A Dynamic Game Framework for Verification and Control of Stochastic Hybrid Systems

Jerry Ding
Maryam Kamgarpour
Sean Summers
Alessandro Abate
John Lygeros
Claire Tomlin

Electrical Engineering and Computer Sciences
University of California at Berkeley

Technical Report No. UCB/EECS-2011-101
http://www.eecs.berkeley.edu/Pubs/TechRpts/2011/EECS-2011-101.html

September 7, 2011

A Dynamic Game Framework for Verification and Control of Stochastic Hybrid Systems

Jerry Ding, Maryam Kamgarpour, Sean Summers,
Alessandro Abate, John Lygeros, Claire Tomlin

Abstract

This report develops a framework for analyzing probabilistic reachability and safety problems for discrete time hybrid systems in a stochastic game setting. In particular, we formulate these problems as zero-sum stochastic games between the control, whose objective is to reach a desired target set or remain within a given safe set, and a rational adversary, whose objective is opposed to that of the control. It will be shown that the maximal probability of achieving the reachability and safety objectives subject to the worst-case adversary behavior can be computed through a suitable dynamic programming algorithm. Furthermore, there always exists an optimal control policy which achieves this worst-case probability, regardless of the choice of disturbance strategy, and sufficient conditions for optimality of the policy can be derived in terms of the dynamic programming recursion. We provide several application examples from the domains of air traffic management and robust motion planning to demonstrate our modeling framework and solution approach.

I. Introduction

In application scenarios ranging from air traffic management [1], [2], [3], automotive control [4], systems biology [5], [6], to bipedal walking [7], the behavior of the system one would like to control can be described in terms of a hybrid system abstraction in which the system state evolves both in the discrete and continuous domain. While the discrete state can be used to capture qualitative behavior of the system, for example the operating modes of a flight management system or the foot impact of a bipedal walker, the continuous state can be used to capture quantitative characteristics such as the velocity and heading of the aircraft or the joint angles of

These authors contributed equally to this work.

a biped. When the evolution of the discrete and continuous state can be modeled probabilistically, for example through analysis of statistical data, then a natural modeling framework is that of a stochastic hybrid system (SHS) [8], [9], [10].

For a controlled SHS, the performance of the closed-loop system can be measured in terms of the probability that the system trajectory obeys certain desired specifications. Of interest to safety-critical applications are probabilistic safety and reachability problems where the control objective is to maximize the probability of remaining within a certain safe set or of reaching a desired target set. Early contributions in this domain for continuous-time SHS include [11] and [12]. A slight generalization of the safety and reachability problem is considered in [13], called the *reach-avoid*, in which the control objective is to reach the desired target set, while remaining within a safe set. The probability of achieving this objective is shown to be the solution of an appropriate Hamiltion-Jacobi-Bellman equation. To address the computational issues associated with probabilistic reachability analysis, the authors in [14], [15] propose a Markov chain approximation of the SHS [16], and apply the results to air traffic control studies, while in [17], the authors propose an efficient method for estimating an upper bound for the safety probability for an autonomous SHS using barrier certificates. In the discrete time case, a theoretical framework for the study of probabilistic safety problems is established in [18] for discrete-time stochastic hybrid systems (DTSHS). These results are generalized in [19] to address the reach-avoid problem in discrete time, with considerations for time-varying and stochastic target sets and safe sets given in [20] and [21].

The main contribution of this work is the extension of results for probabilistic safety and reachability of DTSHS, as studied in [18] and [19], to a zero-sum stochastic game setting. In particular, we consider a scenario where the evolution of the system state is affected not only by the actions of the control (as in previous work), but also by the actions of a rational adversary, whose objectives are opposed to that of the control. This is motivated by practical applications such as conflict resolution between pairs of aircraft in air traffic management [2] and control of networked systems subject to external attacks [22], in which the decisions of the external agent may not obey any *a priori* known probability distribution, but rather depend in a rational fashion on the current state of the system and possibly also on the actions of the control. In such cases, a robust control design must take into account the worst-case behavior of the external agent.

This technical report builds upon our recent work in [23], where a stochastic game formulation of the reach-avoid problem was proposed and a theoretical result was given stating that under certain standard continuity/compactness assumptions [24], [25] on the underlying stochastic kernels and player action spaces, there exists: 1) a dynamic programming algorithm for determining the maximal probability of satisfying the reach-avoid objective, subject to the worst-case adversary behavior, called the max-min reach-avoid probability; 2) a max-min control policy which achieves the max-min reach-avoid probability under the worst-case adversary strategy. Due to space limitations, the proof was omitted from [23]. In this report, we provide a detailed proof of this result and, in the process, derive sufficient conditions of optimality for both the control and the adversary in terms of the dynamic programming recursion. Furthermore, we demonstrate how the reach-avoid problem can be specialized to address the safety problem and provide two equivalent dynamic programming algorithms for computing the maximal safety probability in a stochastic game setting. Finally, we provide a detailed analysis of several application examples in order to illustrate the value and implications of the proposed framework. This includes a tutorial example in which both the max-min safety probability and max-min control policy can be calculated in an analytic fashion, as well as a pairwise aircraft collision avoidance example from air traffic management, in which a stochastic wind model is used to account for wind effects on aircraft motion.

It is worth noting that although there is a large number of previous results in the field of non-cooperative stochastic games [26], [24], [25], [27], [28], we found the direct application of these results to our problem difficult, for several reasons. First, the pay-off functions for the safety and reach-avoid problems are sum-multiplicative, which prevents the use of results from the more common additive cost problems [24], [27]. Second, although there is previous work on more general utility functions which depend on the entire history of the game [25], [28], the results are primarily for the existence of randomized policies under a symmetric information pattern. Due to practical implementation and also robustness concerns, we are more interested in the existence of nonrandomized policies under a non-symmetric information pattern. Finally, an important feature of hybrid systems is that the dynamics in the continuous state space can change abruptly across certain switching boundaries. This requires a relaxation of the continuity assumptions in the continuous state space such as those given in [26].

The report is organized as follows. In Section II, we describe the model for a discrete-time stochastic hybrid game (DTSHG), along with some technical assumptions necessary for our results. In Section III, we give a formal stochastic game formulation of the probabilistic reach-avoid problem and discuss how the safety problem can be considered a special case of the reach-avoid problem. In Section IV, we state and prove our main result from [23] for computing the max-min reach-avoid probability, and give sufficient conditions of optimality for both the control policy and the adversary strategy. This is followed by a discussion of some practical implications, along with an analytic example to illustrate the dynamic programming algorithm. The specialization of this result to the safety problem is given in Section IV-B. We demonstrate the application of the proposed modeling and analysis framework in Section V through several practical examples. Finally, some concluding remarks along with directions for future work are given in Section VI.

II. Discrete-time Stochastic Hybrid Dynamic Game Model

In this section, we describe the model proposed in [23] for a Discrete-time Stochastic Hybrid Dynamical Game (DTSHG), as an extension of the discrete-time stochastic hybrid systems (DTSHS) model proposed in [18], [19] to a two-player stochastic game setting. Following standard conventions, we will refer to the control as Player I and to the adversary as Player II and denote by $\mathcal{B}(\cdot)$ the Borel $\sigma-$algebra on a topological space.

Definition 1 (DTSHG)**.** A discrete-time stochastic hybrid dynamical game between two players is a tuple $\mathcal{H} = (\mathcal{Q}, n, \mathcal{A}, \mathcal{D}, \tau_v, \tau_q, \tau_r)$ as described below.

- *Discrete state space* $\mathcal{Q} := \{q^1, q^2, ..., q^m\}$, where $m \in \mathbb{N}$;
- *Dimension of continuous state space* $n : \mathcal{Q} \to \mathbb{N}$: a map which assigns to each discrete state $q \in \mathcal{Q}$ the dimension of the continuous state space $\mathbb{R}^{n(q)}$. The hybrid state space is given by $X := \bigcup_{q \in \mathcal{Q}} \{q\} \times \mathbb{R}^{n(q)}$;
- *Player I control space* $\mathcal{A}$: a nonempty, compact Borel space;
- *Player II control space* $\mathcal{D}$: a nonempty, compact Borel space;
- *Continuous state transition kernel* $\tau_v : \mathcal{B}(\mathbb{R}^{n(\cdot)}) \times X \times \mathcal{A} \times \mathcal{D} \to [0, 1]$: a Borel-measurable stochastic kernel on $\mathbb{R}^{n(\cdot)}$ given $X \times \mathcal{A} \times \mathcal{D}$ which assigns to each $x = (q, v) \in X$, $a \in \mathcal{A}$ and $d \in \mathcal{D}$ a probability measure $\tau_v(\cdot|x, a, d)$ on the Borel space $(\mathbb{R}^{n(q)}, \mathcal{B}(\mathbb{R}^{n(q)}))$;
- *Discrete state transition kernel* $\tau_q : \mathcal{Q} \times X \times \mathcal{A} \times \mathcal{D} \to [0, 1]$: a discrete stochastic kernel on $\mathcal{Q}$ given $X \times \mathcal{A} \times \mathcal{D}$ which assigns to each $x \in X$ and $a \in \mathcal{A}$, $d \in \mathcal{D}$ a probability

distribution $\tau_q(\cdot|x, a, d)$ over $\mathcal{Q}$;

- *Reset transition kernel* $\tau_r : \mathcal{B}(\mathbb{R}^{n(\cdot)}) \times X \times \mathcal{A} \times \mathcal{D} \times \mathcal{Q} \to [0, 1]$: a Borel-measurable stochastic kernel on $\mathbb{R}^{n(\cdot)}$ given $X \times \mathcal{A} \times \mathcal{D} \times \mathcal{Q}$ which assigns to each $x \in X$, $a \in \mathcal{A}$, $d \in \mathcal{D}$ and $q' \in \mathcal{Q}$ a probability measure $\tau_r(\cdot|x, a, d, q')$ on the Borel space $(\mathbb{R}^{n(q')}, \mathcal{B}(\mathbb{R}^{n(q')}))$.

We note briefly that the measurability requirements given here are necessary for the formal characterization of the probability that the system state remains within or reaches certain desired subsets of the state space, under the semantics of a DTSHG. In the following, we will provide a detailed discussion of this semantics.

Within a non-cooperative dynamical game setting, it is important to first define the *information pattern*, namely the knowledge that each player has about the state of the system and the actions of the other player. With different information patterns, one may arrive at different formulations of the stochastic game, along with correspondingly different algorithms for computing the payoff functions for each player [29]. Motivated by robust control applications, we consider here an information pattern which gives an advantage to player II: at each time step, Player I is allowed to select inputs based upon the current state of the system, while Player II is allowed to select inputs based upon both the system state and the control input of Player I. A mathematical description of this is given below.

Definition 2 (Markov Policy)**.** A Markov policy for player I is a sequence $\mu = (\mu_0, \mu_1, ..., \mu_{N-1})$ of Borel measurable maps $\mu_k : X \to \mathcal{A}$, $k = 0, 1, ..., N - 1$. The set of all admissible Markov policies for player I is denoted by $\mathcal{M}_a$.

Definition 3 (Markov Strategy)**.** A Markov strategy for player II is a sequence $\gamma = (\gamma_0, \gamma_1, ..., \gamma_{N-1})$ of Borel measurable maps $\gamma_k : X \times \mathcal{A} \to \mathcal{D}$, $k = 0, 1, ..., N-1$. The set of all admissible Markov strategies for player II is denoted by Γ_d.

For a given initial condition $x_0 = (q_0, v_0) \in X$, player I policy $\mu \in \mathcal{M}_a$, and player II strategy $\gamma \in \Gamma_d$, the semantics of a DTSHG can be described as follows. At the beginning of each time step k, each player obtains a measurement of the current system state $x_k = (q_k, v_k) \in X$. Using this information, player I selects his/her controls as $a_k = \mu_k(x_k)$, following which player II selects his/her controls as $d_k = \gamma_k(x_k, a_k)$. The discrete state is then updated according to the discrete transition kernel as $q_{k+1} \sim \tau_q(\cdot|x_k, a_k, d_k)$. If the discrete state remains the same, namely

$q_{k+1} = q_k$, then the continuous state is updated according to the continuous state transition kernel as $v_{k+1} \sim \tau_v(\cdot|x_k, a_k, d_k)$. On the other hand, if there is a discrete jump, the continuous state is instead updated according to the reset transition kernel as $v_{k+1} \sim \tau_r(\cdot|x_k, a_k, d_k, q_{k+1})$.

Following this description, we can use a similar approach as in [18] to compose the transition kernels τ_v, τ_q, and τ_r to form a hybrid state transition kernel $\tau : \mathcal{B}(X) \times X \times \mathcal{A} \times \mathcal{D} \to [0, 1]$ which describes the evolution of the hybrid state under the influence of player I and player II inputs. Specifically, let $x = (q, v) \in X$, then

$$\tau((q', dv')|(q, v), a, d, q') = \begin{cases} \tau_v(dv'|(q, v), a, d)\tau_q(q|(q, v), a, d), & \text{if } q' = q \\ \tau_r(dv'|(q, v), a, d, q')\tau_q(q'|(q, v), a, d), & \text{if } q' \neq q. \end{cases}$$

Using the hybrid transition kernel τ, we can now give a formal definition for the executions of a DTSHG.

Definition 4 (DTSHG Execution)**.** Let $\mathcal{H}$ be a DTSHG and $N \in \mathbb{N}$ be a finite time horizon. A stochastic process $\{x_k,\ k = 0, ..., N\}$ with values in X is an execution of $\mathcal{H}$ associated with a Markov policy $\mu \in \mathcal{M}_a$, a Markov strategy $\gamma \in \Gamma_d$, and an initial condition $x_0 \in X$ if its sample paths are obtained according to Algorithm II.1.

Algorithm II.1 DTSHG Execution

Input Initial hybrid state $x_0 \in X$, Markov policy $\mu = (\mu_0, \mu_1, ..., \mu_{N-1}) \in \mathcal{M}_a$, Markov strategy $\gamma = (\gamma_0, \gamma_1, \ldots, \gamma_{N-1}) \in \Gamma_d$

Output Sample Path $\{x_k, k = 0, ..., N\}$

```
Set k = 0;
while k < N do
    Set a_k = μ_k(x_k);
    Set d_k = γ_k(x_k, a_k);
    Extract from X a value x_{k+1} according to τ(·|x_k, a_k, d_k);
    Increment k;
end while
```

As the player I policy μ and player II strategy γ are in general time-varying, the execution $\{x_k, k = 0, \ldots, N\}$ of the DTSHG is a time inhomogeneous stochastic process on the sample space $\Omega = X^{N+1}$, endowed with the canonical product topology $\mathcal{B}(\Omega) := \prod_{k=1}^{N+1} \mathcal{B}(X)$. In particular, the evolution of the closed-loop hybrid state trajectory can be described in terms of the transition kernels $\tau^{\mu_k,\gamma_k}(\cdot|x) := \tau(\cdot|x, \mu_k(x), \gamma_k(x, \mu_k(x)))$, $k = 0, \dots, N$. By Proposition 7.28 of [30], for a given $x_0 \in X$, $\mu \in \mathcal{M}_a$, $\gamma \in \Gamma_d$, these stochastic kernels induce a unique probability measure $P_{x_0}^{\mu,\gamma}$ on Ω as defined by

$$P_{x_0}^{\mu,\gamma}(X_0 \times X_1 \times \cdots \times X_N) = \int_{X_0} \int_{X_1} \cdots \int_{X_N} \tau^{\mu_{N-1},\gamma_{N-1}}(dx_N|x_{N-1}) \times \cdots \times \tau^{\mu_0,\gamma_0}(dx_1|x_0') \delta_{x_0}(dx_0'), \tag{1}$$

where $X_0, ..., X_k \in \mathcal{B}(X)$ are Borel sets and δ_{x_0} denotes the probability measure on X which assigns mass one to the point $x_0 \in X$.

A. Examples

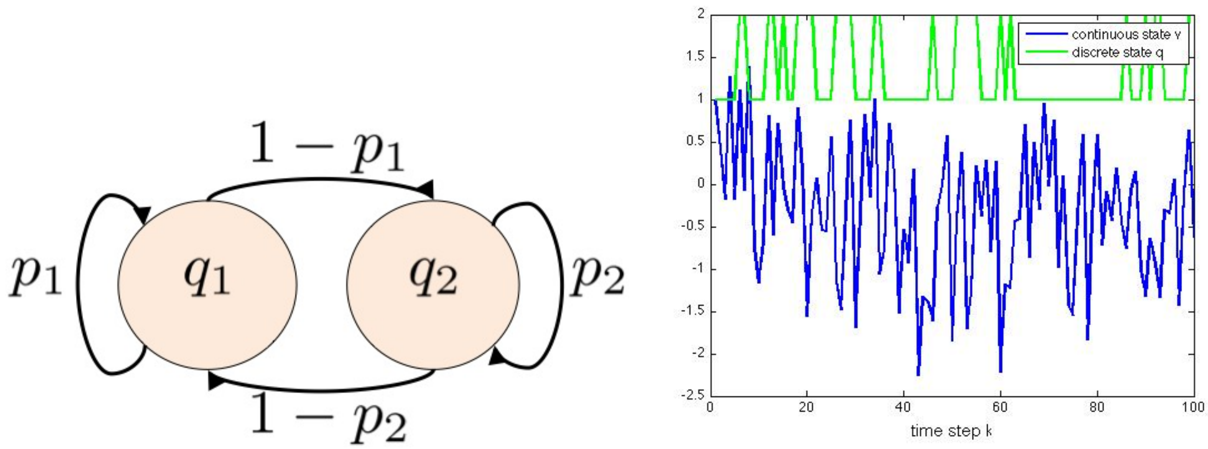
In order to illustrate the definitions given so far, we provide two concrete examples for which DTSHG models are constructed from the given problem descriptions.

1) 2-mode DTSHG: Consider a simple discrete-time stochastic hybrid system with two modes of operation $\mathcal{Q} = \{q^1, q^2\}$, as shown in Fig. 1a. The transitions between the discrete modes are modeled probabilistically, with the probability of dwelling in mode q^i given by p_i, $i = 1, 2$. While in mode q^i, a continuous state $v \in \mathbb{R}$ evolves according to a stochastic difference equation $v_{k+1} = f_i(v_k, a_k, d_k, \eta_k)$, defined as follows

$$\begin{aligned} v_{k+1} &= f_1(v_k, a_k, d_k, \eta_k) = 2v_k + a_k + d_k + \eta_k, \\ v_{k+1} &= f_2(v_k, a_k, d_k, \eta_k) = \frac{1}{2}v_k + a_k + d_k + \eta_k, \end{aligned} \tag{2}$$

where a_k and d_k are the inputs of player I and player II, respectively, and η_k is a random variable modeling the effect of noise upon the system dynamics. It is assumed that the players have identical capabilities, with a_k and d_k taking values in $[-1, 1]$. The noise is modeled by a uniform distribution $\eta_k \sim U(-1, +1)$. A sample execution of this system, with initial condition $x_0 = (q_0, v_0) = (1, 1)$, $\mu_k = -\text{sgn}(v_k)$ and $\gamma_k = \frac{v_k a_k}{|2 v_k a_k|}$ is shown in Fig. 1b.

Under the formal modeling framework defined previously, the hybrid state space is $X = \{q^1, q^2\} \times \mathbb{R}$, and the player input spaces are $\mathcal{A} = \mathcal{D} = [-1, 1]$. The discrete transition kernel τ_q is derived from the mode transition diagram Fig. 1a as $\tau_q(q^1|(q^1, v), a, d) = p_1$, $\tau_q(q^2|(q^1, v), a, d) =$



(a) Discrete modes and transitions

(b) An execution of the DTSHG

Fig. 1: Example 1 - A discrete-time stochastic hybrid game (DTSHG) with two modes, linear dynamics and uniformly distributed noise for each mode.

$1-p_1$, $\tau_q(q^1|(q^2,v),a,d) = 1-p_2$, $\tau_q(q^2|(q^2,v),a,d) = p_2$. The continuous transition kernel τ_v can be derived from the continuous state dynamics (2) as $\tau_v(dv'|(q^1,v),a,d) \sim U(2v+a+d-1, 2v+a+d+1)$, $\tau_v(dv'|(q^2,v),a,d) \sim U(\frac{1}{2}v+a+d-1, \frac{1}{2}v+a+d+1)$. With the assumption that the continuous state v is not reset during a discrete mode transition, the reset kernel is given by $\tau_r(dv'|(q,v),a,d,q') = \tau_v(dv'|(q,v),a,d)$. We observe that the stochastic kernels τ_v and τ_r are continuous in a and d, while τ_q is independent of the players' inputs. Thus, the conditions of Assumption 1 are satisfied.

2) Pairwise Aircraft Collision Avoidance: Now we consider a more practical application scenario which arises in the context of air traffic management. Specifically, the scenario involves two aircraft with possibly intersecting nominal trajectories. From the perspective of the first aircraft, the task is to generate a conflict-free trajectory subject to the worst-case behavior of the second aircraft. This problem has been studied with significant detail in [2] within a deterministic setting. Motivated by practical concerns of accounting for wind effects on aircraft trajectories [31], we consider here a stochastic formulation of the problem using a stochastic wind model.

Let $(v^1, v^2, v^3) \in \mathbb{R}^2 \times [0, 2\pi]$ denote, respectively, the x-position, y-position, and heading of aircraft 2 in the reference frame of aircraft 1. By performing an Euler discretization of the kinematics model given in [2] and augmenting the resulting discrete time dynamics with process noise, we obtain the following stochastic model for the relative motion between the two aircraft.

$$v_{k+1} = \begin{bmatrix} v_{k+1}^1 \\ v_{k+1}^2 \\ v_{k+1}^3 \end{bmatrix} = \begin{bmatrix} v_k^1 + \Delta t(-s^1 + s^2\cos(v_k^3) + \omega_k^1 v_k^2) \\ v_k^2 + \Delta t(s^2 \sin(v_k^3) - \omega_k^1 v_k^1) \\ v_k^3 + \Delta t(\omega_k^2 - \omega_k^1) \end{bmatrix} + \begin{bmatrix} \eta_k^1 \\ \eta_k^2 \\ \eta_k^3 \end{bmatrix} \tag{3}$$
$$= f(v_k, \omega_k^1, \omega_k^2) + \eta_k,$$

where Δt is the discretization step, s^i is the speed of aircraft i (assumed to be constant), ω^i is the angular turning rate of aircraft i (assumed to be time-varying), and $\eta_k = (\eta_k^1, \eta_k^2, \eta_k^3)$ is a stochastic noise vector. The random variables (η_k^1, η_k^2) account for the effect of wind. As per previous work on stochastic wind models [32], they are modeled as normally distributed with a position-dependent covariance matrix $\Sigma_1(v_1, v_2) \in \mathbb{R}^{2\times 2}$. On the other hand, the random variable η_k^3 captures the effect of actuator noise on the turning rate of either aircraft. It is assumed to have a Gaussian distribution $\eta_k^3 \sim \mathcal{N}(0, \Sigma_\omega)$. Taken together, η_k has the distribution $\mathcal{N}(0, \Sigma(v))$ where $\Sigma(v) = \text{diag}(\Sigma_1(v^1, v^2), \Sigma_\omega)$ is a block diagonal covariance matrix.

Consider a scenario in which at any given time, each aircraft can be in one of three flight maneuvers: straight, right turn, or left turn, corresponding to the angular turning rates $\omega^i = 0$, $\omega^i = -\omega$, and $\omega^i = \omega$, respectively. Here, $\omega \in \mathbb{R}$ is assumed to be a constant. The discrete states are then given by the flight maneuvers of aircraft 1: $\mathcal{Q} = \{q_S, q_R, q_L\}$. We associate with this the discrete command set $\mathcal{A} = \{\sigma_0^1, \sigma_-^1, \sigma_+^1\}$. On the other hand, the angular turning rate of aircraft 2 becomes the disturbance input set $\mathcal{D} = \{0, -\omega, \omega\}$ for aircraft 2. From this description, we obtain the hybrid state update equations

$$q_{k+1} = \delta(a_k),\ \ v_{k+1} = \tilde{f}(q_{k+1}, v_k, d_k, \eta_k) = f(v_k, \omega^1(q_{k+1}), d_k) + \eta_k, \tag{4}$$

where the discrete transition function δ is specified as follows: $\delta(a) = q_S$, if $a = \sigma_0$; $\delta(a) = q_R$, if $a = \sigma_-$; $\delta(a) = q_L$, if $a = \sigma_+$.

Using (4), we can derive in a straightforward manner the corresponding stochastic kernels of a DTSHG model. Specifically, the discrete state transition kernel τ_q is determined by the discrete transition function δ, the continuous state transition kernel is given by $\tau_v(dv'|(q,v),a,d) \sim \mathcal{N}(f(v,\omega^1(q),d),\Sigma(v))$, and the reset transition kernel is specified as $\tau_r(dv'|(q,v),a,d,q') = \tau_v(dv'|(q',v),a,d)$. It is clear that these stochastic kernels also satisfy Assumption 1.

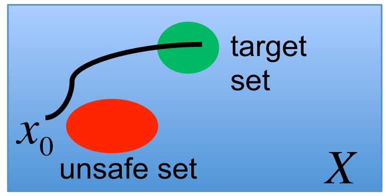


Fig. 2: The reach-avoid problem is concerned with optimizing the probability that the trajectory starting from initial condition $x_0 \in X$ will reach the target set while avoiding the unsafe set.

III. Reach-Avoid Problem Formulation

In the setting of the DTSHG, the reach-avoid problem as considered in [19] becomes a stochastic game in which the objective of player I (the control) is to steer the hybrid system state into a desired target set, while avoiding a set of unsafe states, as shown in Fig. 2. On the other hand, the objective of player II (the adversary) is to either steer the state into the unsafe set or prevent it from reaching the target set.

Suppose that Borel sets $K, K' \in \mathcal{B}(X)$ are given as the desired target set and safe set, respectively, with $K \subseteq K'$. Then the probability that the state trajectory $(x_0, x_1, ..., x_N)$ reaches K while staying inside K' under fixed choices of $\mu \in \mathcal{M}_a$ and $\gamma \in \Gamma_d$ is given by

$$\begin{aligned} r_{x_0}^{\mu,\gamma}(K, K') &:= P_{x_0}^{\mu,\gamma}\left(\bigcup_{j=0}^{N}(K' \setminus K)^j \times K \times X^{N-j}\right) \\ &= \sum_{j=0}^{N} P_{x_0}^{\mu,\gamma}((K' \setminus K)^j \times K \times X^{N-j}), \end{aligned} \tag{5}$$

where the second equality in (5) follows by the fact that the union is disjoint. By (1), this probability can be computed as

$$r_{x_0}^{\mu,\gamma}(K, K') = E_{x_0}^{\mu,\gamma}\left[\mathbf{1}_K(x_0) + \sum_{j=1}^{N}\left(\prod_{i=0}^{j-1}\mathbf{1}_{K'\setminus K}(x_i)\right)\mathbf{1}_K(x_j)\right], \tag{6}$$

where $E_{x_0}^{\mu,\gamma}$ denotes the expectation with respect to the probability measure $P_{x_0}^{\mu,\gamma}$. Now define the worst-case reach-avoid probability under a choice of Markov policy μ as

$$r_{x_0}^{\mu}(K, K') = \inf_{\gamma \in \Gamma_d} r_{x_0}^{\mu,\gamma}(K, K'). \tag{7}$$

Our control objective is to maximize this worst-case probability over the set of Markov policies. The precise problem statement is as follows:

Problem 1. Given a DTSHG $\mathcal{H}$, target set $K \in \mathcal{B}(X)$, and safe set $K' \in \mathcal{B}(X)$ such that $K \subseteq K'$:

(I) Compute the max-min value function $r^*_{x_0}(K, K') := \sup_{\mu \in \mathcal{M}_a} r^\mu_{x_0}(K, K')$, $\forall x_0 \in X$;

(II) Find a max-min policy $\mu^* \in \mathcal{M}_a$, whenever it exists, such that $r^*_{x_0}(K, K') = r^{\mu^*}_{x_0}(K, K')$, $\forall x_0 \in X$.

Similarly as in the single player case, the above framework can be readily modified to account for time-varying [20] and stochastic [21] safe sets. For simplicity in notation, here we focus on static and deterministic target and safe sets. In the following, we describe stochastic game formulations of the probabilistic safety and target hitting problems as special cases of the reach-avoid problem given above.

First, consider the target hitting time problem [19], in which the objective of player I is to drive the system state into a desired target set $B \in \mathcal{B}(X)$ within some finite time horizon $[0, N]$, while the objective of player II is to prevent player I from doing so. Clearly, the problem of computing the optimal target hitting probability under the worst-case player II behavior can be formulated exactly as in Problem 1, by taking $K = B$, $K' = X$.

Second, consider the probabilistic safety problem [18], in which the objective of player I is to keep the system state within a given safe set $S \in \mathcal{B}(X)$ over some finite time horizon $[0, N]$, while the objective of player II is again opposed to that of player I. Similarly as before, the probability that the hybrid state trajectory $(x_0, x_1, ..., x_N)$ remains in S under fixed choices of $\mu \in \mathcal{M}_a$ and $\gamma \in \Gamma_d$ is given by

$$p^{\mu,\gamma}_{x_0}(S) := P^{\mu,\gamma}_{x_0}(S^{N+1}) = E^{\mu,\gamma}_{x_0}\left[\prod_{k=0}^{N} \mathbf{1}_S(x_k)\right].$$

The connection between the safety problem and reach-avoid problem is established by the observation that the hybrid state remains inside a set S for all $k = 0, 1, ..., N$ if and only if it does not reach $X \backslash S$ for any $k = 0, 1, ..., N$. Mathematically speaking, for any $\mu \in \mathcal{M}_a$ and $\gamma \in \Gamma_d$, we have

$$p^{\mu,\gamma}_{x_0}(S) = 1 - r^{\mu,\gamma}_{x_0}(X \backslash S, X). \tag{8}$$

Thus, a safety problem where the objective of player I is to maximize $p^{\mu,\gamma}_{x_0}(S)$ over $\mu \in \mathcal{M}_a$ and the objective of player II is to minimize $p^{\mu,\gamma}_{x_0}(S)$ over $\gamma \in \Gamma_d$ is equivalent to a reach-avoid problem where the objective of player I is to minimize $r_{x_0}^{\mu,\gamma}(X\backslash S, X)$ over $\mu \in \mathcal{M}_a$ and the objective of player II is to maximize $r_{x_0}^{\mu,\gamma}(X\backslash S, X)$ over $\gamma \in \Gamma_d$.

IV. Reach-Avoid Probability Computation and Controller Synthesis

In this section, we provide a detailed proof of our main result from [23], as the solution to Problem 1. In particular, it will be shown that under mild technical assumptions, the max-min probability $r_{x_0}^{*}(K, K')$ can be computed using an appropriate dynamic programming, and that there exists a max-min Markov policy $\mu^* \in \mathcal{M}_a$ for player I which achieves this probability under the worst-case player II strategy. Following the proof, we will discuss some practical implications of the theorem and specialize the results to a stochastic game formulation of the probabilistic safety problem. Finally, a concrete example will be provided to illustrate the procedure for computing $r_{x_0}^{*}(K, K')$, as well as the max-min policy $\mu^* \in \mathcal{M}_a$ for player I and the worst-case strategy $\gamma^* \in \Gamma_d$ for player II.

For our theoretical derivations, we require the following assumptions on the stochastic kernels, as inspired by [24], [25].

Assumption 1.

(a) For each $x = (q, v) \in X$ and $E_1 \in \mathcal{B}(\mathbb{R}^{n(q)})$, the function $(a, d) \to \tau_v(E_1|x, a, d)$ is continuous on $\mathcal{A} \times \mathcal{D}$;

(b) For each $x = (q, v) \in X$ and $q' \in \mathcal{Q}$, the function $(a, d) \to \tau_q(q'|x, a, d)$ is continuous on $\mathcal{A} \times \mathcal{D}$;

(c) For each $x = (q, v) \in X$, $q' \in \mathcal{Q}$, and $E_2 \in \mathcal{B}(\mathbb{R}^{n(q')})$, the function $(a, d) \to \tau_r(E_2|x, a, d, q')$ is continuous on $\mathcal{A} \times \mathcal{D}$.

It should be noted that we only assume continuity of the stochastic kernels in the actions of Player I and Player II, but not necessarily in the system state. Thus, our Borel-measurable model still allows for stochastic hybrid systems where transition probabilities change abruptly with changes in the system state. Furthermore, if the action spaces $\mathcal{A}$ and $\mathcal{D}$ are finite or countable, then the above assumptions are clearly satisfied under the discrete topology on $\mathcal{A}$ and $\mathcal{D}$. Also, if $\tau_v(\cdot|(q, v), a, d)$ has a density function $f_v(v'|(q, v), a, d), v' \in \mathbb{R}^{n(q)}$ for every $q \in \mathcal{Q}$, and $f_v(v'|(q, v), a, d)$ is continuous in a and d, it can be checked that the assumption for τ_v is satisfied. A similar condition can also be formulated for the reset kernel τ_r.

For the statement of the dynamic programming result, we define an operator T which takes as its argument a Borel measurable function $J : X \to [0, 1]$:

$$T(J)(x) = \sup_{a\in\mathcal{A}} \inf_{d\in\mathcal{D}} \mathbf{1}_K(x) + \mathbf{1}_{K'\setminus K}(x) H(x, a, d, J) \quad x \in X, \tag{9}$$

where $H(x, a, d, J) := \int_X J(y)\tau(dy|x, a, d)$.

Our main result (Theorem 1 of [23]) is as follows.

Theorem 1. *Let $\mathcal{H}$ be a DTSHG satisfying the Assumption 1. Let $K, K' \in \mathcal{B}(X)$ be Borel sets such that $K \subseteq K'$. Let the operator T be defined as in (9). Then the composition $T^N = T \circ T \circ \cdots \circ T$ (N times) is well-defined and*

(a) $r^*_{x_0}(K, K') = T^N(\mathbf{1}_K)(x_0), \forall x_0 \in X$;

(b) There exists a player I policy $\mu^* \in \mathcal{M}_a$ and player II strategy $\gamma^* \in \Gamma_d$ satisfying

$$r^{\mu,\gamma^*}_{x_0}(K, K') \le r^*_{x_0}(K, K') \le r^{\mu^*,\gamma}_{x_0}(K, K'), \tag{10}$$

for every $x_0 \in X$, $\mu \in \mathcal{M}_a$, and $\gamma \in \Gamma_d$. In particular, μ^* is a max-min policy for player I.

(c) If $\mu^* \in \mathcal{M}_a$ is a Markov policy which satisfies

$$\mu^*_k(x) \in \arg \sup_{a\in\mathcal{A}} \inf_{d\in\mathcal{D}} H(x, a, d, J_{k+1}),\ x \in K' \setminus K, \tag{11}$$

where $J_k = T^{N-k}(\mathbf{1}_K)$, $k = 0, 1, ..., N$, then μ^* is a max-min policy for Player I. If $\gamma^* = (\gamma^*_0, \gamma^*_1, ..., \gamma^*_{N-1}) \in \Gamma_d$ is a Markov strategy which satisfies

$$\gamma^*_k(x, a) \in \arg \inf_{d\in\mathcal{D}} H(x, a, d, J_{k+1}),\ x \in K' \setminus K,\ a \in \mathcal{A}, \tag{12}$$

$k = 0, 1, ..., N$, then γ^* is a worst-case strategy for Player II.

The proof of this theorem, omitted from [23] due to space limitations, proceeds through a sequence of lemmas and propositions which generalize the dynamic programming algorithms given in [18] and [19] for the single player case. First, it is shown that for fixed $\mu \in \mathcal{M}_a$ and $\gamma \in \Gamma_d$, the reach-avoid probability $r^{\mu,\gamma}_{x_0}(K, K')$ can be computed using a recursive formula as per statement (a). Second, the operator T is shown to preserve measurability properties, and so the sequential composition of T is well-defined. Furthermore, using the continuity assumptions given in Assumption 1, it is shown that there exist Borel measurable functions which achieve the supremum and infimum in (9) at each step of the dynamic programming recursion. Finally, using properties of the recursive formula for $r^{\mu,\gamma}_{x_0}(K, K')$ and the operator T, the function $T^N(\mathbf{1}_K)$

is shown to simultaneously upper bound and lower bound $r^*_{x_0}(K,K')$ and hence is equal to $r^*_{x_0}(K,K')$. In the course of proving this last result, the existence of a max-min policy for Player I and a worst-case strategy for Player II is also established, along with the sufficient conditions for optimality (11) and (12).

As a first step in the proof, motivated by the expressions for $r^{\mu,\gamma}_{x_0}(K,K')$ in (6), we define for fixed $\mu \in \mathcal{M}_a$ and $\gamma \in \Gamma_d$ the utility-to-go functions $V^{\mu,\gamma}_k : X \to [0,1]$, $k = 0,\dots,N$

$$\begin{aligned} V^{\mu,\gamma}_N(x) =& \mathbf{1}_K(x) \ , \\ V^{\mu,\gamma}_k(x) =& \mathbf{1}_K(x) + \mathbf{1}_{K'\setminus K}(x) \int_{X^{N-k}} \sum_{j=k+1}^{N} \prod_{i=k+1}^{j-1} \mathbf{1}_{K'\setminus K}(x_i)\mathbf{1}_K(x_j) \\ & \prod_{j=k+1}^{N-1} \tau^{\mu_j,\gamma_j}(dx_{j+1}|x_j)\tau^{\mu_k,\gamma_k}(dx_{k+1}|x). \end{aligned} \tag{13}$$

From this definition, it is clear that $r^{\mu,\gamma}_{x_0}(K,K') = V^{\mu,\gamma}_0(x_0), \forall x_0 \in X$. The task then becomes formulating a recursive procedure for computing $V^{\mu,\gamma}_k(x)$.

For this purpose, consider a recursion operator $T_{f,g}$, parameterized by Borel measurable functions $f : X \to \mathcal{A}$ and $g : X \times \mathcal{A} \to \mathcal{D}$, and operating on the set of Borel measurable functions from X to $[0,1]$:

$$T_{f,g}(J)(x) = \mathbf{1}_K(x) + \mathbf{1}_{K'\setminus K}(x) H(x, f(x), g(x, f(x)), J), \ x \in X. \tag{14}$$

where H is as given in the definition of T in (9).

The following result shows that the functions $V^{\mu,\gamma}_k$ can be computed using backwards recursion under the operator $T_{f,g}$.

Lemma 1. *Let $\mu \in \mathcal{M}_a$ and $\gamma \in \Gamma_d$. Then for $k = 0, 1, ..., N-1$, the following identity holds*

$$V^{\mu,\gamma}_k = T_{\mu_k,\gamma_k}(V^{\mu,\gamma}_{k+1}), \tag{15}$$

where $V^{\mu,\gamma}_N = \mathbf{1}_K$.

Proof: For the case of $k = N-1$, the choice of the terminal cost $V^{\mu,\gamma}_N$ implies that for any $x \in X$,

$$\begin{aligned} V^{\mu,\gamma}_{N-1}(x) &= \mathbf{1}_K(x) + \mathbf{1}_{K'\setminus K}(x) \int_X \mathbf{1}_K(x_N) \tau^{\mu_{N-1},\gamma_{N-1}}(dx_N|x) \\ &= T_{\mu_{N-1},\gamma_{N-1}}(V^{\mu,\gamma}_N). \end{aligned}$$

For the case of $k < N-1$, we have by the expression for $V_k^{\mu,\gamma}$ given in (13) that for any $x \in X$,

$$\begin{aligned} V_k^{\mu,\gamma}(x) =& \mathbf{1}_K(x) + \mathbf{1}_{K'\setminus K}(x) \int_X \mathbf{1}_K(x_{k+1}) + \mathbf{1}_{K'\setminus K}(x_{k+1}) \\ & \left(\int_{X^{N-k-1}} \sum_{j=k+2}^{N} \prod_{i=k+2}^{j-1} \mathbf{1}_{K'\setminus K}(x_i) \mathbf{1}_K(x_j) \right) \prod_{j=k+1}^{N-1} \tau^{\mu_j,\gamma_j}(dx_{j+1}|x_j) \tau^{\mu_k,\gamma_k}(dx_{k+1}|x) \\ =& \mathbf{1}_K(x) + \mathbf{1}_{K'\setminus K}(x) \int_X V_{k+1}^{\mu,\gamma}(x_{k+1}) \tau^{\mu_k,\gamma_k}(dx_{k+1}|x). \end{aligned}$$

It follows from definition of $T_{f,g}$ that the last expression above is $T_{\mu_k,\gamma_k}(V_{k+1}^{\mu,\gamma})$ and this concludes the proof. ■

Next, we will prove some properties of T that are required for the dynamic programming results. First, we state a special case of Corollary 1 given in [33]. This result allows us to show that the operator T preserves Borel measurability and that it is sufficient to consider Borel measurable selectors.

Lemma 2. *Let X, Y be complete separable metric spaces such that Y is compact, and f be a real-valued Borel measurable function defined on $X \times Y$ such that $f(x,\cdot)$ is lower semicontinuous with respect to the topology on Y. Define $f^* : X \to \mathbb{R} \cup \{\pm\infty\}$ by*

$$f^*(x) = \inf_{y \in Y} f(x,y).$$

(a) The set

$$I = \{x \in X : \text{ for some } y \in Y, f(x,y) = f^*(x)\},$$

is Borel measurable.

(b) For every $\epsilon > 0$, there exists a Borel measurable function $\phi : X \to Y$, satisfying, for all $x \in X$,

$$f(x,\phi(x)) = f^*(x), \text{ if } x \in I,$$

$$f(x,\phi(x)) \leq \begin{cases} f^*(x) + \epsilon, & \text{if } x \notin I,\ f^*(x) > -\infty, \\ -1/\epsilon, & \text{if } x \notin I,\ f^*(x) = -\infty. \end{cases}$$

In order to prove that the supremum and infimum in the expression for T is achieved, we will need the operator H to produce functions continuous in $\mathcal{A}$ and $\mathcal{D}$. For this purpose, we introduce the following technical result from [24] (stated as Fact 3.9).

Lemma 3. *Let f be a bounded real-valued Borel measurable function on a Borel space Y, and t be a Borel measurable transition probability from a Borel space X into Y such that $t(B|\cdot)$ is continuous on X for each $B \in \mathcal{B}(Y)$. Then the function $x \to \int f(y)t(dy|x)$ is continuous on X.*

In the following proposition, we prove that the operator T preserves Borel measurability, and that the infimum and supremum in (9) can be achieved by Borel measurable selectors. Let $\mathcal{F}$ be the set of Borel measurable functions from X to $[0,1]$. For notational convenience, we introduce an operator G which takes a real-valued Borel measurable function on X and produces a real-valued function on $X \times \mathcal{A}$:

$$G(J)(x,a) = \inf_{d\in\mathcal{D}} \mathbf{1}_K(x) + \mathbf{1}_{K'\setminus K}(x)H(x,a,d,J). \tag{16}$$

Proposition 1.

(a) $\forall J \in \mathcal{F}$, $T(J) \in \mathcal{F}$.

(b) For any $J \in \mathcal{F}$, there exists a Borel measurable function $g^* : X \times \mathcal{A} \to \mathcal{D}$ such that, for all $(x,a) \in X \times \mathcal{A}$,

$$g^*(x,a) \in \arg\inf_{d\in\mathcal{D}} \mathbf{1}_K(x) + \mathbf{1}_{K'\setminus K}(x)H(x,a,d,J). \tag{17}$$

(c) For any $J \in \mathcal{F}$, there exists a Borel measurable function $f^* : X \to \mathcal{A}$, such that for all $x \in X$,

$$f^*(x) \in \arg\sup_{a\in\mathcal{A}} \inf_{d\in\mathcal{D}} \mathbf{1}_K(x) + \mathbf{1}_{K'\setminus K}(x)H(x,a,d,J). \tag{18}$$

Proof: For any $J \in \mathcal{F}$, define a function $F_J : X \times \mathcal{A} \times \mathcal{D} \to \mathbb{R}$ as

$$F_J(x,a,d) = H(x,a,d,J).$$

From the definition of H, the range of F_J lies in $[0,1]$. By the Borel measurability of J and Q, Proposition 7.29 of [30] implies that F_J is Borel measurable. Furthermore, for each $x \in X$, Lemma 3 implies that $F_J(x,a,d)$ is continuous in a and d. Now consider $\tilde{F}_J : X \times \mathcal{A} \times \mathcal{D} \to \mathbb{R}$,

$$\tilde{F}_J(x,a,d) = \mathbf{1}_K(x) + \mathbf{1}_{K'\setminus K}(x)F_J(x,a,d).$$

Clearly, $0 \leq \tilde{F}_J \leq 1$. Furthermore, given that Borel measurability is preserved under summation and multiplication (see for example Proposition 2.6 of [34]), $\tilde{F}_J$ is also Borel measurable. Finally, it is clear that $\tilde{F}_J(x,a,d)$ is continuous in a and d for each $x \in X$. We observe that

$$G(J)(x,a) = \inf_{d\in\mathcal{D}} \tilde{F}_J(x,a,d). \tag{19}$$

Since the range of $\tilde{F}_J$ lies in $[0,1]$, the range of $G(J)$ also lies in $[0,1]$. By assumption, $\mathcal{A}$ and $\mathcal{D}$ are Borel spaces and hence metrizable. Thus, $\mathcal{A}$ can be endowed with a metric d_1 consistent with the topology on $\mathcal{A}$, while $\mathcal{D}$ can be endowed with a metric d_2 consistent with the topology on $\mathcal{D}$. Furthermore, as shown in [35], the hybrid state space X can be endowed with a metric equivalent to the standard Euclidean metric when restricted to each continuous domain $\mathbb{R}^{n(q)}$, $q \in \mathcal{Q}$. Under the assumptions on the DTSHG model, the spaces X, $\mathcal{A}$, and $\mathcal{D}$ are also complete and separable.

Now for each $(x,a) \in X \times \mathcal{A}$, we have by the previous derivations that $\tilde{F}_J(x,a,\cdot)$ is continuous on $\mathcal{D}$. By the compactness of $\mathcal{D}$, the infimum in (19) is achieved for each fixed (x,a) (see for example Theorem 4.16 in [36]). Thus, applying Lemma 2, we have that there exists a Borel measurable function $g^*: X \times \mathcal{A} \to \mathcal{D}$ for which (17) holds. Since the composition of Borel measurable functions remains Borel measurable, $G(J)$ is a Borel measurable function. From the fact that the infimum is achieved and the fact that $\tilde{F}_J$ is continuous on $\mathcal{A}$, it is not difficult to see that $G(J)$ is also continuous on $\mathcal{A}$.

For the outer supremum, we observe that

$$T(J)(x) = -\inf_{a \in \mathcal{A}} -G(J)(x,a), \; x \in X. \tag{20}$$

By the compactness of $\mathcal{A}$, the infimum in (20) is achieved for each $x \in X$. Thus, a repeated application of Lemma 2 shows that there exists a Borel measurable function $f^*: X \to \mathcal{A}$ such that $-T(J)(x) = -G(J)(x, f^*(x)), \forall x \in X$. By the composition of Borel measurable functions, this implies that $T(J)$ is Borel measurable. Finally, it is clear that the range of $T(J)$ lies in $[0,1]$, and so $T(J) \in \mathcal{F}$. ■

For the proofs of the dynamic programming results, we make use of the fact that the operator $T_{f,g}$ satisfies a monotonicity property: for any Borel measurable functions J, J' from X to $[0,1]$ such that $J(x) \leq J'(x), \forall x \in X$, $T_{f,g}(J)(x) \leq T_{f,g}(J')(x), \forall x \in X$. This property can be checked in a straightforward manner using the definition of H and the properties of integrals.

We are now ready for our first dynamic programming result, which provides a lower bound for $r^*_{x_0}(K,K')$, $x_0 \in X$.

Proposition 2.

(a) $\forall x_0 \in X$, $T^N(\mathbf{1}_K)(x_0) \leq r^*_{x_0}(K,K')$.

(b) There exists $\mu^* \in \mathcal{M}_a$ *such that, for any* $\gamma \in \Gamma_d$, $T^N(\mathbf{1}_K)(x_0) \leq r^{\mu^*,\gamma}_{x_0}(K,K'), \forall x_0 \in X$.

Proof: For notational convenience, we define $J_{N-k} := T^k(\mathbf{1}_K)$, $k = 0, 1, ..., N$. First, we prove the following claim by induction on k: there exists $\mu^*_{N-k\to N} = (\mu^*_{N-k}, \mu^*_{N-k+1}, ..., \mu^*_{N-1}) \in \mathcal{M}_a$ such that, for any $\gamma = (\gamma_{N-k}, \gamma_{N-k+1}, ..., \gamma_{N-1}) \in \Gamma_d$, $J_{N-k}(x) \le V^{\mu^*_{N-k\to N},\gamma}_{N-k}(x)$, $\forall x \in X$.

Let $\gamma = (\gamma_0, \gamma_1, ..., \gamma_{N-1}) \in \Gamma$ be arbitrary. The case of $k = 0$ is trivial. For the inductive step, assume that this holds for $k = h$. By the induction hypothesis, there exists a policy $\mu^*_{N-h\to N} = (\mu^*_{N-h}, \mu^*_{N-h+1}, ..., \mu^*_{N-1}) \in \mathcal{M}_a$ such that, for any $\gamma \in \Gamma_d$, $J_{N-h}(x) \le V^{\mu^*_{N-h\to N},\gamma}_{N-h}(x), \forall x \in X$. Furthermore, by Proposition 1(c), there exists a Borel measurable function $f^* : X \to \mathcal{A}$ such that $G(J_{N-h})(x, f^*(x)) = T(J_{N-h})(x), \forall x \in X$. Choose a policy $\mu^*_{N-h-1\to N} = (f^*, \mu^*_{N-h}, \mu^*_{N-h+1}, ..., \mu^*_{N-1})$. Then by the monotonicity of the operator $T_{f,g}$ and Lemma 1, we have for each $x \in X$:

$$\begin{aligned}
V^{\mu^*_{N-h-1\to N},\gamma}_{N-h-1}(x) =& T_{f^*,\gamma_{N-h-1}}(V^{\mu^*_{N-h\to N},\gamma}_{N-h})(x) \\
\ge& T_{f^*,\gamma_{N-h-1}}(J_{N-h})(x) \\
=& \mathbf{1}_K(x) + \mathbf{1}_{K'\setminus K}(x) H(x, f^*(x), \gamma_{N-h-1}(x, f^*(x)), J_{N-h}) \\
\ge& \inf_{d\in\mathcal{D}} \mathbf{1}_K(x) + \mathbf{1}_{K'\setminus K}(x) H(x, f^*(x), d, J_{N-h}) \\
=& G(J_{N-h})(x, f^*(x)) \\
=& T(J_{N-h})(x) = J_{N-h-1}(x),
\end{aligned}$$

which concludes the proof of the claim.

This result implies that there exists $\mu^*_{0\to N} = (\mu^*_0, \mu^*_1, ..., \mu^*_{N-1}) \in \mathcal{M}_a$ such that, for any $\gamma = (\gamma_0, \gamma_1, ..., \gamma_{N-1}) \in \Gamma_d$, $T^N(\mathbf{1}_K)(x_0) = J_0(x_0) \le V^{\mu^*_{0\to N},\gamma}_0(x_0) = r^{\mu^*_{0\to N},\gamma}_{x_0}(K, K')$, $\forall x_0 \in X$. This shows that $\mu^*_{0\to N}$ is the Markov policy satisfying statement (b). Furthermore, since γ is arbitrary, $T^N(\mathbf{1}_K)(x_0) \le \inf_{\gamma\in\Gamma_d} r^{\mu^*_{0\to N},\gamma}_{x_0}(K, K')$, $\forall x_0 \in X$. Thus, $T^N(\mathbf{1}_K)(x_0) \le r^*_{x_0}(K, K')$, $\forall x_0 \in X$, proving statement (a). ■

We now turn to our second dynamic programming result, which provides an upper bound for $r^*_{x_0}(K, K')$, $x_0 \in X$.

Proposition 3.

(a) $\forall x_0 \in X$, $r^*_{x_0}(K, K') \le T^N(\mathbf{1}_K)(x_0)$.

(b) There exists $\gamma^* \in \Gamma_d$ *such that, for any* $\mu \in \mathcal{M}_a$, $r^{\mu,\gamma^*}_{x_0}(K, K') \le T^N(\mathbf{1}_K)(x_0), \forall x_0 \in X$.

Proof: As in the proof of Proposition 2, we define $J_{N-k} := T^k(\mathbf{1}_K)$, $k = 0, 1, ..., N$. First, we prove the following claim by induction on k: there exists $\gamma^*_{N-k\to N} = (\gamma^*_{N-k}, \gamma^*_{N-k+1}, ..., \gamma^*_{N-1}) \in \Gamma_d$ such that, for any $\mu = (\mu_{N-k}, \mu_{N-k+1}, ..., \mu_{N-1}) \in \mathcal{M}_a$, $V^{\mu,\gamma^*_{N-k\to N}}_{N-k}(x) \le J_{N-k}(x)$, $\forall x \in X$.

Let $\mu = (\mu_0, \mu_1, ..., \mu_{N-1}) \in \mathcal{M}_a$ be arbitrary. The case of $k = 0$ is trivial. For the inductive step, assume that this holds for $k = h$. By the induction hypothesis, there exists a strategy $\gamma^*_{N-h\to N} = (\gamma^*_{N-h}, \gamma^*_{N-h+1}, ..., \gamma^*_{N-1}) \in \Gamma_d$ such that, for any $\mu \in \mathcal{M}_a$, $V^{\mu,\gamma^*_{N-h\to N}}_{N-h}(x) \le J_{N-h}(x)$, $\forall x \in X$. Furthermore, by Proposition 1(b), there exists a Borel measurable function $g^* : X \times \mathcal{A} \to \mathcal{D}$ such that for all $(x, a) \in X \times \mathcal{A}$ the following holds

$$\mathbf{1}_K(x) + \mathbf{1}_{K'\setminus K}(x)H(x, a, g^*(x, a), J_{N-h}) = G(J_{N-h})(x, a).$$

Choose a Markov strategy

$$\gamma^*_{N-h-1\to N} = (g^*, \gamma^*_{N-h}, \gamma^*_{N-h+1}, ..., \gamma^*_{N-1}).$$

Then by the monotonicity of the operator $T_{f,g}$ and Lemma 1, we have for each $x \in X$:

$$\begin{aligned}
&V^{\mu,\gamma^*_{N-h-1\to N}}_{N-h-1}(x) = T_{\mu_{N-h-1},g^*}(V^{\mu,\gamma^*_{N-h\to N}}_{N-h})(x) \\
&\le T_{\mu_{N-h-1},g^*}(J_{N-h})(x) \\
&= \mathbf{1}_K(x) + \mathbf{1}_{K'\setminus K}(x)H(x, \mu_{N-h-1}(x), g^*(x, \mu_{N-h-1}(x)), J_{N-h}) \\
&= G(J_{N-h})(x, \mu_{N-h-1}(x)) \\
&\le \sup_{a\in\mathcal{A}} G(J_{N-h})(x, a) \\
&= T(J_{N-h})(x) = J_{N-h-1}(x),
\end{aligned}$$

which concludes the proof of the claim.

This result implies that there exists $\gamma^*_{0\to N} = (\gamma^*_0, \gamma^*_1, ..., \gamma^*_{N-1}) \in \Gamma_d$ such that, for any $\mu = (\mu_0, \mu_1, ..., \mu_{N-1}) \in \mathcal{M}_a$, $r^{\mu,\gamma^*_{0\to N}}_{x_0}(K, K') = V^{\mu,\gamma^*_{0\to N}}_0(x_0) \le J_0(x_0) = T^N(\mathbf{1}_K)(x_0)$, $\forall x_0 \in X$. This shows that $\gamma^*_{0\to N}$ is the Markov strategy satisfying statement (b) and that $r^\mu_{x_0}(K, K') = \inf_{\gamma\in\Gamma_d} r^{\mu,\gamma}_{x_0}(K, K') \le T^N(\mathbf{1}_K)(x_0)$, for any $\mu \in \mathcal{M}_a$ and $x_0 \in X$. Since μ is arbitrary, $r^*_{x_0}(K, K') \le T^N(\mathbf{1}_K)(x_0)$, $\forall x_0 \in X$, proving statement (a). ■

Combining the results of Proposition 2 and 3, we can now prove Theorem 1.

Proof: Statement (a) of Theorem 1 follows directly from the inequalities in Proposition 2(a) and Proposition 3(a).

By Proposition 2(b) and statement (a), there exists a Markov policy $\mu^* \in \mathcal{M}_a$ such that, for any $\gamma \in \Gamma_d$, $r^*_{x_0}(K, K') \leq r^{\mu^*,\gamma}_{x_0}(K, K')$, $\forall x_0 \in X$. This implies that $r^*_{x_0}(K, K') \leq r^{\mu^*}_{x_0}(K, K')$, $\forall x_0 \in X$. On the other hand, the reverse inequality always holds: $r^{\mu^*}_{x_0}(K, K') \leq r^*_{x_0}(K, K')$, $\forall x_0 \in X$. This shows that μ^* is a max-min policy. Similarly, by Proposition 3(b) and statement (a), there exists a Markov strategy $\gamma^* \in \Gamma_d$ such that, for any $\mu \in \mathcal{M}_a$, $r^{\mu,\gamma^*}_{x_0}(K, K') \leq r^*_{x_0}(K, K'), \forall x_0 \in X$. Thus, we have statement (b).

Finally, statement (c) follows directly from the proof of Proposition 2 and Proposition 3. ■

A. *Implications of the Main Theorem*

1) Robust optimal policy: By statement (b) of Theorem 1, if the control were to choose μ^* and the adversary were to deviate from the worst-case γ^*, then the reach-avoid probability will be at least $r^*_{x_0}(K, K')$. On the other hand, if the control were to deviate from the max-min policy and the adversary were to choose the worst-case Markov strategy, then the reach-avoid probability will be at most $r^*_{x_0}(K, K')$. Thus, μ^* can be interpreted as a robust control policy which optimizes the worst-case probability for achieving the reach-avoid objective.

2) Controller synthesis: Equations (11) and (12) provides us with sufficient conditions for optimality of the players' policies and strategies. In particular, this can be used to synthesize a maxmin control policy for player I from the value functions computed through the dynamic programming recursion. To illustrate, suppose that the input ranges $\mathcal{A}$ and $\mathcal{D}$ along with the state space X has been appropriately discretized, for example according to the method suggested in [37], then for each system state $x \in K' \setminus K$ at the k-th iteration of the dynamic programming algorithm, we can compute and store an optimal control input

$$a^* \in \arg\sup_{a \in \mathcal{A}} \inf_{d \in \mathcal{D}} H(x, a, d, J_{k+1}).$$

This provides us with a discretized representation of the one-step maxmin control policy μ^*_{N-k} on a grid of the continuous state space within each mode. Storing these values as lookup tables then allows us to select control inputs in an optimal fashion as state measurements are received.

3) Probabilistic reach-avoid set: Consider the case in which it is required from the system designer perspective to have a reach-avoid probability greater than some threshold $(1 - \epsilon)$, for $\epsilon \in [0, 1)$. The set of initial conditions X_ϵ for which this specification is feasible, under the

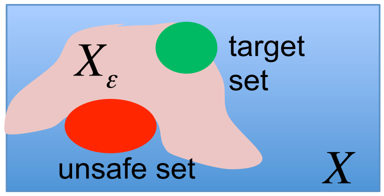


Fig. 3: X_ϵ is the set of states which can reach the target set while avoiding the unsafe set with probability of at least $1-\epsilon$, using the optimal policy.

worst-case adversary behavior, can be derived from the max-min reach-avoid probability as:

$$X_\epsilon = \{x_0 \in X : r^*_{x_0}(K, K') \geq (1-\epsilon)\}.$$

In other words, X_ϵ is the $(1-\epsilon)$-sublevel set of the reach-avoid probability map $r^*_{x_0}(K, K')$, $x_0 \in X$. A conceptual illustration of such a set is shown in Fig. 3.

B. Specialization to Stochastic Game Formulation of Safety Problem

As discussed in section III, the solution to the probabilistic safety problem can be obtained from a complementary reach-avoid problem. In particular, consider a reach-avoid problem with the value function

$$\bar{r}^*_{x_0}(X \setminus S, X) = \inf_{\mu \in \mathcal{M}_a} \sup_{\gamma \in \Gamma_d} r^{\mu,\gamma}_{x_0}(X \backslash S, X), \;\; x_0 \in X.$$

Then the max-min probability of safety is given by

$$p^*_{x_0}(S) = \sup_{\mu \in \mathcal{M}_a} \inf_{\gamma \in \Gamma_d} p^{\mu,\gamma}_{x_0}(S) = 1 - \bar{r}^*_{x_0}(X \setminus S, X), \;\; x_0 \in X. \tag{21}$$

By minor modifications of the proof for Theorem 1, it is not difficult to see that $\bar{r}^*_{x_0}(X \setminus S, X)$ can be computed by the dynamic programming recursion

$$\bar{r}^*_{x_0}(X \setminus S, X) = T^N_S(\mathbf{1}_{X \backslash S})(x_0), \;\; x_0 \in X,$$

where the operator T_S is defined as

$$T_S(J)(x) = \inf_{a \in \mathcal{A}} \sup_{d \in \mathcal{D}} \mathbf{1}_{X \backslash S}(x) + \mathbf{1}_S(x) H(x, a, d, J), \;\; x \in X. \tag{22}$$

The corresponding max-min probability of safety can be then obtained through (21).

For completeness, we note that there exists an equivalent dynamic programming recursion to compute the safety probability, similar to the one given in [18] for the single player case. Specifically, consider an operator $\tilde{T}_S$ defined as

$$\tilde{T}_S(J)(x) = \sup_{a \in \mathcal{A}} \inf_{d \in \mathcal{D}} \mathbf{1}_S(x) H(x, a, d, J), \ x \in X. \tag{23}$$

The relation between $\tilde{T}_S$ and T_S is established through the following lemma.

Lemma 4. *For every* $x \in X$ *and* $k = 0, 1, ..., N$,

$$\tilde{T}_S^k(\mathbf{1}_S)(x) = 1 - T_S^k(\mathbf{1}_{X \setminus S})(x)$$

Proof: We prove this result by induction on k. The case of $k = 0$ is established by the fact that $\mathbf{1}_S = 1 - \mathbf{1}_{X\setminus S}$. Now suppose the identity holds for $k = h$, then we have for every $x \in X$,

$$\begin{aligned}
&\tilde{T}_S^{h+1}(\mathbf{1}_S)(x) = \tilde{T}_S(\tilde{T}_S^h(\mathbf{1}_S))(x) = \tilde{T}_S(1 - T_S^h(\mathbf{1}_{X\setminus S}))(x) \\
&= \sup_{a \in \mathcal{A}} \inf_{d \in \mathcal{D}} \mathbf{1}_S(x) H(x, a, d, 1 - T_S^h(\mathbf{1}_{X\setminus S})) \\
&= \sup_{a \in \mathcal{A}} \inf_{d \in \mathcal{D}} \mathbf{1}_S(x)(1 - H(x, a, d, T_S^h(\mathbf{1}_{X\setminus S}))) \\
&= \mathbf{1}_S(x) + \sup_{a \in \mathcal{A}} \inf_{d \in \mathcal{D}} -\mathbf{1}_S(x) H(x, a, d, T_S^h(\mathbf{1}_{X\setminus S})).
\end{aligned}$$

It then follows that for every $x \in X$

$$\begin{aligned}
&1 - \tilde{T}_S^{h+1}(\mathbf{1}_S)(x) \\
&= 1 - \mathbf{1}_S(x) - \sup_{a \in \mathcal{A}} \inf_{d \in \mathcal{D}} -\mathbf{1}_A(x) H(x, a, d, T_S^h(\mathbf{1}_{X\setminus S})) \\
&= \mathbf{1}_{X\setminus S}(x) + \inf_{a \in \mathcal{A}} \sup_{d \in \mathcal{D}} \mathbf{1}_S(x) H(x, a, d, T_S^h(\mathbf{1}_{X\setminus S})) \\
&= T_S(T_S^h(\mathbf{1}_{X\setminus S}))(x) = T_S^{h+1}(\mathbf{1}_{X\setminus S})(x),
\end{aligned}$$

which completes the proof. ■

Thus, an equivalent dynamic programming recursion for computing the max-min safety probability is given by

$$p_{x_0}^*(S) = \tilde{T}_S^N(\mathbf{1}_S)(x_0), \ x_0 \in X. \tag{24}$$

Using either the operator T_S or the operator $\tilde{T}_S$, we can also derive sufficient conditions of optimality for player I and II, similar to those given in (11) and (12).

C. Analytic Reach-avoid Problem Example

In order to illustrate the procedure for computing the reach-avoid probability and the optimal player I policy and player II strategy, we descibe here a simple reach-avoid problem for which an analytic solution can be obtained. Specifically, consider the system dynamics given in Example 1 of Section II, and a regulation problem where the objective of player I is to drive the continuous state into a neighborhood of the origin, while staying within some safe operating region. In this case, the target set and safe set are chosen to be $K = \{q^1, q^2\} \times [-\frac{1}{4}, \frac{1}{4}]$ and $K' = \{q^1, q^2\} \times [-2, 2]$, respectively, and the time horizon is chosen to be $N = 1$.

First, we characterize the operator H for this particular example. Suppose we are given a function $J : X \to \mathbb{R}$, then the value of $H(x, a, d, J)$ for a hybrid state $x = (q^1, v)$ can be computed as follows:

$$\begin{aligned} H((q^1, v), a, d, J) = &\int_X J(x')\tau(dx'|(q^1, v), a, d) \\ =&\tau_q(q^1|(q^1, v), a, d) \int_{\mathbb{R}} J(q^1, v')\tau_v(dv'|(q^1, v), a, d) + \\ &\tau_q(q^2|(q^1, v), a, d) \int_{\mathbb{R}} J(q^2, v')\tau_r(dv'|(q^1, v), a, d, q^2) \\ =&p_1 \int_{-1}^{1} J(q^1, 2v + a + d + \eta)d\eta + (1 - p_1) \int_{-1}^{1} J(q^2, 2v + a + d + \eta)d\eta. \end{aligned} \tag{25}$$

Similarly, we can derive $H(x, a, d, J)$ for $x = (q^2, v)$. Given the form of the target set K, the dynamic programming recursion is initialized by the function

$$\mathbf{1}_K(q, v) = \begin{cases} 1, & |v| \le \frac{1}{4},\ q = q^1, q^2 \\ 0, & |v| > \frac{1}{4},\ q = q^1, q^2 \end{cases}$$

By Theorem 1, the reach-avoid probability $r^*_{x_0}(K, K')$ for an initial condition $x_0 = (q_0, v_0)$ can be computed as

$$T(\mathbf{1}_K)(q_0, v_0) = \begin{cases} 1, & |v_0| \le \frac{1}{4},\ q_0 = q^1, q^2 \\ 0, & |v_0| > 2,\ q_0 = q^1, q^2 \\ \sup_{a \in \mathcal{A}} \inf_{d \in \mathcal{D}} H((q_0, v_0), a, d, \mathbf{1}_K), & \frac{1}{4} < |v_0| \le 2,\ q_0 = q^1, q^2 \end{cases} \tag{26}$$

It can be observed that the dynamic programming step only needs to be carried out on the set $K' \setminus K = \{q^1, q^2\} \times [-2, -\frac{1}{4}) \cup (\frac{1}{4},\ 2]$. From equation (25), it can be verified that for $q_0 = q^1$,

$$H((q^1, v_0), a, d, \mathbf{1}_K) = \begin{cases} \frac{1}{4}, & 0 \leq |2v_0 + a + d| \leq \frac{3}{4} \\ \frac{5}{8} - \frac{1}{2}|2v_0 + a + d|, & \frac{3}{4} < |2v_0 + a + d| \leq \frac{5}{4} \\ 0, & |2v_0 + a + d| > \frac{5}{4}. \end{cases} \tag{27}$$

Combining equations (26) and (27), the max-min reach-avoid probability for an initial condition $x_0 = (q^1, v_0)$ can be derived as

$$r^*_{x_0}(K, K') = T(\mathbf{1}_K)(q^1, v_0) = \begin{cases} 1, & |v_0| \leq \frac{1}{4} \\ \frac{1}{8}, & \frac{1}{4} < |v_0| \leq \frac{1}{2} \\ \frac{5}{8} - |v_0|, & \frac{1}{2} < |v_0| \leq \frac{5}{8} \\ 0, & |v_0| > \frac{5}{8}. \end{cases}$$

In the process of performing the dynamic programming step in (26), we also obtain a max-min player I policy μ_0^* and a worst-case player II strategy γ_0^* in mode q^1 satisfying the sufficient conditions for optimality in (11) and (12):

$$\mu_0^*(q^1, v_0) = \begin{cases} 1, & |v_0| > \frac{1}{2} \\ -2v_0, & |v_0| \leq \frac{1}{2}, \end{cases} \qquad \gamma_0^*((q^1, v_0), a) = \begin{cases} -1, & 2v_0 + a < 0 \\ 1, & 2v_0 + a \geq 0. \end{cases}$$

Using a similar procedure, we can compute the max-min reach-avoid probability for an initial condition $x_0 = (q^2, v_0)$ as

$$r^*_{x_0}(K, K') = T(\mathbf{1}_K)(q^2, v_0) = \begin{cases} 1, & |v_0| \leq \frac{1}{4} \\ \frac{1}{8}, & \frac{1}{4} \leq |v_0| \leq 2 \\ 0, & |v_0| > 2. \end{cases}$$

Furthermore, a max-min player I policy and a worst-case player II strategy satisfying the sufficient conditions for optimality in mode q^2 can be derived as follows:

$$\mu_0^*(q^2, v_0) = \begin{cases} 1, & |v_0| > 2 \\ -\frac{1}{2}v_0, & |v_0| \leq 2, \end{cases} \qquad \gamma_0^*((q^2, v_0), a) = \begin{cases} -1, & \frac{1}{2}v_0 + a < 0 \\ 1, & \frac{1}{2}v_0 + a \geq 0. \end{cases}$$

As we consider more complicated scenarios which arise in practical applications, there may not be a closed-form expression for the operator T. In such problems, one would then have to perform the dynamic programming recursion of Theorem 1 numerically through a discretization of the continuous state space and player input spaces. In [38], it is shown that, for a single player probabilistic safety problem, piecewise constant approximations of the value function on a grid of the continuous state space converge uniformly to the optimal value function at a rate that is linear in the grid size parameter. We anticipate that a similar result can be shown for the recursion $T^N(\mathbf{1}_K)$ and that approximations of the optimal strategies can be constructed using equations (11) and (12). However, it can be observed that the computational cost of such an approach scales exponentially with the dimensions of the continuous state space and player input spaces, which currently limits the application of our approach to problems with continuous state dimensions of $n \leq 4$. The reduction in computation time is a topic of ongoing research [39].

V. Applications

In this section, we discuss applications of the DTSHG framework to two problems of practical interest. The first is that of pairwise aircraft collision avoidance as described in Section II, formulated as a probabilistic safety problem with two competing players. This is followed by a probabilistic reach-avoid problem in the context of robust motion planning. In particular, we consider a target tracking application where the control objective is to drive a quadrotor helicopter to a hover position over a moving ground vehicle while satisfying certain velocity constraints.

A. *Pairwise Aircraft Collision Avoidance*

In the pairwise aircraft collision avoidance scenario described in Section II, the objective of the controlled aircraft (aircraft 1) is to minimize the probability of collision, under the worst-case assumption that the uncontrolled aircraft (aircraft 2) would try to maximize the probability of collision. This then becomes a probabilistic safety problem involving two players (aircraft 1 and 2), with the safe set being the set of all relative aircraft states outside the collision zone. A more precise formulation of the problem is given below.

First, recall that the relative motion of the two aircraft is given by the equation

$$v_{k+1} = f(v_k, \omega_k^1, \omega_k^2) + \eta_k,$$

where $f(v_k, \omega_k^1, \omega_k^2)$ is the deterministic component of the model, specified as in (3), and η_k is a stochastic noise vector modeling wind effects. In order to develop a realistic model of the wind, we take into account correlation of aircraft motion due to presence of stochastic wind. Our model is based on [32] in which the time integral of the stochastic wind component is modeled, in continuous time, as a time-dependent random field over the $2D$ space. At each planar position $(v_1, v_2) \in \mathbb{R}^2$, the stochastic wind component has the distribution $\sigma dB(v_1, v_2, t)$ in which B is a position-dependent Brownian motion and σ is a positive constant. It is then shown that the wind in relative coordinates has the distribution

$$\eta^1(t) = \sigma\sqrt{2(1 - h(\|(v_1, v_2)\|))}W^1(t)$$
$$\eta^2(t) = \sigma\sqrt{2(1 - h(\|(v_1, v_2)\|))}W^2(t)$$

where $h : \mathbb{R} \to \mathbb{R}$ is a continuous decreasing function with $h(0) = 1$ and $\lim_{c\to\infty} h(c) = 0$ and $W(t) = (W^1(t), W^2(t))$ is a standard Brownian motion. The function h is referred to as the spatial correlation function and is chosen to be $h(c) = \exp(-\beta c)$, where β is a positive constant (see [32] for details). As such, the wind model in discrete time has the distribution

$$(\eta_k^1, \eta_k^2) \sim \mathcal{N}\big(0, 2(\Delta t\sigma)^2(1 - h(\|(v_1, v_2)\|))I_2\big)$$
$$\eta_k^3 \sim \mathcal{N}(0, (\Delta t\sigma_\omega)^2)$$

where Δt is the sampling time. For the aircraft conflict resolution scenario, a collision is defined as the event where aircraft 2 enters a protected zone around aircraft 1. The protected zone in two dimensions is a disk with radius R_c centered on aircraft 1. Hence, the safe set can be defined as

$$S = \{(v^1, v^2) \in \mathbb{R}^2 \ s.t. \ \|(v^1, v^2)\|_2 \geq R_c\}.$$

The probabilistic safety problem is then to compute the max-min probability of safety $p^*_{x_0}(S)$ for aircraft 1, as well as an optimal control policy $\mu^* \in \mathcal{M}_a$ which achieves this probability under the worst-case aircraft 2 behavior.

As discussed in Section IV-B, the solution to this problem can be obtained from a complementary reach-avoid problem where the objective of aircraft 1 is to minimize the probability of entering the collision set $X \setminus S$, and the objective of player II is to maximize this probability. This latter problem has the min-max value function $\bar{r}^*_{x_0}(X \setminus S, X)$.

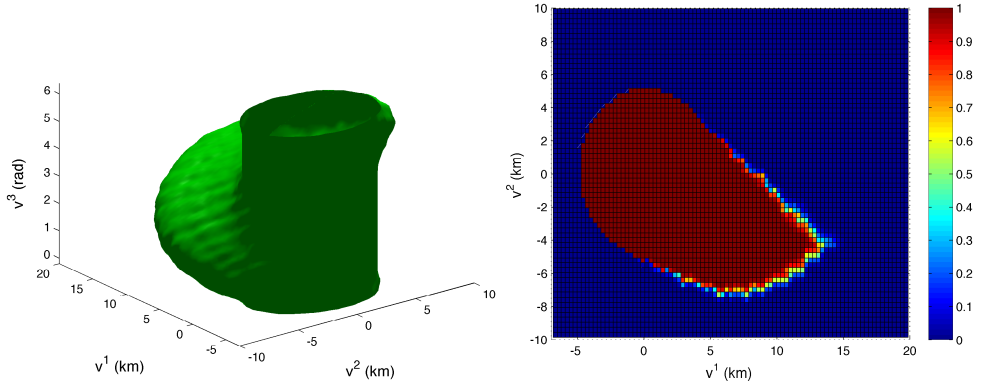
For the numerical results, the parameters of the problem are chosen as follows: the sampling time is set to $\Delta t = 0.1$ minute, the time horizon to 2.5 minutes, $R_c = 5$ km, the aircraft speed to $s = 5$ km per minute and the angular turning rate to $\omega = 1$ radian per minute. The noise covariance is set to $\sigma = 0.73$ and $\sigma_w = 0.26$. The constant β in function h is chosen as 0.1. Computation is performed over a subset of $\mathbb{R}^3$ given by $[-7, 20] \times [-10, 10] \times [0, 2\pi]$, on a grid size of $90 \times 67 \times 65$.

The set of initial conditions $x_0 \in X$ for which the worst-case probability of collision is at most 2% (namely, where $\bar{r}^*_{x_0}(X \setminus S, X) \geq 0.02$) is shown in Fig. 4a. The worst-case probability of collision $\bar{r}^*_{x_0}(X \setminus S, X)$ for the set of initial conditions with initial relative heading of 2.05 radians is shown in Fig. 4b. The interpretation of this probability map is as follows. Consider an initial condition of $(10.55\text{km}, -6.85\text{km}, 2.05\text{rad})$, then from the value function we obtain $\bar{r}^*_{(q,v_0)}(X \setminus S, X) \approx 0.01$, $\forall q \in \mathcal{Q}$. This means that if aircraft 1 selects flight manuevers according to the max-min policy μ^* and aircraft 2 selects maneuvers according to the worst-case strategy γ^*, then the probability of collision within a 2.5 minute time horizon is approximately 1%. Furthermore, if aircraft 2 were to deviate from the worst-case strategy γ^*, while aircraft 1 selected maneuvers according to μ^*, then by the results in Section IV, the probability of collision would remain at most 1%. On the other hand, if aircraft 1 were to deviate from the max-min policy μ^*, while aircraft 2 selected headings according to γ^*, then the probability of collision may be greater than 1%. Thus, aircraft 1 has an incentive for choosing the max-min policy as a robust control policy to counter the worst-case behavior by aircraft 2.

B. Target Tracking

Now consider a motion planning application where the task specification is to drive an autonomous quadrotor helicopter into a neighborhood of planar positions over a moving ground vehicle, without exceeding certain velocity limits. This problem was previously considered in [40] within a continuous time robust control framework, with experimental tests carried out on the Stanford Testbed of Autonomous Rotorcraft for Multi-Agent Control (STARMAC), an unmanned aerial vehicle platform consisting of six quadrotor helicopters each equipped with onboard computation, sensing, and control capabilities [41].

In this work, we consider a stochastic formulation of the problem where the uncertainties within the system are characterized through a mixture of deterministic bounds and stochastic



(a) Set of initial conditions with at most 2% worst-case probability of collision.

(b) Slice of $\bar{r}^*_{x_0}(X \setminus S, X)$ at $x_0^3 = 2.05$ radians

Fig. 4: Worst-case probability of collision for pairwise aircraft conflict resolution example.

noise. The motivation for this is that in an aerial robotics platform such as STARMAC, the effects of higher order dynamics and actuator noise can often be difficult to characterize through a deterministic model [42]. Under a robust control approach, one tends to put conservative bounds on the effects of these disturbances, thus resulting in conservative control laws or sometimes even the lack of a control law which satisfies the desired motion planning specifications. This conservatism can be alleviated through a probabilistic model with a correspondingly modified task specification of achieving the desired objective with a high level of confidence.

The model of the system dynamics is as follows. Let x^1, x^2, y^1, y^2 denote the position and velocity of the quadrotor relative to the ground vehicle in the $x-$axis and $y-$axis, respectively. Then from the point of view of a high-level controller, the position-velocity dynamics of the quadrotor in the planar x and y directions can be modeled as decoupled double integrator, controlled in the x-direction by the roll angle ϕ and in the y-direction by the pitch angle θ angle. The corresponding equations of motion in discrete time is given by

$$\begin{aligned}
x_{k+1}^1 &= x_k^1 + \Delta t x_k^2 + \frac{\Delta t^2}{2}(g\sin(\phi_k) + d_k^x) + \eta_k^1 \\
x_{k+1}^2 &= x_k^2 + \Delta t(g\sin(\phi_k) + d_k^x) + \eta_k^2 \\
y_{k+1}^1 &= y_k^1 + \Delta t y_k^2 + \frac{\Delta t^2}{2}(g\sin(-\theta_k) + d_k^y) + \eta_k^3 \\
y_{k+1}^2 &= y_k^2 + \Delta t(g\sin(-\theta_k) + d_k^y) + \eta_k^4
\end{aligned}$$

In the above, Δt is the discretization step, g is the gravitational acceleration constant, and d^x and d^y are bounded uncertainty terms corresponding to the acceleration of the ground vehicle. The variables η_k^i, for $i = 1, \ldots, 4$ are stochastic uncertainty terms arising from unmodeled dynamics and actuator noise. The noise variables are assumed to have a Gaussian distribution, with $\eta^i \sim \mathcal{N}(0, (\sigma^i \Delta t)^2)$.

Based upon experimental trials, the bounds for the acceleration d^x and d^y of the ground vehicle are chosen to be $[-0.4, 0.4]$ m/s^2 corresponding to about 30% of the maximum allowable acceleration of the quadrotor. For this example the roll and pitch commands ϕ and θ are selected from a quantized input range due to digital implementation. Specifically, they are selected from the input range $[-10°, 10°]$ at a $2.5°$ quantization step. These quantization levels can be viewed as the discrete states of the system, similar to the discrete flight maneuvers of the previous example. From this description, we can derive a DTSHG model for this model using a similar procedure as given for the pairwise aircraft conflict resolution example in Section II.

For the specification of the reach-avoid problem, the target set is chosen to be a square-shaped hover region centered on the ground vehicle, specified in (x^1, x^2) coordinates as

$$K_x = [-0.2, 0.2]m \times [-0.2, 0.2]m/s.$$

The safe set in this case is chosen to be the set of all states within the domain of interest for which the relative position remains within a desired bound and a desired velocity bound is satisfied, specified in (x^1, x^2) coordinates as

$$K'_x = [-1.2, 1.2]m \times [-1, 1]m/s.$$

The corresponding sets in K_y and K'_y in y^1, y^2 coordinates are chosen identically as above. The target and safe sets in two dimensions are then defined as $K = K_x \times K_y$ and $K' = K'_x \times K'_y$ respectively. Under a stochastic game formulation of the motion planning problem, the objective of the quadrotor (player I) is to reach the hover region K within finite time, while staying within the safe set K', subject to the worst-case acceleration inputs of the ground vehicle (player II).

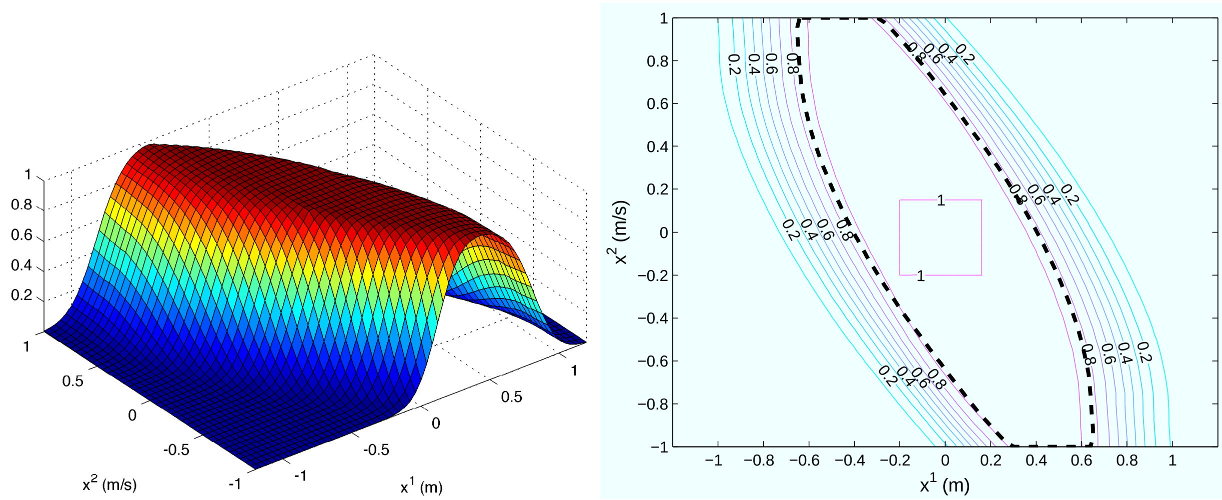
Given that the dynamics, target set, and safe set in the x and y directions are decoupled and identical, the problem reduces to a two dimensional probabilistic reach-avoid calculation in the position-velocity space. For the numerical results to be shown here, we set the noise variance to $\sigma^i = 0.4$, the sampling time to $\Delta t = 0.1s$, and the time horizon to one second ($N = 10$). The disturbance input was discretized at $0.1m/s^2$ intervals for numerical computation. The numerical computation is performed over the safe set K'_x, on a grid size of 61×41.

The max-min probability of satisfying the desired motion planning objectives is shown in Fig. 5a over the safe set K'_x. The corresponding contours of this probability map are shown in Fig. 5b, with the target set K_x in the center. As a comparison, we also plot in the same figure the result of a deterministic reachability calculation from [40], characterizing the set of feasible initial conditions under the assumption that the noise obeys certain deterministic bounds. The advantages of the stochastic model then becomes apparent: although the uncertainties present in the system may satisfy certain bounds with high probability, there may exist realizations of the noise variables which violate the bounds, albeit at low probability. Accounting for all such realizations through deterministic bounds could lead to conservative feasible sets, outside of which a control law satisfying the motion planning objectives does not exist. On the other hand, if one were to resort to a stochastic formulation of the problem, the behavior of the disturbances can be characterized through Gaussian distributions, and the specification can be relaxed to one where the desired objectives are satisfied with a high level of confidence, for example with 70% or 80% probability, resulting in a larger set of feasible initial conditions. The desired control laws can then be obtained for such initial conditons using the sufficient conditions of optimality as given in Section IV.

VI. Conclusion

In this technical report, we discussed a framework for studying probabilistic safety and reachability problems for discrete-time stochastic hybrid systems in a zero-sum stochastic game setting. It was shown that, under certain assumptions on the underlying stochastic kernels and action spaces, there exists a robust control policy which guarantees a worst-case probability of satisfying the safety and reachability objectives, regardless of the adversary strategy. Furthermore, this worst-case probability can be computed via an appropriate dynamic programming recursion, from which sufficient conditions for optimality can be derived.

On the theoretical side, there are several possible directions for future work. First, we would like to establish results, similar to those given in [38], on the approximation of the value function and optimal strategies. Second, it would be interesting to explore whether the consideration of randomized strategies and non-Markov policies confers an advantage to either player. Third,



(a) max-min reach-avoid probability function $r^*_{x_0}(K, K')$

(b) max-min reach-avoid probability contours

Fig. 5: Probability of reach-avoid for the relative position and velocity of the quadrotor with respect to the ground vehicle.

we are working on extensions of the results given in this report for the finite horizon reach-avoid problem to the infinite horizon case, as motivated by [19]. Finally, for applications beyond robust control, it may be necessary to consider alternative stochastic game formulations with less conservative assumptions on the information pattern.

REFERENCES

[1] S. Sastry, G. Meyer, C. Tomlin, J. Lygeros, D. Godbole, and G. Pappas, "Hybrid control in air traffic management systems," in *IEEE Conference on Decision and Control*, vol. 2, 1995, pp. 1478–1483.

[2] C. Tomlin, G. Pappas, and S. Sastry, "Conflict resolution for air traffic management: A study in multiagent hybrid systems," *IEEE Transactions on Automatic Control*, vol. 43, no. 4, pp. 509–521, 2002.

[3] M. Soler, A. Olivares, and E. Staffed, "Hybrid Optimal Control Approach to Commercial Aircraft Trajectory Planning," *Journal of Guidance, Control and Dynamics*, vol. 33, no. 3, pp. 985–991, 2010.

[4] A. Balluchi, L. Benvenuti, M. Di Benedetto, C. Pinello, and A. Sangiovanni-Vincentelli, "Automotive engine control and hybrid systems: Challenges and opportunities," *Proceedings of the IEEE*, vol. 88, no. 7, pp. 888–912, 2000.

[5] R. Ghosh and C. Tomlin, "Symbolic reachable set computation of piecewise affine hybrid automata and its application to biological modelling: Delta-Notch protein signalling," *Systems Biology*, vol. 1, no. 1, pp. 170–183, 2004.

[6] P. Lincoln and A. Tiwari, "Symbolic systems biology: Hybrid modeling and analysis of biological networks," in *Hybrid Systems: Computation and Control*, ser. Lecture Notes in Computer Science, R. Alur and G. Pappas, Eds. Springer, 2004, pp. 147–165.

[7] A. Ames, R. Sinnet, and E. Wendel, "Three-dimensional kneed bipedal walking: A hybrid geometric approach," in *Hybrid Systems: Computation and Control*, ser. Lecture Notes in Computer Science, R. Majumdar and P. Tabuada, Eds. Springer, 2009, pp. 16–30.

[8] J. P. Hespanha, "Stochastic hybrid systems: Application to communication networks," in *Hybrid Systems: Computation and Control*, ser. Lecture Notes in Computer Science, R. Alur and G. J. Pappas, Eds. Springer Berlin / Heidelberg, 2004, vol. 2993, pp. 47–56.

[9] W. Glover and J. Lygeros, "A stochastic hybrid model for air traffic control simulation," in *Hybrid Systems: Computation and Control*, ser. Lecture Notes in Computer Science, R. Alur and G. J. Pappas, Eds. Springer Berlin / Heidelberg, 2004, vol. 2993, pp. 372–386.

[10] R. Alur, C. Belta, F. Ivani, V. Kumar, M. Mintz, G. Pappas, H. Rubin, and J. Schug, "Hybrid modeling and simulation of biomolecular networks," in *Hybrid Systems: Computation and Control*, ser. Lecture Notes in Computer Science, M. Di Benedetto and A. Sangiovanni-Vincentelli, Eds. Springer Berlin / Heidelberg, 2001, vol. 2034, pp. 19–32.

[11] J. Hu, J. Lygeros, and S. Sastry, "Towards a theory of stochastic hybrid systems," in *Hybrid Systems: Computation and Control*, ser. Lecture Notes in Computer Science, N. A. Lynch and B. H. Krogh, Eds. Springer, 2000, vol. 1790, pp. 160–173.

[12] M. L. Bujorianu and J. Lygeros, "Reachability questions in piecewise deterministic Markov processes," in *Hybrid Systems: Computation and Control*, ser. Lecture Notes in Computer Science, O. Maler and A. Pnueli, Eds. Springer, 2003, pp. 126–140.

[13] X. D. Koutsoukos and D. Riley, "Computational methods for reachability analysis of stochastic hybrid systems," in *Hybrid Systems: Computation and Control*, ser. Lecture Notes in Computer Science, J. P. Hespanha and A. Tiwari, Eds. Springer, 2006, pp. 377–391.

[14] J. Hu, M. Prandini, and S. Sastry, "Probabilistic safety analysis in three dimensional aircraft flight," in *IEEE Conference on Decision and Control*, vol. 5, December 2003, pp. 5335–5340.

[15] ——, "Aircraft conflict prediction in the presence of a spatially correlated wind field," *IEEE Transactions on Intelligent Transportation Systems*, vol. 6, no. 3, pp. 326–340, September 2005.

[16] H. J. Kushner and P. G. Dupuis, *Numerical methods for stochastic control problems in continuous time*. London, UK: Springer-Verlag, 1992.

[17] S. Prajna, A. Jadbabaie, and G. Pappas, "A framework for worst-case and stochastic safety verification using barrier certificates," *Automatic Control, IEEE Transactions on*, vol. 52, no. 8, pp. 1415 –1428, aug. 2007.

[18] A. Abate, M. Prandini, J. Lygeros, and S. Sastry, "Probabilistic reachability and safety for controlled discrete time stochastic hybrid systems," *Automatica*, vol. 44, no. 11, pp. 2724 – 2734, 2008.

[19] S. Summers and J. Lygeros, "Verification of discrete time stochastic hybrid systems: A stochastic reach-avoid decision problem," *Automatica*, vol. 46, no. 12, pp. 1951 – 1961, 2010.

[20] A. Abate, S. Amin, M. Prandini, J. Lygeros, and S. Sastry, "Probabilistic reachability and safe sets computation for discrete time stochastic hybrid systems," in *IEEE Conference on Decision and Control*, pp. 258–263.

[21] S. Summers, M. Kamgarpour, J. Lygeros, and C. Tomlin, "A Stochastic Reach-Avoid Problem with Random Obstacles," in *Hybrid Systems: Computation and Control*. ACM, 2011, pp. 251–260.

[22] S. Amin, A. Cardenas, and S. Sastry, "Safe and secure networked control systems under denial-of-service attacks," in *Hybrid Systems: Computation and Control*, ser. Lecture Notes in Computer Science, R. Majumdar and P. Tabuada, Eds. Springer Berlin / Heidelberg, 2009, vol. 5469, pp. 31–45.

[23] M. Kamgarpour, J. Ding, S. Summers, A. Abate, J. Lygeros, and C. Tomlin, “Discrete time stochastic hybrid dynamical games: Verification & controller synthesis,” in *IEEE Conference on Decision and Control*, Dec 2011, accepted.

[24] A. S. Nowak, “Universally measurable strategies in zero-sum stochastic games,” *The Annals of Probability*, vol. 13, no. 1, pp. pp. 269–287, 1985.

[25] U. Rieder, “Non-cooperative dynamic games with general utility functions,” in *Stochastic Games and Related Topics*, T. Raghavan, T. S. Ferguson, T. Parthasarathy, and O. J. Vrieze, Eds. Kluwer Academic Publishers, 1991, pp. 161 –174.

[26] J. I. Gonzalez-Trejo, O. Hernandez-Lerma, and L. F. Hoyos-Reyes, “Minimax control of discrete-time stochastic systems,” *SIAM Journal on Control and Optimization*, vol. 41, no. 5, pp. 1626–1659, 2002.

[27] P. R. Kumar and T. H. Shiau, “Existence of value and randomized strategies in zero-sum discrete-time stochastic dynamic games,” *SIAM Journal on Control and Optimization*, vol. 19, no. 5, pp. pp. 617–634, 1981.

[28] A. Maitra and W. Sudderth, “Finitely additive stochastic games with Borel measurable payoffs,” *International Journal of Game Theory*, vol. 27, pp. 257–267, 1998.

[29] T. Başar and G. Olsder, *Dynamic noncooperative game theory*. Society for Industrial Mathematics, 1999.

[30] D. P. Bertsekas and S. E. Shreve, *Stochastic Optimal Control: The Discrete Time Case*. Academic Press, 1978.

[31] J. Lygeros and M. Prandini, “Aircraft and weather models for probabilistic collision avoidance in air traffic control,” in *IEEE Conference on Decision and Control*, vol. 3, 2002, pp. 2427–2432.

[32] J. Hu, M. Prandini, and S. Sastry, “Aircraft conflict prediction in the presence of a spatially correlated wind field,” *Intelligent Transportation Systems, IEEE Transactions on*, vol. 6, no. 3, pp. 326–340, 2005.

[33] L. D. Brown and R. Purves, “Measurable selections of extrema,” *The Annals of Statistics*, vol. 1, no. 5, pp. 902–912, 1973.

[34] G. B. Folland, *Real Analysis*. John Wiley & Sons, 1999.

[35] M. Davis, *Markov Models and Optimization*. London: Chapman & Hall, 1993.

[36] W. Rudin, *Principles of Mathematical Analysis*, 3rd ed. McGraw-Hill, 1976.

[37] A. Abate, S. Amin, M. Prandini, J. Lygeros, and S. Sastry, “Computational approaches to reachability analysis of stochastic hybrid systems,” in *Hybrid Systems: Computation and Control*, ser. Lecture Notes in Computer Science, A. Bemporad, A. Bicchi, and G. C. Buttazzo, Eds. Springer, 2007, vol. 4416, pp. 4–17.

[38] A. Abate, J. Katoen, J. Lygeros, and M. Prandini, “Approximate model checking of stochastic hybrid systems,” *European Journal of Control*, no. 6, pp. 624–641, 2010.

[39] S. Esmaeil Sadegh Soudjani and A. Abate, “Adaptive gridding for abstraction and verification of stochastic hybrid systems,” in *Proceedings of the 8th International Conference on Quantitative Evaluation of SysTems*, Aachen, DE, September 2011.

[40] J. Ding, E. Li, H. Huang, and C. J. Tomlin, “Reachability-based synthesis of feedback policies for motion planning under bounded disturbances,” in *2011 IEEE International Conference on Robotics and Automation (ICRA)*, may 2011, pp. 2160 –2165.

[41] G. Hoffmann, H. Huang, S. Waslander, and C. J. Tomlin, “Quadrotor helicopter flight dynamics and control: Theory and experiment,” in *AIAA Conference on Guidance, Navigation and Control*, Aug. 2007.

[42] H. Huang, G. Hoffmann, S. Waslander, and C. Tomlin, “Aerodynamics and control of autonomous quadrotor helicopters in aggressive maneuvering,” in *IEEE International Conference on Robotics and Automation*, 2009, pp. 3277–3282.