threshold ε

Output: strategy σ maximising $ExpTot^{\sigma,\mathbf{k}}_{\mathcal{M}}[\mathbf{w} \cdot \mathbf{r}]$, $\mathbf{g} := (ExpTot^{\sigma,k_i}_{\mathcal{M}}[\mathbf{r}_i])_{1 \le i \le n}$

begin
- $\mathbf{x} := 0$; $\mathbf{x}^1 := 0$; …; $\mathbf{x}^n := 0$; $\mathbf{y} := 0$; $\mathbf{y}^1 := 0$; …; $\mathbf{y}^n := 0$;
- $\sigma^\infty(s) := \bot$ for all $s \in S$
- **while** $\delta > \varepsilon$ **do**
  - **foreach** $s \in S$ **do**
    - $y_s := \max_{a \in \mathcal{A}(s)} (\sum_{\{\, i \mid k_i = \infty \,\}} w_i \cdot \mathbf{r}_i(s,a) + \min_{\mathfrak{h}^a_s \in \mathcal{H}^a_s} \sum_{s' \in S} \mathfrak{h}^a_s(s') \cdot x_{s'})$;
    - $\sigma^\infty(s) := \arg\max_{a \in \mathcal{A}(s)} (\sum_{\{\, i \mid k_i = \infty \,\}} w_i \cdot \mathbf{r}_i(s,a) + \min_{\mathfrak{h}^a_s \in \mathcal{H}^a_s} \sum_{s' \in S} \mathfrak{h}^a_s(s') \cdot x_{s'})$
    - $\bar{\mathfrak{h}}^{\sigma^\infty(s)}_s(s') := \arg\min_{\mathfrak{h}^a_s \in \mathcal{H}^a_s} \sum_{s' \in S} \mathfrak{h}^a_s(s') \cdot x_{s'}$
  - $\delta := \max_{s \in S}(y_s - x_s)$; $\mathbf{x} := \mathbf{y}$;
- **while** $\delta > \varepsilon$ **do**
  - **foreach** $s \in S$ *and* $i \in \{1, \ldots, n\}$ **where** $k_i = \infty$ **do**
    - $y^i_s := \mathbf{r}_i(s, \sigma^\infty(s)) + \sum_{s' \in S} \bar{\mathfrak{h}}^{\sigma^\infty(s)}_s(s') \cdot x^i_{s'}$;
  - $\delta := \max_{i=1}^n \max_{s \in S}(y^i_s - x^i_s)$; $\mathbf{x}^1 := \mathbf{y}^1$; …; $\mathbf{x}^n := \mathbf{y}^n$;
- **for** $j = \max\{\, k_b < \infty \mid b \in \{1, \ldots, n\} \,\}$ **down to** 1 **do**
  - **foreach** $s \in S$ **do**
    - $y_s := \max_{a \in \mathcal{A}(s)} (\sum_{\{\, i \mid k_i \ge j \,\}} w_i \cdot \mathbf{r}_i(s,a) + \min_{\mathfrak{h}^a_s \in \mathcal{H}^a_s} \sum_{s' \in S} \mathfrak{h}^a_s(s') \cdot x_{s'})$;
    - $\sigma^j(s) := \arg\max_{a \in \mathcal{A}(s)} (\sum_{\{\, i \mid k_i \ge j \,\}} w_i \cdot \mathbf{r}_i(s,a) + \min_{\mathfrak{h}^a_s \in \mathcal{H}^a_s} \sum_{s' \in S} \mathfrak{h}^a_s(s') \cdot x_{s'})$;
    - $\bar{\mathfrak{h}}^{\sigma^j(s)}_s(s') := \arg\min_{\mathfrak{h}^a_s \in \mathcal{H}^a_s} \sum_{s' \in S} \mathfrak{h}^a_s(s') \cdot x_{s'}$;
    - **foreach** $i \in \{1, \ldots, n\}$ *where* $k_i \ge j$ **do**
      - $y^i_s := \mathbf{r}_i(s, \sigma^j(s)) + \sum_{s' \in S} \bar{\mathfrak{h}}^{\sigma^j(s)}_s(s') \cdot x^i_{s'}$;
    - $\mathbf{x} := \mathbf{y}$; $\mathbf{x}^1 := \mathbf{y}^1$; …; $\mathbf{x}^n := \mathbf{y}^n$;
- **for** $i = 1$ **to** n **do**
  - $g_i := y^i_s$;
- σ acts as σ^j in j^{th} step when $j < \max_{i \in \{1,\ldots,n\}} k_i$ and as σ^∞ afterwards;
- **return** $\sigma, \mathbf{g}$

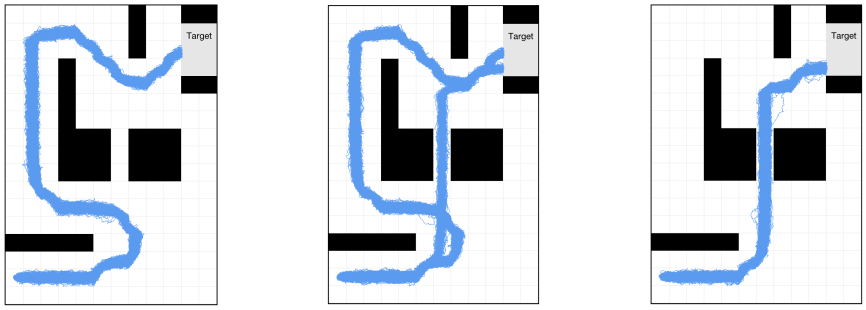
such a motion planning problem for a noisy robot with continuous dynamics in an environment with obstacles and a target region, as depicted in Fig. 3a. The robot's motion model was a single integrator with additive Gaussian noise. The initial state of the robot was on the bottom-left of the environment. The objectives were to reach the target safely while reducing the energy consumption, which is proportional to the travelled distance.

We approached this problem by first abstracting the motion of the noisy robot in the environment as an *IMDP* $\mathcal{M}$ and then computing strategies on $\mathcal{M}$ as in [24–26]. The abstraction was achieved by partitioning the environment into a grid and computing local (continuous) controllers to allow transitions from



(a) Robot Environment (b) Pareto Curve

Fig. 3: Robotic Scenario. (a) Environment map, with black obstacles and gray target area. (b) Pareto curve for the property $([\mathtt{r}_p]^{\leq\infty}_{\max}, [\mathtt{r}_d]^{\leq\infty}_{\min})$.



(a) φ_1 (b) φ_2 (c) φ_3

Fig. 4: Robot sample paths under strategies for φ_1, φ_2, and φ_3

every cell to each of its neighbours. The cells and the local controllers were then associated to the states and actions of the *IMDP*, respectively, resulting in 204 states (cells) and 4 actions per state. The boundaries of the environment were also associated with a state. Note that the transition probabilities between cells were raised by the noise in the dynamics and their ranges were due to variation of the possible initial robot (continuous) state within each cell.

The *IMDP* states corresponding to obstacles (including boundaries) were given deterministic self-transitions, modelling robot termination as the result of a collision. To allow for the computation of the probability of reaching target, we included an extra state in the *IMDP* with a deterministic self-transition and then added incoming deterministic transitions to this state from the target states. A reward structure $\mathtt{r}_p$, which assigns a reward of 1 to these transitions and 0 to all the others, in fact, computes the probability of reaching the target. To capture the travelled distance, we defined a reward structure $\mathtt{r}_d$ assigning a reward of 0 to the state-action pairs with self-transitions and 1 to the other pairs.

The two robot objectives then can be expressed as: $([\mathtt{r}_p]^{\leq\infty}_{\max}, [\mathtt{r}_d]^{\leq\infty}_{\min})$ – see [17, Appendix C] for Pareto queries. We first computed the Pareto curve for the property, which is shown in Fig. 3b, to find the set of all achievable values (optimal trade-offs) for the reachability probability and expected travelled distance. The Pareto curve shows that there is clearly a trade-off between the two objectives. To achieve high probability of reaching target safely, the robot needs to travel a longer distance, i.e., spend more energy, and vice versa. We chose three points on the curve and computed the corresponding robust strategies for

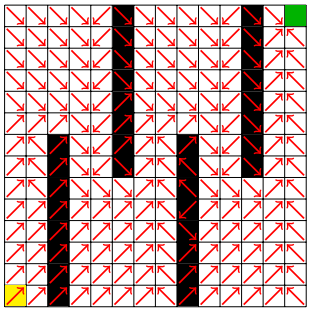
$$\varphi_1 = ([\mathtt{r}_p]^{\leq\infty}_{\geq 0.95}, [\mathtt{r}_d]^{\leq\infty}_{\leq 50}), \quad \varphi_2 = ([\mathtt{r}_p]^{\leq\infty}_{\geq 0.90}, [\mathtt{r}_d]^{\leq\infty}_{\leq 45}), \quad \varphi_3 = ([\mathtt{r}_p]^{\leq\infty}_{\geq 0.66}, [\mathtt{r}_d]^{\leq\infty}_{\leq 25}).$$

We then simulated the robot under each strategy 500 times. The statistical results of these simulations are consistent with the bounds in φ_1, φ_2, and φ_3. The collision-free robot trajectories are shown in Fig. 4. These trajectories illustrate that the robot is conservative under φ_1 and takes a longer route with open spaces around it to go to target in order to be safe (Fig. 4a), while it becomes reckless under φ_3 and tries to go through a narrow passage with the knowledge that its motion is noisy and could collide with the obstacles (Fig. 4c). This risky behaviour, however, is required in order to meet the bound on the expected travelled distance in φ_3. The sample trajectories for φ_2 (Fig. 4b) demonstrate the stochastic nature of the strategy. That is, the robot probabilistically chooses between being safe and reckless in order to satisfy the bounds in φ_2.

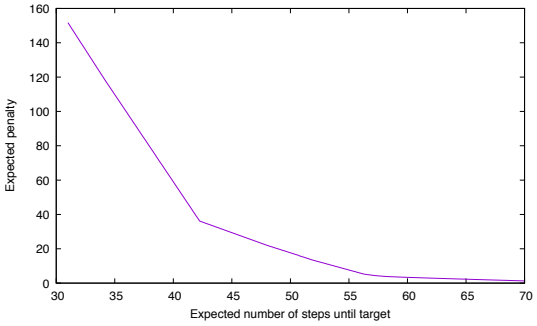
4.2 The Model of Autonomous Nondeterministic Tour Guides

Our second case study is inspired by "Autonomous Nondeterministic Tour Guides" (ANTG) in [4,18], which models a complex museum with a variety of collections. We note that the model introduced in [4] is an *MDP*. In this case study, we use an *IMDP* model by inserting uncertainties into the *MDP*. Due to the popularity of the museum, there are many visitors at the same time. Different visitors may have different preferences of arts. We assume the museum divides all collections into different categories so that visitors can choose what they would like to visit and pay tickets according to their preferences. In order to obtain the best experience, a visitor can first assign certain weights to all categories denoting their preferences to the museum, and then design the best strategy for a target. However, the preference of a sort of arts to a visitor may depend on many factors including price or length of queue at that moment etc., hence it is hard to assign fixed values to these preferences. In our model we allow uncertainties of preferences such that their values may lie in an interval.

For simplicity we assume all collections are organised in an $n \times n$ square with $n \geq 10$, with $(0,0)$ being the south-west corner of the museum and $(n-1, n-1)$ the north-east one. Let $c = \frac{n-1}{2}$; note that (c,c) is at the centre of the museum. We assume all collections at (x,y) are assigned with a weight interval $[3,4]$ if $\max\{|x-c|, |y-c|\} \leq \frac{n}{10}$, with a weight 2 if $\frac{n}{10} < \max\{|x-c|, |y-c|\} \leq \frac{n}{5}$, and a weight 1 if $\max\{|x-c|, |y-c|\} > \frac{n}{5}$. In other words, we expect collections in the centre to be more popular and subject to more uncertainties



(a) The ANTG model for $n = 14$. The yellow, black and green cells represent the entrance, closed and exit parts of the museum, respectively. The red arrows indicate an example strategy.



(b) The Pareto Curve

Fig. 5: The ANTG case study: model and analysis

than others. Furthermore, we assume that people at each location (x, y) have four nondeterministic choices of moving to (x', y') in the north east, south east, north west, and south west of (x, y) (limited to the boundaries of the museum). The outcome of these choices, however, is not deterministic. That is, deciding to go to (x', y') takes the visitor to either (x, y') or (x', y) depending on the weight intervals of (x, y') and (x', y). Thus, the actual outcome of the move is probabilistic to north, south, east or west. To obtain an *IMDP*, weights are normalised. For instance, if the visitor chooses to go to the north east and on $(x, y + 1)$ there is a weight interval of $[3, 4]$ and on $(x + 1, y)$ there is a weight interval of $[2, 2]$, it will go to $(x, y + 1)$ with probability interval $[\frac{3}{3+2}, \frac{4}{4+2}]$ and to $(x + 1, y)$ with probability interval $[\frac{2}{2+4}, \frac{2}{2+3}]$. Therefore a model with parameter n has n^2 states in total and roughly $4n^2$ transitions, a few of which are associated with uncertain transition probabilities. An instance of the museum model for $n = 14$ is depicted in Fig. 5a. In this instantiation, we assume that the visitor starts in the lower left corner (marked yellow) and wants to move to the upper right corner (marked green) with as few steps as possible. On the other hand, it wants to avoid moving to the black cells, because they correspond to exhibitions which are closed. For closed exhibitions located at $x = 2$, the visitor receive a penalty of 2, for those at $x = 5$ it receives a penalty of 4, for $x = 8$ one of 16 and for $x = 11$ one of 64. Therefore, there is a tradeoff between leaving the museum as fast as possible and minimising the penalty received. With $\mathtt{r}_s$ being the reward structure for the number of steps and $\mathtt{r}_p$ denoting the penalty accumulated, $([\mathtt{r}_s]^{\leq\infty}_{\leq 40}, [\mathtt{r}_p]^{\leq\infty}_{\leq 70})$ requires that we leave the museum within 40 steps but with a penalty of no more than 70. The red arrows indicate a strategy which has been used when computing the Pareto curve by our tool. Here, the tourist mostly ignores closed exhibitions at $x = 2$ but avoids them later. In [17, Appendix D], we provide a few more strategies occurring during the computation. We provide the Pareto curve for this situation in Fig. 5b. With an increasing step bound considered

acceptable, the optimal accumulated penalty decreases. This is expected, since with a larger step bound, the visitor has more time to walk around more of the closed exhibitions, thus facing a lower penalty.

5 Concluding Remarks

In this paper, we have analysed *IMDP*s under controller synthesis semantics in a dynamic setting; we discussed the multi-objective robust strategy synthesis problem for *IMDP*s, aiming for strategies that satisfy a given multi-objective predicate under all resolutions of the uncertainty in the transition probabilities. We showed that this problem is **PSPACE**-hard and introduced a value iteration-based decision algorithm to approximate the Pareto set. We finally presented the effectiveness of the proposed algorithms on several real-world case studies.

Even though we focused here on *IMDP*s with multi-objective reachability and reward properties, the proposed robust synthesis algorithm can also handle *MDP*s with convex uncertain sets and any ω-regular properties such as LTL. For future work, we aim to explore the upper bound of the time complexity of the multi-objective robust strategy synthesis which is left open in this paper.

References

1. N. Basset, M. Kwiatkowska, and C. Wiltsche. Compositional controller synthesis for stochastic games. In *CONCUR*, pages 173–187. Springer, 2014.
2. M. Benedikt, R. Lenhardt, and J. Worrell. LTL model checking of interval Markov chains. In *TACAS*, pages 32–46, 2013.
3. S. Boyd and L. Vandenberghe. *Convex optimization*. Cambridge Univ. Press, 2004.
4. A. S. Cantino, D. L. Roberts, and C. L. Isbell. Autonomous nondeterministic tour guides: improving quality of experience with TTD-MDPs. In *AAMAS*, page 22, 2007.
5. K. Chatterjee, R. Majumdar, and T. A. Henzinger. Markov decision processes with multiple objectives. In *STACS*, volume 3884 of *LNCS*, pages 325–336, 2006.
6. K. Chatterjee, K. Sen, and T. A. Henzinger. Model-checking ω-regular properties of interval Markov chains. In *FoSSaCS*, pages 302–317, 2008.
7. T. Chen, V. Forejt, M. Kwiatkowska, A. Simaitis, and C. Wiltsche. On stochastic games with multiple objectives. In *MFCS*, pages 266–277. Springer, 2013.
8. T. Chen, T. Han, and M. Kwiatkowska. On the complexity of model checking interval-valued discrete time Markov chains. *Inf. Proc. Lett.*, 113(7):210–216, 2013.
9. M. Ehrgott. *Multicriteria optimization*. Springer Science & Business Media, 2006.
10. M.-A. Esteve, J.-P. Katoen, V. Y. Nguyen, B. Postma, and Y. Yushtein. Formal correctness, safety, dependability and performance analysis of a satellite. In *ICSE*, pages 1022–1031, 2012.
11. K. Etessami, M. Kwiatkowska, M. Y. Vardi, and M. Yannakakis. Multi-objective model checking of Markov decision processes. In *TACAS*, pages 50–65, 2007.
12. H. Fecher, M. Leucker, and V. Wolf. Don't know in probabilistic systems. In *SPIN*, volume 3925 of *LNCS*, pages 71–88. Springer, 2006.
13. V. Forejt, M. Kwiatkowska, G. Norman, D. Parker, and H. Qu. Quantitative multi-objective verification for probabilistic systems. In *TACAS*, pages 112–127, 2011.

14. V. Forejt, M. Kwiatkowska, and D. Parker. Pareto curves for probabilistic model checking. In *ATVA*, pages 317–332. Springer, 2012.
15. R. Givan, S. M. Leach, and T. L. Dean. Bounded-parameter Markov decision processes. *AI*, 122(1-2):71–109, 2000.
16. E. M. Hahn, T. Han, and L. Zhang. Synthesis for PCTL in parametric Markov decision processes. In *NFM*, volume 6617 of *LNCS*, pages 146–161, 2011.
17. E. M. Hahn, V. Hashemi, H. Hermanns, M. Lahijanian, and A. Turrini. Multi-objective robust strategy synthesis for interval Markov decision processes. Available at `http://arxiv.org/abs/1706.06875`, 2017.
18. V. Hashemi, H. Hermanns, and L. Song. Reward-bounded reachability probability for uncertain weighted MDPs. In *VMCAI*, pages 351–371. Springer, 2016.
19. B. Jonsson and K. G. Larsen. Specification and refinement of probabilistic processes. In *LICS*, pages 266–277. IEEE Computer Society, 1991.
20. I. Kozine and L. V. Utkin. Interval-valued finite Markov chains. *Reliable Computing*, 8(2):97–113, 2002.
21. M. Kwiatkowska, G. Norman, D. Parker, and H. Qu. Compositional probabilistic verification through multi-objective model checking. *I&C*, 232:38–65, 2013.
22. M. Lahijanian, S. B. Andersson, and C. Belta. Formal verification and synthesis for discrete-time stochastic systems. *IEEE Tr. Autom. Contr.*, 60(8):2031–2045, 2015.
23. M. Lahijanian and M. Kwiatkowska. Specification revision for Markov decision processes with optimal trade-off. In *CDC*, pages 7411–7418, 2016.
24. R. Luna, M. Lahijanian, M. Moll, and L. E. Kavraki. Asymptotically optimal stochastic motion planning with temporal goals. In *WAFR*, pages 335–352, 2014.
25. R. Luna, M. Lahijanian, M. Moll, and L. E. Kavraki. Fast stochastic motion planning with optimality guarantees using local policy reconfiguration. In *ICRA*, pages 3013–3019, 2014.
26. R. Luna, M. Lahijanian, M. Moll, and L. E. Kavraki. Optimal and efficient stochastic motion planning in partially-known environments. In *AAAI*, pages 2549–2555, 2014.
27. A. Mouaddib. Multi-objective decision-theoretic plan problem. In *ICRA*, pages 2814–2819, 2004.
28. A. Nilim and L. El Ghaoui. Robust control of Markov decision processes with uncertain transition matrices. *Operations Research*, 53(5):780–798, 2005.
29. W. Ogryczak, P. Perny, and P. Weng. A compromise programming approach to multiobjective Markov decision processes. *IJITDM*, 12(5):1021–1054, 2013.
30. P. Perny, P. Weng, J. Goldsmith, and J. P. Hanna. Approximation of Lorenz-optimal solutions in multiobjective Markov decision processes. In *AAAI*, pages 92–94, 2013.
31. A. Puggelli. *Formal Techniques for the Verification and Optimal Control of Probabilistic Systems in the Presence of Modeling Uncertainties*. PhD thesis, UC Berkeley, 2014.
32. A. Puggelli, W. Li, A. L. Sangiovanni-Vincentelli, and S. A. Seshia. Polynomial-time verification of PCTL properties of MDPs with convex uncertainties. In *CAV*, pages 527–542, 2013.
33. M. Randour, J.-F. Raskin, and O. Sankur. Percentile queries in multi-dimensional Markov decision processes. In *CAV*, pages 123–139. Springer, 2015.
34. E. M. Wolff, U. Topcu, and R. M. Murray. Robust control of uncertain Markov decision processes with temporal logic specifications. In *CDC*, pages 3372–3379, 2012.
35. D. Wu and X. D. Koutsoukos. Reachability analysis of uncertain systems using bounded parameter Markov decision processes. *AI*, 172(8-9):945–954, 2008.