ScienceDirect

IFAC PapersOnLine 51-16 (2018) 271–276

Temporal Logic Control of POMDPs via Label-based Stochastic Simulation Relations

S. Haesaert [*] P. Nilsson [*] C.I. Vasile [**] R. Thakker [***]
A. Agha-mohammadi [***] A.D. Ames [*] R.M. Murray [*]

[*] *California Institute of Technology, Pasadena, CA 91125 USA*
[**] *Massachusetts Institute of Technology, Cambridge, MA 02139 USA*
[***] *Jet Propulsion Laboratory, Pasadena, CA 91109 USA*

Abstract: The synthesis of controllers guaranteeing linear temporal logic specifications on partially observable Markov decision processes (POMDP) via their belief models causes computational issues due to the continuous spaces. In this work, we construct a finite-state abstraction on which a control policy is synthesized and refined back to the original belief model. We introduce a new notion of label-based approximate stochastic simulation to quantify the deviation between belief models. We develop a robust synthesis methodology that yields a lower bound on the satisfaction probability, by compensating for deviations a priori, and that utilizes a less conservative control refinement.

Keywords: Temporal properties, control synthesis, partially observable, Markov decision processes

1. INTRODUCTION

Emerging applications in robotics necessitate control systems capable of autonomously performing complex tasks in a safe manner. These systems are often deployed in (partially) unknown environments where they have to maximize the probability of task completion. Temporal logics have emerged as a principled formalism for expressing behavior together with associated control synthesis techniques (Wongpiromsarn et al., 2009). However, work to date in this area has mostly focused on specifications expressed in terms of the system state; such specifications are less suitable for expressing properties about system uncertainty. For instance, in a Mars rover exploration mission the accuracy of pose estimates must be of much higher quality during the crucial sample extraction phase, as compared to when the rover is traversing a “safe” area. In this work, we develop a control synthesis technique that yields guarantees for properties that quantify such notions of uncertainty.

For probabilistic temporal logic properties over finite-state Markov decision processes (MDPs), there exist several tools for policy synthesis and verification, such as PRISM (Kwiatkowska et al., 2011) and Storm (Dehnert et al., 2017). However, when the MDP state space is uncountable the characterization of these properties cannot in general be attained analytically (Abate et al., 2008). An alternative is to approximate these models by simpler processes such as finite-state MDPs. By quantifying the approximation accuracy via (approximate) simulation relations (Girard et al., 2010; Haesaert et al., 2017b,a) guarantees on the resulting controller can be obtained.

Without full-state observations, control synthesis and verification becomes more challenging. For finite-state partially observable Markov decision processes (POMDPs), verification and policy synthesis has been considered for PCTL properties (Norman et al., 2017; Chatterjee et al., 2015). For POMDPs over continuous spaces, results have been focused on reachability and safety (Ding et al., 2013; Lesser and Oishi, 2014). As in (Vasile et al., 2016; Jones et al., 2013), we consider properties defined on the belief space of a POMDP—the space of probability distributions over states. For the resulting belief models, the inherent complexity of the belief space makes the application of model simplification and quantification more difficult. Therefore, it is also more difficult to compute guarantees for these synthesized control systems.

In this paper, we give a robust synthesis methodology tailored to belief models that is guaranteed to be correct-by-construction. The contributions of this paper are as follows: Firstly, we tailor the notion of an approximate stochastic simulation relation towards belief models by omitting distance measures on the belief space and introducing non-determinism in the proposition labeling. Secondly, we propose a method to construct the associated control refinement directly from the value function. Thirdly, we explicitly compute this robust synthesis methodology for non-stationary Kalman-filtered Linear Time-Invariant (LTI) systems. The results are illustrated on a Mars exploration scenario in which a finite environment POMDP is combined with an LTI POMDP.

In the next section, belief models, POMDPs and the associated temporal logic specifications are introduced. The computationally intractable exact control synthesis is given in Sec. 3, before introducing our robust synthesis method for abstractions in Sec. 4. This is clarified further for LTI Gaussian POMDPs in the subsequent section. Finally an illustrative case study is given.

Notation: For a Polish[1] space $\mathbb{Y}$, we denote by $\mathcal{B}(\mathbb{Y})$ its Borel σ-field. Then $\mathcal{P}(\mathbb{Y})$ is the set of probability measures on the Borel-measurable space $(\mathbb{Y}, \mathcal{B}(\mathbb{Y}))$ whose elements $\mathbf{P}$ have realizations denoted as $y \sim \mathbf{P}$. The indicator function of a set A is written $\mathbf{1}_A(x)$ and is equal to 1 if $x \in A$, and to 0 otherwise. For a relation $\mathcal{R} \subset \mathbb{X}_1 \times \mathbb{X}_2$, denote the associated mappings

[⋆] This research was carried out at JPL and Caltech under a contract with the NASA and funded through the President’s and Director’s Fund Program.

[1] A Polish space is a complete and separable metrizable space (Bogachev, 2007). All spaces in this work are assumed to be Polish.

10.1016/j.ifacol.2018.08.046

$\mathcal{R}(X_1) := \{x_2 : x_1\mathcal{R}x_2, x_1 \in \mathbb{X}_1\}$ and $\mathcal{R}^{-1}(X_2) := \{x_1 : x_1\mathcal{R}x_2, x_2 \in \mathbb{X}_2\}$ for $X_1 \subseteq \mathbb{X}_1$ and $X_2 \subseteq \mathbb{X}_2$.

2. POMDPS AND TEMPORAL LOGIC SPECIFICATIONS

2.1 POMDPs and belief models

We define a Markov decision process as follows.

Definition 1. A discrete-time *Markov decision process* (MDP) is a tuple $\mathsf{M} = (\mathbb{X}, \rho, t, \mathbb{U})$ where $\mathbb{X}$ is a (Polish) state space with states $x \in \mathbb{X}$; $\rho \in \mathcal{P}(\mathbb{X})$ is an initial probability distribution; $\mathbb{U}$ is a (Polish) input space with inputs $u \in \mathbb{U}$; $t : \mathbb{X} \times \mathbb{U} \times \mathcal{B}(\mathbb{X}) \to [0, 1]$ is a Borel-measurable stochastic kernel that assigns to each state $x \in \mathbb{X}$ and control $u \in \mathbb{U}$ a probability measure $t(\cdot \,|\, x, u)$ over $(\mathbb{X}, \mathcal{B}(\mathbb{X}))$. □

An execution of M is a state-input sequence $(x_0, u_0)(x_1, u_1)\dots$ where $x_0 \sim \rho$ and $x_{k+1} \sim t(\cdot \,|\, x_k, u_k)$ for inputs $u_k \in \mathbb{U}$. While MDPs capture uncertainty in state transitions, full knowledge is assumed about the state of the system. We therefore augment MDPs with a model for partial observations.

Definition 2. A *partially observable Markov decision process* (POMDP) $\mathsf{M}_{\mathbb{Z}}$ is an MDP $\mathsf{M} = (\mathbb{X}, \rho, t, \mathbb{U})$ together with an observation model $(\mathbb{Z}, r)$ where $\mathbb{Z}$ is a (Polish) output space with outputs $z \in \mathbb{Z}$, and $r : \mathbb{X} \times \mathcal{B}(\mathbb{Z}) \to [0, 1]$ is a conditional stochastic observation kernel that assigns to each state $x \in \mathbb{X}$ a probability measure $r(\cdot \,|\, x) \in (\mathbb{Z}, \mathcal{B}(\mathbb{Z}))$. □

An execution of the POMDP up to time K is a sequence

$$(x_0, u_0, z_0)(x_1, u_1, z_1)\dots(x_K, u_K, z_K) \tag{1}$$

where $(x_0, u_0)\dots(x_K, u_K)$ is an execution of M and $z_k \sim r(\cdot|x_k)$. In a POMDP, control actions u_k can be chosen as a function of the available information. For this purpose, define the *k-th information space* as $\mathbb{I}_k := (\mathbb{Z} \times \mathbb{U})^k \times \mathbb{Z}$ with elements $i_k = (z_0, u_0, z_1, u_1, \dots, z_{k-1}, u_{k-1}, z_k)$ referred to as the *k-th information vector*. Based on the notion of information, an *observation-based* policy for $\mathsf{M}_{\mathbb{Z}}$ is a sequence $\boldsymbol{\mu} = (\mu_0, \dots, \mu_{K-1})$ such that for each k, $\mu_k(\mathrm{d}u_k|\rho, i_k)$ is a universally measurable stochastic kernel on $\mathbb{U}$ given $\mathcal{P}(\mathbb{X}) \times \mathbb{I}_k$. We say that $\boldsymbol{\mu}$ is *non-randomized* if for all ρ, k, and i_k, $\mu_k(\cdot|\rho, i_k)$ is a Dirac distribution. Given an observation-based policy $\boldsymbol{\mu}$ and an initial distribution ρ, there exists a unique probability measure $\mathbf{P}^{\boldsymbol{\mu}}_{\rho}$ over the space of executions (see theorem by Ionescu Tulcea (Hernández-Lerma and Lasserre, 1996)).

The probability distribution $b_k(\mathrm{d}x) = \mathbf{P}(x_k \in \mathrm{d}x|\rho, i_k) \in \mathcal{P}(\mathbb{X})$ expresses the state knowledge at time k. This is referred to as the *belief state* and is an element of the *belief space* $\mathbb{B} \subset \mathcal{P}(\mathbb{X})$. The state b_k evolves based on the stochastic kernel

$$b_{k+1} \sim t_b(\cdot|b_k, u_k), \quad b_0 \sim \rho_b, \tag{2}$$

for an initial belief ρ_b. Since (2) is completely observable, it follows that a POMDP $\mathsf{M}_{\mathbb{Z}}$ can equivalently be expressed by

$$\mathsf{B}(\mathsf{M}_{\mathbb{Z}}) = (\mathbb{B}, \rho_b, t_b, \mathbb{U}), \tag{3}$$

an MDP over the belief space. For this we implicitly assume that Borel measurability of the stochastic kernels is preserved (Bertsekas, 1976). In the sequel we will omit $\mathsf{M}_{\mathbb{Z}}$ and simply write $\mathsf{B} = \mathsf{B}(\mathsf{M}_{\mathbb{Z}})$ for the belief model. The information vector for B at time k is given as $i_k = (b_0, u_0, b_1, u_1, \dots, b_{k-1}, u_{k-1}, b_k)$. Thus a policy for B is a sequence $\boldsymbol{\mu} = (\mu_0, \dots, \mu_{K-1})$ such that for all k, $\mu_k(\mathrm{d}u_k|\rho_b, i_k)$ is a universally measurable stochastic kernel on $\mathbb{U}$. We say that a policy $\boldsymbol{\mu}$ is a *Markov policy* if for each k it depends only on the current state, i.e., $\mu_k(\mathrm{d}u_k|\rho_b, i_k) = \mu_k(\mathrm{d}u_k|b_k)$. Furthermore $\boldsymbol{\mu}$ is also *stationary* if there exists a policy μ such that $\mu_k(\mathrm{d}u|b) = \mu(\mathrm{d}u|b)$ for all k and b.

2.2 Linear temporal logic for belief models

Similar to Jones et al. (2013), we construct specifications over the belief space. Atomic propositions are the basic building blocks from which temporal specifications (Pnueli, 1977) are constructed. In this work we associate atomic propositions p_i to measurable subsets of the belief space $\mathbb{B}$. While belief spaces are generally infinite-dimensional, Gaussian distributions $\mathcal{N}(\hat{x}, P)$ are uniquely characterized by the mean $\hat{x}$ and variance P. Examples of atomic propositions over a Gaussian belief space are: (1) a position-based proposition $p_1 \Leftrightarrow \hat{x} \in A$; (2) an uncertainty-based proposition $p_2 \Leftrightarrow \det(P) \leq c$ with $\det()$ the determinant; (3) a proposition $p_3 \Leftrightarrow \int_A \mathcal{N}(\mathrm{d}x \,|\, \hat{x}, P) \leq c$.

Consider a set $AP = \{p_1, \dots, p_L\}$ of atomic propositions; it defines an *alphabet* $\Sigma := 2^{AP}$ where each *letter* π of the alphabet is a subset of AP. An infinite string of letters is a *word* $\boldsymbol{\pi} = \pi_0\pi_1\pi_2\dots \in \Sigma^{\mathbb{N}}$. A labeling function $\mathsf{L} : \mathbb{B} \to \Sigma$ maps belief states to letters in the alphabet, such that a belief trajectory $\mathbf{b} = b_0b_1b_2\dots$ generates a word as $\boldsymbol{\pi} := \mathsf{L}(b_0)\mathsf{L}(b_1)\mathsf{L}(b_2)\dots$ with respect to which system properties can be expressed. We trivially require that L is measurable, i.e. that $\{b|p \in \mathsf{L}(b)\} \in \mathcal{B}(\mathbb{B})$ for all $p \in AP$. Since Borel measurability is preserved by standard linear operations (Azoff, 1974), the uncertainty-based properties are also Borel-measurable.

Properties are formulas composed of atomic propositions and operators. In the sequel, we focus on a fragment of linear temporal logic (Belta et al., 2017).

Definition 3. Formulas in the *syntactically co-safe LTL* (scLTL) fragment are constructed according to the grammar

$$\psi := \top \mid p \mid \neg p \mid \psi_1 \vee \psi_2 \mid \psi_1 \wedge \psi_2 \mid \psi_1 \,\mathcal{U}\, \psi_2 \mid \bigcirc \psi, \tag{4}$$

where $p \in AP$ is an atomic proposition.

Definition 4. The *semantics* of scLTL are defined recursively over suffix sequences $\boldsymbol{\pi}_i := \pi_i\pi_{i+1}\pi_{i+2}\dots$ as $\boldsymbol{\pi}_i \models \top$; $\boldsymbol{\pi}_i \models p$ iff $p \in \pi_i$; $\boldsymbol{\pi}_i \models \psi_1 \wedge \psi_2$ iff $(\boldsymbol{\pi}_i \models \psi_1) \wedge (\boldsymbol{\pi}_i \models \psi_2)$; $\boldsymbol{\pi}_i \models \psi_1 \vee \psi_2$ iff $(\boldsymbol{\pi}_i \models \psi_1) \vee (\boldsymbol{\pi}_i \models \psi_2)$; $\boldsymbol{\pi}_i \models \psi_1 \,\mathcal{U}\, \psi_2$ iff $\exists j \geq i$ s.t. $(\boldsymbol{\pi}_j \models \psi_2)$ and $\boldsymbol{\pi}_k \models \psi_1, \forall k \in \{i, \dots j-1\}$; $\boldsymbol{\pi}_i \models \bigcirc\psi$ iff $\boldsymbol{\pi}_{i+1} \models \psi$.

We say that a belief trajectory $\mathbf{b} = b_0b_1b_2\dots$ satisfies a specification ψ, written $\mathbf{b} \models \psi$, if the generated word $\boldsymbol{\pi} = \mathsf{L}(b_0)\mathsf{L}(b_1)\mathsf{L}(b_2)\dots$ satisfies ψ, i.e. $\boldsymbol{\pi} \models \psi$.

The objective of this work is to design a policy $\boldsymbol{\mu}$ such that a specification ψ is satisfied with a given probability.

Problem 1. Consider belief model B (3), labeling function $\mathsf{L} : \mathbb{B} \to \Sigma$ and an scLTL formula ψ. Construct a policy $\boldsymbol{\mu}$ such that

$$\mathbf{P}^{\boldsymbol{\mu}}_{\rho}(\mathbf{b} \models \psi) \geq p, \tag{5}$$

where p is either given or to be maximized.

3. CONTROL SYNTHESIS FOR SCLTL FORMULAE

In this section, we give the exact, but intractable, computation of policies over belief MDPs.

Definition 5. A *deterministic finite-state automaton* (DFA) is a tuple $\mathcal{A} = (Q, q_0, \Sigma, \delta_{\mathcal{A}}, Q_f)$, with Q a finite set of states, $q_0 \in Q$ the initial state, Σ the input alphabet, $\delta_{\mathcal{A}} : Q \times \Sigma \to Q$ the transition function, and $Q_f \subseteq Q$ the accepting states.

A word $\boldsymbol{\pi} = \pi_0\pi_1\pi_2\ldots$ *is accepted* by the DFA if there exists a sequence $q_0q_1q_2\ldots q_f$ with $q_f \in Q_f$, that starts with the initial state q_0 and for which $q_{k+1} = \delta_{\mathcal{A}}(q_k, \pi_k)$. Denote the set of words accepted by a DFA $\mathcal{A}$ as $\mathbf{Lang}(\mathcal{A})$.

For every scLTL property ψ (4), there exists a DFA $\mathcal{A}_\psi$ for which $\boldsymbol{\pi} \models \psi$ if and only if $\boldsymbol{\pi} \in \mathbf{Lang}(\mathcal{A}_\psi)$, see inter alia (Belta et al., 2017). We can therefore reason about satisfaction of probabilistic properties on B by analyzing its product MDP $\mathsf{B} \otimes \mathcal{A}_\psi$ (Tkachev et al., 2013).

Definition 6. Given belief model B (3), labeling function $\mathsf{L} : \mathbb{B} \to \Sigma$ and DFA $\mathcal{A}_\psi$, the *product* of M and $\mathcal{A}_\psi$ is an MDP

$$\mathsf{B} \otimes \mathcal{A}_\psi = (\mathbb{B} \times Q, \bar{\rho}, \bar{t}, \mathbb{U}), \qquad (6)$$

where $\bar{\rho}(db, q) = \rho_b(db)$ if $q = \delta_{\mathcal{A}}(q_0, \mathsf{L}(b))$ and $\bar{\rho}(db, q) = 0$ otherwise, and the transition kernel is similarly given as

$$\bar{t}(db' \times \{q'\}|b, q, u) = \begin{cases} t(db'|b, u) & \text{if } q' = \delta_{\mathcal{A}}(q, \mathsf{L}(b')), \\ 0 & \text{otherwise.} \end{cases}$$

Any policy $\boldsymbol{\mu}$ for $\mathsf{B} \otimes \mathcal{A}_\psi$ induces a policy for B. Thus Problem 1 can be converted into the problem of constructing a reachability-enforcing policy on $\mathsf{B} \otimes \mathcal{A}_\psi$. In the following we review the reachability problem on MDPs.

Given policy $\boldsymbol{\mu}$ for $\mathsf{B} \otimes \mathcal{A}_\psi$, define the time-dependent value function $\mathbf{V}^K_{\boldsymbol{\mu}} : \mathbb{B} \times Q \to [0, 1]$ as

$$\mathbf{V}^K_{\boldsymbol{\mu}}(b, q) = \mathbf{E}_{\boldsymbol{\mu}}\left[\sum_{i=0}^{K} \mathbf{1}_{Q_f}(q_i) \prod_{j=0}^{i-1} \mathbf{1}_{Q \setminus Q_f}(q_j) \middle| (b_0, q_0) = (b, q)\right]. \qquad (7)$$

Since $\mathbf{V}^K_{\boldsymbol{\mu}}(x)$ expresses the probability that a trajectory generated by $\boldsymbol{\mu}$ starting from (b, q) will reach the target set Q_f within a time horizon K (Abate et al., 2008), it also expresses the probability that the ψ will be satisfied in the time horizon K. Next express the associated Bellman operator $\mathbf{T}_\mu$ as

$$\mathbf{T}_\mu(\mathbf{V})(b, q) = \int_{\mathbb{B}} \max\left(\mathbf{1}_{Q_f}(q'), \mathbf{V}(b', q')\right) t(\mathrm{d}b'|b, \mu(b, q)) \qquad (8)$$

with the implicit DFA transitions $q' = \delta_{\mathcal{A}}(q, \mathsf{L}(b'))$, Consider a policy $\boldsymbol{\mu}_i = (\mu_{i+1}, \ldots \mu_K)$ with time horizon $K - i$, then it follows that $\mathbf{V}^{K-i+1}_{\boldsymbol{\mu}_{i-1}} = \mathbf{T}_{\mu_i}\mathbf{V}^{K-i}_{\boldsymbol{\mu}_i}$. Thus if $\mathbf{V}^{K-i}_{\boldsymbol{\mu}_i}$ expresses the probability of reaching Q_f within $K - i$ steps, then $\mathbf{T}_{\mu_i}\mathbf{V}^{K-i}_{\boldsymbol{\mu}_i}$ expresses the probability of reaching Q_f within $K - i + 1$ steps with policy $\boldsymbol{\mu}_{i-1}$. It follows that for a stationary policy $\boldsymbol{\mu}$, the infinite-horizon value function can be computed as $\mathbf{V}^\infty_{\boldsymbol{\mu}} = \lim_{K\to\infty} \mathbf{T}^K_\mu \mathbf{V}^0$ with $\mathbf{V}^0 \equiv 0$.

Instead of defining the recursions for a given policy $\boldsymbol{\mu}$, we can also optimize with respect to the set $\mathbf{D}_\mu$ of universally measurable deterministic policies. This yields the policy-optimal Bellman recursion as

$$\mathbf{T}_*(\mathbf{V}) = \sup\nolimits_{\mu \in \mathbf{D}_\mu} \mathbf{T}_\mu(\mathbf{V}). \qquad (9)$$

From (Abate et al., 2008), we know that there exists a policy $\boldsymbol{\mu}_*$ optimizing the value functions (7), or equivalently, the reachability recursions (8) that is a stationary, universally measurable, and deterministic policy. For the converged value function $\mathbf{V}^\infty_* := \lim_{K\to\infty}[\mathbf{T}_*]^K\mathbf{V}^0$, the probability of satisfaction is

$$\sup_{\boldsymbol{\mu}} \mathbf{P}^{\boldsymbol{\mu}}_\rho(\mathbf{b} \models \psi) = \int_{\mathbb{B}} \max\left(\mathbf{1}_{Q_f}(q'), \mathbf{V}^\infty_*(b_0, q')\right) \rho_b(\mathrm{d}b_0), \qquad (10)$$

and the associated policy $\boldsymbol{\mu}_* = (\mu_*, \mu_*, \ldots)$ is defined for the product MDP $\mathsf{B} \otimes \mathcal{A}_\psi$ and maps a state $(b, q) \in \mathbb{B} \times Q$ to a control action, and can be translated to a non-stationary policy for the original belief MDP B that uses $\mathcal{A}_\psi$ as a memory model. Even though this section outlines a solution method to Problem 1, the required recursions are in general computationally intractable.

4. REFINEMENT-BASED APPROXIMATE CONTROL SYNTHESIS

Let a belief model B and an abstract model $\tilde{\mathsf{B}}$ be given. In this section, we give a robust synthesis methodology tailored to abstractions of belief models that yields a lower bound on the satisfaction probability and we give a specification-based policy refinement.

4.1 Approximate label-based stochastic simulation

Let $\mathbb{B}_1, \mathbb{B}_2$ be two sets for which the relation $\mathcal{R} \subset \mathbb{B}_1 \times \mathbb{B}_2$ is a set that captures pairwise similarity between $x_1 \in \mathbb{B}_1$ and $x_2 \in \mathbb{B}_2$, then also similarity between $\mathbf{P}_1 \in \mathcal{P}(\mathbb{B}_1)$ and $\mathbf{P}_2 \in \mathcal{P}(\mathbb{B}_2)$ can be quantified as follows.

Definition 7. (δ-lifting (Haesaert et al., 2017b)). For a given relation $\mathcal{R} \in \mathcal{B}(\mathbb{B}_1 \times \mathbb{B}_2)$, we say that $\mathbf{P}_1$ and $\mathbf{P}_2$ are in the corresponding *δ-lifted relation $\bar{\mathcal{R}}_\delta$*, written $\mathbf{P}_1\bar{\mathcal{R}}_\delta\mathbf{P}_2$, if there exists a *lifting* $\mathbb{W} \in \mathcal{P}(\mathbb{B}_1 \times \mathbb{B}_2)$ such that

L1. for all $X_1 \in \mathcal{B}(\mathbb{B}_1)$: $\mathbb{W}(X_1 \times \mathbb{B}_2) = \mathbf{P}_1(X_1)$;
L2. for all $X_2 \in \mathcal{B}(\mathbb{B}_2)$: $\mathbb{W}(\mathbb{B}_1 \times X_2) = \mathbf{P}_2(X_2)$;
L3. $\mathbb{W}(\mathcal{R}) \geq 1 - \delta$, i.e., $b_1\mathcal{R}b_2$ with probability at least $1 - \delta$.

For stochastic kernels, we add a measurability condition.

Definition 8. (δ-Lifting). Stochastic kernels $t_1(\cdot|x)$ and $t_2(\cdot|x)$, as defined in Def. 1, are *in a δ-lifted relation $\bar{\mathcal{R}}_\delta$*, i.e., $t_1(\cdot|x)\,\bar{\mathcal{R}}_\delta\, t_2(\cdot|x)$, if there exists a lifting $\mathbb{W}_t : \mathbb{X}_1 \times \mathbb{X}_2 \times \mathcal{B}(\mathbb{X}_1 \times \mathbb{X}_2) \to [0, 1]$ that is Borel-measurable.

We next apply this lifting concept to quantify the difference between two belief models $\tilde{\mathsf{B}}$ and B. To allow for approximate relations where atomic proposition can be ambiguous, we newly consider set-valued labelings $\mathbb{B} \to 2^\Sigma$. For the belief model B, define the set-valued extension of the labeling function as $\mathcal{L} : \mathbb{B} \to 2^\Sigma$ with $\mathcal{L}(b) = \{\mathsf{L}(b)\}$.

Definition 9. Consider a concrete belief MDP $\mathsf{B} = (\mathbb{B}, \rho, t_b, \mathbb{U})$ and an abstract MDP $\tilde{\mathsf{B}} = (\tilde{\mathbb{B}}, \tilde{\rho}, \tilde{t_b}, \tilde{\mathbb{U}})$, with (set-valued) labeling maps $\mathcal{L}$ and $\tilde{\mathcal{L}}$. We say that $\tilde{\mathsf{B}}$ is *δ-stochastically simulated* by B with respect to $(\tilde{\mathcal{L}}, \mathcal{L})$, denoted as $\tilde{\mathsf{B}} \preceq^\delta_{\tilde{\mathcal{L}},\mathcal{L}} \mathsf{B}$, if there exists a Borel-measurable interface function $\mathcal{U}_v : \tilde{\mathbb{U}} \times \tilde{\mathbb{B}} \times \mathbb{B} \to \mathcal{P}(\mathbb{U})$ and a Borel-measurable relation $\mathcal{R} \subseteq \tilde{\mathbb{B}} \times \mathbb{B}$, s.t. for all $\tilde{u} \in \tilde{\mathbb{U}}$ and for all $(\tilde{b}, b) \in \mathcal{R}$:

$$\tilde{\rho}_b\bar{\mathcal{R}}_\delta\rho_b \text{ and } \tilde{t_b}(\cdot|\tilde{b}, \tilde{u})\,\bar{\mathcal{R}}_\delta\, t_b(\cdot|b, \mathcal{U}_v(\tilde{u}, \tilde{b}, b)), \qquad \textbf{(SR 1\&2)}$$

$$\mathcal{L}(b) \subseteq \tilde{\mathcal{L}}(\tilde{b}). \qquad \textbf{(SR } \boldsymbol{\mathcal{L}}\textbf{)}$$

Conditions (**SR 1&2**) enforce δ-probabilistic similarity between the initial distributions and the transition kernels, while (**SR $\boldsymbol{\mathcal{L}}$**) guarantees that any label $\mathsf{L} \in \mathcal{L}(b)$ of a concrete state b is also present in the abstract label collection $\mathcal{L}(\tilde{b})$. The simulation relation is transitive: if $\mathsf{B}_1 \preceq^{\delta_1}_{\mathcal{L}_1,\mathcal{L}_2} \mathsf{B}_2$ and $\mathsf{B}_2 \preceq^{\delta_2}_{\mathcal{L}_2,\mathcal{L}_3} \mathsf{B}_3$, then $\mathsf{B}_1 \preceq^{\delta_1+\delta_2}_{\mathcal{L}_1,\mathcal{L}_3} \mathsf{B}_3$.

4.2 Robust policy synthesis

Given an approximate belief model $\tilde{\mathsf{B}}$, we define a robust quantification and policy synthesis for an scLTL property. We compensate for both non-determinism in the labeling function, as well as the difference δ in probability. The former can be dealt with by considering the worst-case resolution of non-determinism in the Bellman operator (8) :

$$\mathbf{T}_\mu^{\tilde{\mathcal{L}}}(\mathbf{W})(b,q) = \int_{\mathbb{B}} \min_{q' \in \delta_{\mathcal{A}}(q, \tilde{\mathcal{L}}(b'))} \max\left(\mathbf{1}_{Q_f}(q'), \mathbf{W}(b',q')\right) \times t(db'|b, \mu(b,q)). \quad (11)$$

The robust Bellman operator follows by compensating for the difference in probability δ as

$$\mathbf{R}_\mu^{(\tilde{\mathcal{L}},\delta)}(\mathbf{W})(b,q) = \max\left(0, \mathbf{T}_\mu^{\tilde{\mathcal{L}}}(\mathbf{W})(b,q) - \delta\right). \quad (12)$$

The policy-optimal robust Bellman operator $\mathbf{R}_*^{(\tilde{\mathcal{L}},\delta)}$ follows analogously. The robustified optimal probability that an execution $\tilde{\mathbf{b}}$ of $\tilde{\mathsf{B}}$ satisfies ψ, is therefore given as

$$\mathbf{P_{rob}}^*_\rho(\tilde{\mathbf{b}} \vDash \psi; \tilde{\mathcal{L}}, \delta) = \quad (13)$$
$$\int_{\mathbb{B}} \min_{\bar{q} \in \delta_{\mathcal{A}}(q_0, \tilde{\mathcal{L}}(b_0))} \max\left(\mathbf{1}_{Q_f}(\bar{q}), \mathbf{W}_*^\infty(b_0, \bar{q})\right) \rho_b(\mathrm{d}b_0) - \delta$$

with $\mathbf{W}_*^\infty := \lim_{K\to\infty}[\mathbf{R}_*^{(\mathcal{L},\delta)}]^K(\mathbf{W}^0)$ and $\mathbf{W}^0 \equiv 0$, and with the associated policy $\tilde{\boldsymbol{\mu}}_*$. Based on the following proposition, the original belief model inherits these results as follows.

Proposition 10. (Haesaert et al. (2017a)). Suppose that $\tilde{\mathsf{B}} \preceq^\delta_{\tilde{\mathcal{L}},\mathcal{L}} \mathsf{B}$, then a control policy $\tilde{\boldsymbol{\mu}}_*$ for $\tilde{\mathsf{B}}$ can be refined to a control policy $\boldsymbol{\mu}_*$ for B such that

$$\mathbf{P_{rob}}^*_{\tilde{\rho}}(\tilde{\mathbf{b}} \vDash \psi; \tilde{\mathcal{L}}, \delta) \leq \mathbf{P}_\rho^{\boldsymbol{\mu}_*}(\mathbf{b} \vDash \psi). \quad (14)$$

In Haesaert et al. (2017b), this refined policy is constructed via composition of the conditional lifting (see Definition 8) and the abstract policy $\tilde{\boldsymbol{\mu}}_*$. Here we improve this construction in a way that is independent of the lifting but strongly dependent on the value function. Given $\mathbf{W}_*^\infty$ and an abstract policy $\tilde{\boldsymbol{\mu}}_*$, the refined concrete policy over the product MDP $\mathsf{B} \otimes \mathcal{A}_\psi$ is obtained as

$$\boldsymbol{\mu}_*(b,q) := \mathcal{U}_v(\tilde{\boldsymbol{\mu}}_*(\tilde{b}^*, q), \tilde{b}^*, b), \ \tilde{b}^* := \underset{\tilde{b} \in \mathcal{R}^{-1}(b)}{\arg\max} \mathbf{W}_*^\infty(\tilde{b}, q). \quad (15)$$

The following theorem now summarizes this contribution.

Theorem 11. Given a finite-state abstraction $\tilde{\mathsf{B}}$ for which $\tilde{\mathsf{B}} \preceq^\delta_{\tilde{\mathcal{L}},\mathcal{L}} \mathsf{B}$ with interface $\mathcal{U}_v$. Let $\mathbf{W}_*^\infty$ be the value function associated to specification ψ and let $\tilde{\boldsymbol{\mu}}_*$ be a policy for the abstract model. If the refined policy $\boldsymbol{\mu}_*$ is defined by (15), then (14) holds.

Proof. At each time instant, finding a control action u is associated to finding the state update for the abstract system such as to find an abstract input that can be refined via the interface in Definition 9. Based on proposition 10, we know that for every policy $\tilde{\boldsymbol{\mu}}_*$ there exists a refined policy $\tilde{\boldsymbol{\mu}}_*$ such that for all pairs of belief states $(\tilde{b}, b) \in \mathcal{R}$, it holds that $\mathbf{W}_*^\infty(\tilde{b}, q) \leq \mathbf{V}_{ref}^\infty(b,q)$, where $\mathbf{V}_{ref}^\infty(b,q)$ is the exact infinite-horizon value of (7) for the refined policy starting from b. The standard control refinement hinges on selecting the next abstract state as a realization of the lifted kernel (c.f., Def. 8); by instead selecting the next state as the maximizer of $\mathbf{W}_*^\infty$ a strict improvement of the lower bound is achieved. □

In contrast to the work in Haesaert et al. (2017b,a) developed for standard MDPs, this work newly facilitates working with belief models as it incorporates non-determinism in the labeling instead of leveraging a metric error. Furthermore, we have given a different specification-dependent policy refinement.

5. ABSTRACTIONS FOR GAUSSIAN LTI POMDPS

Consider an LTI system

$$\begin{aligned} x_{k+1} &= Ax_k + Bu_k + w_k, & w_k &\sim \mathcal{N}(0, \mathcal{W}), \\ z_k &= Cx_k + v_k, & v_k &\sim \mathcal{N}(0, \mathcal{V}), \end{aligned} \quad (16)$$

we say that it defines POMDP with state space $\mathbb{X} \subseteq \mathbb{R}^n$, initial distribution $\rho := \mathcal{N}(\hat{x}_\rho, P_\rho)$, control inputs $\mathbb{U} \subseteq \mathbb{R}^m$, and transition kernel $t(\cdot|x,u) = \mathcal{N}(Ax+Bu, \mathcal{W})$, together with observation model $r(\cdot|x) = \mathcal{N}(Cx, \mathcal{V})$ generating measurements z_k. It is well known that the belief state $(\hat{x}_{k|k}, P_{k|k})$ for this type of system evolves over the space of Gaussian distributions, and is given by the Kalman filter equations (Bertsekas, 1976):

$$\hat{x}_{k|k} = A\hat{x}_{k-1|k-1} + Bu_{k-1} + L_k e_k, \ e_k \sim \mathcal{N}(0, S_k), \quad (17)$$
$$P_{k|k} = (I - L_k C)P_{k|k-1}, \ P_{k|k-1} = AP_{k-1|k-1}A^T + \mathcal{W}, \quad (18)$$
$$L_k = P_{k|k-1}C^T S_k^{-1}, \quad S_k = CP_{k|k-1}C^T + \mathcal{V}.$$

These equations define a belief model B. As a first abstraction, we choose to reduce the complexity and dimensionality of the dynamics by neglecting the variations of the covariance matrix P. Therefore, abstract model $\tilde{\mathsf{B}}$ is obtained by replacing the stochastic transitions in Eq. (17) by

$$\tilde{x}_k = A\tilde{x}_{k-1} + B\tilde{u}_{k-1} + \tilde{P}C^T\tilde{s}_k, \quad (19)$$

with $\tilde{s}_k \sim \mathcal{N}(0, \tilde{S}_{inv})$ and a matrix $\tilde{S}_{inv}$ such that $\tilde{S}_{inv} \preceq S_k^{-1}$ for all k, and where $\tilde{P}$ defines the steady state value of $P_{k|k-1}$, i.e., the solution of the Kalman equations. We say that the MDP $\tilde{\mathsf{B}}$ has state space $\mathbb{B}_x$. The next step is to show that there exists a value δ such that $\tilde{\mathsf{B}} \preceq^\delta_{\tilde{\mathcal{L}},\mathcal{L}} \mathsf{B}$. If this holds, then we can grid $\mathbb{B}_x$ as in (Haesaert et al., 2017b,a) to obtain a finite-state MDP $\tilde{\mathsf{B}}_{grid}$ over which we can do the control synthesis. Via transitivity of Def. 9, the latter finite-state MDP will also be in a δ-approximate stochastic simulation with B. The details of the gridding step can be found in (Haesaert et al., 2017a). We restrict attention to simulation relations $\mathcal{R}$ between states $b = (\hat{x}, P)$ of B and states $\tilde{x}$ of $\tilde{\mathsf{B}}$, and interfaces $\mathcal{U}_v$, of the forms

$$\mathcal{R} = \left\{(\tilde{x}, (\hat{x}, P)) \middle| (\hat{x} - \tilde{x})^T M (\hat{x} - \tilde{x}) \leq \epsilon^2, \ P^- \preceq P \preceq P^+\right\},$$
$$\mathcal{U}_v(\tilde{u}, \tilde{x}, \hat{x}) = K(\hat{x} - \tilde{x}) + \tilde{u}, \quad (20)$$

for some matrices M, P^-, P^+ and K. Next, we specify this relation, interface and the labeling such that they define a label-based δ-stochastic simulation relation for $\tilde{\mathsf{B}}$ and B.

Labeling requirement. We construct a set-valued labeling function $\tilde{\mathcal{L}} : \mathbb{B}_x \to 2^\Sigma$ satisfying (**SR** $\boldsymbol{\mathcal{L}}$). For a position-based proposition p_i consider, without loss of generality, a labeling $\mathsf{L}_{p_i} : \mathbb{B} \to \{\{p_i\}, \emptyset\}$ for the concrete belief MDP defined by $p_i \in \mathsf{L}_p((\hat{x}, P)) \Leftrightarrow \hat{x} \in A$. The set-valued extension to $2^{\{\{p_i\},\emptyset\}}$ for the abstract MDP is defined as

$$\tilde{\mathcal{L}}_{p_i}(\tilde{x}) = \begin{cases} \{\{p_i\}\} & \text{if } \forall b \in \mathcal{R}(\tilde{x}) : p_i \in \mathsf{L}_{p_i}(b), \\ \{\emptyset\} & \text{if } \forall b \in \mathcal{R}(\tilde{x}) : p_i \notin \mathsf{L}_{p_i}(b), \\ \{\{p_i\}, \emptyset\} & \text{otherwise.} \end{cases} \quad (21)$$

The labeling $\tilde{\mathcal{L}}_{p_i}$ can be easily computed by shrinking or expanding the set A, and the extension to multiple atomic propositions is straightforward. Similarly, for propositions involving the variance of the current belief state any atomic property that is monotonic[2] in P can be mapped to the abstract model. However, atomic propositions that include probability require caution since the quantification of the probability of an event is

[2] Here monotonicity of a function in P is defined based on the preorder $\succeq$ for the matrices P with $A \succeq B$ if $x^T(A-B)x \geq 0 \, \forall x \in \mathbb{R}^n$.

in general not monotonic with respect to variance.

Probability requirements. To show satisfaction of condition (**SR 1&2**), we need to show that there exists a δ-lifting. First we require that P^+ (resp. P^-) is an upper (resp. lower) bound for $P_{k|k}$ of the belief MDP (17)-(18). We say that P^- is a lower bound if it is a lower bound for the initial condition (see above) and if it is monotonically increasing with respect to the Riccati equations (Bitmead et al., 1985). For the upper bound, we require, *mutatis mutandis*, a monotonically decreasing P^+. Consider a choice of noise sources $\tilde{s}_{k+1}$ and s^{Δ}_{k+1}, such that the difference between the states in (17)-(19) evolves according to

$$\begin{aligned}\hat{x}_{k+1|k+1} - \tilde{x}_{k+1} = (A+BK)(\hat{x}_{k|k} - \tilde{x}_k)\\ + P_{k+1|k}C^T s^{\Delta}_{k+1} + \Delta_{k+1}\tilde{s}_{k+1},\end{aligned} \tag{22}$$

with $\Delta_{k+1} := (P_{k+1|k}C^T - \tilde{P}C^T)$, $\tilde{s}_{k+1} \sim \mathcal{N}(0, \tilde{S}_{inv})$, and $s^{\Delta}_{k+1} \sim \mathcal{N}(0,\ S^{-1}_{k+1} - \tilde{S}_{inv})$.

We can now quantify the δ-difference between B and $\tilde{\mathsf{B}}$ by verifying that for all $(\tilde{x}_k, \hat{x}_{k|k}) \in \mathcal{R}$, with probability at least $1-\delta$ it holds that $(\tilde{x}_{k+1}, \hat{x}_{k+1|k+1}) \in \mathcal{R}$. Furthermore, we can bound the norm of the noise terms s^{Δ}_{k+1} and $\tilde{s}_{k+1}$ using probabilistic guarantees computed via inter alia the chi-square distribution. The computation of the values of ϵ and the values of the matrices in $\mathcal{R}$ and in the interface (20), can now be performed as an optimization problem. Though we did not explicitly write out the lifted probability space $\mathbb{W}_t$ of Def. 8, it can easily be obtained from the probability measure induced by joint evolution of (17) and (19) driven by noise sources s^{Δ} and $\tilde{s}$.

Combined, these requirements constitute sufficient conditions for the simulation relation to apply for a Kalman belief model B and an abstraction $\tilde{\mathsf{B}}$ on the form (19). As mentioned above, a grid-based abstraction can then be constructed from $\tilde{\mathsf{B}}$ to obtain a finite abstraction $\tilde{\mathsf{B}}_{grid}$ such that $\tilde{\mathsf{B}}_{grid} \preceq^{\delta_{grid}}_{\tilde{\mathcal{L}}_{grid},\tilde{\mathcal{L}}} \tilde{\mathsf{B}} \preceq^{\delta}_{\tilde{\mathcal{L}},\mathcal{L}} \mathsf{B}$ with $\delta_{grid} = 0$. While B is of dimension $n + n^2$, $\tilde{\mathsf{B}}$ is of dimension n, therefore this intermediate step is crucial to mitigate the curse of dimensionality in the gridding procedure.

6. CASE STUDY

We consider a rover tasked with identifying and collecting scientific samples in a partially unknown Mars environment.

Rover model. We consider a simple rover modeled as a point mass $x_k \in \mathbb{R}^2$ affected by stochastic disturbances and modeled as a partially-observable LTI system (16) with $(A, B, C, \mathcal{W}, \mathcal{V}) = (I, I, I, \left[\begin{smallmatrix} .4 & -.2 \\ -.2 & .4 \end{smallmatrix}\right], I)$. The corresponding belief space model B is defined by the Kalman filter equations (17)-(18) as in Section 5. We assume that initially the belief distribution of x is $\mathcal{N}(\hat{x}_0, P_0)$ with $P_0 = \left[\begin{smallmatrix} 0.74446 & -0.2862 \\ -0.2862 & 0.74446 \end{smallmatrix}\right]$.

Environment model. From the overhead imagery, we consider two target regions T_1 and T_2 where the probability of encountering a desired sample has been assessed as 0.5 and 0.6, respectively. In addition, there are two risk regions R_1 and R_2 with probabilities of 0.9 and 0.7 that the rover can traverse them safely, respectively. For both target and risk regions, we assume that the true nature of the region can be determined via measurements from on-board sensors when the rover is within a certain distance of the regions. The regions are illustrated in Fig. 3, along with the regions from where measurements can be acquired (dashed lines). To each region we associate a discrete belief state $b_r \in \{p_{r,0}, 1, 0\}$ where $p_{r,0}$ are the belief probabilities from satellite imagery.

With B_{env} as the combined environment model with state $(b_{T_1}, b_{T_2}, b_{R_1}, b_{R_2})$, the overall system is $\mathsf{B} \otimes \mathsf{B}_{env}$. This product is defined similarly to the product in Definition 6: B gives inputs to B_{env} according to the position in the state space. If the state estimate $\hat{x}$ of B is inside a measurement region (dashed lines in Fig. 3), a measurement is performed in the corresponding environment MDP.

Specification. The objective is to collect a sample while avoiding unsafe regions. This is expressed by the scLTL specification

$$\psi = \neg\texttt{fail}\ \mathcal{U}\ \texttt{sample}, \tag{23}$$

where the atomic proposition `sample` is true if the rover is in a target region that contains a sample, and the atomic proposition `fail` is defined as being true if the rover is in a risk region that contains an obstacle.

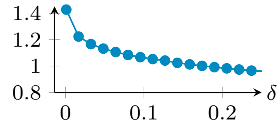

Fig. 1. Trade-off between state error ϵ from Eq. (20) and probability error δ from Definition 9.

Solution. We first construct a finite abstraction of B as outlined in the previous section; first we compute an approximate LTI belief model and then we discretize the state space $[-10, 10]^2$ with grids of width and height $(0.76481, 0.64426)$. The input space $[-1, 1]^2$ is sampled with nine discrete inputs $\{0, 1, -1\}^2$. Fig. 1 illustrates the trade-off between the ϵ error in the relation (20) and the δ-difference in Definition 9 for these choices of discretization parameters. Combined with a relaxation of the labeling as introduced in the previous subsection, we obtain an abstract model $\tilde{\mathsf{B}}_{grid}$ with the property that $\tilde{\mathsf{B}}_{grid} \preceq^{0.01}_{\tilde{\mathcal{L}}_{grid},\mathcal{L}} \mathsf{B}$. The upper bound P^+ and the lower bound P^- on the uncertainty have been selected based on the steady state Kalman gain and on the initial covariance P_0, respectively.

In order to analyze the combined system $\mathsf{B} \otimes \mathsf{B}_{env}$, we create $\tilde{\mathsf{B}}_{grid} \otimes \mathsf{B}_{env}$ as a nondeterministic product and treat it analogously to the nondeterminism in the connection between an MDP and a DFA in Eq. (13). Effectively, when there is uncertainty about whether a measurement of a region can be expected, the worst-case is considered in the value iteration which by Theorem 11 ensures that the obtained probabilities are indeed lower bounds. With this method we compute a value function and an associated control policy for $\tilde{\mathsf{B}}_{grid} \otimes \mathsf{B}_{env}$ and refine it to the concrete system. The resulting policy takes inputs from $\mathbb{B} \times \prod_{i=1}^{4}\{p_{0,i}, 0, 1\}$—the combined rover and environment state—and has internal memory dynamics given by the specification DFA. The value function is illustrated in Fig. 2 for two different environment states. Observe that the knowledge of a sample in T_2 and an obstacle in R_2 yields a more polarized value function in the right figure.

The plots in Fig. 3 depict the evolution of the estimated position for executions generated by the control policy, and the associated probability bounds. The probabilities reflect multiple factors: the probability of encountering samples, the probability of encountering obstacles, and the probability of falling out of the simulation relation (as captured by $\delta = 0.01$). Positive jumps in probability occur when samples are detected or regions are deemed obstacle-free, and negative jumps result from not finding samples in target regions, and from finding obstacles. As can be seen, the policy synthesis technique elicits intelligent behaviors where the agents explore the most promising regions

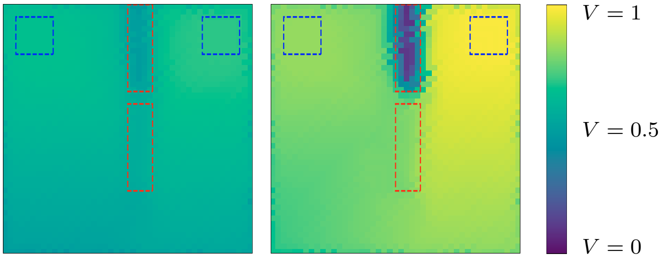

Fig. 2. Illustration of value function for two different states of the environment $(0.5, 0.6, 0.9, 0.7)$ [left] and $(0.5, 1, 0.9, 0)$ [right], with regions of interest plotted with dashed lines.

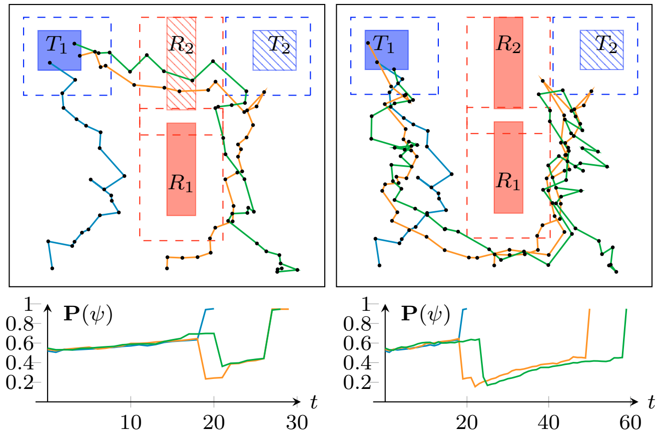

Fig. 3. Above: regions of interest (red/blue), associated measurement regions (dashed lines), and trajectories of noisy state estimates starting at three different initial conditions. Below: estimated probability to satisfy the specification over time for the same trajectories. The left and right plots illustrate two different configurations of the true environment. In both cases, a sample exists at T_1 but not at T_2, and R_1 cannot be traversed due to an obstacle. However, in the left figure, R_2 does not contain an obstacle, so the rover can traverse R_2 after confirming that it is safe. In the right figure, both risk regions contain obstacles.

to satisfy the specification, while remaining robust to noise effects.

7. CONCLUSIONS

In this paper, we have developed a robust control synthesis methodology for POMDPs and specifications expressed over the belief space of the POMDP. Our approach is based on principled abstractions via a label-based type of stochastic simulation relation that is tailored to belief models. In addition, we have introduced a specification-based way to refine control policies. As a result, we have obtained policies with associated guarantees on the probability of specification satisfaction. For the special case of linear systems with Gaussian noise and observation models, we have given a concrete abstraction construction and an associated simulation relation.

REFERENCES

Abate, A., Prandini, M., Lygeros, J., and Sastry, S. (2008). Probabilistic Reachability and Safety for Controlled Discrete Time Stochastic Hybrid Systems. *Automatica*, 44(11), 2724–2734.

Azoff, E.A. (1974). Borel measurability in linear algebra. *Proceedings of the American Mathematical Society*, 42(2), 346–350.

Belta, C., Yordanov, B., and Gol, E.A. (2017). *Formal Methods for Discrete-Time Dynamical Systems*. Springer.

Bertsekas, D.P. (1976). *Dynamic programming and stochastic control*. Academic Press.

Bitmead, R.R., Gevers, M.R., Petersen, I.R., and Kaye, R.J. (1985). Monotonicity and stabilizability-properties of solutions of the Riccati difference equation. *Systems & Control Letters*, 5(5), 309–315.

Bogachev, V.I. (2007). *Measure theory*. Springer.

Chatterjee, K., Chmelik, M., Gupta, R., and Kanodia, A. (2015). Qualitative analysis of POMDPs with temporal logic specifications for robotics applications. In *Proc. IEEE ICRA*, 325–330.

Dehnert, C., Junges, S., Katoen, J.P., and Volk, M. (2017). A storm is coming: A modern probabilistic model checker. In *Proc. CAV*, 592–600.

Ding, J., Abate, A., and Tomlin, C. (2013). Optimal control of partially observable discrete time stochastic hybrid systems for safety specifications. In *Proc. ACC*, 6231–6236.

Girard, A., Pola, G., and Tabuada, P. (2010). Approximately bisimilar symbolic models for incrementally stable switched systems. *IEEE Transactions on Automatic Control*, 55(1), 116–126. doi:10.1109/TAC.2009.2034922.

Haesaert, S., Soudjani, S., and Abate, A. (2017a). Temporal logic control of general Markov decision processes by approximate policy refinement. *arXiv:1712.07622 [cs.SY]*.

Haesaert, S., Soudjani, S., and Abate, A. (2017b). Verification of general Markov decision processes by approximate similarity relations and policy refinement. *SIAM Journal on Control and Optimization*, 55(4), 2333–2367.

Hernández-Lerma, O. and Lasserre, J.B. (1996). *Discrete-time Markov control processes*, volume 30 of *Applications of Mathematics*. Springer.

Jones, A., Schwager, M., and Belta, C. (2013). Distribution temporal logic: Combining correctness with quality of estimation. In *Proc. IEEE CDC*, 4719–4724.

Kwiatkowska, M., Norman, G., and Parker, D. (2011). PRISM 4.0: Verification of probabilistic real-time systems. In *Proc. CAV*, 585–591.

Lesser, K. and Oishi, M. (2014). Reachability for partially observable discrete time stochastic hybrid systems. *Automatica*, 50(8), 1989–1998.

Norman, G., Parker, D., and Zou, X. (2017). Verification and control of partially observable probabilistic systems. *Real-Time Systems*, 53(3), 354–402.

Pnueli, A. (1977). The temporal logic of programs. In *Foundations of Computer Science, 1977., 18th Annual Symposium on*, 46–57. IEEE.

Tkachev, I., Mereacre, A., Katoen, J., and Abate, A. (2013). Quantitative automata-based controller synthesis for non-autonomous stochastic hybrid systems. In *Proc. HSCC*, 293–302.

Vasile, C.I., Leahy, K., Cristofalo, E., Jones, A., Schwager, M., and Belta, C. (2016). Control in belief space with Temporal Logic specifications. In *Proc. IEEE CDC*, 7419–7424.

Wongpiromsarn, T., Topcu, U., and Murray, R.M. (2009). Receding Horizon Temporal Logic Planning for Dynamical Systems. In *Proc. IEEE CDC*, 5997 –6004.