

Optimal Control of Stochastic Hybrid Systems Based on Locally Consistent Markov Decision Processes

Xenofon D. Koutsoukos

Abstract—This paper applies a known approach for approximating controlled stochastic diffusion to hybrid systems. Stochastic hybrid systems are approximated by locally consistent Markov decision processes that preserve local mean and covariance. A randomized switching policy is introduced for approximating the dynamics on the switching boundaries. The validity of the approximation is shown by solving the optimal control problem of minimizing a cost until a target set is reached using dynamic programming. It is shown that using the randomized switching policy, the solution obtained based on the discrete approximation converges to the solution of the original problem.

I. INTRODUCTION

This paper develops a systematic way to approximate stochastic hybrid systems that is amenable to computational methods. We extend the approach presented in [9] to hybrid systems. The basic idea is to approximate the original processes by appropriate Markov Decision Processes (MDPs) defined on a discrete state space. The approximation is achieved by constructing locally consistent MDPs that preserve local mean and covariance. Based on the discrete approximation, the stochastic optimal control problem is solved using dynamic programming. The main advantage of the approach is that the control based on the discrete approximation is directly related to the original processes through the notion of local consistency and further, it converges to the solution of the original problem.

Although the approach has been already applied to several classes of stochastic systems [9], to our knowledge, the application to stochastic hybrid systems is novel. The extension of the approach to hybrid systems faces the significant challenge of approximating the dynamics in the neighborhood of the switching boundaries. The main contribution of this paper is the introduction of a randomized switching policy that guarantees under appropriate conditions continuity of the switching times. Based on this idea, convergence of the approximating processes to the stochastic hybrid process can be shown using an extension of the techniques presented in [9].

Several modeling paradigms for stochastic hybrid systems have been already proposed [8], [2], [11]. Applications include air traffic management systems [10], [4] and communication networks [7]. In this paper, we consider a model similar to that given in [8] but we assume that the

stochastic differential equations that describe the continuous dynamics are controlled diffusions.

To investigate the validity of the approximation we study the optimal control problem of minimizing a cost until a target set is reached. Optimal control of hybrid systems has attracted considerable attention; for example, see [13] and the references therein. There are also extensions to stochastic optimal control methods [1] but computational methods have not been proposed. Sufficient and necessary conditions for the stochastic optimal control problem of switching diffusions have been presented in [5]. These conditions require the solution of a partial differential equation (PDE) that cannot be solved analytically but only in simple cases. A dynamic programming method based on discretization has been also proposed in [6]. Our approach is also based on discretization but it provides a significant advantage. The solution based on the approximating model is directly related to the solution of the original problem through the notion of local consistency and it converges as the discretization becomes finer.

The main research challenge that arises is the scalability of the proposed computational methods. The state space of the approximating MDP increases exponentially with the dimension of the state space. This limits the application of the approach to low-dimensional systems. Development of efficient computational methods for analysis and design based on the locally consistent MDPs is currently under investigation.

The paper is organized as follows. Section II presents the stochastic hybrid system model and Section III the stochastic optimal control problem. The approximating method is described in Section IV. The discretized optimal control problem is presented in Section V. Section VI contains the convergence results. Finally, the approach is illustrated in Section VII with a simplified 3-dimensional example of a car with two gears.

II. STOCHASTIC HYBRID SYSTEMS

Definition 1 A stochastic hybrid system (SHS) is defined as $(X, Q, U, \Omega, A, f, \sigma, \delta, R, (x_0, q_0))$ where

- $X \subseteq \mathbb{R}^d$ is the continuous state space,
- $Q, |Q| = N$ is a finite set of discrete states,
- $U = \{U_q\}_{q \in Q}, U_q \subseteq \mathbb{R}^{m_q}$ is a collection of continuous control input sets,
- $\Omega = \{\Omega_q\}_{q \in Q}, \Omega_q \subseteq \mathbb{R}^d$ is a partition of X ,
- $A = \{A_q\}_{q \in Q}, A_q \subseteq \partial\Omega_q$ is a collection of autonomous switching sets,

This work was supported by NSF CAREER Award CNS-0347440

X. Koutsoukos is with the Department of Electrical Engineering and Computer Science, Vanderbilt University, Nashville, TN 37235, USA
xenofon.koutsoukos@vanderbilt.edu

- $f : X \times Q \times U_q \rightarrow X$ and $\sigma : X \rightarrow \mathbb{R}^{d \times p}$ are the controlled drift vectors and dispersion matrices respectively,
- $\delta : Q \times A \rightarrow Q$ is the autonomous switching map,
- $R : Q \times A \rightarrow \mathcal{P}(X)^1$ is a reset map which assigns to each q and $x \in A_q$ a reset probability kernel on X concentrated on $\Omega_{q'}$ where $q' = \delta(q, x)$,
- (x_0, q_0) is an initial probability measure on $X \times Q$.

To define the *execution* of the SHS, we consider an \mathbb{R}^p -valued Wiener process (Brownian motion) $w(t)$ and a sequence of stopping times $\{t_0 = 0, t_1, t_2, \dots\}$ that represent the times when the continuous and discrete dynamics interact. Let the state at time t_i be $(x_i, q_i) = (x(t_i), q(t_i))$ with $x_i \in \Omega_{q_i}^0$ ². While the continuous state stays in $\Omega_{q_i}^0$, $x(t)$ is evolving according to the stochastic differential equation (SDE)

$$dx = f(x, q, u)dt + \sigma(x)dw \quad (1)$$

where the discrete state $q(t) = q_i$ remains constant and the solution is understood using the Itô stochastic integral. Let $t_{i+1} = \inf\{t > t_i : x(t) \notin \Omega_{q_i}^0\}$. At t_{i+1} an autonomous discrete transition and a reset of the continuous state occur. The new discrete state is $q_{i+1} = \delta(q_i, x(t_{i+1}^-))$. The new continuous state $x(t_{i+1})$ is selected randomly according to the probability measure $R(q_i, x(t_{i+1}^-))(\Xi)$ where $\Xi \subset \Omega_{q_{i+1}}$ is a measurable set. The evolution of $x(t)$ is then described by the SDE (1) with $q(t) = q_{i+1}$ and initial condition $x(t_{i+1})$ until the next switching time.

It is assumed that the functions $f(x, q, u)$ and $\sigma(x)$ are Lipschitz continuous in x , then the SDE (1) has a unique solution. We also assume that every point $x \in A_q$ is a *regular* point for the autonomous switching set A_q . Note that if $x \in A_q$ is regular for A_q , then a sample path of (1) which starts at x will not remain in A_q for a nonempty time interval. If x is regular for A_q then it is also regular for a neighborhood of A_q around x and we can conclude that

$$\lim_{\epsilon \rightarrow 0} \int_0^t P[x(s) \in N_\epsilon(A_q)] ds = 0$$

where $N_\epsilon(A_q) = \{x : d(x, A_q) \leq \epsilon\}$ and $d(x, A_q)$ is the Euclidean distance between x and A_q . The regularity assumption ensures that the sample paths would transverse the autonomous switching sets and therefore, the autonomous switchings occur instantaneously. A sufficient condition for the regularity assumption is that the set A_q has dimension $d - 1$ and the diffusion $a(x) = \sigma(x)\sigma^T(x)$ is non-degenerate. Let $x(t_{i+1})$ be the continuous state after a discrete transition. We also assume that for every $x(t_{i+1}) \in \Xi$, $d(x(t_{i+1}), A) \geq \epsilon > 0$ and $\exists \delta > 0$ such that $P(\inf\{t > t_{i+1}, x(t) \in A\} \geq \delta) = 1$ and therefore, $t_{i+1} - t_i > \delta, i = 1, 2, \dots$, with probability 1. Finally, the continuous control is a measurable stochastic process $u(t)$ taking values in a compact set. The control policy $u(t)$ is

¹ $\mathcal{P}(X)$ denotes the family of probability measures on X .

² Ω^0 denotes the interior of the set Ω

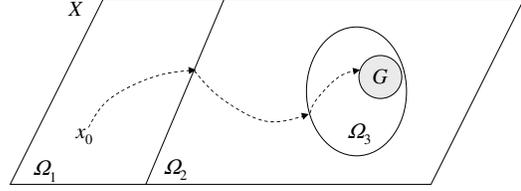


Fig. 1. Optimal control problem

said to be *admissible* if (i) it is non-anticipative with respect to the Wiener process $w(t)$, (i.e. $u(t)$ is independent of $w(s) - w(t), \forall s > t$).

III. STOCHASTIC OPTIMAL CONTROL

In this section, we describe the problem of minimizing a cost until a target set is reached. Figure 1 illustrates the optimal control problem. The target set $G \subset \mathbb{R}^d$ is assumed to be a compact set with a smooth boundary ∂G which satisfies the same regularity conditions as the autonomous switching sets. Further, we assume that $x_0 \notin G$ and $G \subset \Omega_q$ for some $q \in Q$. We define the *stopping time* τ by $\tau = \inf\{t : x(t) \in \partial G\}$. If the stopping time is not defined then the value of τ is set to infinity.

Given a stochastic hybrid system, a target set G , an initial state (x_0, q_0) at $t_0 = 0$, and a discount factor $\beta \geq 0$ the optimal control problem is formulated as the minimization of the cost

$$W(x_0, q_0, u) = E \left[\int_0^\tau e^{-\beta s} k(x(s), q(s), u(s)) ds + e^{-\beta \tau} g(x(\tau)) \right] \quad (2)$$

with respect to the admissible controls $u(t)$.

Next, we describe an approach for stochastic optimal control based on dynamic programming. To ensure that the stopping time and the cost (2) are well-defined and bounded, we assume that if $\beta = 0$ then for every initial state (x_0, q_0) there exists an admissible control policy so that the state will reach the target set G . If $\beta > 0$ the cost will be bounded even if the stopping time is not.

The value function is defined by $V(x_0, q_0) = \inf_u W(x_0, q_0, u), x_0 \in \Omega_{q_0}$. Based on a standard dynamic programming argument, we can formulate the following result. Since the initial condition can be arbitrary we will denote the value function by $V(x, q)$.

Theorem 1: Given a SHS and the cost (2), an optimal admissible control policy $u(x)$ must satisfy the conditions

$$\inf_u \left[\nabla V(x, q) f(x, q, u) + \frac{1}{2} \text{tr}(\nabla^2 V(x, q) a(x)) \right] = 0,$$

$$\forall q \in Q, \forall x \in \Omega_q^0$$

$$V(x', q') \leq V(x, q), \forall q \in Q, x \in A_q, q' = \delta(q, x), x' \sim R(q, x)(\Xi)$$

$$V(x, q) = g(x), \forall x \in \partial G, q \in Q : G \subset \Omega_q.$$

In addition, the following verification theorem can be proved in a straightforward manner.

Theorem 2: Suppose that there exist $V(x, q)$ continuous from the right, twice differentiable, and bounded in Ω_q^0 and a feedback control $\bar{u}(x)$ such that the conditions of Theorem 1

hold and $W(x, q, \bar{u})$ is bounded. Then $V(x, q)$ is the optimal cost and $\bar{u}(x)$ the optimal control.

In practice, computing the optimal value function $V(x, q)$ can be very difficult and usually requires computational methods based on discretization of the state space. In this paper, we employ a discretization method for the approximation of stochastic hybrid systems by appropriately chosen MDPs [9]. The SDE at every location q of the hybrid system is approximated by a controlled Markov process that evolves in a state space that is a discretization of the region Ω_q . The criterion which must be satisfied by the approximating MDP is *local consistency*. Local consistency means that the conditional mean and covariance of the MDP are proportional to the local mean and covariance of the original process. An approximation parameter h analogous to a "finite element size" parameterizes the approximating Markov process. As h goes to zero, the local properties of the MDP resemble the local properties of the original stochastic process.

IV. LOCALLY CONSISTENT MDPs

A. Background Material

Consider the SDE (1) evolving in Ω_q^0 . The local mean and covariance on the interval $[0, \delta]$ are

$$E[x(\delta) - x] = f(x(t), q(t), u(t))\delta + o(\delta)$$

$$E[(x(\delta) - x)(x(\delta) - x)^T] = \sigma(x(t))\sigma^T(x(t))\delta + o(\delta).$$

Let $\{\xi_n^h\}$ be an MDP on a discrete state space $S_q^h \subset \Omega_q$ with transition probabilities denoted by $p((x, q), (y, q)|u)$. A locally consistent MDP must satisfy

$$E[\Delta\xi_n^h] = f(x, q, u)\Delta t^h(x, q, u) + o(\Delta t^h(x, q, u))$$

$$E[(\Delta\xi_n^h - E[\Delta\xi_n^h])(\Delta\xi_n^h - E[\Delta\xi_n^h])^T] = \sigma(x)\sigma^T(x)\Delta t^h(x, q, u) + o(\Delta t^h(x, q, u))$$

where $\Delta\xi_n^h = \xi_{n+1}^h - \xi_n^h$, $\xi_n^h = x$ and $\Delta t^h(x, q, u)$ are appropriate interpolation intervals (or the "holding times") for the MDP. We say that a control policy $\{u_n^h, n < \infty\}$ is *admissible* if the chain has the Markov property under this policy.

The transition probabilities $p((x, q), (y, q)|u)$ and the interpolation intervals can be computed systematically from the parameters of the SDE (details can be found in [9]). In the case the diffusion matrix $a(x) = \sigma(x)\sigma^T(x)$ is diagonal and $v_i(x) = e_i$ where e_i is unit vector in the i^{th} direction, the transition probabilities are

$$p((x, q), (x \pm he_i, q)|u) = \frac{a_{ii}(x)/2 + hf_i^\pm(x, q, u)}{Q(x, q, u)} \quad (3)$$

and the interpolation interval is

$$\Delta t(x, q, u) = \frac{h^2}{Q(x, q, u)} \quad (4)$$

where $Q(x, q, u) = \sum_i [a_{ii}(x) + h|f_i(x, q, u)|]$ and $a^+ = \max\{a, 0\}$ and $a^- = \max\{-a, 0\}$ denote the positive and negative parts of a real number.

It should be noted that an MDP that is locally consistent with the SDE (1) is not unique. Any reasonable approximation that satisfies the local consistency conditions can be used. Optimization algorithms for MDPs employ iteration in policy/value space. To perform efficiently the minimization over the admissible controls at every iteration (see Section V) it is desirable to eliminate the control dependence u in the denominators of the transition probabilities and the interpolation interval. This is always possible if the SDE (1) is affine in the controls [9] and can be accomplished by defining $\bar{Q}(x, q) = \max_{u \in U_q} Q(x, q, u)$ and replace $Q(x, q, u)$ by $\bar{Q}(x, q)$ in equations (3) and (4). To ensure that the transition probabilities sum to one for each x and u , we introduce

$$p((x, q), (x, q)|u) = 1 - \sum_{y, y \neq x, q', q' \neq q} p((x, q), (y, q')|u).$$

B. Switching Boundaries

The method discussed above approximates the continuous dynamics by discrete MDPs only in the interior of the regions Ω_q . Approximating stochastic hybrid systems requires defining the MDP in the neighborhood of the switching boundaries in a way that preserves local consistency. There are two particular forms of switching boundaries of interest, smooth hypersurfaces and boundaries of polyhedral sets that have "corners". Here, we consider the case of smooth hypersurfaces. The method can be extended for the case of switching boundaries with "corners" in a straightforward manner and details are omitted due to length limitations. The main idea in this paper consists of the following steps: (i) transform the partition of the SHS to a cover, (ii) define appropriate random switching functions for approximating the behavior at the boundaries.

First, the partition $\Omega = \{\Omega_q\}$ is transformed to a cover. Consider the region Ω_q and denote its boundary A_q as

$$A_q = \{x \in \mathbb{R}^d : a_q(x) = 0\}$$

where $a_q : \mathbb{R}^d \rightarrow \mathbb{R}$ is assumed to be a smooth functional. The functional a_q must satisfy the condition $\nabla a_q(x) \neq 0$, $\forall x \in A_q$ which ensures that the boundary is an $(d-1)$ -dimensional hypersurface separating the state space. Assume without loss of generality that $\forall x \in \Omega_q^0$ we have $a_q(x) < 0$. The region Ω_q is expanded to Ω'_q defined by

$$\Omega'_q = \{x \in \mathbb{R}^d : a_q(x) - \gamma(h) = 0\}$$

where $\gamma(h) > h > 0$ for every $h > 0$ and $\gamma(h) \rightarrow 0$ as $h \rightarrow 0$. By expanding Ω_q to Ω'_q we obtain $\Omega' = \{\Omega'_q\}$, $q \in Q$. Since $\bigcup_q \Omega_q \subseteq \bigcup_q \Omega'_q$ for every q , Ω' is a cover of the state space X of the SHS.

Let $\{(\xi_n^h, q_n^h)\}$ be an MDP on a discrete state space $S = \{(x, q) \in S^h \times Q : x \in \Omega'_q\}$ with transition probabilities denoted by $\tilde{p}((x, q), (x', q')|u)$. For all states

(x, q) such that $a_q(x) < 0$ (interior of Ω_q), the system cannot switch and the transition probabilities are computed so that the MDP is locally consistent with the corresponding SDE. Hence, $\forall(x, q) \in S : a_q(x) < 0$ we have

$$\begin{aligned}\tilde{p}((x, q), (x \pm he_i, q)|u) &= p((x, q), (x \pm he_i, q)|u), \\ \tilde{p}((x, q), (x, q)|u) &= p((x, q), (x, q)|u)\end{aligned}$$

and $\tilde{p}((x, q), (x, q')|u) = 0$, if $q \neq q'$.

The switching behavior of the SHS is approximated by introducing random switching times and discretizing the reset maps. For each boundary $A'_q = \{x \in \mathbb{R}^d : a_q(x) = \gamma(h)\}$, we define the *switching rate function* $\lambda_q(x)$ such that $\lambda_q(x)$ is continuous on $O_q = \{x \in \mathbb{R}^d : 0 \leq a_q(x) \leq \gamma(h)\}$, $\lambda_q(x) = 0$ if $a_q(x) = 0$, and $\lambda_q(x) \rightarrow \infty$ as $x \rightarrow A'_q$. We also approximate the reset map $R(q, x)(\Xi)$ by a discrete transition probability kernel. If a discrete transitions $q \rightarrow q'$ occurs, the next continuous state is selected randomly from the grid points that belong to Ξ according to a uniform distribution. Let $x_i \in \Xi, i = 1, 2, \dots, \zeta$, then

$$p((x, q), (x', q')|u) = \begin{cases} 1/\zeta & \text{if } x' \in \Xi \\ 0 & \text{otherwise} \end{cases}$$

Consider the interpolation intervals $\Delta\tau_n^h = \Delta t^h(\xi_n^h, q_n^h, u_n^h)$, we define the process $\{q_n^h\}$

$$P[q_{n+1}^h \text{ changes in } \Delta\tau_n^h | q_n^h, \xi_n^h, u_n^h] = \begin{cases} 1, & \text{if } \xi_n^h \in A'_{q_n^h} \\ 1 - e^{-\lambda_{q_n^h}(\xi_n^h)\Delta t^h(\xi_n^h, q_n^h, u_n^h)}, & \text{if } \xi_n^h \in N_h(A'_{q_n^h}) \cap \Omega_{q_n^h}^0 \\ 0, & \text{otherwise} \end{cases}$$

Based on the random switching times and the discretization of the reset maps, $\forall(x, q) \in S$ such that $0 \leq a_q(x) \leq \gamma(h)$ the transition probabilities of the approximating MDP for states $(x, q), x \in O_q$ are defined as

$$\tilde{p}((x, q), (x', q')|u) = \begin{cases} (1 - e^{-\lambda_q(x)\Delta t^h(x, q, u)})p((x, q), (x', q')|u), & \text{if } q \neq q' \text{ and } x' \in \Xi \\ e^{-\lambda_q(x)\Delta t^h(x, q, u)}p((x, q), (x', q')), & \text{if } q = q' \text{ and } x = x \pm he_i \end{cases}$$

By the construction of the switching rate function, as $h \rightarrow 0$, the cover $\{\Omega'_q\}$ converges to the original partition $\{\Omega_q\}$ and the approximating process preserves local consistency.

Remark 1: We also define a random stopping rule when the state approaches the boundary of the target set $G \subset \Omega_{q_f}$. The process stops at step n with probability $1 - e^{-\lambda_G(\xi_n^h)\Delta t^h(\xi_n^h, q_n^h, u_n^h)}$ if $\xi_n^h \in N_h(\partial G) \cap \Omega_{q_f}^0$, and with probability 1 if $\xi_n^h \in \partial G$.

Remark 2: In practical applications, the physical process is usually constrained in a bounded state space. Reflective boundaries are introduced to approximate such constraints. For the approximating MDPs, the constraints are modeled as reflective (or constrained) boundaries equipped with reflection directions that point into the state space. The process is reflected back when it tries to violate the constraints.

V. COMPUTATIONAL METHODS

Consider the approximating MDP $\{\xi_n^h, q_n^h\}$ with transition probabilities $\tilde{p}((x, q), (y, q')|u)$ and denote ν_h the stopping time representing that ξ_n^h reaches the target set G . Then, assuming that the discounting is constant in the intervals $[t_n^h, t_{n+1}^h)$ the cost (2) can be approximated by

$$W^h(x_0, q_0, u) = E \left[\sum_{n=0}^{\nu_h} e^{-\beta t_n^h} c(\xi_n^h, q_n^h, u_n^h) + e^{-\beta t_{\nu_h}^h} g(\xi_{\nu_h}^h) \right]$$

where $c(\xi_n^h, q_n^h, u_n^h) = k(\xi_n^h, q_n^h, u_n^h)\Delta\tau_n^h$.

Assuming that the above sum is well-defined and bounded, minimizing the cost is a discrete problem that can be solved using standard dynamic programming algorithms based on policy or value iteration methods.

We define the optimal value function $V^h(x_0, q_0) = \inf_u W^h(x_0, q_0, u)$ and using a standard dynamic programming argument, we can derive the equation

$$V^h(x, q) = \min_u \left[\sum_{y, q'} \tilde{p}((x, q), (y, q')|u) V^h(y, q') + c(x, q, u) \right]$$

if $x \in X \setminus G$ and $V^h(x, q) = g(x)$, if $x \in G$

For the reflective boundaries, we have considered that the transition probabilities are independent of the control and the reflections are instantaneous. Let the reflections be defined by a set of vectors of unit length $r_q(x)$, a cost $c_r^T(x)E[\Delta\xi_n^h]$ is associated with the reflective boundary such that $c_r^T(x)r_q(x) \geq 0$ and $c_r^T(x) \geq 0$ elementwise to approximate the cost of the unconstrained process. The equation for the optimal value for point on the reflective boundaries is

$$V^h(x, q) = \sum_{y, q'} \tilde{p}((x, q), (y, q')|u) V^h(y, q') + c_r^T(x)E[\Delta x].$$

The optimal value function can be computed by the value iteration

$$V_{n+1}^h(x, q) = \min_u \left[\sum_{y, q'} \tilde{p}((x, q), (y, q')|u) V_n^h(y, q') + c(x, q, u) \right]$$

with the boundary conditions described above. The value function $V^h(x, q)$ is a discrete approximation of the optimal cost for the hybrid system. Finally, given the discrete optimal value function, a feedback control scheme can be designed for computing $u(x)$. For (x, q) with $x \in \Omega_q$, the control law is given by

$$u(x) = \arg \min_u \left\{ \frac{\partial V^h(x, q)}{\partial x} f(x, q, u) + k(x, q, u) \right\} \quad (5)$$

The gradient of $V^h(x, q)$ can be approximated as a weighted function of the differences ΔV^h of the values at the grid points. The advantage of the discretization method based on the locally consistent MDPs is that the cost of the approximating discrete-time process converges weakly to the original cost as shown in Section VI.

VI. CONVERGENCE

Consider the locally consistent approximating process $\{\xi_n^h, q_n^h\}$ and the optimal control input $\{u_n^h\}$ and denote $\{t_i^h\}$ the sequence of switching times. First, a continuous time interpolation $\{\psi^h(t), q^h(t)\}$ is constructed so that $\{\psi^h(t)\}$ is a Markov process. This will allow the construction of the Wiener process $w(t)$ as $h \rightarrow 0$. Denote the moments of change of $\psi^h(t)$ by $\tau_n^h, n < \infty$ with $\tau_0^h = 0$. To ensure that $\psi^h(t)$ is a Markov process, the interpolation intervals $\Delta t^h(x, q, u)$ are considered not to be deterministic but they are described by an exponential distribution with mean $\Delta t^h(x, u)$, i.e.,

$$P[\Delta\tau_n^h \leq t | \xi_n^h = x, q_n^h = q, u_n^h = u] = 1 - e^{-\frac{t}{\Delta t^h(x, q, u)}}$$

$$E[\Delta\tau_n^h | \xi_n^h = x, q_n^h = q, u_n^h = u] = \Delta t^h(x, q, u).$$

Using these new intervals we define

$$\psi^h(\tau_n^h) = \xi_n^h, n < \infty \quad (6)$$

$$\psi^h(t) = \sum_{i: \tau_{i+1}^h \leq t} \Delta \xi_i^h + \xi_0^h \quad (7)$$

and

$$q^h(t) = q_n^h, t \in [\tau_n^h, \tau_{n+1}^h)$$

$$u^h(t) = u_n^h, t \in [\tau_n^h, \tau_{n+1}^h).$$

From (7) we can write

$$\psi^h(t) = \xi_0^h + \sum_{i: \tau_{i+1}^h \leq t} [E[\Delta \xi_n^h] + \Delta \xi_n^h] - E[\Delta \xi_n^h]$$

$$= \xi_0^h + \sum_{i: \tau_{i+1}^h \leq t} E[\Delta \xi_n^h] + B_n^h$$

where $B_n^h = \sum_{i: \tau_{i+1}^h \leq t} [\Delta \xi_n^h - E[\Delta \xi_n^h]]$ is an \mathbb{R}^d -valued discrete martingale. Denote $\Delta t_i^h = \Delta t^h(\xi_i^h, q_i^h, u_i^h)$, then by local consistency $\sum_{i: \tau_{i+1}^h \leq t} E[\Delta \xi_n^h] = \sum_{i: \tau_{i+1}^h \leq t} f(\xi_i^h, q_i^h, u_i^h) \Delta t_i^h + o(\Delta t_i^h)$ and B_n^h has quadratic variation $\sum_{i: \tau_{i+1}^h \leq t} a(\xi_i^h) \Delta t_i^h + o(\Delta t_i^h)$. As $h \rightarrow 0$, we get

$$\psi^h(t) = \psi^h(0) + \int_{t_0}^t f(\psi^h(s), q^h(s), u^h(s)) ds + \delta_1^h(t) + B^h(t) \quad (8)$$

where

$$B^h(t) = \int_{t_0}^t a(\psi^h(s)) ds + \delta_2^h(t)$$

with $E[\sup_{s \leq t} \delta_1^h(s)] \rightarrow 0$ and $E[\sup_{s \leq t} \delta_2^h(s)] \rightarrow 0$.

The computational methods will give an optimal control sequence $\{u_n^h\}$. The optimal control may not exist as $h \rightarrow 0$. To show convergence, a relaxed control representation [12] is employed. Denote the space of the relaxed control as \mathcal{A} and $\mathcal{B}_{\mathcal{A}}$ and $\mathcal{B}_{\mathcal{A} \times [0, \infty)}$ the Borel σ -algebras on \mathcal{A} and $\mathcal{A} \times [0, \infty)$ respectively. A relaxed control representation

can be obtained by defining probability measures μ_t on $\mathcal{B}_{\mathcal{A}}$ and μ on $\mathcal{B}_{\mathcal{A} \times [0, \infty)}$ as

$$\mu_t(A) = I_A(\alpha(t))$$

$$\mu(A \times [0, t]) = \int_0^t \mu_s(A) ds$$

where $\alpha(t) \in A \subseteq \mathcal{A}$ and I_A is the characteristic function for the set A .

Denote μ_t^h and μ^h the corresponding probability measures for the sequence $\psi^h(t)$, then (8) can be written as

$$\psi^h(t) = \psi^h(0) + \int_{t_0}^t \int_{\mathcal{A}} f(\psi^h(s), q^h(s), \alpha^h(s)) \mu_s^h(d\alpha) ds + \delta_1^h(t) + B^h(t) \quad (9)$$

In the following, we prove that $\{\psi^h(t), q^h(t)\}$ converges weakly to the execution of the SHS. Let E denote a metric space and $D_E[0, \infty)$ the set of functions that are continuous from the right and have limits from the left. The $\psi^h(t)$ and $q^h(t)$ are viewed as elements of $D_E[0, \infty)$ for $E = \mathbb{R}^d$ and \mathbb{R} respectively. The difference with the results of [9] is that we show convergence of the switching times and then assuming finitely many switchings in a bounded interval we show convergence for the hybrid process based on the weak convergence results for $D_E[0, \infty)$ in [3] (Thm 7.8).

Theorem 3: Consider the locally consistent approximating process $\{\xi_n^h, q_n^h\}$ with an admissible control sequence $\{u_n^h\}$ and a sequence of (random) switching times $\{t_i^h\}$. Let $\{\psi^h(t), q^h(t)\}$ be the continuous time Markov interpolation and $\mu^h(\cdot)$ a relaxed control representation of $\{u_n^h\}$ for $\psi^h(t)$. Then $\{\psi^h(t), q^h(t)\}$ converges weakly to the execution of the SHS.

The next theorem shows convergence of the optimal cost. The proof follows from Theorem 3 and the results in [9]. The cost for the continuous time interpolation $\{\psi^h(t), q^h(t)\}$ with control sequence $\mu^h(t)$ is

$$W^h(x_0, q_0, \mu^h) = E \left[\int_{t_0}^{\tau^h} e^{-\beta s} k(\psi^h(s), q^h(s), \alpha) \mu_s^h(d\alpha) ds + e^{-\beta \tau^h} g(\psi^h(\tau_h)) \right].$$

Denote $V^h(x, q) = \inf_{\mu^h} W^h(x, q, \mu^h)$ and let $V^h(x, q)$ denote the optimal value function for the process $\{\xi_n^h, q_n^h\}$ with an admissible control sequence $\{u_n^h\}$. Then $|V^h(x, q) - V^h(x, q)| \rightarrow 0$ as $h \rightarrow 0$ and by abuse of notation we will use $V(h, q)$ to denote the optimal value function for both processes.

Theorem 4: Consider the locally consistent approximating process $\{\xi_n^h, q_n^h\}$ with an admissible control sequence $\{u_n^h\}$ and a sequence of (random) switching times $\{t_i^h\}$. Let $\{\psi^h(t), q^h(t)\}$ be the continuous time Markov interpolation and $\mu^h(\cdot)$ a relaxed control representation of $\{u_n^h\}$ for $\psi^h(t)$. If $\beta > 0$ or if $\beta = 0$ and the sequence τ_h is uniformly integrable then $W^h(x_0, q_0, \mu^h) \rightarrow W(x_0, q_0, \mu)$ and $V^h(x, q) \rightarrow V(x, q)$.

VII. EXAMPLE

We illustrate the proposed approach using a simplified model of a truck with flexible transmission presented in [6]. The system is described by

$$\begin{aligned} dx_1 &= x_2 dt \\ dx_2 &= -x_2 + x_3 \\ dx_3 &= -x_2 + g_q(x_2)u dt + \sigma dw, \\ q &= 1, 2, -0.1 \leq u \leq 1.1 \sigma = 0.01 \end{aligned}$$

where x_1, x_2 and x_3 are the position, velocity, and the rotational displacement of its transmission shaft respectively. The efficiency for gear q is $g_q(x)$ shown in Figure 2(a), u is the throttle, and dw is a scalar Wiener process. We have modified the model of [6] by assuming that gears switches occur at the speed of equal efficiency between the gears ($x_2 = 0.5$).

The objective is to drive the state (x_0, q_0) to the target set $G = \{x \in \mathbb{R}^2 : \frac{1}{2}x^T x \leq 0.25\}$ while minimizing the cost

$$W(x_0, q_0, u) = E \left[\int_0^\tau k(x(s), q(s), u(s)) ds + g(x(\tau)) \right]$$

where $k(x, q, u) = 1$ and $g(x) = \frac{1}{2}x^T x$. First, we approximate the system by an MDP over the region

$$X = \{x : -5 \leq x_1 \leq 1, 0 \leq x_2 \leq 1.5, -0.5 \leq x_3 \leq 1.5\}$$

using a uniform grid with approximation parameter $h = 0.25$. The reflective boundary is defined as an outer approximation of X by expanding by h in all directions. For the corner points, we select the reflection direction $r(x)$ as the vector of length h in the direction of the diagonal and we assume that the transition probabilities are independent of the control u . For the remaining points on the reflective boundary, we select $r(x)$ as the normal vector of length h pointing inside X . We transform the partition of the state space to a cover by defining two new boundaries $A = \{x : x_2 = 0.5 \pm 2h\}$ and the switching rate functions by $\lambda(x) = \frac{0.5}{\ln 0.5} \ln(1 \mp \frac{x_2 - 0.5}{2h})$ respectively. Everytime a switching occurs, we reset the continuous state to guarantee finitely many switchings. We define the transition probabilities for local consistency as defined in Section IV. The optimal value function $V(x)$ is computed using an iteration method in value space. The results shown in Figure 2 are obtained by simulating the SHS model in continuous-time (using Simulink) where the control law is computed by (5) using multilinear interpolation. Except the stochastic nature of the state trajectory, the results are very similar to those presented in [6]. The advantage of the approach is that this solution which is based on a discrete approximation that preserves the local mean and variance of the original system.

VIII. CONCLUSIONS AND FUTURE WORK

The paper employs an approximation method for solving the optimal control problem for stochastic hybrid systems

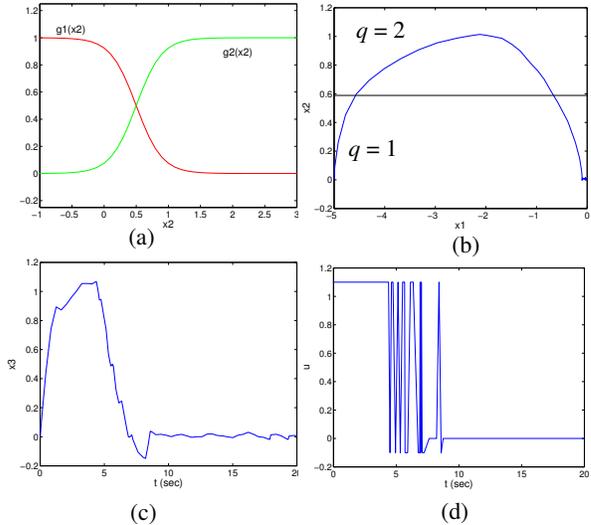


Fig. 2. (a) Gear profiles, (b),(c), and (d) Simulation results

based on locally consistent Markov decision processes that preserve the local mean and covariance of the original system. The approach gives rise to several significant problems. A fundamental challenge is to develop scalable numerical methods that can be applied to large systems. Towards this goal, currently we are investigating methods based on variable resolution grids as well as methods based on value function approximation.

REFERENCES

- [1] A. Bensoussan and J.L. Menaldi. Stochastic hybrid control. *Journal of Mathematical Analysis and Applications*, 249:261-288, 2000.
- [2] M.L. Bujorianu. Extended Stochastic Hybrid Systems and Their Reachability Problem. In *Hybrid Systems: Computation and Control - HSCC 2004*, LNCS 2993, pp. 234-249, Springer, 2004.
- [3] S.N. Ethier and T.G. Kurtz. *Markov Processes: Characterization and Convergence*. Wiley, 1986.
- [4] W. Glover and J. Lygeros. A Stochastic Hybrid Model for Air Traffic Control Simulation. In *Hybrid Systems: Computation and Control - HSCC 2004*, LNCS 2993, pp. 372-386, Springer, 2004.
- [5] M.K. Ghosh, A. Arapostathis, and S.I. Marcus. Optimal Control of Switching Diffusions with Application to Flexible Manufacturing Systems. *SIAM J. Control and Optimization*, 30(6), 1-23, 1992.
- [6] S. Hedlund and A. Rantzer. Convex dynamic programming for hybrid systems. *IEEE Trans. on Automatic Control*, 47(9):1536-1540, 2002.
- [7] J.P. Hespanha. Stochastic Hybrid Systems: Application to Communication Networks. In *Hybrid Systems: Computation and Control - HSCC 2004*, LNCS 2993, pp. 387-401, Springer, 2004.
- [8] J. Hu, J. Lygeros, and S. Sastry. Towards a Theory of Stochastic Hybrid Systems. *Hybrid Systems: Computation and Control - HSCC 2000*, LNCS 1790, pp. 160-173, Springer, 2000.
- [9] H.J. Kushner and P. Dupuis. *Numerical Methods for Stochastic Control Problems in Continuous Time*. Springer, 2001.
- [10] G. Pola, M. Bujorianu, J. Lygeros, and M. Di Benedetto. Stochastic Hybrid Models: An Overview with Application to Air Traffic Management. In *IFAC ADHS03*, June 2003.
- [11] S.N. Strubbe, A.A. Julius, and A.J. van der Schaft. Communicating piecewise deterministic Markov processes. In *IFAC ADHS03*, June 2003.
- [12] L.C. Young. *Optimal Control Theory*. AMS Chelsea, 1980.
- [13] X. Xu and P.J. Antsaklis. Results and Perspectives on Computational Methods for Optimal Control of Switched Systems, In *Hybrid Systems: Computation and Control - HSCC 2003*, LNCS 2623, pp. 540-555, Springer.