

Reachability for Partially Observable Discrete Time Stochastic Hybrid Systems [★]

Kendra Lesser ^{*}, Meeko Oishi

Dept. of Electrical and Computer Engineering, University of New Mexico, Albuquerque, USA

Abstract

When designing optimal controllers for any system, it is often the case that the true state of the system is unknown to the controller, for example due to noisy measurements or partially observable states. Incomplete state information must be taken into account in the controller's design in order to preserve its optimality. The same is true when performing reachability calculations. To estimate the probability that the state of a stochastic system reaches, or stays within, some set of interest in a given time horizon, it is necessary to find a controller (or at least prove one exists) that drives the system to that set with maximum probability. This controller, however, does not have access to the true state of the system. To date, little work has been done on stochastic reachability calculations with partially observable states. What work has been done relies on converting the reachability optimization problem to one with an additive cost function, for which theoretical results are well known. Our approach is to preserve the multiplicative cost structure when deriving a sufficient statistic that reduces the problem to one of perfect state information. Our transformation includes a change of measure that simplifies the distribution of the sufficient statistic conditioned on its previous value. We develop a dynamic programming recursion for the solution of the equivalent perfect information problem, proving that the recursion is valid, an optimal solution exists, and results in the same solution as to the original problem. We also show that our results are equivalent to those for the reformulated additive cost problem, and so such a reformulation is not required.

Key words: Hybrid Systems; Reachability; Optimal Control; Stochastic Control.

1 Introduction

The concept of stochastic reachability enables calculation of the probability that the state of a dynamical system will reach a desired set within a given time horizon. Alternately, the safety of the system may be considered by examining the probability that the system remains within some safe region. This problem has gained particular attention in the area of stochastic hybrid systems, which are composed of both continuous and discrete co-evolving states.

The reachability problem has been studied mainly for de-

terministic hybrid systems. In particular, level set methods have been used to approximate the solution to an appropriate Hamilton-Jacobi-Bellman equation, as in Tomlin et al. (2003), and Mitchell (2008). Interest in stochastic systems led to developments in the area of reachability for continuous time hybrid systems whose dynamics update stochastically. Preliminary results focused on stochastic differential equations combined with deterministic jumps between discrete states in Hu et al. (2000), and was extended to allow for random discrete jumps as well in Prandini and Hu (2006). More recently, Bujorianu (2010) examined the continuous time reachability problem as an optimal stopping problem, characterized in terms of variational inequalities. Esfahani et al. (2011) also considered the reachability problem as a stochastic optimal control problem with discontinuous payoff functions, and developed a weak dynamic programming principle for its value function. However, measurability complications that make solution strategies for the reachability problem more difficult to characterize and derive have led to interest in the discrete time equivalent, which circumvents many of those issues (see Abate et al. (2008)).

[★] This research was funded by National Science Foundation (NSF) Career Award CMMI-1254990, NSF Award CPS-1329878, NSA Science of Security Lablet at North Carolina State University (subaward to the University of New Mexico), and start-up funding from the Department of Electrical and Computer Engineering, University of New Mexico

^{*} Corresponding Author. Tel.: +1 505 353 2424; Fax: +1 505 277 0299.

Email addresses: lesser@unm.edu (Kendra Lesser), oishi@unm.edu (Meeko Oishi).

The discrete time stochastic hybrid system (DTSHS) is presented in Abate et al. (2008), and the reachability problem solved as a stochastic optimal control problem using dynamic programming, based on the theory and techniques presented in Bertsekas and Shreve (1996). Summers and Lygeros (2010) extend Abate et al. (2008) to the *reach-avoid* problem, where the objective is to avoid an unsafe region while ultimately arriving at a target set, and establishes its solution using the techniques of Bertsekas and Shreve (1996). Finally, Kamgarpour et al. (2011) build upon both of the above works to allow for a disturbance acting in opposition to the controller (as in a two-player dynamical game).

The goal of our presented work is to extend Abate et al. (2008), Summers and Lygeros (2010), and Kamgarpour et al. (2011) to the case of a partially observable system, where the controller only has access to noisy (possibly incomplete) measurements of the state. We specifically examine the reachability problem as presented in Abate et al. (2008), which focuses on the safety problem of keeping the state within a known safe region. Imperfect state information could ultimately lead to suboptimal control inputs as compared to the case in which the true state of the system is known. In the case of safety verification and reach-avoid set calculations, if the controller is falsely assumed to have perfect knowledge of the state, the reachability probability may be overestimated. This is certainly undesirable in the context of safety verification.

There has been extensive work on hybrid estimation (see for instance Hofbaur and Williams (2004), Liu and Hwang (2012), Koutsoukos et al. (2003)), but its application to the reachability problem for hybrid systems is limited. In particular, Verma and del Vecchio (2009), Verma and del Vecchio (2010) examine a continuous time hybrid system subject to continuous control inputs and both continuous and discrete disturbance inputs, where the discrete mode of the system is unknown. Verma and del Vecchio (2010) assumes separation between state estimation and control, and reduces the problem of hidden discrete modes to one of perfect state information, by redefining the state. However, in the case of reachability for hybrid systems, separation between state estimation and control cannot be assumed optimal, and so state estimation cannot be directly applied to the reachability problem.

Only very recently has there been a surge in work on the reachability problem for partially observable DTSHS (see Ding et al. (2013)). Although the reachability problem was originally presented in terms of a multiplicative cost function (Abate et al. (2008)), Ding et al. (2013) rewrites it as a terminal cost function, by appending to the state of the hybrid system a binary variable representing whether the state has remained within the desired region up to the previous time step. The partially observed control problem can be reformulated in

terms of a sufficient statistic, which encapsulates and condenses all necessary information for the control of a system. Thus the problem is recast as one of perfect information, for which solution strategies are well known.

Ding et al. (2013) make use of the fact that for an additive cost function, the posterior distribution, or probability density of the state given all available information (observations, control inputs) up to the present, provides sufficient information for control of the system (this result is derived in, e.g., Bertsekas and Shreve (1996)). Indeed, inspired by this approach, Tkachev et al. (2013) reformulate the reachability problem more generally as an additive cost optimal control problem, although they do not discuss the partially observed case.

While we examine the same problem as Ding et al. (2013), our derivations preserve the multiplicative cost structure of Abate et al. (2008). For a nonadditive cost function the posterior distribution is no longer sufficient (see Shiryaev (1964)), and a different sufficient statistic for reducing the problem to one of perfect information must be derived. However, we will show that while our approach differs from that of Ding et al. (2013), our results are in fact nearly identical. Although we preserve the multiplicative cost function, leading to a seemingly more complex problem, the additive cost formulation effectively moves the complexity from the cost function to the modified state of the system. The posterior distribution of the new state is actually the same as the distribution produced by the sufficient statistic we will derive, so that ultimately the only advantage to the additive cost formulation is its familiar and well-studied form. Further, we make use of a change of measure in formulating the sufficient statistic that enables easier calculation (by simplifying the distribution of the state of the sufficient statistic conditioned on its previous value).

To derive a sufficient statistic while preserving the multiplicative cost formulation of the reachability problem, we draw mainly upon theoretical work done in the context of partially observable risk-sensitive stochastic optimal control problems. The risk-sensitive control problem minimizes the *exponential* of a sum of costs, rather than a sum of costs, so that the cost objective is in fact nonadditive. In particular, James et al. (1994) derived a sufficient statistic for a partially observable discrete-time nonlinear system, which was further analyzed and extended in the context of a partially observable Markov decision process (POMDP) in Fernandez-Gaucherand and Marcus (1997). In the latter, the state, observation, and control took values from finite, discrete sets, whereas in the former all values were continuous. As an aside, this highlights how such problems can be regarded often interchangeably as a stochastic optimal control problem or as a Markov decision process (MDP), and results from one field usually carry over to the other, assuming the system dynamics follow the Markov property.

We present here the reachability problem using a control theory framework, and relate the controlled stochastic hybrid system to an MDP.

Motivated mainly by James et al. (1994), we derive a) a sufficient statistic for the multiplicative reachability cost function with hybrid state dynamics, and b) the dynamic programming (DP) equations to solve the reachability problem in terms of the sufficient statistic. We also introduce a change of measure to the hybrid space, so that the observations are independent of the state of the system (and are in fact independent and identically distributed). This makes for simpler dynamic programming equations, and should aid in computation and simulations. The novelty of this work is therefore 1) preservation of a *multiplicative* cost function for the reachability problem to verify the safety of a partially observable DTSHS, 2) introduction of a change of measure to make the observations independent and identically distributed 3) derivation of a sufficient statistic to convert the partially observed problem to a fully observed one, and 4) validation of a DP recursion to solve the reachability problem in terms of the sufficient statistic. Our main focus is on the theoretical foundations for solving the reachability problem for a partially observable DTSHS. We note that implementation of our technique will require further work in approximation strategies as well as in special classes of systems in which exact solutions are available.

The paper is organized as follows. First, we review the characteristics of a DTSHS, then extend it to include a hybrid observation space in Section 2. We define the reachability problem, as in Abate et al. (2008), and derive a sufficient statistic, a recursion to update the state of the sufficient statistic, and DP equations for the reachability problem in Section 3. Here we will also elaborate upon the technique presented in Ding et al. (2013), and its relation to our own method. In Section 4, we describe two examples of partially observable discrete time stochastic hybrid systems, demonstrate how to reformulate them in terms of our sufficient statistic, and discuss some of the computational challenges as well as possible solution strategies. Concluding remarks are given in Section 5.

2 Discrete Time Stochastic Hybrid Systems

A hybrid system is characterized by a set of both discrete and continuous states with interacting dynamics. The discrete state may affect the evolution of the continuous dynamics, and the continuous dynamics may affect when the discrete state changes. In the case of a discrete time stochastic hybrid system (DTSHS), both the discrete and continuous dynamics may be characterized by stochastic kernels, the product of which determines the stochastic transition kernel governing the combined discrete/continuous state of the system. We present a

slightly modified definition of a DTSHS first introduced in Abate et al. (2008).

Definition 1 (*Discrete Time Stochastic Hybrid System*). A DTSHS is a tuple $\mathcal{H} = (Q, S, \mathcal{U}, T_x, T_q)$ where

- (1) $X \subseteq \mathbb{R}^n$ is a set of continuous states
- (2) $\mathcal{Q} = \{q_1, q_2, \dots, q_{N_q}\}$ is a finite set of discrete states with cardinality N_q , and $S = X \times \mathcal{Q}$ is the hybrid state space
- (3) \mathcal{U} is a compact Borel space which contains all possible control inputs affecting discrete and continuous state transitions
- (4) $T_x : \mathcal{B}(\mathbb{R}^n) \times S \times \mathcal{U} \rightarrow [0, 1]$ is a Borel-measurable stochastic kernel which assigns a probability measure to x_{k+1} given $s_k = (x_k, q_k), u_k, q_{k+1} \forall k$: $T_x(dx_{k+1} \in B \mid q_{k+1}, s_k, u_k)$ where $B \in \mathcal{B}(\mathbb{R}^n)$, the Borel σ -algebra on \mathbb{R}^n .
- (5) $T_q : \mathcal{Q} \times S \times \mathcal{U} \rightarrow [0, 1]$ is a discrete transition kernel assigning a probability distribution to q_{k+1} given $x_k, q_k, u_k, \forall k$.

Kernels T_x and T_q can be combined for ease of notation to produce the hybrid state transition kernel

$$\tau(ds' \mid s, u) = T_x(dx' \mid x, q, u, q') T_q(q' \mid x, q, u) \quad (1)$$

The discrete state q_{k+1} update depends on q_k, x_k and u_k , and the continuous state x_{k+1} update depends on x_k, u_k , and according to the specific problem may also be governed by q_k, q_{k+1} , or both. For ease of notation we assume that the discrete state updates first, and the updated discrete state affects the continuous state, i.e. that $T_x(dx_{k+1} \mid x_k, u_k, q_{k+1})$, although modifying T_x to include q_k would not alter any subsequent results. Hence the hybrid system can also be modeled as an MDP with state space S , control space \mathcal{U} , and transition function $\tau(s' \mid s, u)$.

2.1 Partially Observable DTSHS

We assume the hybrid process $\{x_k, q_k\}$ is not measured, but rather that an observation process $y_k = (y_k^x, y_k^q)$ is available. The observations $y_k^x \in \mathbb{R}^n$ of the continuous state, and observations $y_k^q \in \mathcal{Q}$ of the discrete state, are assumed independent given the true state (x_k, q_k) , so that

$$y_k^x = h(x_k, u_{k-1}) + v_k \quad (2)$$

$$y_k^q \sim Q_{q_k, y^q}(u_{k-1}) \quad (3)$$

The continuous state observation y_k^x is subject to additive noise v_k , which is independent and identically distributed with positive density $\phi(v)$ (i.e. Gaussian), and the function h is assumed bounded and continuous, as in James et al. (1994). The distribution of the discrete state observation y_k^q follows the discrete map $Q_{q, y^q}(u) : \mathcal{Q} \times \mathcal{Q} \times \mathcal{U} \rightarrow [0, 1]$, so that $P[y_k^q = n \mid q_k = q, u_{k-1} = u] =$

$Q_{q,n}(u)$. The filtrations \mathcal{G}_k and \mathcal{Y}_k are generated by the sequences $\{s_0, \dots, s_k, y_1, \dots, y_{k-1}\}$ and $\{y_1, \dots, y_k\}$, respectively. Denote $i_k = (y_1, \dots, y_k, u_0, \dots, u_{k-1}) \in \mathcal{Y}_k \times \mathcal{U}_k = \mathcal{I}_k$, with \mathcal{U}_k the k -times product space of \mathcal{U} , as the vector of information available at time k . The information vector i_k is used to make the control decisions u_k through a control policy $\pi = (\mu_0, \dots, \mu_{N-1})$, where $\mu_k : \mathcal{I}_k \rightarrow \mathcal{U}$ is a function mapping the space of available information, \mathcal{I}_k , into the control space \mathcal{U} for all $k = 0, \dots, N-1$.

We also assume an initial Borel-measurable density on $s_0 = (x_0, q_0)$, $s_0 \sim \rho(x, q) \in P(\mathcal{S})$, i.e. that ρ lies in the space of all probability measures on \mathcal{S} . Finally, based on ρ , τ , ϕ , and $Q(u)$, we obtain a probability measure \mathbb{P}^π induced by the control policy π defined over the full state space Ω , which includes s_k and y_k for all k . We can therefore model the hybrid system with observations as a POMDP.

3 Reachability Control Problem

3.1 Cost Function

We wish to find a control policy that maximizes the probability that the true state of the system stays within some safe or desired set for a finite time horizon. As in Abate et al. (2008), this problem can be formulated as a stochastic optimal control problem. For a Borel set K , terminal time N , and predefined policy π , define the cost function as

$$r_K(\pi) = \mathbb{P}^\pi[s_i \in K \forall i = 0, \dots, N] \quad (4)$$

Since for a random variable X , $\mathbb{P}[x \in A] = \mathbb{E}[\mathbf{1}_A(x)]$, with \mathbb{E} denoting expected value and indicator function $\mathbf{1}_A(x) = 1$ if $x \in A$ and $\mathbf{1}_A(x) = 0$ otherwise, (4) is rewritten as in Abate et al. (2008):

$$r_K(\pi) = \mathbb{E}^\pi \left[\prod_{i=0}^N \mathbf{1}_K(s_i) \right] \quad (5)$$

The expected value is taken with respect to the measure \mathbb{P}^π , hence the notation \mathbb{E}^π . We want to maximize $r_K(\pi)$ with respect to the control policy π . The set Π of admissible policies will be restricted to non-randomized policies, i.e. in which $\mu_k(i_k)$ generates one control input u_k with probability 1. The optimal policy π^* is then given by

$$\pi^* = \arg \sup_{\pi \in \Pi} \{r_K(\pi)\} \quad (6)$$

We can now formally define the partially observable reachability problem we wish to solve.

Definition 2 (Problem Statement) *Consider a DTSMS (defined in Definition 1) with observations (2) - (3) and initial distribution $\rho(x, q) \in P(\mathcal{S})$. Given a safe set K and time horizon N we would like to*

- (1) *Find the maximal probability of remaining within K for N time steps, given by $\max_\pi r_K(\pi)$.*
- (2) *Find the optimal policy π^* such that $\max_\pi r_K(\pi) = r_K(\pi^*)$.*

In the case of perfect state information, where the control u_k is a function of s_k rather than i_k , Definition 2 can be solved via dynamic programming as presented in Abate et al. (2008). When only an observation process is available, however, the standard approach is to reformulate the problem as one with perfect information by redefining the state of the system in terms of a *sufficient statistic* (see, e.g. Bertsekas and Shreve (1996), Aoki (1989)) and then solving the equivalent problem using dynamic programming.

The difficulty in solving Definition 2 is twofold. First, since the cost function is multiplicative, standard sufficient statistics are not valid (i.e. the sufficient statistic cannot be the posterior distribution of the state at time k given all available information up to time k). Second, the hybrid nature of the dynamics complicates the probability space our problem is defined on. A new sufficient statistic must be derived, and its corresponding theoretical results carefully extended.

3.2 Sufficient Statistic

We will first formally define a sufficient statistic in relation to the multiplicative optimal control problem of Definition 2, which is modified from Definition 10.6 in Bertsekas and Shreve (1996).

Definition 3 *A statistic for Definition 2 is a sequence of Borel-measurable functions $(\eta_0, \eta_1, \dots, \eta_N)$ with $\eta_k : P(\mathcal{S}) \times \mathcal{I}_k \rightarrow \Sigma_k$ where Σ_k is a nonempty Borel space, for all $k = 0, \dots, N$. The sequence (η_0, \dots, η_N) is a statistic sufficient for control if*

- (1) *There exist Borel-measurable stochastic kernels $\hat{\tau}_k(d\sigma_{k+1} \mid \sigma_k, u_k)$ on Σ_{k+1} given Σ_k, \mathcal{U} such that*

$$\begin{aligned} \mathbb{P}^\pi[\eta_{k+1}(\rho, i_{k+1}) \in \underline{\Sigma}_{k+1} \mid \eta_k(\rho, i_k) = \sigma_k, u_k = \bar{u}_k] \\ = \hat{\tau}_k(\underline{\Sigma}_{k+1} \mid \sigma_k, \bar{u}_k) \end{aligned}$$

for \mathbb{P}^π almost every (σ_k, \bar{u}_k) (i.e. up to a set of measure zero with respect to measure \mathbb{P}^π).

- (2) *There exist functions $g_k : \Sigma_k \rightarrow [0, \infty)$ such that for all $\rho \in P(\mathcal{S})$, $k = 1, \dots, N$, and $\pi \in \Pi$*

$$\mathbb{E}^\pi \left[\prod_{i=1}^k \mathbf{1}_K(s_i) \mid \eta_k(\rho, i_k) = \sigma_k \right] = g_k(\sigma_k)$$

for \mathbb{P}^π almost every σ_k .

In other words, the distribution of $\sigma_k = \eta_k(\rho, i_k)$ (a specific value of the sufficient statistic which we refer to as

the *information state*) must follow the Markov property, and therefore be updated recursively according to σ_{k-1} and u_{k-1} . There also must exist an equivalent cost function whose argument is σ_k , so that the cost corresponding to a specific policy can be determined solely through the distribution of the information state. The problem presented in Definition 2 can then be redefined according to the information state σ , which itself is defined according to the sufficient statistic η .

We now propose a sufficient statistic for the partially observable reachability problem, and demonstrate that it obeys Definition 3. First, we introduce the concept of a change of measure, which is used in the derivation of our sufficient statistic to facilitate the analysis and subsequent computation. The ability to change probability measures stems from the Radon-Nikodym Theorem.

Definition 4 *The Radon-Nikodym Theorem (see Stein and Shakarchi (2005)) states that given two σ -finite measures ν and μ on a measurable space (Ω, \mathcal{M}) , if μ and ν are absolutely continuous, then there exists a μ -integrable function f on Ω such that*

$$\nu(E) = \int_E f(\omega) d\mu(\omega)$$

The function f is referred to as the Radon-Nikodym derivative, and is written as $\frac{d\nu}{d\mu}$.

Essentially, for two probability measures ν and μ , defined on the same space (Ω, \mathcal{M}) and that satisfy $\nu(E) = 0$ whenever $\mu(E) = 0$ for all $E \in \mathcal{M}$, we know that

$$\mathbb{E}^\nu[h(\omega)] = \mathbb{E}^\mu[f(\omega)h(\omega)]$$

for any \mathcal{M} -measurable function h . We can define a change of measure \mathbb{P}^\dagger from the existing measure \mathbb{P}^π on our space Ω , with \mathcal{M} being the Borel σ -algebra on Ω , so long as the continuous observation process is nowhere zero, and the discrete observation is nowhere zero on $\mathcal{Q} \times \mathcal{Q} \times \mathcal{U}$ (which would occur if certain discrete states were perfectly observable). Following James et al. (1994) and Elliot (1993), we define the Radon-Nikodym derivative Λ_k as

$$\left. \frac{d\mathbb{P}^\pi}{d\mathbb{P}^\dagger} \right|_{\mathcal{G}_k} = \Lambda_k \quad (7)$$

where

$$\Lambda_k = \prod_{l=1}^k \frac{\phi(y_l^x - h(x_l, u_{l-1})) Q_{q_l, y_l^q}(u_{l-1})}{\phi(y_l^x) \frac{1}{N_q}}$$

However, in contrast to James et al. (1994), we must contend with two separate observation processes, one continuous and one discrete. Note that in (7) we restrict the derivative to the filtration \mathcal{G}_k rather than the full state space Ω , which allows us to update the derivative as the hybrid process evolves.

Lemma 5 *Under \mathbb{P}^\dagger , the processes $\{y_k^x\}$ are independent and identically distributed (i.i.d.) with density ϕ , and the processes $\{y_k^q\}$ are i.i.d. with uniform density $\frac{1}{N_q}$.*

The proof of Lemma 5 is in the Appendix. We claim that under the measure \mathbb{P}^\dagger we can define a sufficient statistic for the reachability control problem.

$$\eta_k(\rho, i_k) = \mathbb{E}^\dagger \left[\mathbf{1}_q(q_k) \mathbf{1}_x(x_k) \prod_{i=1}^{k-1} \mathbf{1}_K(s_i) \Lambda_k \middle| i_k \right] \quad (8)$$

The sufficient statistic $(\eta_0, \eta_1, \dots, \eta_N)$ generates a sequence of unnormalized probability densities on the state s_k , conditioned on the information vector i_k , and where the dependence on ρ is implicitly defined in the measure \mathbb{P}^\dagger . The information state σ_k is therefore a modification of the posterior distribution, and represents an unnormalized conditional density of the current state joined with the probability that all previous states are in K , given a specific i_k . In order to show that our sufficient statistic (8) satisfies conditions (1) and (2) of Definition 3, we refer explicitly to the information state σ_k rather than the sufficient statistic η_k , noting that although the sufficient statistic is a function of ρ and i_k , the information state itself is a function of the state s_k , and its dependence on ρ and i_k is not explicitly stated. We first show that σ_k can be defined recursively via a bounded linear operator $T_{u,y}[\sigma]$, and therefore satisfies (1) of Definition 3. The proof can be found in the Appendix.

Lemma 6 *There exists a bounded linear operator $T : L^1(\mathcal{S}) \rightarrow L^1(\mathcal{S})$ such that σ is defined recursively as:*

$$\begin{cases} \sigma_0 = \rho \\ \sigma_k = T_{u_{k-1}, y_k}[\sigma_{k-1}] \end{cases} \quad (9)$$

where $T_{u,y}[\sigma]$ is given by

$$\begin{aligned} T_{u,y}[\sigma] &= \sum_{q^- \in \mathcal{Q}} N_q Q_{q, y^q}(u) \int_{\mathbb{R}^n} \mathbf{1}_K(x^-, q^-) \\ &\times \frac{\phi(y^x - h(x, u))}{\phi(y^x)} \tau(x, q \mid x^-, q^-, u) \sigma(x^-, q^-) dx^- \end{aligned} \quad (10)$$

In addition, $\sigma_k \in L^1(\mathcal{S})$ for all k since $\sigma_0 = \rho \in L^1(\mathcal{S})$ and T maps L^1 into L^1 .

The stochastic kernel $\hat{\tau}_k$ for the distribution of σ_{k+1} given σ_k and u_k can be written in terms of the new measure \mathbb{P}^\dagger :

$$\hat{\tau}_k(\underline{\Sigma}_{k+1} \mid \bar{\sigma}_k, \bar{u}_k) = \sum_{y^q \in \underline{Y}^q} \int_{\underline{Y}^x} \frac{1}{N_q} \phi(y^x) dy^x \quad (11)$$

with $\underline{Y}^q \times \underline{Y}^x = \{(y^q, y^x) : T_{\bar{u}_k, y^q, y^q}[\bar{\sigma}_k] \in \underline{\Sigma}_{k+1}\}$.

Next, we rewrite the cost function (5) in terms of the information state σ , for part (2) of Definition 3. Since the indicator function $\mathbf{1}_K(s)$ is in the space $L^\infty(\mathcal{S})$, the inner product of σ and $\mathbf{1}_K$ is a well defined bounded linear operator, given by

$$\langle \sigma, \mathbf{1}_K \rangle = \sum_{q \in \mathcal{Q}} \int_{\mathbb{R}^n} \sigma(x, q) \mathbf{1}_K(x, q) dx$$

The functions g_k in Definition 3 can be defined as

$$\begin{aligned} g_k(\sigma_k) &= \mathbb{E}^\pi \left[\prod_{i=1}^k \mathbf{1}_K(s_i) \middle| \eta_k(\rho, i_k) = \sigma_k \right] \\ &= \mathbb{E}^\dagger \left[\mathbf{1}_K(s_k) \prod_{i=1}^{k-1} \mathbf{1}_K(s_i) \Lambda_k \middle| i_k \right] \\ &= \langle \sigma_k, \mathbf{1}_K \rangle \end{aligned} \quad (12)$$

3.3 Equivalence to Perfect State Information Problem

Under the measure \mathbb{P}^\dagger , we can rewrite the cost function (5) as $r_K(\pi) = \mathbb{E}^\dagger \left[\Lambda_N \prod_{i=0}^N \mathbf{1}_K(x_i) \right]$. We define

$$\bar{r}_K(\bar{\pi}) = \mathbb{E}^\dagger [\langle \sigma_N, \mathbf{1}_K \rangle] \quad (13)$$

as the equivalent cost function for the reachability problem in terms of σ , using $\bar{\pi}$ to denote a policy in terms of σ (whereas π denotes a policy for the partially observed case in terms of i_k). Note that for a fixed vector of control inputs (i.e. open loop), $\mathbf{u} = [u_0, \dots, u_{N-1}]$, $r_K(\mathbf{u}) = \bar{r}_K(\mathbf{u})$:

$$\begin{aligned} r_K(\mathbf{u}) &= \mathbb{E}^\dagger \left[\Lambda_N \prod_{i=0}^N \mathbf{1}_K(s_i) \right] \\ &= \mathbb{E}^\dagger \left[\mathbb{E}^\dagger \left[\Lambda_N \prod_{i=0}^N \mathbf{1}_K(s_i) \middle| i_N \right] \right] \\ &= \mathbb{E}^\dagger [\langle \sigma_N, \mathbf{1}_K \rangle] \\ &= \bar{r}_K(\mathbf{u}) \end{aligned}$$

where the third line follows from (12). A recursion for $\bar{r}_K(\bar{\pi})$ is given as in James et al. (1994) by

$$\begin{cases} V_N^\pi(\sigma) = \langle \sigma, \mathbf{1}_K \rangle \\ V_k^\pi(\sigma) = \mathbb{E}^\dagger [V_{k+1}^\pi(T_{\bar{\mu}_k(\sigma), y_{k+1}}[\sigma])] \end{cases} \quad (14)$$

where

$$\begin{aligned} \mathbb{E}^\dagger [V_{k+1}^\pi(T_{\bar{\mu}_k(\sigma), y_{k+1}}[\sigma])] \\ = \int_{\Sigma} V_{k+1}(\sigma') \hat{\tau}(\sigma' | \sigma, \bar{\mu}_k(\sigma)) d\sigma' \end{aligned} \quad (15)$$

$$= \sum_{\mathcal{Q}} \int_{\mathbb{R}^n} V_{k+1}(T_{\bar{\mu}_k(\sigma), y_{k+1}}[\sigma]) \frac{1}{N_q} \phi(y_{k+1}^x) dy^x \quad (16)$$

so that $V_0^\pi(\rho) = \bar{r}_K(\bar{\pi})$. Next, we provide two theorems: 1) a dynamic programming algorithm to find the optimal solution to $\sup_{\bar{\pi}} \bar{r}_K(\bar{\pi})$, and the optimal policy $\bar{\pi}^* = \arg \sup_{\bar{\pi}} \bar{r}_K(\bar{\pi})$ as a function of the information state σ , and 2) proof that this optimal policy has the same value as the optimal policy for the partially observed case. The proofs of Theorems 7 and 8 are provided in the Appendix.

Theorem 7 Using the recursion (14), the dynamic programming equations

$$\begin{cases} V_N^*(\sigma) = \langle \sigma, \mathbf{1}_K \rangle \\ V_k^*(\sigma) = \sup_{u \in \mathcal{U}} \mathbb{E}^\dagger [V_{k+1}^*(T_{u, y_{k+1}}[\sigma])] \end{cases} \quad (17)$$

produce $V_0^*(\rho) = \sup_{\bar{\pi} \in \Pi} \bar{r}_K(\bar{\pi})$, where $V_k : \Sigma_k \rightarrow [0, \infty)$. For σ normalized, we have $V_k : \Sigma_k \rightarrow [0, 1]$. Furthermore, setting

$$\bar{\mu}_k^*(\sigma) = \arg \sup_{u \in \mathcal{U}} \mathbb{E}^\dagger [V_{k+1}^*(T_{u, y_{k+1}}[\sigma])] \quad (18)$$

for all $k = 0, \dots, N-1$ gives the optimal policy $\bar{\pi}^* = (\bar{\mu}_0^*, \bar{\mu}_1^*, \dots, \bar{\mu}_{N-1}^*)$, where $\bar{\mu}_k : \Sigma_k \rightarrow \mathcal{U}$.

Theorem 8 If $u_k^* = \bar{\mu}_k^*(\sigma_k)$ is optimal as defined in Theorem 7, then u_k^* is also optimal for the partially observable problem of Definition 2, and can be written as $u_k^* = \mu_k^*(i_k) = \bar{\mu}_k^*(\eta(\rho, i_k)) = \bar{\mu}_k^*(\sigma_k)$. For $\pi^* = (\mu_0^*, \mu_1^*, \dots, \mu_{N-1}^*)$, $\bar{r}_K(\bar{\pi}^*) = \sup_{\pi} r_K(\pi) = r_K(\pi^*)$.

These results guarantee that we can solve (18) as a fully observed problem for each k in terms of the new state σ , and generate a policy $\bar{\pi}$ in which the optimal action at each time step k is given as a function only of the information state at time k . Calculating the optimal policy $\bar{\pi}^*$ and optimal value $\bar{r}_K(\bar{\pi}^*)$ gives us the optimal policy π^* and optimal value $r_K(\pi^*)$.

3.4 Relationship to Additive Cost Formulation

The sufficient statistic (8) modifies the posterior distribution to include the probability of all previous states being in the set K . Were the sufficient statistic (8) derived without the change of measure (7), it would be identical (aside from a normalizing constant) to the sufficient statistic for the additive cost function formulation in Ding et al. (2013) (see equation (14)). In Ding et al. (2013), by extending the state to include a binary variable that represents whether or not the system has remained within K up to the previous time step, the

posterior distribution is also the distribution of the current state s_k , coupled with the distribution of all previous states being in K . The transition kernel for the modified state in Ding et al. (2013), equation (5), incorporates an indicator function that signals whether the state remained within the safe set at the previous time step. The prediction and update steps for a Bayesian filter (see equations (11) - (13) of Ding et al. (2013)) are used to express the sufficient statistic (14), which, aside from the change of measure and normalization, is the same as (8).

Next, the terminal payoff (15) of Ding et al. (2013) expresses the probability that the final state is within set K , and that all previous states are within set K , given the probability distribution of all previous states being in K as well as the current distribution of the final state. Were the terminal payoff written in terms of the original state, it would be identical to (12) for $k = N$. Prop. 3 in Ding et al. (2013), which describes the solution of the terminal payoff, iteratively evaluates the expected value as in Theorem 7 here (although integrating the expected value over $\hat{\tau}_k(\sigma_{k+1} \mid \sigma_k, u_k)$ does not reduce to (16) as ours does). Thus, formulating the cost function as either multiplicative or additive ultimately does not alter the end result.

4 Case Studies and Computational Issues

We provide two examples of partially observable hybrid systems to demonstrate the use of the sufficient statistic in their solution, then discuss computational challenges. Since solving (17) requires looping over all functions $\sigma \in L^1(\mathcal{S})$, an infinite space, we can only hope to use (17) as a practical solution method for special cases in which σ can be defined over a finite subspace of L^1 .

4.1 Temperature Regulation

A stochastic version of the benchmark temperature regulation problem with perfect state information is presented in Abate et al. (2007). We consider the case of one heater, which can either be turned off, or turned on to heat one of n rooms. The average temperature of room i at time k is given by the continuous variable $x_i(k)$, and the discrete state $q(k) = i$ indicates room $i \in \{1, \dots, n\}$ is heated at time k , and $q(k) = 0$ denotes the heater is off. The stochastic difference equation governing the average temperature for room i is given by

$$x_i(k+1) = (1 - b_i)x_i(k) + \sum_{i \neq j} a_{i,j}(x_j(k) - x_i(k)) + c_i h_i + b_i x_a + v_i(k)$$

with constants $a_{i,j}$, b_i , c_i , and x_a , $v_i(k)$ that are i.i.d. normally distributed random variables with mean zero and variance v^2 , and $h_i = 1$ for $q(k) = i$ and $h_i = 0$

otherwise. The control input is given by $u(k) \in \mathcal{U}$ with $\mathcal{U} = \{0, 1, \dots, n\}$, but the chosen control is not always implemented with probability 1. Instead, $q(k)$ is updated probabilistically, dependent on $u(k-1)$ and $q(k-1)$, with transition function $T_q(q(k+1) \mid q(k), u(k))$. So while function $\bar{\mu}_k(\sigma_k)$ deterministically returns a single control input, control input $u(k) = \bar{\mu}_k(\sigma_k)$ may not always be implemented.

The exact average temperature in each room is unknown, and only a noisy measurement of each room's temperature is available to the controller. The controller does, however, know which room is being heated at time k (i.e. $q(k)$ is perfectly observed). Then the observation $y(k) = (y^x(k), y^q(k))$ with $y^x(k) = [y_1^x(k) \dots y_n^x(k)]^T$ is given by

$$\begin{aligned} y_i^x(k) &= x_i(k) + w_i(k) \\ y^q(k) &= q(k) \end{aligned}$$

with $w_i(k)$ i.i.d. normally distributed with mean zero and variance w^2 (so that the distribution $\phi(w)$ is Gaussian). The transition matrix $Q(u)$ is the identity matrix for all u , so that $Q_{q,y^q}(u) = \mathbf{1}_q(y^q)$. Because the discrete state is perfectly observed, we do not keep track of a discrete observation, and it is not included in the sufficient statistic.

It is desirable to keep the temperature of each room between 17.5 and 22 degrees celsius at all times, producing the safe region $K = [17.5, 22] \times \dots \times [17.5, 22]$, which does not depend on the discrete state $q(k)$ (so $\mathbf{1}_K(s) = \mathbf{1}_K(x)$). To find the maximum probability that each room stays within the desired temperature range given that the controller only has access to the observations $y(k)$ we reformulate the problem in terms of the information state σ_k . We then use the dynamic programming equations (17) given in Theorem 7, so that

$$\begin{aligned} V_N^*(\sigma) &= \sum_{q=0}^n \int_{\mathbb{R}^n} \mathbf{1}_K(x) \sigma(x, q) dx \\ V_k^*(\sigma) &= \sup_{u \in \mathcal{U}} \int_{\mathbb{R}^n} V_{k+1}^*(T_{u,y}[\sigma]) \phi(y^x) dy^x \end{aligned}$$

With $x_i(0) \sim \mathcal{N}(\mu_i, s^2)$ for each $i = 1, \dots, n$ and $q(0) = 0$, then

$$\sigma_0(x, q) = \mathbf{1}_0(q) \prod_{i=1}^n \rho_i(x_i)$$

for $\rho_i(x)$ Gaussian with mean μ_i and variance s^2 .

However, even for the trivial case of $n = 1$ (e.g. a one room system), updating σ_k becomes complicated very

quickly. Using Lemma 6, we obtain

$$\sigma_1(x, q) = \frac{\phi(y_1^x - x)}{\phi(y_1^x)} T_q(q \mid q(0) = 0, u(0)) \times \int_K T_x(x \mid x(0), q, u(0)) \sigma_0(x(0), q(0)) dx(0)$$

The main difficulty with solving for σ_1 is the fact that the integral is evaluated over K , as opposed to over \mathbb{R} . Because of the bounds on the integral, we cannot claim σ_k is Gaussian given that σ_{k-1} is. However, because the expression does quite closely resemble a Gaussian distribution, it may be possible to approximate σ_k by an un-normalized Gaussian distribution without losing significant accuracy. We intend to explore this possibility in future works. Further, we note that there may be classes of systems for which such straightforward sufficient statistics may be found.

4.2 Skid-Steered Vehicle

A skid-steered vehicle (SSV), modeled as a switched system, is presented in Caldwell and Murphy (2011). The SSV moves according to lateral sticking and sliding of its four wheels. Caldwell and Murphy (2011) identify four modes associated with the vehicle: In mode 1, front and rear wheels stick laterally; in mode 2, front wheels stick and rear wheels skid laterally; in mode 3, front wheels skid and rear wheels stick; in mode 4, both front and rear wheels skid laterally. For each mode, the vehicle's continuous states X , Y , and θ are governed by a different set of second order ordinary differential equations (ODEs). The states X and Y represent the cartesian coordinates for the vehicle's center of geometry, and θ gives the heading of the vehicle. We can represent the continuous state of the system by $x = (X, \dot{X}, Y, \dot{Y}, \theta, \dot{\theta})$, such that $\dot{x} = f_q(x) = f(x, q)$, with discrete state $q \in \mathcal{Q} = \{1, 2, 3, 4\}$. See Caldwell and Murphy (2011) for the actual expressions for f_q , which are too lengthy to reproduce here. We discretize the ODEs using an Euler approximation method to produce an equivalent discrete time system.

The control input can be expressed as a command informing the vehicle of what mode it should be in. If the vehicle responded perfectly, we would have $q_k = u_k$ for the mode at time k . Instead, let us assume that the mode changes behave similarly to the temperature regulation problem above, where the control command is implemented with a certain probability, dependent on the current mode: $T_q(q_{k+1} \mid q_k, u_{k+1})$. The continuous state is assumed to be deterministic given the mode, so that $T_x(x_{k+1} \mid x_k, q_k, u_k) = \mathbf{1}_{f(x_k, q_k)}(x_{k+1})$. Finally, assume we have a noisy observation of the continuous state, and have an observation of the mode which is not completely reliable:

$$y_k^x = x_k + w_k \\ y_k^q \sim Q_{q, y^q}$$

The vector $w_k \in \mathbb{R}^6$ is an i.i.d. sequence of multivariate Gaussians with $w_k \sim \mathcal{N}(0, W)$. The matrix Q_{q, y^q} is given by

$$Q_{q, y^q} = \begin{bmatrix} .9 & .033 & .033 & .033 \\ .033 & .9 & .033 & .033 \\ .033 & .033 & .9 & .033 \\ .033 & .033 & .033 & .9 \end{bmatrix}$$

Each row corresponds to a given value of q , and the probability that y^q takes on each value one to four. Thus, the probability that the observed mode equals the true mode is 0.9, and if it is not the true mode, is equally likely to be any of the other three modes.

The safe region K can be defined as a path we would like the vehicle to stay on, which will be defined in terms of bounds on X and Y . For instance, we could define K as a rectangular strip $K = \{X, Y : -1 \leq X \leq 1, -10 \leq Y \leq 10\}$. Assuming the initial position of the vehicle, x_0 , is known and equal to $\hat{x}_0 \in K$, and the initial mode is independent of x_0 , uniformly distributed, and represented by $\rho(q_0)$, σ_0 is given by

$$\sigma_0(x, q) = \mathbf{1}_{\hat{x}_0}(x) \rho(q) = \frac{1}{4} \mathbf{1}_{\hat{x}_0}(x)$$

In this case σ_1 is easily calculated:

$$\begin{aligned} \sigma_1(x, q) &= \sum_{q_0=1}^4 \int_K 4Q_{q, y_1^q} T_q(q \mid q_0, u_0) \frac{\phi(y_1^x - x)}{\phi(y_1^x)} \\ &\quad \times \mathbf{1}_{f(x_0, q_0)}(x) \sigma_0(x_0, q_0) dx_0 \\ &= \sum_{q_0=1}^4 Q_{q, y_1^q} T_q(q \mid q_0, u_0) \frac{\phi(y_1^x - f(\hat{x}_0, q_0))}{\phi(y_1^x)} \\ &\quad \times \mathbf{1}_{f(\hat{x}_0, q_0)}(x) \end{aligned}$$

Thus there are four possible values of x for which $\sigma_1(x, q)$ is nonzero. Similarly, given an x_k value, $\sigma_{k+1}(x, q)$ will only be nonzero for four values of x .

$$\begin{aligned} \sigma_{k+1}(x, q) &= \sum_{q_k=1}^4 \int_K 4Q_{q, y_{k+1}^q} T_q(q \mid q_k, u_k) \\ &\quad \times \frac{\phi(y_{k+1}^x - f(x_k, q_k))}{\phi(y_{k+1}^x)} \mathbf{1}_{f(x_k, q_k)}(x) \sigma_k(x_k, q_k) dx_k \end{aligned}$$

Even when σ_k takes the above seemingly simple form, there is no immediately obvious way to avoid evaluating the value functions for all $\sigma \in L^1$ in order to solve (17).

4.3 Computational Challenges

Because of the complexity of the hybrid dynamics and cost function, the sufficient statistic and DP equations are computationally intensive. The DP equations require looping over an infinite state space. No computational work has yet been done on the reachability problem for partially observable DTSHS, and the applicability of current computational work on general DTSHS seems limited. For instance, discretization procedures for continuous state processes, like those presented in Soudjani and Abate (2013) are not immediately applicable because it is much more difficult to grid a continuous function of a continuous state (like our sufficient statistic). Other methods involve reformulating the reachability problem using chance-constrained optimization, so that the safety constraint is enforced with some probability. These chance constraints are often evaluated using sampling-based methods (see, e.g., Vrakopoulou et al. (2013)). Unfortunately, such methods run into the same problems as DP (curse of dimensionality) if formulated as multistage stochastic programs, and it is again not obvious how to extend such methods to the partially observed case, where dimensionality is an even greater issue.

For the two examples presented here, one major challenge is circumventing the evaluation of the value functions for all $\sigma \in L^1(\mathcal{S})$ to solve (17). One possible alternative is using approximate DP to estimate the value functions V_k by sampling from y_k for each k to get sample trajectories of the σ_k . Since via our change of measure the y_k are i.i.d, such sampling should be straightforward. Each y_k^x is sampled from $\phi(\cdot)$, and each y_k^y is sampled from the uniform distribution on $\{1, \dots, N_q\}$. Some work has explored approximate dynamic programming for DTSHS (see Kariotoglou et al. (2013)), in which the value function is approximated using a linear combination of basis functions, and constraints on the value function are evaluated by sampling from the state space. It is possible a similar approach could be applied to the partially observed case, where we must sample from the observation space to obtain instances of σ .

In addition, some work has been done on approximating continuous state POMDPs using point-based value iteration (see Porta et al. (2006)), albeit in the context of additive cost functions with the belief state as a sufficient statistic. The method exploits the structure of the value function, and uses Monte Carlo methods to generate a set of samples from the belief space, in order to approximate the value function at a given starting belief state. Further, Brunskill et al. (2008) have applied this to a system with hybrid dynamics. More recently, there has been a greater focus on solving continuous state POMDPs through approximation and sampling, including a Monte Carlo technique that samples both from the belief space and the state space (Bai et al. (2011)). Although our cost function and sufficient statistic are dif-

ferent, we are currently working to extend these methods to solve the reachability problem.

5 Conclusion

We have presented a statistic sufficient for the control of a partially observable discrete time stochastic hybrid system, when the objective is to maximize the probability of remaining within a safe set for some finite time horizon. By redefining the partially observed optimal control problem as one that is fully observed, with state variable σ (the information state generated by the sufficient statistic), we are able to define an optimal control policy as a function of σ . This control policy is equivalent to the policy defined as a function of the information vector, and leads to the same maximal safety probability. Further, we showed the equivalence between our approach and one that uses an additive cost function.

The major disadvantage of the sufficient statistic is that the dynamic programming equations must be solved for every possible $\sigma \in L^1$ at every time step. As a direct solution method, it is seemingly impractical. However, there may be cases where σ can be limited to a subset of L^1 so that the dynamic programming equations can be solved. Further, our choice of measure in defining the sufficient statistic may lend itself well to approximate dynamic programming techniques that avoid looping over all possible states.

We hope to investigate ways in which these and other partially observable hybrid systems may be solved using our sufficient statistic, via practical solution strategies. While such approximate results would still be suboptimal, they may be more informative and accurate than a suboptimal controller that results from a separation between state estimation and control, or from using the belief state given the observations as an (in)sufficient statistic. We intend to explore approximate solution methods using our sufficient statistic, and compare them to other suboptimal schemes, in future work.

Acknowledgements

The authors would like to thank the reviewers for their thoroughness and insight, and also David Weirich for his help with functional analysis.

6 Appendix

6.1 Proof of Lemma 5

PROOF. The proof follows that of Elliot (1993).

$$\begin{aligned}\mathbb{P}^\dagger[y_k^x \in A, y_k^q = q \mid \mathcal{G}_k] &= \mathbb{E}^\dagger[\mathbf{1}_A(y_k^x) \mathbf{1}_{\{q\}}(y_k^q) \mid \mathcal{G}_k] \\ &= \frac{\mathbb{E}^\dagger[\mathbf{1}_A(y_k^x) \mathbf{1}_{\{q\}}(y_k^q) \Lambda_k^{-1} \mid \mathcal{G}_k]}{\mathbb{E}^\dagger[\Lambda_k^{-1} \mid \mathcal{G}_k]}\end{aligned}$$

Pulling Λ_{k-1}^{-1} outside the expected value from both the numerator and denominator and canceling, since Λ_{k-1} is \mathcal{G}_k measurable, the numerator reduces to

$$\begin{aligned}&\int_A \frac{\phi(y) \frac{1}{N_q}}{\phi(y - h(x_k, u_{k-1})) Q_{q_k, q}(u_{k-1})} \\ &\quad \times \mathbb{P}^\dagger[y_k^x = y, y_k^q = q \mid s_k, u_{k-1}] dy \\ &= \frac{1}{N_q} \int_A \phi(y) dy = \mathbb{P}^\dagger[y_k^q = q] \mathbb{P}^\dagger[y_k^x \in A]\end{aligned}$$

and the denominator becomes

$$\begin{aligned}&\sum_{q=1}^{N_q} \int_{\mathbb{R}^n} \frac{\phi(y) \frac{1}{N_q}}{\phi(y - h(x_k, u_{k-1})) Q_{q_k, q}(u_{k-1})} \\ &\quad \times \mathbb{P}^\dagger[y_k^x = y, y_k^q = q \mid s_k, u_{k-1}] dy \\ &= \sum_{q=1}^{N_q} \frac{1}{N_q} \int_{\mathbb{R}^n} \phi(y) dy = 1\end{aligned}$$

Hence,

$$\mathbb{P}^\dagger[y_k^x \in A, y_k^q = q \mid \mathcal{G}_k] = \mathbb{P}^\dagger[y_k^q = q] \mathbb{P}^\dagger[y_k^x \in A]$$

■

6.2 Proof of Lemma 6

PROOF. We first show that T is a bounded linear operator mapping L^1 into itself. We then show that σ_k can be defined recursively using T . Linearity follows obviously from the properties of integrals. For any function $\nu \in L^1(\mathcal{S})$, $u \in \mathcal{U}$, $y \in \mathcal{Y}$,

$$\begin{aligned}\|T_{u,y}[\nu]\|_{L^1} &= \sum_{q \in \mathcal{Q}} \int_{\mathbb{R}^n} \left| \frac{\phi(y^x - h(x, u))}{\phi(y^x)} N_q Q_{q,y^q}(u) \right. \\ &\quad \times \left[\sum_{q^- \in \mathcal{Q}} \int_{\mathbb{R}^n} \mathbf{1}_K(x^-, q^-) \tau(x, q \mid x^-, q^-, u) \right. \\ &\quad \times \left. \left. \nu(x^-, q^-) dx^- \right] \right| dx\end{aligned}$$

$$\begin{aligned}\|T_{u,y}[\nu]\|_{L^1} &\leq \sum_{q \in \mathcal{Q}} \frac{N_q Q_{q,y^q}(u)}{|\phi(y^x)|} \int_{\mathbb{R}^n} |\phi(y^x - h(x, u))| dx \\ &\quad \times \left[\sum_{q^- \in \mathcal{Q}} \int_{\mathbb{R}^n} |\nu(x^-, q^-)| dx^- \right] \quad (19) \\ &\leq \sum_{q \in \mathcal{Q}} \frac{N_q Q_{q,y^q}(u)}{|\phi(y^x)|} \sum_{q^- \in \mathcal{Q}} \int_{\mathbb{R}^n} |\nu(x^-, q^-)| dx^- \\ &= M \|\nu\|_{L^1} \quad (20)\end{aligned}$$

Equation (19) follows because $\mathbf{1}_K(x^-, q^-) \tau(x, q \mid x^-, q^-, u) \leq 1$ for all x, q, x^-, q^- , and (20) follows because ϕ is a distribution, and $\int \phi(x) dx = 1$, therefore $\int \phi(y - h(x)) dx \leq 1$ for h a bounded continuous function. Hence for any $\nu \in L^1(\mathcal{S})$, T is a bounded linear operator, with $T_{u,y}[\nu] \in L^1(\mathcal{S})$.

Induction shows that $\sigma_k = T_{u_{k-1}, y_k}[\sigma_{k-1}]$. Given $\sigma_0 = \rho$,

$$\begin{aligned}T_{u_0, y_1}[\rho](x, q) &= \sum_{q_0 \in \mathcal{Q}} \int_{\mathbb{R}^n} \mathbf{1}_K(x_0, q_0) \frac{\phi(y_1^x - h(x, u_0))}{\phi(y_1^x)} \\ &\quad \times N_q Q_{q, y_1^q}(u_0) \tau(x, q \mid x_0, q_0, u_0) \rho(x_0, q_0) dx_0 \\ &= \mathbb{E}^\dagger[\mathbf{1}_q(q_1) \mathbf{1}_x(x_1) \mathbf{1}_K(x_0, q_0) \Lambda_1 \mid i_1] \\ &= \sigma_1(x, q)\end{aligned}$$

Given $\sigma_l = T_{u_{l-1}, y_l}[\sigma_{l-1}] \forall l = 1, \dots, k$,

$$\begin{aligned}T_{u_k, y_{k+1}}[\sigma_k](x, q) &= \sum_{q_k \in \mathcal{Q}} \int_{\mathbb{R}^n} \mathbf{1}_K(x_k, q_k) \\ &\quad \times \frac{\phi(y_{k+1}^x - h(x, u_k))}{\phi(y_{k+1}^x)} N_q Q_{q, y_{k+1}^q}(u_k) \\ &\quad \times \tau(x, q \mid x_k, q_k, u_k) \sigma_k(x_k, q_k) dx_k \\ &= \sum_{q_k \in \mathcal{Q}} \int_{\mathbb{R}^n} \mathbf{1}_K(x_k, q_k) \frac{\phi(y_{k+1}^x - h(x, u_k))}{\phi(y_{k+1}^x)} \\ &\quad \times N_q Q_{q, y_{k+1}^q}(u_k) \tau(x, q \mid x_k, q_k, u_k) \\ &\quad \times \left[\sum_{q_0, \dots, q_{k-1}} \int_{\mathbb{R}^n \times \dots \times \mathbb{R}^n} \prod_{i=1}^{k-1} \mathbf{1}_K(x_i, q_i) \right. \\ &\quad \times \tau(x_i, q_i \mid x_{i-1}, q_{i-1}, u_{i-1}) \mathbf{1}_K(x_0, q_0) \\ &\quad \times \left. \rho(x_0, q_0) \Lambda_{k-1} dx_0, \dots, dx_{k-1} \right] dx_k\end{aligned}$$

$$\begin{aligned}
&= \mathbb{E}^\dagger \left[\mathbf{1}_q(q_{k+1}) \mathbf{1}_x(x_{k+1}) \prod_{i=1}^k \mathbf{1}_K(x_i, q_i) \right. \\
&\quad \left. \times \Lambda_{k+1} \middle| i_{k+1} \right] \\
&= \sigma_{k+1}(x, q)
\end{aligned}$$

■

6.3 Proof of Theorem 7

The following proofs are based on those appearing in Bertsekas and Shreve (1996), chapters 6 and 11. To facilitate the connection between these proofs and those appearing in Bertsekas and Shreve (1996) we first reformulate the recursion (17) as a minimization

$$\sup_{\bar{\pi}} V_0^{\bar{\pi}}(\sigma_0) = -\inf_{\bar{\pi}} -V_0^{\bar{\pi}} = -\inf_{\bar{\pi}} J_0^{\bar{\pi}}$$

Let $J_k^{\bar{\pi}}(\sigma) = -V_k^{\bar{\pi}}(\sigma)$ and $\bar{\Pi} = \{\bar{\pi} = (\bar{\mu}_0, \bar{\mu}_1, \dots) : \bar{\mu}_i(\sigma_i) \in \mathcal{U} \forall i\}$. In the following we drop the bar notation over $\bar{\pi}$, and use π to denote a policy with respect to the sufficient statistic. The recursion for $J_k^*(\sigma)$ is identical to that of $V_k^*(\sigma)$ in (17) except that $J_N^*(\sigma) = -\langle \sigma, \mathbf{1}_K \rangle$.

Next we define the operators

$$\begin{aligned}
H_\mu[J] &= \mathbb{E}^\dagger [J(T_{\mu(\sigma), y}[\sigma])] \\
&= \sum_{y^q} \int_{\mathbb{R}^n} J(T_{\mu(\sigma), y}[\sigma]) \phi(y^x) \frac{1}{N_q} dy^x \\
H[J] &= \inf_{\mu(\sigma) \in \mathcal{U}} H_\mu[J]
\end{aligned}$$

The operator $H_\mu[J]$ preserves the linearity and boundedness of value function $J(\sigma)$ for all σ in L^1 , which can be seen from a straightforward induction argument. Because $J_N^\pi(\sigma)$ is a bounded linear functional, this then implies that $J_k^\pi(\sigma)$ is a bounded linear functional for all $k = 0, \dots, N$ and for all σ in L^1 .

Lemma 9 *For all bounded linear functionals $J, \sigma \in L^1$, $\pi \in \Pi$, and $r \in \mathbb{R}^+$*

$$H_\mu[J] \leq H_\mu[J + r] \leq H_\mu[J] + r$$

PROOF. Because $J \leq J + r$ when $r \geq 0$, we get the

following:

$$\begin{aligned}
H_\mu[J](y) &= \sum_{y^q} \int_{\mathbb{R}^n} J(T_{\mu(\sigma), y}[\sigma]) \frac{1}{N_q} \phi(y^x) dy^x \\
&\leq \sum_{y^q} \int_{\mathbb{R}^n} (J(T_{\mu(\sigma), y}[\sigma]) + r) \frac{1}{N_q} \phi(y^x) dy^x \\
&= H_\mu[J + r](y) \\
&= \sum_{y^q} \int_{\mathbb{R}^n} J(T_{\mu(\sigma), y}[\sigma]) \frac{1}{N_q} \phi(y^x) dy^x + r \\
&= H_\mu[J] + r
\end{aligned}$$

■

Proposition 10 *For any $M \in \mathbb{N}$, where $J_0^*(\sigma) = \inf_{\pi \in \Pi} J_0^\pi(\sigma)$,*

$$J_0^*(\sigma) = H^M[J_M^*(\sigma)]$$

Further, for any $\epsilon > 0$ there exists an M -stage ϵ -optimal policy π_ϵ , defined as

$$J_0^* \leq J_0^{\pi_\epsilon} \leq J_0^* + \epsilon$$

PROOF. (By backwards induction on M)

For $M = N$,

$$J_N^*(\sigma) = H^0[J_N^*(\sigma)]$$

(since $H^0[J] = J$). Also, because J_N does not depend on a control input, $J_N^*(\sigma) = J_N^\pi(\sigma)$ for any policy $\pi \in \Pi$. Therefore, for any $\epsilon > 0$, $J_N^{\pi_\epsilon}(\sigma) = J_N^*(\sigma) \leq J_N^*(\sigma) + \epsilon$.

Assume for $M = k + 1$ that $J_{k+1}^*(\sigma) = H^{N-k-1}[J_N^*(\sigma)]$ and that for all $\epsilon > 0$ there exists a policy π_ϵ such that $J_{k+1}^{\pi_\epsilon}(\sigma) \leq J_{k+1}^*(\sigma) + \epsilon$. Then by Lemma 9, for any $\mu \in \Pi$

$$H_\mu[J_{k+1}^{\pi_\epsilon}] \leq H_\mu[J_{k+1}^* + \epsilon] \leq H_\mu[J_{k+1}^*] + \epsilon$$

By aggregating μ with the control function π_ϵ to get $\hat{\pi}_\epsilon = (\mu, \pi_\epsilon)$, we then have

$$\begin{aligned}
\inf_{\pi} J_k^\pi(\sigma) &\leq J_k^{\hat{\pi}_\epsilon}(\sigma) = H_\mu[J_{k+1}^{\pi_\epsilon}] \\
&\leq H_\mu[J_{k+1}^*] + \epsilon
\end{aligned}$$

Since the above holds for any $\mu \in \Pi$,

$$\begin{aligned}
\inf_{\pi} J_k^\pi(\sigma) &\leq H[J_{k+1}^*(\sigma)] \\
&= H[H^{N-k-1}[J_N^*(\sigma)]] = H^{N-k}[J_N^*(\sigma)]
\end{aligned}$$

By definition $H^{N-k}[J_N^*(\sigma)] \leq J_k^*(\sigma)$, hence $H^{N-k}[J_N^*(\sigma)] = J_k^*(\sigma)$.

Next, by the induction argument, for any $\hat{\epsilon} > 0$, let $\hat{\pi}$ be such that

$$J_{k+1}^{\hat{\pi}}(\sigma) \leq J_{k+1}^*(\sigma) + \frac{\hat{\epsilon}}{2}$$

Let $\hat{\mu} \in \Pi$ be such that

$$H_{\hat{\mu}}[J_{k+1}^*] \leq H[J_{k+1}^*] + \frac{\bar{\epsilon}}{2}$$

Define $\hat{\pi}_\epsilon = (\hat{\mu}, \hat{\pi})$. Then

$$\begin{aligned} J_k^{\hat{\pi}_\epsilon} &= H_{\hat{\mu}}[J_{k+1}^{\hat{\pi}}] \leq H_{\hat{\mu}}[J_{k+1}^*] + \frac{\hat{\epsilon}}{2} \\ &\leq H[J_{k+1}^*] + \frac{\hat{\epsilon}}{2} + \frac{\hat{\epsilon}}{2} \\ &= J_k^* + \hat{\epsilon} \end{aligned}$$

It follows from induction that $J_0^* \leq J_0^{\pi_\epsilon} \leq J_0^* + \epsilon$ for any M . \blacksquare

We also use the result from Bertsekas and Shreve (1996) on the existence of a uniformly N -stage optimal policy $\pi^* = (\mu_0^*, \mu_1^*, \dots)$, which we give without proof (see Ch. 6), since the proof does not change in our context.

Proposition 11 *A policy is uniformly N -stage optimal if and only if $H_{\mu_k^*}[H^{N-k-1}[J_N^*]] = H^{N-k}[J_N^*]$ for all $k = 0, \dots, N$, and this policy exists if and only if the infimum of*

$$H^{N-k}[J_N^*] = \inf_{u \in \mathcal{U}} H_u[H^{N-k-1}[J_N^*]]$$

is attained for all $\sigma \in L^1$ and $k = 0, \dots, N$. A sufficient condition for the infimum to be attained is that

$$U_k(\sigma, \lambda) = \{u \in \mathcal{U} : H_u[H^{N-k-1}[J_N^*]] \leq \lambda\}$$

is compact for all $\sigma \in L^1$, $\lambda \in \mathbb{R}$, and $k = 0, \dots, N$.

Substituting $V_k^* = -J_k^*$, it is clear that Prop. 10 validates the dynamic programming algorithm (17), and proves the existence of at least an ϵ -optimal policy, and so the first part of Theorem 7 is proved. Finally, using Prop. 11, because \mathcal{U} is defined as a compact (i.e. closed and bounded) Borel set, and J_k (and so V_k) is bounded for all $\sigma \in L^1$ and for each $u \in \mathcal{U}$, then there exists some $u \in \mathcal{U}$ such that the infimum in $\inf_{u \in \mathcal{U}} H_u[H^{N-k-1}[J_N^*]] = \inf_{u \in \mathcal{U}} \mathbb{E}[J_k(T_{u,y}[\sigma])]$ is attained for all k (and likewise the supremum of V_k is achieved for all k). Therefore, for (17), there always exists an optimal policy π given by (18).

6.4 Proof of Theorem 8

PROOF. For a vector $\bar{u} = [u_0, u_1, \dots, u_{N-1}]$ with each $u_i \in \mathcal{U}$, we have by definition that

$$r_K(\bar{u}) = \bar{r}_K(\bar{u}) \quad \forall \bar{u} \in \mathcal{U}_N$$

Since $\sigma_k = \eta_k(\rho, i_k)$, the control policy $\bar{\pi} = (\bar{\mu}_0(\sigma_0), \bar{\mu}_1(\sigma_1), \dots)$ can be rewritten as a function of the information vector i_k , where $\bar{\mu}_k(\sigma_k) = \bar{\mu}_k(\eta_k(\rho, i_k)) = \mu_k(i_k)$. Then by defining the policy π in terms of μ , we have that $r_K(\pi) = \bar{r}_K(\bar{\pi})$ for all $\bar{\pi} \in \bar{\Pi}$. If $\bar{\pi}^*$ is optimal for $\bar{r}_K(\bar{\pi})$, it then must be optimal for $r_K(\pi)$ as well, and further,

$$r_K(\pi^*) = \bar{r}_K(\bar{\pi}^*) \quad \blacksquare$$

References

- Abate, A., Amin, S., Prandini, M., Lygeros, J., Sastry, S., 2007. Computational approaches to reachability analysis of stochastic hybrid systems, in: Hybrid Systems: Computation and Control.
- Abate, A., Prandini, M., Lygeros, J., Sastry, S., 2008. Probabilistic reachability and safety for controlled discrete time stochastic hybrid systems. *Automatica* 44, 2724–2734.
- Aoki, M., 1989. Optimization of Stochastic Systems: Topics in Discrete-Time Dynamics. Academic Press, Inc.
- Bai, H., Hsu, D., Lee, W., Ngo, V.A., 2011. Monte carlo value iteration for continuous-state POMDPs, in: Algorithmic Foundations of Robotics IX. volume 68, pp. 175–191.
- Bertsekas, D.P., Shreve, S.E., 1996. Stochastic Optimal Control: The Discrete-Time Case. Athena Scientific.
- Brunskill, E., Kaelbling, L., Lozano-Perez, T., Roy, N., 2008. Continuous-state POMDPs with hybrid dynamics, in: Symposium on Artificial Intelligence and Mathematics.
- Bujorianu, M.L., 2010. Variational inequalities for the stochastic reachability problem, in: IEEE Conference on Decision and Control, pp. 1854–1859.
- Caldwell, T., Murphy, T., 2011. Switching mode generation and optimal estimation with application to skid-steering. *Automatica* 47, 50–64.
- Ding, J., Abate, A., Tomlin, C., 2013. Optimal control of partially observable discrete time stochastic hybrid systems for safety specifications, in: American Control Conference, pp. 6231–6236.
- Elliot, R.J., 1993. A general recursive discrete-time filter. *Journal of Applied Probability* 30, 575–588.
- Esfahani, P.M., Chatterjee, D., Lygeros, J., 2011. On a problem of stochastic reach-avoid set characterization, in: IEEE Conference on Decision and Control and European Control Conference, pp. 7069–7074.

- Fernandez-Gaucherand, E., Marcus, S.I., 1997. Risk-sensitive optimal control of hidden markov models: Structural results. *IEEE Transactions on Automatic Control* 42, 1418–1422.
- Hofbaur, M.W., Williams, B.C., 2004. Hybrid estimation of complex systems. *IEEE Transactions on Systems, Man, and Cybernetics* 34, 2178–2191.
- Hu, J., Lygeros, J., Sastry, S., 2000. Towards a theory of stochastic hybrid systems. *Hybrid Systems: Computation and Control*, 160–173.
- James, M.R., Baras, J.S., Elliot, R.J., 1994. Risk-sensitive control and dynamic games for partially observed discrete-time nonlinear systems. *IEEE Transactions on Automatic Control* 39, 780–792.
- Kamgarpour, M., Ding, J., Summers, S., Abate, A., Lygeros, J., Tomlin, C., 2011. Discrete time stochastic hybrid dynamical games: Verification and controller synthesis, in: *IEEE Conference on Decision and Control*.
- Kariotoglou, N., Summers, S., Summers, T., Kamgarpour, M., Lygers, J., 2013. Approximate dynamic programming for stochastic reachability, in: *European Control Conference*, pp. 584 – 589.
- Koutsoukos, X., Kurien, J., Zhao, F., 2003. Estimation of distributed hybrid systems using particle filtering methods, in: *Hybrid Systems: Computation and Control*, pp. 298–313.
- Liu, W., Hwang, I., 2012. A stochastic approximation based state estimation algorithm for stochastic hybrid systems, in: *American Control Conference*, pp. 312–317.
- Mitchell, I., 2008. The flexible, extensible and efficient toolbox of level set methods. *Journal of Scientific Computing* 35, 300–329. 10.1007/s10915-007-9174-4.
- Porta, J.M., Vlassis, N., Spain, M.T., Poupart, P., 2006. Point-based value iteration for continuous POMDPs. *Journal of Machine Learning Research* 7, 2329–2367.
- Prandini, M., Hu, J., 2006. *Stochastic Reachability: Theoretical Foundations and Numerical Approximation*. Springer Verlag. *Lecture Notes in Control and Information Sciences*, pp. 107–139.
- Shiryaev, A., 1964. On markov sufficient statistics in nonadditive Bayes problems of sequential analysis. *Theory of Probability and its Applications* 9, 604–618.
- Soudjani, S., Abate, A., 2013. Adaptive and sequential gridding procedures for the abstraction and verification of stochastic processes. *SIAM Journal on Applied Dynamical Systems* 12, 921–956.
- Stein, E., Shakarchi, R., 2005. *Real Analysis: Measure Theory, Integration, and Hilbert Spaces*. Princeton Lectures in Analysis, Princeton University Press.
- Summers, S., Lygeros, J., 2010. Verification of discrete time stochastic hybrid systems: A stochastic reach-avoid decision problem. *Automatica* 46, 1951–1961.
- Tkachev, I., Katoen, J.P., Mereacre, A., Abate, A., 2013. Quantitative automata-based controller synthesis for non-autonomous stochastic hybrid systems, in: *Hybrid Systems: Computation and Control*, pp. 293–302.
- Tomlin, C., Mitchell, I., Bayen, A., Oishi, M., 2003. Computational techniques for the verification and control of hybrid systems, in: *Proceedings of the IEEE*, pp. 986–1001.
- Verma, R., del Vecchio, D., 2009. Continuous control of hybrid automata with imperfect mode information assuming separation between state estimation and control, in: *IEEE Conference on Decision and Control*, Shanghai.
- Verma, R., del Vecchio, D., 2010. Control of hybrid automata with hidden modes: Translation to a perfect state information problem, in: *IEEE Conference on Decision and Control*.
- Vrakopoulou, M., Margellos, K., Lygeros, J., Andersson, G., 2013. Probabilistic guarantees for the n-1 security of systems with wind power generation, in: *Reliability and Risk Evaluation of Wind Integrated Power Systems*. Springer India, pp. 59–73.