Computational Approaches to Reachability Analysis of Stochastic Hybrid Systems

Alessandro Abate[1], Saurabh Amin[1], Maria Prandini[2], John Lygeros[3], and Shankar Sastry[1]

[1] University of California, at Berkeley - Berkeley, CA, USA
{aabate,saurabh,sastry}@eecs.berkeley.edu
[2] Politecnico di Milano - Milano, Italy
prandini@elet.polimi.it
[3] ETH Zurich - Zurich, Switzerland
lygeros@control.ee.ethz.ch

Abstract. This work investigates some of the computational issues involved in the solution of probabilistic reachability problems for discrete-time, controlled stochastic hybrid systems. It is first argued that, under rather weak continuity assumptions on the stochastic kernels that characterize the dynamics of the system, the numerical solution of a discretized version of the probabilistic reachability problem is guaranteed to converge to the optimal one, as the discretization level decreases. With reference to a benchmark problem, it is then discussed how some of the structural properties of the hybrid system under study can be exploited to solve the probabilistic reachability problem more efficiently. Possible techniques that can increase the scale-up potential of the proposed numerical approximation scheme are suggested.

1 Introduction

This paper addresses the problem of determining the control policy that maximizes the probability that a stochastic system will remain within a safe set over some look-ahead time horizon (finite-time probabilistic reachability problem). We focus on the discrete time controlled stochastic hybrid system (DTSHS) model introduced in [1], and consider the case when the control input to be applied at a certain time is selected based only on the value of the state at that same time (Markov policy). Following the approach in [1,2], the stochastic reachability problem of interest can be formulated as a finite-horizon optimal control problem with a multiplicative cost function to be maximized. This optimal control problem, in turn, can be solved by dynamic programming (DP). This requires to introduce a cost-to-go function and to determine the value of the control input maximizing the cost-to-go function along the reference time horizon for all values of the state within the safe set. Since an analytic solution to the DP equation is generally hard to find, the computational aspects of the problem are of key importance to its actual implementation. This is the main motivation of the present work.

A. Bemporad, A. Bicchi, and G. Buttazzo (Eds.): HSCC 2007, LNCS 4416, pp. 4–17, 2007.

There are two approaches to the problem: the first is to resort to a numerical approximation scheme relying on the discretization of the continuous state and control input spaces (gridding approach). Alternatively, one can introduce a family of finitely parameterized functions, and then look for the cost-to-go function within that family (neuro-dynamic programming approach [9]).

Here, we study a gridding procedure for the numerical solution to the DP equation of the stochastic reachability problem. We assess the convergence of the numerical solution to the actual solution as the grid size goes to zero, and derive explicit bounds on the level of approximation introduced for a given small but nonzero grid size. The study is inspired by the reference work [3], discussing discretization procedures for the numerical solution to DP in the additive cost case and for stochastic –non hybrid– systems: we extend this approach to a hybrid system setting with multiplicative cost and general disturbances. A numerical approximation scheme was proposed in [6] for estimating the probability of remaining within a safe set for a certain class of autonomous, continuous time stochastic hybrid systems. The discretization process in that case involved gridding the system both in time and in space. Convergence of the estimate to the true probability as the grid size goes to zero was proven, but no bounds were provided for assessing the quality of the estimates derived for a small but nonzero grid size.

Furthermore, we reinterpret some ideas proposed in [4] and [5] within the hybrid systems framework to suggest that structural properties of the problem, such as its decentralized nature, may be exploited to obtain a more compact state representation and efficient implementation of the computations involved in the solution to DP. This feature may partly mitigate the curse of dimensionality that affects DP as well as other approaches proposed in the literature to address the reachability problem [8].

The rest of the paper is organized as follows. We first briefly recall the DTSHS model of [1] in Section 2 and describe the DP algorithm to solve the probabilistic reachability problem in Section 3. Section 4 proposes a numerical approximation scheme for solving the DP algorithm. Based on some regularity assumptions on the stochastic kernels that characterize the system dynamics, convergence of the numerical scheme and explicit bounds assessing the quality of the approximated solution to DP are shown in Section 5. Section 6 illustrates the convergence properties and scaling issues of the proposed numerical approximation scheme with reference to a multi-room heating benchmark. Possible extensions regarding efficient representations and computation of solutions are finally outlined in Section 7.

2 Stochastic Hybrid System Model

In this section we briefly recall the discrete time stochastic hybrid system (DTSHS) model first introduced in [1].

Definition 1. *A discrete time stochastic hybrid system (DTSHS) is a tuple* $\mathcal{H} = (\mathcal{Q}, n, \mathcal{A}, T_x, T_q, R)$*, where*

- $\mathcal{Q} := \{q_1, q_2, \ldots, q_m\}$, *for some* $m \in \mathbb{N}$, *represents the discrete state space;*
- $n : \mathcal{Q} \to \mathbb{N}$ *assigns to each discrete state value* $q \in \mathcal{Q}$ *the dimension of the continuous state space* $\mathbb{R}^{n(q)}$. *The hybrid state space is then given by* $\mathcal{S} := \cup_{q \in \mathcal{Q}} \{q\} \times \mathbb{R}^{n(q)}$;
- $\mathcal{A}$ *is a compact Borel space representing the control space;*
- $T_x : \mathcal{B}(\mathbb{R}^{n(\cdot)}) \times \mathcal{S} \times \mathcal{A} \to [0,1]$ *is a Borel-measurable stochastic kernel on* $\mathbb{R}^{n(\cdot)}$ *given* $\mathcal{S} \times \mathcal{A}$, *which assigns to each* $s = (q,x) \in \mathcal{S}$ *and* $a \in \mathcal{A}$ *a probability measure on the Borel space* $(\mathbb{R}^{n(q)}, \mathcal{B}(\mathbb{R}^{n(q)}))$: $T_x(dx|(q,x),a)$
- $T_q : \mathcal{Q} \times \mathcal{S} \times \mathcal{A} \to [0,1]$ *is a discrete stochastic kernel on* $\mathcal{Q}$ *given* $\mathcal{S} \times \mathcal{A}$, *which assigns to each* $s \in \mathcal{S}$ *and* $a \in \mathcal{A}$, *a probability distribution over* $\mathcal{Q}$: $T_q(q|(q,x),a)$;
- $R : \mathcal{B}(\mathbb{R}^{n(\cdot)}) \times \mathcal{S} \times \mathcal{A} \times \mathcal{Q} \to [0,1]$ *is a Borel-measurable stochastic kernel on* $\mathbb{R}^{n(\cdot)}$ *given* $\mathcal{S} \times \mathcal{A} \times \mathcal{Q}$, *that assigns to each* $s = (q,x) \in \mathcal{S}$, $a \in \mathcal{A}$, *and* $q' \in \mathcal{Q}$, *a probability measure on the Borel space* $(\mathbb{R}^{n(q')}, \mathcal{B}(\mathbb{R}^{n(q')}))$: $R(dx|(q,x),a,q')$. □

The system initialization at time $k = 0$ is specified through some probability measure $\pi : \mathcal{B}(\mathcal{S}) \to [0,1]$ on the Borel space $(\mathcal{S}, \mathcal{B}(\mathcal{S}))$, where $\mathcal{B}(\mathcal{S})$ is the σ-field generated by the subsets of $\mathcal{S}$ of the form $\cup_q \{q\} \times B_q$, with B_q denoting a Borel set in $\mathbb{R}^{n(q)}$. With reference to the time horizon $[0, N]$, we next define the notion of Markov policy.

Definition 2. *Consider a DTSHS* $\mathcal{H} = (\mathcal{Q}, n, \mathcal{A}, T_x, T_q, R)$. *A Markov policy for* $\mathcal{H}$ *is a sequence* $\mu = (\mu_0, \mu_1, \ldots, \mu_{N-1})$ *of universally measurable maps* $\mu_k : \mathcal{S} \to \mathcal{A}$, $k = 0, 1, \ldots, N-1$. *We denote the set of Markov policies as* $\mathcal{M}_m$. □

For conciseness sake, we can introduce the Borel-measurable stochastic kernel $T_s : \mathcal{B}(\mathcal{S}) \times \mathcal{S} \times \mathcal{A} \to [0,1]$ on $\mathcal{S}$ given $\mathcal{S} \times \mathcal{A}$, which assigns to each $s = (q,x)$, $s' = (q',x') \in \mathcal{S}$, $a \in \mathcal{A}$ a probability measure on the Borel space $(\mathcal{S}, \mathcal{B}(\mathcal{S}))$ as follows:

$$T_s(ds'|s,a) = \begin{cases} T_x(dx'|(q,x),a) T_q(q'|s,a), & \text{if } q' = q \\ R(dx'|(q,x),a,q') T_q(q'|s,a), & \text{if } q' \neq q. \end{cases} \tag{1}$$

Definition 3. *An execution for a DTSHS* $\mathcal{H} = (\mathcal{Q}, n, \mathcal{A}, T_x, T_q, R)$ *associated with a policy* $\mu = (\mu_0, \mu_1, \ldots, \mu_{N-1}) \in \mathcal{M}_m$ *and an initial distribution* π *is a stochastic process* $\{\mathbf{s}(k), k \in [0, N]\}$ *with values in* $\mathcal{S}$ *whose sample paths are obtained according to the following algorithm:*

extract from $\mathcal{S}$ a value s_0 for $\mathbf{s}(0)$ according to π;

for $k = 0$ to $N-1$

set $a_k = \mu_k(s_k)$;

extract from $\mathcal{S}$ a value s_{k+1} for $\mathbf{s}(k+1)$ according to $T_s(\cdot\,|s_k, a_k)$;

end □

A DTSHS $\mathcal{H}$ can then be described as a controlled Markov process with state space $\mathcal{S}$, control space $\mathcal{A}$, and controlled transition probability function $T_s : \mathcal{B}(\mathcal{S}) \times \mathcal{S} \times \mathcal{A} \to [0,1]$ defined in (1). Thus, the execution $\{\mathbf{s}(k), k \in [0,N]\}$ associated with $\mu \in \mathcal{M}_m$ and π is a time inhomogeneous stochastic process defined on the canonical sample space $\Omega = \mathcal{S}^{N+1}$, endowed with its product topology $\mathcal{B}(\Omega)$, with probability measure P_π^μ uniquely defined by the initial probability measure π on $(\mathcal{S}, \mathcal{B}(\mathcal{S}))$ and one-step transition kernels $T_s^{\mu_k}(ds'|s) := T_s(ds'|s, \mu_k(s))$, $k = 0, 1, \ldots, N-1$. When π is concentrated at $\{s\}$, $s \in \mathcal{S}$, that is $\pi(ds) = \delta_s(ds)$, we shall write simply P_s^μ.

3 Probabilistic Reachability Problem

Given a stochastic hybrid system $\mathcal{H}$, a Borel compact set $D \in \mathcal{B}(\mathcal{S})$, and a Markov policy $\mu \in \mathcal{M}_m$, let

$$p_\pi^\mu(D) := P_\pi^\mu(\mathbf{s}(k) \in D \text{ for all } k \in [0,N])$$

denote the probability that the execution of $\mathcal{H}$ associated with policy μ and with the initial state distribution π will stay within set D over the time horizon $[0,N]$. If π is concentrated at $\{s\}$, $s \in \mathcal{S}$, we use the notation $p_s^\mu(D)$. If set D represents a safe set for $\mathcal{H}$, by computing $p_s^\mu(D)$, we shall evaluate the safety level for system $\mathcal{H}$ when it starts from $s \in D$ and is subject to policy μ. The objective is to determine the Markov policy that maximizes the probability $p_\pi^\mu(D)$.

Let $\mathbf{1}_C : \mathcal{S} \to \{0,1\}$ denote the indicator function of a set $C \subseteq \mathcal{S}$: $\mathbf{1}_C(s) = 1$, if $s \in C$, and 0, if $s \notin C$. Observe that

$$\prod_{k=0}^{N} \mathbf{1}_D(s_k) = \begin{cases} 1, & \text{if } s_k \in D \text{ for all } k \in [0,N] \\ 0, & \text{otherwise,} \end{cases}$$

where $s_k \in \mathcal{S}$, $k \in [0,N]$. Then,

$$p_\pi^\mu(D) = P_\pi^\mu\left(\prod_{k=0}^{N} \mathbf{1}_D(\mathbf{s}(k)) = 1\right) = E_\pi^\mu\left[\prod_{k=0}^{N} \mathbf{1}_D(\mathbf{s}(k))\right]. \tag{2}$$

One can then introduce functions $V_k^\mu : \mathcal{S} \to [0,1]$, $k = 0, 1, \ldots, N$, associated with a Markov policy μ:

$$V_k^\mu(s) := \mathbf{1}_D(s) \int_{\mathcal{S}^{N-k}} \prod_{l=k+1}^{N} \mathbf{1}_D(s_l) \prod_{h=k+1}^{N-1} T_s(ds_{h+1}|s_h, \mu_h(s_h)) T_s(ds_{k+1}|s, \mu_k(s)),$$

$s \in \mathcal{S}$, where T_s is the controlled transition function of the embedded controlled Markov process, and $\int_{\mathcal{S}^0}(\ldots) = 1$. These functions are known as *cost-to-go functions* because they satisfy $V_k^\mu(s) = E_\pi^\mu[\prod_{h=k}^{N} \mathbf{1}_D(\mathbf{s}(h))|\mathbf{s}(k) = s]$ for any $s \in \mathcal{S}$ within the support of the distribution of $\mathbf{s}(k)$. Thus, $V_k^\mu(s)$ returns the

value of the probability of remaining within D over the (residual) time horizon $[k, N]$ starting from s at time k, under policy $\mu \in \mathcal{M}_m$ applied from π.

For any policy $\mu \in \mathcal{M}_m$, the cost-to-go functions $V_k^\mu : \mathcal{S} \to [0,1]$, $k = 0, 1 \ldots, N$, can be computed by the backward recursion:

$$V_k^\mu(s) = \mathbf{1}_D(s) \int_{\mathcal{S}} V_{k+1}^\mu(s_{k+1}) T_s(ds_{k+1}|s, \mu_k(s)),\ s \in \mathcal{S}, \tag{3}$$

initialized with $V_N^\mu(s) = \mathbf{1}_D(s)$, $s \in \mathcal{S}$, [1].

From equation (2) we have that

$$p_\pi^\mu(D) = \int_{\mathcal{S}} E_\pi^\mu\big[\prod_{k=0}^{N} \mathbf{1}_D(\mathbf{s}(k)) |\ s(0) = s\big] \pi(ds) = \int_{\mathcal{S}} V_0^\mu(s) \pi(ds). \tag{4}$$

Moreover, given that $\mu \in \mathcal{M}_m$ and that the execution associated with a Markov policy is a Markov process, it is easily seen that $E_\pi^\mu[\prod_{k=0}^{N} \mathbf{1}_D(\mathbf{s}(k)) |\ s(0) = s] = p_s^\mu(D)$; hence, $p_\pi^\mu(D) = \int_{\mathcal{S}} \mathcal{P}_s^\mu(D) \pi(ds)$.

Definition 4. *Let $\mathcal{H} = (\mathcal{Q}, n, \mathcal{A}, T_x, T_q, R)$ be a DTSHS and $D \in \mathcal{B}(\mathcal{S})$ a safe set. A Markov policy μ^* is maximally safe if $p_s^{\mu^*}(D) = \sup_{\mu \in \mathcal{M}_m} p_s^\mu(D)$, $\forall s \in D$.* □

In view of (4), a maximally safe Markov policy in fact maximizes $\mathcal{P}_\pi^\mu(D)$ for any initial state distribution π. The following theorem was shown in [1]:

Theorem 1. *Define functions $V_k^* : \mathcal{S} \to [0,1]$, $k = 0, 1, \ldots, N$, by the following dynamic programming algorithm:*

$$V_k^*(s) = \sup_{a \in \mathcal{A}} \mathbf{1}_D(s) \int_{\mathcal{S}} V_{k+1}^*(s_{k+1}) T_s(ds_{k+1}|s, a),\ s \in \mathcal{S}, \tag{5}$$

initialized with $V_N^*(s) = \mathbf{1}_D(s)$, $s \in \mathcal{S}$.

Then, $V_0^*(s) = \sup_{\mu \in \mathcal{M}_m} \mathcal{P}_s^\mu(D)$ for all $s \in \mathcal{S}$. Moreover, if $U_k(s, \lambda) = \{a \in \mathcal{A} | \mathbf{1}_D(s) \int_{\mathcal{S}} V_{k+1}^*(s_{k+1}) T_s(ds_{k+1}|s, a) \geq \lambda\}$ is compact for all $s \in \mathcal{S}$, $\lambda \in \mathbb{R}$, $k \in [0, N-1]$, then there exists a maximally safe policy $\mu^* = (\mu_0^*, \ldots, \mu_{N-1}^*)$, with $\mu_k^* : \mathcal{S} \to \mathcal{A}$, $k \in [0, N-1]$, given by

$$\mu_k^*(s) = \arg\sup_{a \in \mathcal{A}} \mathbf{1}_D(s) \int_{\mathcal{S}} V_{k+1}^*(s_{k+1}) T_s(ds_{k+1}|s, a),\ \forall s \in \mathcal{S}, \tag{6}$$

and $V_k^{\mu^*}(s) = V_k^*(s)$, $s \in \mathcal{S}$, $k = 0, 1, \ldots, N$.

In the sequel, we consider the case when

Assumption 1. *The control space $\mathcal{A}$ is a finite set.* □

Under this assumption, the compactness condition in Theorem 1 for the existence of a maximally safe Markov policy is not required. The results illustrated next can be extended to the case when $\mathcal{A}$ is a compact uncountable set in an Euclidean space following a similar line of reasoning.

4 Numerical Approximation Scheme

In this section we describe a numerical scheme for determining an approximately maximally safe policy based on the characterization of a maximally safe policy given in Theorem 1.

For the purpose of numerical approximation, it is important to note that the DP algorithm (5) as well as the optimal argument in equation (6) can be restricted to the compact set D of the state space as follows:

$$V_k^*(s) = \max_{a\in\mathcal{A}} \int_D V_{k+1}^*(s_{k+1}) T_s(ds_{k+1}|s,a),\ s \in D, \tag{7}$$

initialized with $V_N^*(s) = 1$, $s \in D$, and

$$\mu_k^*(s) = \arg\max_{a\in\mathcal{A}} \int_D V_{k+1}^*(s_{k+1}) T_s(ds_{k+1}|s,a),\ \forall s \in D. \tag{8}$$

This is quite intuitive, since for values of the state outside D the cost-to-go function is identically zero for any μ and the optimal policy $\mu^* : \mathcal{S} \to \mathcal{A}$ can be set arbitrarily. Thus, we just have to consider the values for the state within the compact set D. The advantage of restricting the state space to the compact set D is that we can adopt a *finite* discretization in the numerical approximation scheme for solving the dynamic programming algorithm and determining the optimal policy μ^*. Moreover, under suitable regularity conditions on the transition kernels defining the DTSHS, the optimal cost-to-go functions can be shown to be Lipschitz continuous over D in the continuous state component. This property (valid only within D, given the discontinuity when passing from a safe state within D to an unsafe state outside D) is used for determining bounds to the numerical approximated solution.

4.1 Discretization Procedure

State discretization. As discussed before, we can restrict computations to the compact safe set D. Thus we only need to discretize D. The set $D \subset \mathcal{S}$ is given by $D = \cup_{q\in\mathcal{Q}}\{q\} \times X^q$. The size of the continuous state space within D is measured by $\lambda := \max_{q\in\mathcal{Q}} \mathcal{L}(X_q)$, where $\mathcal{L}(X_q)$ denotes the Lebesgue measure of the set $X_q \subset \mathbb{R}^{n(q)}$. For simplicity, we assume that the compact set X^q is not empty, for all $q \in \mathcal{Q}$. Let us introduce a partition of cardinality m_q of the set $X^q \subset \mathbb{R}^{n(q)}$, $q \in \mathcal{Q}$: $X^q = \cup_{i=1}^{m_q} X_i^q$, where X_i^q, $i = 1, \ldots, m_q$, are pairwise disjoint Borel sets $X_i^q \in \mathcal{B}(\mathbb{R}^{n(q)})$, $X_i^q \cap X_j^q = \emptyset$, $\forall i \neq j$. For any q and i, pick a hybrid state value $v_i^q \in \{q\} \times X_i^q$. The set of all discrete values for the hybrid state is $\mathcal{G} := \{v_i^q, i = 1, \ldots, m_q,\ q \in \mathcal{Q}\}$. Notice that the compactness assumption on D ensures the finiteness of the cardinality of $\mathcal{G}$. Denote with d_i^q the diameter of the set X_i^q, $d_i^q = \sup\{\|x - x'\| : x, x' \in X_i^q\}$. Then, $\Delta := \max_{i=1,\ldots,m_q, q\in\mathcal{Q}} d_i^q$ represents the *grid size parameter*.

Note that, differently from [3] where the system dynamics is described through a difference nonlinear equation affected by a stochastic disturbance taking value

in a finite set, we do not have any disturbance input appearing explicitly. The definition of the dynamics of the system via stochastic kernels incorporates both the disturbance effect and the deterministic contribution to the system evolution (see the example in Section 6). As a consequence, by discretizing the state space, we implicitly define a discretization of the disturbance space.

Dynamic Programming approximation. With reference to the finite state grid $\mathcal{G}$, we introduce a discretized version of the dynamic programming algorithm (7). For $k = 0, 1, \ldots, N-1$, compute the approximated optimal cost-to-go functions as follows

$$\begin{aligned} \hat{V}_k^*(v_i^q) &= \max_{a \in \mathcal{A}} \int_D \hat{V}_{k+1}^*(s) T_s(ds|v_i^q, a), \quad \text{if } v_i^q \in \mathcal{G} \\ \hat{V}_k^*(s) &= \hat{V}_k^*(v_i^q), \text{ if } s \in \{q\} \times X_i^q, \text{ for some } i \in \{1, \ldots, m_q\}, \, q \in \mathcal{Q}, \end{aligned} \tag{9}$$

with $\hat{V}_N^*(s) = 1$, $s \in D$.

Note that due to the piecewise constant approximation of the optimal cost-to-go function and to the definition of T_s in Eqn. (1), the integral in equation (9) can be rewritten as

$$\begin{aligned} \hat{V}_k^*(v_i^q) = \max_{a \in \mathcal{A}} \Big\{ & \sum_{j=1,\ldots,m_q} \hat{V}_{k+1}^*(v_j^q) T_q(q|v_i^q, a) \int_{X_j^q} T_x(dx|v_i^q, a) \\ & + \sum_{\substack{j = 1, \ldots, m_{\bar{q}}, \\ \bar{q} \neq q \in \mathcal{Q}}} \hat{V}_{k+1}^*(v_j^{\bar{q}}) T_q(\bar{q}|v_i^q, a) \int_{X_j^{\bar{q}}} R(dx|v_i^q, a, \bar{q}) \Big\}, \end{aligned}$$

which explicitly shows that (9) consists of a computation on the finite grid $\mathcal{G}$.

Based on the approximated optimal cost-to-go $\hat{V}_k^*$, we define a Markov policy $\hat{\mu}^* = (\hat{\mu}_0^*, \ldots, \hat{\mu}_{N-1}^*)$, $\hat{\mu}_k^* : \mathcal{S} \to \mathcal{A}$, $k \in [0, N-1]$, as follows:

$$\begin{aligned} \hat{\mu}_k^*(v_i^q) &= \arg\max_{a \in \mathcal{A}} \int_D \hat{V}_{k+1}^*(s) T_s(ds|v_i^q, a), \quad \text{if } v_i^q \in \mathcal{G}, \\ \hat{\mu}_k^*(s) &= \hat{\mu}^*(v_i^q), \quad \text{if } s \in \{q\} \times X_i^q, \text{ for some } i \in \{1, \ldots, m_q\}, \, q \in \mathcal{Q}. \end{aligned} \tag{10}$$

As for any other policy, $\hat{\mu}^*$ can be arbitrarily selected outside D.

The performance of such policy $\hat{\mu}^*$ is given by the corresponding values for the cost-to-go functions $V_k^{\hat{\mu}^*}$, $k = 0, 1, \ldots, N$, that can be computed by the recursion in (3). In particular, $V_0^{\hat{\mu}^*}(s)$, $s \in D$, provides the value of the probability that the system will remain within D in the time horizon $[0, N]$ starting from $s \in D$ under policy $\hat{\mu}^*$.

In the following section, we shall show that, under proper assumptions, the performance of policy $\hat{\mu}^*$ tends to the one of a maximally safe policy, as the grid size parameter Δ goes to zero.

5 Convergence Analysis

We suppose that the stochastic kernels T_x and R on the continuous component of the hybrid state in Definition 1 of the DTSHS admit density t_x and r. We further assume that t_x and r as well as the stochastic kernel T_q satisfy the following Lipschitz condition.

Assumption 2

1. $|T_q(\bar{q}|s,a) - T_q(\bar{q}|s',a)| \leq k_1\|x - x'\|$, *for all* $s = (q,x), s' = (q,x') \in D$, $a \in \mathcal{A}$, *and* $\bar{q} \in \mathcal{Q}$,
2. $|t_x(\bar{x}|s,a) - t_x(\bar{x}|s',a)| \leq k_2\|x - x'\|$, *for all* $s = (q,x), s' = (q,x') \in D$, $a \in \mathcal{A}$, *and* $(q,\bar{x}) \in D$,
3. $|r(\bar{x}|s,a,\bar{q}) - r(\bar{x}|s',a,\bar{q})| \leq k_3\|x - x'\|$, *for all* $s = (q,x), s' = (q,x') \in D$, $a \in \mathcal{A}$, $(\bar{q},\bar{x}) \in D$, *and* $\bar{q} \neq q$,

where k_1, k_2 *and* k_3 *are suitable Lipschitz constants.* □

Based on this assumption, we can prove that the optimal cost-to-go functions satisfy some Lipschitz condition over D. This property will be fundamental in proving the convergence result. Due to space limitations, proofs are omitted.

Theorem 2. *Under Assumption 2 the optimal cost-to-go functions satisfy the following Lipschitz condition over* D*:*

$$|V_k^*(s) - V_k^*(s')| \leq \mathcal{K}\|x - x'\|, \forall\, s = (q,x),\ s' = (q,x') \in D, \tag{11}$$

for any $k \in [0,N]$. *The constant* $\mathcal{K}$ *is given by* $\mathcal{K} = mk_1 + \lambda\big(k_2 + (m-1)k_3\big)$.

Based on Theorem 2, the following convergence result can be proven.

Theorem 3. *Under Assumption 2, there exist positive constants* γ_k, $k = 0,\ldots,N$, *such that the solutions* $\hat{V}_k^*$ *to the approximated dynamic programming equations* (9) *and the cost-to-go functions of the corresponding Markov policy* $\hat{\mu}^*$ *defined in* (10) *satisfy:*

$$\big|V_k^*(s) - \hat{V}_k^*(s)\big| \leq \gamma_k\Delta,\ s \in D,$$
$$\big|V_k^*(s) - V_k^{\hat{\mu}^*}(s)\big| \leq \nu_k\Delta,\ s \in D.$$

where $\gamma_k = \gamma_{k+1} + \mathcal{K}$, $k = 1,2\ldots,N-1$, *initialized with* $\gamma_N = 0$, $\nu_k = \gamma_k + \gamma_{k+1} + \mathcal{K} + \nu_{k+1}$, $k = 1,2\ldots,N-1$, *initialized with* $\nu_N = 0$, *and* $\mathcal{K} = mk_1 + \lambda\big(k_2 + (m-1)k_3\big)$.

From this theorem it follows that the quality of the approximation by the numerical procedure described in equations (9) and (10) improves as the grid size parameter Δ decreases. The rate of convergence is linear in Δ with a constant that depends on the Lipschitz constants k_1, k_2, and k_3 in Assumption 2 through the $\mathcal{K}$ constant defined in Theorem 2. This is not surprising because we are using

a piecewise constant approximation of the optimal cost-to-go function and we expect that the optimal cost-to-go function is smoother as k_1, k_2, and k_3 are smaller. As the time horizon grows, the approximation error propagates. This is taken into account by the constants γ_k and ν_k in Theorem 3 that grow linearly as k decreases from N to 0, where N is the length of the time-horizon.

6 Computational Study

In this section we present the results of a computational study for a multi-room heating benchmark inspired by [1,7]. We numerically analyze the convergence of quantities computed by the discretization scheme proposed in Section 4.1. We also propose possible improvements in the implementation when the underlying structure of the DTSHS can be exploited to implement the DP algorithm in a computationally efficient manner.

The benchmark in [7] deals with the problem of regulating temperature in a house with n rooms via m heaters. In this report we focus on $m = 1$, the single heater case. The state of the system can be described as a hybrid state with discrete state component described by the position and status of the heater. The continuous state component can be described by the average temperature in each of the rooms. Let Δt be the time step and N be the total number of time intervals. Let $\mathbf{x}_i(k)$ denote the average temperature in room i at time k, x_a the ambient temperature and h_i a boolean vector of size equal to the number of rooms with components equal to 1 if the heater is present and in the "on" status in the corresponding room, and 0 otherwise. The average temperature in room i is governed by the following linear stochastic difference equation:

$$\mathbf{x}_i(k+1) = \mathbf{x}_i(k) + \Big(b_i(x_a - \mathbf{x}_i(k)) + \sum_{i \neq j} a_{i,j}(\mathbf{x}_j(k) - \mathbf{x}_i(k)) + c_i h_i\Big)\Delta t + \mathbf{n}_i(k) \quad (12)$$

where, $a_{i,j}, b_i, c_i$ are constants and $\{\mathbf{n}_i(k), k = 0, \ldots, N\}$ is a sequence of i.i.d Gaussian random variables with zero mean and variance ν^2. For $i \neq j$, $E[\mathbf{n}_i \mathbf{n}_j^T] = \mathbf{0}$. The heater is controlled by a thermostat that is prone to delay and failures in switching the heater between one room to another and between the "on" and "off" status: the effect of these control actions on the discrete state transitions is specified by a finite-state, finite-action, controlled Markov chain which is independent of the continuous state, that is, $T_q : \mathcal{Q} \times \mathcal{Q} \times \mathcal{A} \to [0, 1]$. One can easily check that the number of possible discrete states is $n + 1$ and the maximum number of available control actions is $n(n + 1) + 1$. We define the compact safe set to be

$$D = \cup_{q \in \{1,\ldots,(n+1)\}} \cup_{i \in \{1,\ldots,n\}} \{q\} \times \{i\} \times [x^q_{li}, x^q_{ui}],$$

where x^q_{ui} and x^q_{ui} specify the lower and upper limits for the desired temperature in room i for discrete state q. For simplicity, these are assumed to be independent of i and q. We now describe the discretization procedure as follows: we adopt a uniform partitioning of the set $[x^q_{li}, x^q_{ui}]$ into m disjoint intervals each of size $\varkappa = (x^q_{ui} - x^q_{li})/m$. Therefore, $[x^q_{li}, x^q_{ui}] = [x^q_{li}, x^q_{li} + \varkappa) \cup \ldots \cup [x^q_{li} + (m-1)\varkappa, x^q_{ui}]$. The value of the temperature in room i for the discrete state q is defined by

$e^q_{r_i i} = x^q_{li} + (r_i - 1)\varkappa$, where $r_i \in \{1, \ldots, m\}$. Define $r = [r_1, \ldots, r_n]^T$. We pick $v^q_r = (q, [e^q_{r_1 1}, \ldots, e^q_{r_n n}]^T)$ as hybrid state value. Thus, the set of all discrete values for the hybrid state is $\mathcal{G} = \{v^q_r, r = [r_1, \ldots, r_n]^T; r_i = 1, \ldots, m; i = 1, \ldots, n; q = 1, \ldots, (n+1)\}$. Let $\mathcal{N}(\cdot; \eta, \sigma^2)$ denote the probability measure over $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ associated with a Gaussian density function with mean η and variance σ^2. Then the stochastic kernel $T_s(ds'|v^q_r, a)$ that is used in the discretized dynamic programming equations (9) can be defined as follows:

$$T_s(ds'|v^q_r, a) = T_x(dx'|v^q_r, a) T_q(q'|q, a),$$

for $v^q_r \in \mathcal{G}$, $a \in \mathcal{A}$, and $s' \in \mathcal{S}$. Here, $T_x(\cdot|v^q_r, a) = \mathcal{N}(\cdot; \mu^q_r, \nu^2 I_n)$, I_n being the identity matrix of size n, $\mu^q_r = [\mu^q_{r_1}, \ldots, \mu^q_{r_n}]^T$ and $\mu^q_{r_i} = e^q_{r_i} + \Big(b_i(x_a - e^q_{r_i}) + \sum_{i \neq j} a_{i,j}(e^q_{r_j} - e^q_{r_i}) + c_i h_i\Big)\Delta t$. It is easy to check that $T_x(dx'|v^q_r, a)$ and $T_q(q'|q, a)$ satisfy the Assumption 2.

6.1 Convergence Properties

We first analyze the convergence properties of the discretization scheme for the case when $n = 2$ (two rooms). The number of modes is 3 and maximum number of allowable control actions is 7, as shown in Figure 1(a). The computations are performed for the safe set $D = \cup_{q \in \{1,2,3\}} \cup_{i \in \{1,2\}} \{q\} \times \{i\} \times [17.5, 22]^o$C. The size of time interval is $\Delta t = 1/15$ and the number of intervals is $N = 60$. The parameters values in equation (12) are: $x_a = 6$, $b_1 = b_2 = 0.25, a_{12} = a_{21} = 0.33, c_1 = 12, c_2 = 14$ and $\nu^2 = 0.9$. All the parameters should be interpreted in appropriate units. For each control action by the thermostat that elicits a transition between two different modes of the heater, the transition happens with probability 0.8. The remaining 0.2 probability is divided evenly between the "do nothing" transition that models the delay and the transition to the third, non-recommended mode that models a faulty behavior.

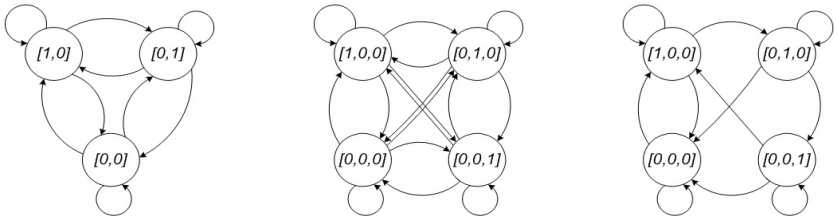

Fig. 1. (a) Maximum available control actions for $n = 2$. (b) Maximum available control actions for $n = 3$. (c) Reduced number of available control actions for $n = 3$. The discrete states are assigned numbers clockwise starting from the top-left state.

The computations of the solutions $\hat{V}^*_0$ to the approximated DP equations in (9) were performed for four discretization levels: $m \in \{9, 18, 36, 45\}$. Inspired by [1], we define the *approximately maximal probabilistic safe sets* $\hat{S}^*(\epsilon)$ with

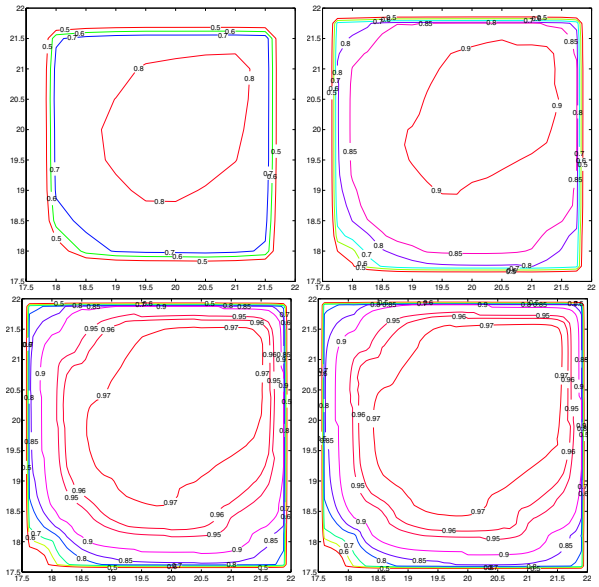

Fig. 2. Maximal probabilistic safe sets corresponding to safety levels: $0.5, 0.6, 0.7, 0.8, 0.85, 0.9, 0.95, 0.96$ and 0.97 for the case $n = 2$ and initial discrete state "off'. In going from left-to-right and top-to-bottom, the plots correspond to discretization levels of $9, 18, 36$ and 45 respectively.

safety level $(1 - \epsilon)$ as $\hat{S}^*(\epsilon) = \{s \in \mathcal{S} : \hat{V}_0^*(s) \geq (1 - \epsilon)\}$. Figure 2 shows the approximately maximal safe sets when the initial discrete state is "off", and corresponding to different safety levels. As expected, the maximal safe sets get smaller as the required safety level increases. Furthermore, as the discretization level decreases, the maximal safe sets tend to graphically converge: this visually confirms the numerical convergence of the proposed discretization scheme.

The optimal control actions for the case when the initial discrete state is "off" are plotted in Figure 3 for the four discretization levels and $k = 1$. The optimal actions at finer resolution were obtained from that of coarser resolution by nearest neighbor interpolation. It can be noticed that the regions of optimal recommended actions become more well-formed and again visually converge as the discretization step decreases.

6.2 Scaling to Higher Dimensions

We now present the results from the three-room, one heater benchmark case. For this case, the number of continuous states is $n = 3$, the number of discrete states is 4 and maximum number of allowable actions is 13, as shown in Figure 1(b). The safe set is specified to be $D = \cup_{q\in\{1,2,3,4\}} \cup_{i\in\{1,2,3\}} \{q\} \times \{i\} \times [17.5, 22]^o\mathrm{C}$. The size of time interval is $\Delta t = 1/15$ and the number of intervals is $N = 60$. The parameters values in equation (12) are: $a_{12} = a_{21} = 0.80$, $a_{13} = a_{31} = 0.60$,

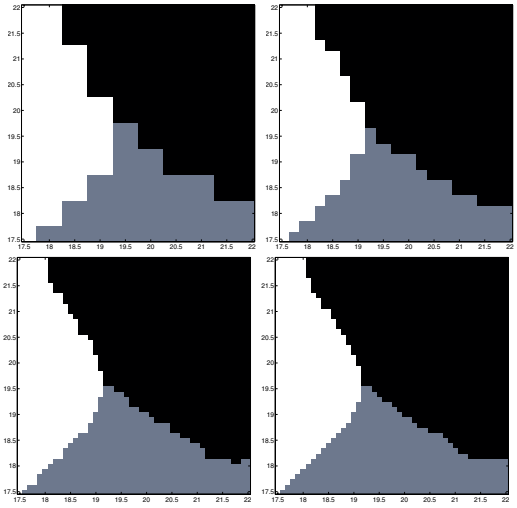

Fig. 3. Maximally safe actions for the case $n = 2$, initial discrete state "off" and $k = 1$. In going from left-to-right and top-to-bottom, the plots correspond to discretization levels of $9, 18, 36$ and 45 respectively. The colors *black*, *white* and *grey* respectively stand for "do nothing", "switch heater to room 1" and "switch heater to room 2" actions.

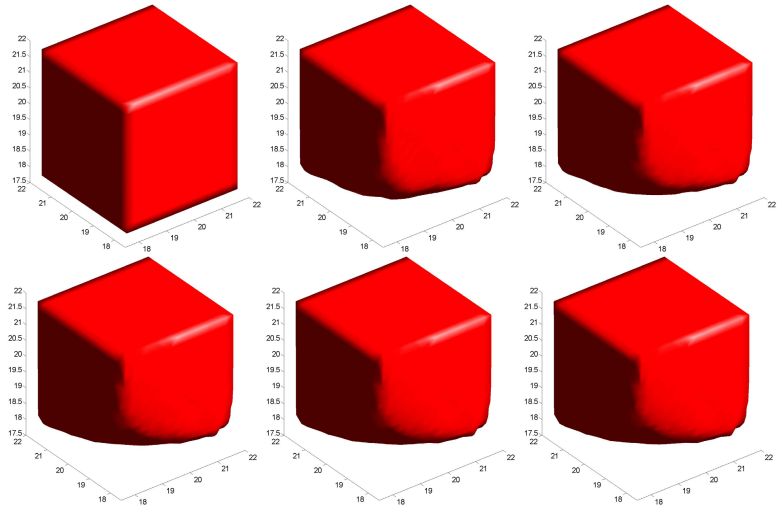

Fig. 4. Maximal probabilistic safe sets corresponding to a safety level of 0.95 for the case when $n = 3$ and initial discrete state is "off'. Available control actions are shown in Figure 1(b). In going from left to right and top to bottom, the plots correspond to $k = 60, 55, 50, 40, 20$ and 1.

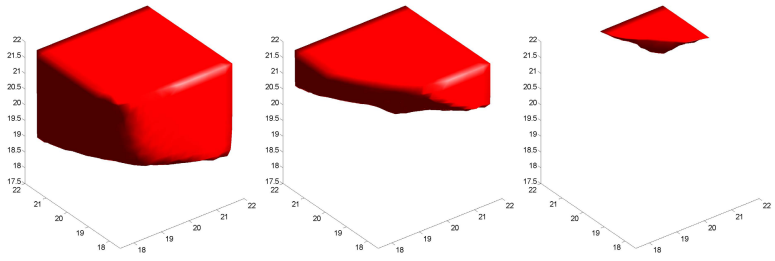

Fig. 5. Maximal probabilistic safe sets corresponding to the safety level 0.95 for the case $n = 3$ and initial discrete state "off'. The reduced set of available control actions is shown in Figure 1(c). In going from left-to-right, the plots correspond to $k = 55, 50$ and 45. The safe set for $k = 60$ is same as the corresponding safe set in Figure 4.

$a_{23} = a_{32} = 0.70$, $x_a = 6$, $b = [0.30, 0.20, 0.30]^T$, $c = [12.00, 14.00, 12.00]^T$ and $\nu^2 = 0.33$. Similar to the two-room case, the effect of control actions is described by a controlled Markov chain. The exact details are omitted due to space limitations. The computation of the DP algorithm was performed for the discretization level $m = 18$. Figure 4 shows the maximal safe sets corresponding to the safety level $1 - \epsilon = 0.95$, at different times. As expected, as the number of steps-to-go increases, the size of the safe sets also decreases. It is of interest to compare the effect of number of available control actions on the size of the maximal safe set. In order to study this, we performed the DP computations for the three-room, one heater example for the reduced set of actions shown in Figure 1(c). The resulting maximal safe sets corresponding to safety level $1 - \epsilon = 0.95$ are shown in Figure 5. We observe that the maximal safe set is becomes very small and eventually decreases to the empty set as the number of steps-to-go increases.

We finally notice an important structural property of the benchmark, namely the conditional independence of the continuous stochastic kernel: $T_x(\bar{x}|v_r^q, a) = T_x(\bar{x}_1|v_r^q, a) \times \ldots \times T_x(\bar{x}_n|v_r^q, a)$. This enables us to efficiently compute the state transition probabilities.

7 Possible Extensions and Future Work

The above discretization schemes can be directly extended to the case of uncountable, but compact control space (see Assumption 1) in a similar way. We shall include the details of this in a future work.

Even in the presence of the conditional independence property of the continuous transitions kernel, increasing the problem size further will make the computation of approximate cost-to-go value functions prohibitively expensive. This motivates the study of more efficient approaches to solve the DP algorithm in hybrid state space; the literature suggests some methods to attack this problem. One technique exploits some decentralization in the structure of the state-space in order to distribute the computations: the HS structure naturally

yields itself to this distributed approach according to the topology of the underlying graph consisting of the modes and the edges of the HS. In [4], an approach to asynchronously perform in parallel the computations with proven convergence is suggested. A second more recent approach is to solve large-scale Markov Decision Processes (MDPs) by approximating the optimal value function by a linear combination of basis functions and finding the associated optimal weights by linear programming [5]. The authors are currently investigating and experimenting these methods that leverage on the structure of the DP to achieve computationally attractive performances for the proposed schemes, to be further tested on the benchmark [7] and compared to other approaches in the literature.

References

1. Amin, S., Abate, A., Prandini, M., Lygeros, J., Sastry, S.: *Reachability analysis for controlled discrete time stochastic hybrid systems.* In J. Hespanha and A. Tiwari, eds.: Hybrid Systems: Computation and Control. Lecture Notes in Computer Science 3927. Springer Verlag, pp. 49-63 (April 2006).
2. Abate, A., Amin, S., Prandini, M., Lygeros, J., Sastry, S.: *Probabilistic Reachability for Safety and Regulation of Controlled Discrete-Time Stochastic Hybrid Systems.* In the Proceedings of the 45th IEEE Conference on Decision and Control, San Diego, CA, pp. 258-263 (December 2006).
3. Bertsekas, D.: *Convergence of Discretization Procedures in Dynamic Programming.* IEEE Transactions on Automatic Control, vol. 20, 3, pp. 415-419 (June 1975).
4. Bertsekas, D.: *Distributed Dynamic Programming.* IEEE Transactions on Automatic Control, vol. 27, 3, pp. 610-616 (June 1982).
5. Kveton, B., Hauskrecht, M., Guestrin, C.,: *Solving factored MDPs with Hybrid State and Action Variables.* Journal of Artificial Intelligence Research, vol. 27, pp. 1-49 (September 2006).
6. Prandini, M., and Hu, J.: *Stochastic reachability: Theoretical foundations and numerical approximation.* In C. Cassandras and J. Lygeros, eds.: Stochastic Hybrid Systems. Taylor & Francis Group/CRC Press (2006).
7. Fehnker, A., Ivancic F.: *Benchmarks for Hybrid Systems Verifications.* In R. Alur and G.J. Pappas, eds.: Hybrid Systems: Computation and Control. Lecture Notes in Computer Science 2993. Springer Verlag, pp. 326-341 (April 2004).
8. Mitchell, I., Templeton J.: *A Toolbox of Hamilton-Jacobi Solvers for Analysis of Nondeterministic Continuous and Hybrid Systems.* In M. Morari and R. Tiele, eds.: Hybrid Systems: Computation and Control. Lecture Notes in Computer Science 3414. Springer Verlag, pp. 480-494 (March 2005).
9. Bertsekas, D.P., Tsitsiklis, J.N.: *Neuro-Dynamic Programming.* Athena Scientific, Belmont, MA (1996).