Automatica 94 (2018) 94–101

Contents lists available at ScienceDirect

Automatica

journal homepage: www.elsevier.com/locate/automatica

Brief paper

A dynamic game approach to distributionally robust safety specifications for stochastic systems[☆]

Insoon Yang

Department of Electrical and Computer Engineering, Automation and Systems Research Institute, Seoul National University, Seoul, 08826, South Korea

ARTICLE INFO

Article history:
Received 30 January 2017
Received in revised form 26 October 2017
Accepted 14 March 2018
Available online 4 May 2018

Keywords:
Stochastic control
Stochastic systems
Verification
Safety analysis
Reachability
Game theory
Duality
Dynamic programming

ABSTRACT

This paper presents a new safety specification method that is robust against errors in the probability distribution of disturbances. Our proposed distributionally robust safe policy maximizes the probability of a system remaining in a desired set for all times, subject to the worst possible disturbance distribution in an ambiguity set. We propose a dynamic game formulation of constructing such policies and identify conditions under which a non-randomized Markov policy is optimal. Based on this existence result, we develop a practical design approach to safety-oriented stochastic controllers with limited information about disturbance distributions. However, an associated Bellman equation involves infinite-dimensional minimax optimization problems since the disturbance distribution may have a continuous density. To alleviate computational issues, we propose a duality-based reformulation method that converts the infinite-dimensional minimax problem into a semi-infinite program that can be solved using existing convergent algorithms. We prove that there is no duality gap, and that this approach thus preserves optimality. The results of numerical tests confirm that the proposed method is robust against distributional errors in disturbances, while a standard stochastic safety verification tool is not.

1. Introduction

For safety-critical systems subject to uncertain disturbances, reachability-based safety specification techniques have been used to compute the reachable sets and safe sets, which allow one to verify that a system is evolving within a safe range of operation and to synthesize controllers to satisfy safety constraints (e.g., Althoff, Le Guernic, and Krogh, 2011; Bertsekas and Rhodes, 1971; Cardaliaguet, Quincampoix, and Saint-Pierre, 1999; Chen, Herbert, Vashishtha, Bansal, and Tomlin, 2016; Ghaemi and Del Vecchio, 2014; Girard, 2005; Kurzhanski and Varaiya, 2002; Lygeros, Tomlin, and Sastry, 1999; Margellos and Lygeros, 2011; Mitchell, Bayen, and Tomlin, 2005; Rakovic, Kerrigan, Mayne, and Lygeros, 2006). These methods assume that disturbances lie in a compact set, and thus require information only about the support of disturbances. However, these techniques often produce conservative results as no additional information about uncertain disturbances is used. These deterministic methods are a natural choice when the data of disturbances are not continuously collected, and thus a reliable stochastic model is unavailable for them. Advances in sensing, communication, and computing technologies as well as statistical learning and estimation tools make it possible to shift this paradigm; sensors, data storage and computing infrastructure can now provide data to help estimate disturbance distributions. Stochastic reachability analysis tools are based on the assumption that the probability distribution of disturbances is available and can be used to reduce the conservativeness of their deterministic counterpart. However, this assumption is often restrictive in practice because obtaining an accurate distribution requires large-scale high-resolution sensor measurements over a long training period or multiple periods. Furthermore, the accuracy of the distribution obtained by computational methods is often unreliable as it is subject to the quality of the collected data, statistical learning or filtering methods, and prior knowledge. Thus, probabilistic safety specification tools can lead to the design of an unreliable controller that may violate safety constraints.

This paper aims to bridge the gap between the two methods by proposing a *distributionally robust safety specification* tool. Our approach assumes that the distribution of disturbances is not fully known but lies in a so-called *ambiguity set* of probability distributions. The proposed *distributionally robust safe policy* maximizes the probability of a system remaining within a desired set for all times subject to the worst possible disturbance distribution in the ambiguity set. Therefore, the probabilistic safe set of the closed-loop system is robust against distributional errors within

[☆] Supported in part by NSF under ECCS 1708906 and CNS 1657100, and in part by Research Resettlement Fund for the new faculty of Seoul National University (0668-20180071). The material in this paper was not presented at any conference. This paper was recommended for publication in revised form by Associate Editor Gurdal Arslan under the direction of Editor Ian R. Petersen.

E-mail address: insoonyang@snu.ac.kr.

https://doi.org/10.1016/j.automatica.2018.04.022
0005-1098/© 2018 Elsevier Ltd. All rights reserved.

the ambiguity set. We propose a dynamic game formulation of constructing distributionally robust safe policies and safe sets (in Section 2). Specifically, it is a two-player zero-sum dynamic game in which Player I selects a policy by which the controller can maximize the probability of safety, while (fictitious) Player II determines a strategy for the probability distribution of disturbances to minimize the same probability. Player II's action space is generally infinite dimensional since the disturbances may have a continuous density function. Therefore, an associated Bellman equation involves infinite-dimensional optimization problems that are computationally challenging. Furthermore, the existence of an optimal control policy is not guaranteed.

The contributions of this work are threefold. First, we characterize conditions for the existence and optimality of a non-randomized Markov control policy for Player I (in Section 3). This characterization helps greatly reduce the control strategy space we need to search for because it is enough to restrict our attention to non-randomized Markov policies. Second, we develop a design approach to a safety-oriented controller with limited information about disturbance distributions. This control method can be used to minimize another cost function while guaranteeing that the probability for a system being safe for all remaining stages is greater than or equal to a pre-specified threshold. Third, we propose a duality-based reformulation method for the Bellman equation in cases with moment uncertainty (in Section 4). We show that there is no duality gap in the inner minimization problem of the Bellman equation, which is an infinite-dimensional minimax problem. Using the strong duality result, we reformulate the Bellman equation as a semi-infinite program without sacrificing optimality. This reformulation alleviates the computational issue arising from the infinite dimensionality of the original Bellman equation because the reformulated Bellman equation can be solved via existing convergent algorithms for semi-infinite programs. The distributional robustness of the proposed tool is illustrated through examples (in Section 5).

We summarize related studies as follows. A probabilistic reachability tool using a Markov chain approximation has been proposed in Hu, Prandini, and Sastry (2005) and Prandini and Hu (2006). In Prajna, Jadbabaie, and Pappas (2007), barrier certificates are employed to calculate an upper bound of the probability that a system will reach a target set. Additionally, Mitchell and Templeton (2005) propose a toolbox that supports expectation-based reachability problems by extending the celebrated Hamilton–Jacobi–Isaacs reachability analysis (Mitchell et al., 2005). For discrete-time stochastic hybrid systems, an elegant dynamic programming approach has been proposed to compute the maximal probability of safety (Abate, Prandini, Lygeros, & Sastry, 2008). This method has been extended to stochastic reach–avoid problems (Summers & Lygeros, 2010), stochastic hybrid games (Ding, Kamgarpour, Summers, Abate, Lygeros, & Tomlin, 2013), and partially observable hybrid systems (Lesser & Oishi, 2014). However, all the aforementioned methods are based on the possibly restrictive assumption that the probability distribution of disturbances is completely known. This paper is also closely related to *distributionally robust control*; it minimizes the worst-case cost, assuming that the probability distribution of uncertain variables lies within an ambiguity set of distributions. A distributionally robust Markov decision process (MDP) formulation has recently been developed while focusing on finite-state, finite-action MDPs (Xu & Mannor, 2012; Yang, 2017a; Yu & Xu, 2016). For cases with moment uncertainty, Van Parys, Kuhn, Goulart, & Morari (2016) investigate linear feedback strategies in linear–quadratic settings with risk constraints and propose a semidefinite programming approach. We extend the theory of distributionally robust control to the case of continuous state spaces and apply it to safety specifications.

We use the following notation throughout the paper. Given a Borel space X, $\mathcal{B}(X)$ and $\mathcal{P}(X)$ represent its Borel σ-algebra and the set of Borel probability measures on X, respectively. The set $\mathbb{S}_+^l$ denotes the space of $l \times l$ symmetric positive semidefinite matrices. We also let $\mathcal{T} := \{0, 1, \ldots, T-1\}$ and $\bar{\mathcal{T}} := \{0, 1, \ldots, T\}$.

2. Distributionally robust safe sets and policies

Consider a discrete-time stochastic system of the form

$$x_{t+1} = f(x_t, u_t, w_t) \quad \forall t \in \mathcal{T}, \quad x_0 = \boldsymbol{x}, \tag{1}$$

where $x_t \in \mathbb{R}^n$ is the state, $u_t \in \mathbb{R}^m$ is the control input, $w_t \in \mathbb{R}^l$ is the stochastic disturbance, and $f : \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^l \to \mathbb{R}^n$ is a measurable function. We assume that the disturbance process $\{w_t\}_{t=0}^{T-1}$ is defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, and that w_s and w_t are independent for any $s \neq t$. As mentioned in Section 1, it is often difficult to obtain full information about the probability distribution μ_t of w_t. To mathematically model distributional ambiguity, we assume that μ_t is not fully known but contained in a so-called *ambiguity set* of distributions, denoted by $\mathbb{D}_t \subseteq \mathcal{P}(\mathbb{R}^l)$.

We now briefly discuss admissible control and disturbance distribution strategies. Let H_t be the set of histories up to stage t, whose element takes the form $h_t := (x_0, u_0, w_0, \ldots, x_{t-1}, u_{t-1}, w_{t-1}, x_t)$.[1] The set of admissible control strategies is chosen as $\Pi := \{\pi := (\pi_0, \ldots, \pi_{T-1}) \mid \pi_t(\mathbb{U}(x_t)|h_t) = 1 \ \forall h_t \in H_t\}$, where π_t is a stochastic kernel from H_t to $\mathbb{R}^m$ and $\mathbb{U}(x_t)$ is the set of admissible actions given state x_t. Note that this strategy space is sufficiently broad to contain randomized non-Markov policies. Considering an adversarial player who chooses the disturbance's probability distribution μ_t, the set of admissible disturbance distribution strategies is defined as $\Gamma := \{\gamma := (\gamma_0, \ldots, \gamma_{T-1}) \mid \gamma_t(\mathbb{D}_t|h_t^e) = 1 \ \forall h_t^e \in H_t^e\}$, where H_t^e is a set of extended histories up to stage t, whose element is of the form $h_t^e := (x_0, u_0, w_0, \mu_0, \ldots, x_{t-1}, u_{t-1}, w_{t-1}, \mu_{t-1}, x_t, u_t)$. Note that the distributional constraints in the ambiguity set $\mathbb{D}_t$ is encoded in the strategy space Γ.

2.1. Distributionally robust safety specifications

Our goal is to compute the worst-case probability of a system remaining in a desired set for all times when the distribution of w_t is not fully known but lies within an ambiguity set, $\mathbb{D}_t$. To formulate a concrete safety specification problem, we consider a desired set A for safety, which is an arbitrary compact Borel set in the state space $\mathbb{R}^n$. We also introduce the following definitions:

Definition 1 (*Probability of Safety*)**.** We define the probability that the system (1) is safe for all $t \in \bar{\mathcal{T}}$ given the strategy pair (π, γ) and the initial value $\boldsymbol{x}$ as

$$P_{\boldsymbol{x}}^{\text{safe}}(\pi, \gamma; A) := \mathbb{P}^{\pi,\gamma}\{x_t \in A \ \ \forall t \in \bar{\mathcal{T}} \mid x_0 = \boldsymbol{x}\},$$

which we call the *probability of safety* for A. We also define the *probabilistic safe set* with probability α under (π, γ) as $S_\alpha(\pi, \gamma; A) := \{\boldsymbol{x} \in \mathbb{R}^n \mid P_{\boldsymbol{x}}^{\text{safe}}(\pi, \gamma; A) \geq \alpha\}$.

This set contains all the initial states such that the probability that the system stays in the set A is greater than or equal to α given the strategy pair (π, γ). This definition generalizes the probabilistic safe set introduced in Abate et al. (2008) to the case

[1] To make it practically sound, we assume that the controller is unable to observe the disturbance's probability distribution. However, due to Theorem 1, all of our results and analyses remain valid even when h_t includes $(\mu_0, \ldots, \mu_{t-1})$.

with ambiguous disturbance distributions. Using these notions, we now define a distributionally robust safe policy and set as follows:

Definition 2 (*Distributionally Robust Safe Set*). A control strategy $\pi^\star \in \Pi$ is said to be a *distributionally robust safe policy* given $x_0 = \boldsymbol{x}$ if it satisfies

$$\inf_{\gamma\in\Gamma} P_{\boldsymbol{x}}^{\text{safe}}(\pi^\star, \gamma; A) \geq \inf_{\gamma'\in\Gamma} P_{\boldsymbol{x}}^{\text{safe}}(\pi, \gamma'; A) \quad \forall \pi \in \Pi.$$

The set $S_\alpha^\star(A)$ is said to be the *distributionally robust safe set* for A with probability α if

$$S_\alpha^\star(A) = \left\{\boldsymbol{x} \in \mathbb{R}^n \mid \sup_{\pi\in\Pi} \inf_{\gamma\in\Gamma} P_{\boldsymbol{x}}^{\text{safe}}(\pi, \gamma; A) \geq \alpha\right\}.$$

The distributionally robust safe policy $\pi^\star$ maximizes the worst-case probability of safety under distributional ambiguity characterized by the constraints in the set $\mathbb{D}_t$. No matter what form the strategy γ takes so that the realized distribution lies in the ambiguity set $\mathbb{D}_t$, the probability that the system starting from $\boldsymbol{x} \in S_\alpha^*(A)$ stays safe is greater than or equal to α under the policy $\pi^\star$. Once we obtain $\pi^\star$ and $P_{\boldsymbol{x}}^{\text{safe}}$ for each $\boldsymbol{x}$, we can calculate $S_\alpha^\star(A)$ through simple thresholding.

2.2. A dynamic game formulation

We first note that $P_{\boldsymbol{x}}^{\text{safe}}(\pi, \gamma; A) = \mathbb{E}^{\pi,\gamma}[\prod_{t=0}^T \mathbf{1}_A(x_t) \mid x_0 = \boldsymbol{x}]$, where $\mathbf{1}_A : \mathbb{R}^n \to \{0, 1\}$ is the indicator function of the set A such that $\mathbf{1}_A(\boldsymbol{x}) = 1$ if $\boldsymbol{x} \in A$ and $\mathbf{1}_A(\boldsymbol{x}) = 0$ otherwise. Here, $\mathbb{E}^{\pi,\gamma}$ is the expectation taken with respect to the probability measure $\mathbb{P}^{\pi,\gamma}$ induced by the strategy pair (π, γ). The problem of constructing a distributionally robust safe policy can then be formulated as the following dynamic game problem:

$$\sup_{\pi\in\Pi} \inf_{\gamma\in\Gamma} P_{\boldsymbol{x}}^{\text{safe}}(\pi, \gamma; A) = \mathbb{E}^{\pi,\gamma}\left[\prod_{t=0}^{T} \mathbf{1}_A(x_t) \mid x_0 = \boldsymbol{x}\right]. \qquad (2)$$

In this two-player zero-sum game, Player I determines a control policy π to maximize the probability of safety assuming that Player II selects a disturbance distribution strategy γ to minimize the probability of safety. In the next section, we establish some analytical results about the dynamic game problem. In particular, we show that under mild conditions a non-randomized Markov control policy is optimal.

3. Dynamic programming

3.1. Existence and optimality of Markov policies

Let $\mathbb{K}_t \in \mathcal{B}(\mathbb{R}^n \times \mathbb{R}^m \times \mathcal{P}(\mathbb{R}^l))$ be the collection of elements such that each $(\boldsymbol{x}, \boldsymbol{u}, \boldsymbol{\mu}) \in \mathbb{K}_t$ satisfies (*i*) $\boldsymbol{u} \in \mathbb{U}(\boldsymbol{x})$ and (*ii*) $\boldsymbol{\mu} \in \mathbb{D}_t$.

Assumption 1. The following conditions hold:

(i) For any bounded continuous function $g : \mathbb{R}^n \to \mathbb{R}$, the function $\hat{g}_t(\boldsymbol{x}, \boldsymbol{u}, \boldsymbol{\mu}) := \int_{\mathbb{R}^l} g(f(\boldsymbol{x}, \boldsymbol{u}, \boldsymbol{w}))\mathrm{d}\boldsymbol{\mu}(\boldsymbol{w})$ is continuous on $\mathbb{K}_t$ for each $t \in \mathcal{T}$;

(ii) The set $\mathbb{U}(\boldsymbol{x})$ is compact for each $\boldsymbol{x} \in \mathbb{R}^n$. Furthermore, the set-valued mapping $\boldsymbol{x} \mapsto \mathbb{U}(\boldsymbol{x})$ is upper semi-continuous;

(iii) The desired set A for safety is compact.

These conditions are standard *measurable selection conditions* for semi-continuous stochastic control models (e.g., Dubins and Savage, 1965; González-Trejo, Hernández-Lerma, and Hoyos-Reyes, 2003; Hernández-Lerma and Lasserre, 2012) and will be used to ensure the existence of a distributionally robust safe policy. For each measurable function $\mathbf{v}$ on $\mathbb{R}^n$, we define a dynamic programming operator, denoted by $\mathbf{T}_t$, as $\mathbf{T}_t\mathbf{v}(\boldsymbol{x}) := \sup_{\boldsymbol{u}\in\mathbb{U}(\boldsymbol{x})}\inf_{\boldsymbol{\mu}\in\mathbb{D}_t} \mathbf{1}_A(\boldsymbol{x}) \int_{\mathbb{R}^l} \mathbf{v}(f(\boldsymbol{x}, \boldsymbol{u}, \boldsymbol{w}))\mathrm{d}\boldsymbol{\mu}(\boldsymbol{w})$. We also define the value function of the distributionally robust safe control problem (2) as $v_t(\boldsymbol{x}) := \mathbf{T}_t \circ \mathbf{T}_{t+1} \circ \cdots \circ \mathbf{T}_{T-1}\mathbf{1}_A(\boldsymbol{x})$. It represents the maximal worst-case probability of the system being safe from stage t to T when $x_t = \boldsymbol{x}$. For $t = T$, the value function is defined as $v_T(\boldsymbol{x}) := \mathbf{1}_A(\boldsymbol{x})$. Under Assumption 1, we can show that the problem (2) admits a non-randomized Markov policy, which is optimal.

Theorem 1. *Suppose that Assumption 1 holds. For each $t \in \mathcal{T}$, there exists a measurable function $\pi_t^\star : \mathbb{R}^n \to \mathbb{R}^m$ such that $\pi_t^\star(\boldsymbol{x}) \in \mathbb{U}(\boldsymbol{x})$ and*

$$v_t(\boldsymbol{x}) = \inf_{\boldsymbol{\mu}\in\mathbb{D}_t}\left[\mathbf{1}_A(\boldsymbol{x}) \int_{\mathbb{R}^l} v_{t+1}(f(\boldsymbol{x}, \pi_t^\star(\boldsymbol{x}), \boldsymbol{w}))\mathrm{d}\boldsymbol{\mu}(\boldsymbol{w})\right]$$

for all $\boldsymbol{x} \in \mathbb{R}^n$. The non-randomized Markov policy $\pi^\star := (\pi_0^\star, \ldots, \pi_{T-1}^\star) \in \Pi$ is a distributionally robust safe policy, i.e.,

$$v_0(\boldsymbol{x}) = \inf_{\gamma\in\Gamma} P_{\boldsymbol{x}}^{safe}(\pi^\star, \gamma; A).$$

Furthermore, the value function v_t is upper semi-continuous for each $t \in \bar{\mathcal{T}}$.

Its proof is contained in Appendix A. The key idea is to show that the upper semi-continuity of the value function is preserved through the dynamic programming operator. This theorem greatly reduces the control strategy space we need to search for because it suffices to restrict our attention to non-randomized Markov policies. The Markov policy $\pi^\star$ maximizes the worst-case probability of safety no matter how the disturbance (Player II) chooses its probability distribution μ_t in the ambiguity set $\mathbb{D}_t$. Furthermore, the dynamic programming principle allows us to obtain the following Bellman equation:

Proposition 1. *Suppose that Assumption 1 holds. The value function v_t solves the following Bellman equation:*

$$v_t(\boldsymbol{x}) = \max_{\boldsymbol{u}\in\mathbb{U}(\boldsymbol{x})} \inf_{\boldsymbol{\mu}\in\mathbb{D}_t} \mathbf{1}_A(\boldsymbol{x}) \int_{\mathbb{R}^l} v_{t+1}(f(\boldsymbol{x}, \boldsymbol{u}, \boldsymbol{w}))\mathrm{d}\boldsymbol{\mu}(\boldsymbol{w}) \qquad (3)$$

for $t \in \mathcal{T}$ with $v_T(\boldsymbol{x}) = \mathbf{1}_A(\boldsymbol{x})$.

Note that “max” is used instead of “sup” in the outer problem as its optimal solution exists due to Theorem 1.

3.2. Constructing distributionally robust safe sets and safety-oriented controllers

To obtain distributionally robust control policies and safe sets, we evaluate the value function $\{v_t\}_{t=0}^T$ by solving the Bellman equation (3) backward in time.[2] A distributionally robust safe policy can be constructed as $\pi_t^\star(x_t) \in \operatorname{argmax}_{\boldsymbol{u}\in\mathbb{U}(x_t)}\inf_{\boldsymbol{\mu}\in\mathbb{D}_t} \mathbf{1}_A(x_t) \int_{\mathbb{R}^l} v_{t+1}\,(f(x_t, \boldsymbol{u}, \boldsymbol{w}))\mathrm{d}\boldsymbol{\mu}(\boldsymbol{w})$. The distributionally robust safe set with probability α can then be computed as $S_\alpha^\star(A) = \{\boldsymbol{x} \in \mathbb{R}^n \mid v_0(\boldsymbol{x}) \geq \alpha\}$ because $v_0(\boldsymbol{x}) = \max_{\pi\in\Pi}\inf_{\gamma\in\Gamma} P_{\boldsymbol{x}}^{\text{safe}}(\pi, \gamma; A)$.

Define the t-distributionally robust safe set as $S_{\alpha,t}^\star(A) := \{\boldsymbol{x} \in \mathbb{R}^n \mid \forall\gamma \in \Gamma\ \exists\pi \in \Pi \text{ s.t. } \mathbb{P}^{\pi,\gamma}(x_s \in A\ \forall s = t, \ldots, T \mid x_t = \boldsymbol{x}) \geq \alpha\}$. If the system state lies in $S_{\alpha,t}^\star(A)$ at stage t, there exists a control policy such that the probability of the system being safe during the remaining stages is greater than or equal to α. Note that $S_{\alpha,0}^\star(A) = S_\alpha^\star(A)$ under Assumption 1 which guarantees the

[2] When numerically solving continuous-state dynamic programs, the value function is evaluated at a finite number of state points and thus a function approximation (e.g., interpolation) of v_t is required in general (Gordon, 1995). In the example in Section 5, v_t is evaluated at grid points and the standard piecewise linear interpolation is adopted.

existence of a distributionally robust safe policy. Using the value function, we can compute the set as $S^{\star}_{\alpha,t}(A) = \{\boldsymbol{x} \in \mathbb{R}^n \mid v_t(\boldsymbol{x}) \geq \alpha\}$ because $v_t(\boldsymbol{x}) = \max_{\pi\in\Pi}\inf_{\gamma\in\Gamma}\mathbb{E}^{\pi,\gamma}[\prod_{s=t}^{T}\mathbf{1}_A(x_s)]$ with $x_t = \boldsymbol{x}$. Suppose now that we are given the support $\mathbb{W}_t$ of μ_t for each t. Consider the following controller: given x_t at stage t, the control action is determined as

$$u_t^{\star}\begin{cases} \in \mathbb{U}(x_t) & \text{if } x_t \in \{\boldsymbol{x} \mid f(\boldsymbol{x},\boldsymbol{u},\boldsymbol{w}) \in S^{\star}_{\alpha,t+1}(A) \\ & \qquad\qquad \forall(\boldsymbol{u},\boldsymbol{w}) \in \mathbb{U}(\boldsymbol{x}) \times \mathbb{W}_t\} \\ = \pi_t^{\star}(x_t) & \text{otherwise.} \end{cases} \tag{4}$$

This controller chooses an arbitrary admissible control action if it can drive the system into $S^{\star}_{\alpha,t+1}(A)$ at stage $t+1$ for any realization of the disturbance. Otherwise, it uses a distributionally robust safe policy. This procedure is motivated by the safe controller synthesis method of Lygeros et al. (1999). For each $t \in \mathcal{T}$, this controller ensures that the probability for the system being safe for all remaining stages is greater than or equal to α, regardless of how the disturbance distribution is chosen in $\mathbb{D}_t$.

Proposition 2. *Suppose that Assumption 1 holds. If $x_0 \in S^{\star}_{\alpha}(A)$ and the Markov control policy (4) is used, then*

$$x_t \in S^{\star}_{\alpha,t}(A) \quad \forall t \in \bar{\mathcal{T}}.$$

Its proof can be found in Appendix B. If another objective function needs to be minimized in a distributionally robust way while ensuring safety, one may solve $\inf_{\pi\in\Pi}\sup_{\gamma\in\Gamma}\mathbb{E}^{\pi,\gamma}[\sum_{t=0}^{T-1} r(x_t,u_t) + q(x_T)]$ to obtain an optimal distributionally robust policy π^{opt}. Then, one can employ the proposed controller (4) that chooses $\pi_t^{opt}(x_t)$ whenever $x_t \in \{\boldsymbol{x} \mid f(\boldsymbol{x},\boldsymbol{u},\boldsymbol{w}) \in S^{\star}_{\alpha,t+1}(A) \ \forall(\boldsymbol{u},\boldsymbol{w}) \in \mathbb{U}(\boldsymbol{x}) \times \mathbb{W}_t\}$. Note that this controller prioritizes safety and tries to minimize the worst-case cost value whenever there is the flexibility to do so. This *safety-oriented distributionally robust control design* approach can be overly conservative, particularly when $\mathbb{W}_t$ is large. However, it is computationally efficient because the cost-minimizing control problem is decoupled from the safe control problem.

4. Moment uncertainty and duality

4.1. Ambiguity sets with moment uncertainty

Modeling the ambiguity set $\mathbb{D}_t$ may critically affect the proposed distributionally robust safe policies as they maximize the worst-case probability of safety, assuming that the disturbance distribution lies within $\mathbb{D}_t$. Several ambiguity set modeling approaches have been developed in the context of single-stage optimization problems. A moment-based approach employs an ambiguity set of distributions whose moments satisfy certain constraints (Delage & Ye, 2010; El Ghaoui, Oks, & Oustry, 2003; Wiesemann, Kuhn, & Sim, 2014). A statistical distance-based approach takes into account an ambiguity set of probability distributions that are closed to a nominal distribution in terms of a chosen statistical distance, such as ϕ-divergence (Ben-Tal, Den Hertog, De Waegenaere, Melenberg, & Rennen, 2013; Jiang & Guan, 2016), Prokhorov metric (Erdoğan & Iyengar, 2006) and Wasserstein distance (Gao & Kleywegt, 2016; Mohajerin Esfahani & Kuhn, 2017). In this work, we take a moment-based approach since moments are the most common information that can be reliably obtained from data. Suppose that an estimate of the mean and covariance matrix of the disturbance w_t is the only available information. Let $\mathbf{m}_t \in \mathbb{R}^l$ and $\boldsymbol{\Sigma}_t \in \mathbb{R}^{l\times l}$ be the estimate of the mean and covariance matrix, respectively. The set of all the probability distributions that are consistent with these estimates can be modeled as

$$\begin{aligned}\mathbb{D}_t := \{\mu_t \in \mathcal{P}(\mathbb{W}_t) \mid |\mathbb{E}_{\mu_t}[w_t] - \mathbf{m}_t| \leq b_t, \\ \mathbb{E}_{\mu_t}[(w_t - \mathbf{m}_t)(w_t - \mathbf{m}_t)^{\top}] \preceq c_t\boldsymbol{\Sigma}_t\},\end{aligned} \tag{5}$$

where $\mathbb{E}_{\mu_t}$ denotes the expectation taken with respect to the probability distribution μ_t. Here, $b_t \in \mathbb{R}^l_+$ and $c_t \geq 1$ are given constants that depend on one's confidence in the estimates $\mathbf{m}_t$ and $\boldsymbol{\Sigma}_t$. Any probability distribution in this set satisfies the following properties: (i) the support of w_t is $\mathbb{W}_t$; (ii) the mean of $w_{t,i}$ lies in $[-b_{t,i}, b_{t,i}]$; and (iii) the centered second moment matrix of w_t lies in a positive semidefinite cone.

4.2. Dual Bellman equations

In general, solving the Bellman equation (3) to evaluate $v_t(\boldsymbol{x})$ with the ambiguity set (5) is challenging as it involves infinite-dimensional minimax optimization problems. To resolve this difficulty, we propose a duality-based dynamic programming method.[3]

Theorem 2. *Suppose that Assumption 1 holds and $\mathbb{W}_t$ is compact. Then, the following equality holds*[4]:

$$\begin{aligned} v_t(\boldsymbol{x}) = \mathbf{1}_A(\boldsymbol{x}) \times \\ \sup_{\boldsymbol{u},\underline{\lambda},\overline{\lambda},\Lambda,\nu} \quad & -\underline{b}_t^{\top}\underline{\lambda} - \overline{b}_t^{\top}\overline{\lambda} - c_t Tr(\boldsymbol{\Sigma}_t\Lambda) - \nu \\ s.t. \ & \boldsymbol{w}^{\top}(\overline{\lambda} - \underline{\lambda}) + (\boldsymbol{w} - \mathbf{m}_t)^{\top}\Lambda(\boldsymbol{w} - \mathbf{m}_t) + \nu \\ & + v_{t+1}(f(\boldsymbol{x},\boldsymbol{u},\boldsymbol{w})) \geq 0 \quad \forall \boldsymbol{w} \in \mathbb{W}_t \\ & \boldsymbol{u} \in \mathbb{U}(\boldsymbol{x}), \ \underline{\lambda}, \overline{\lambda} \in \mathbb{R}^l_+, \ \Lambda \in \mathbb{S}^l_+, \ \nu \in \mathbb{R}\end{aligned}$$

for all $(t,\boldsymbol{x}) \in \mathcal{T} \times \mathbb{R}^n$ with the terminal condition $v_T(\boldsymbol{x}) = \mathbf{1}_A(\boldsymbol{x})$, where $\underline{b}_t := b_t - \mathbf{m}_t$ and $\overline{b}_t := b_t + \mathbf{m}_t$. Furthermore, the inner minimization problem in the Bellman equation (3) admits an optimal solution.

See Appendix C for a proof. Its key idea is to show that a generalized Slater condition holds for strong duality. The proof is valid even when $b_{t,i} = 0$ for some i, i.e., the mean estimate $\mathbf{m}_{t,i}$ is perfect. Note that the dual Bellman equation involves semi-infinite optimization problems. Each semi-infinite program can be solved by using several convergent methods, such as discretization methods, exchange methods, homotopy methods, primal–dual methods, and constraint sampling methods (e.g., Calafiore and Campi, 2005; Hettich and Kortanek, 1993; López and Still, 2007 and and the reference therein). This is a remarkable advantage of the proposed reformulation because it is unclear if the infinite-dimensional minimax problem in (3) can be directly solved in a convergent way. In Section 5, we employ the discretization method proposed in Reemtsen (1991). This algorithm adaptively generates a grid on $\mathbb{W}_t$ and converges to a locally optimal value of the semi-infinite program in the dual Bellman equation (see Reemtsen, 1991 for a proof).[5] When the semi-infinite program is concave, it converges to the globally optimal value. The following concavity result holds with general ambiguity sets:

[3] The proposed method does not resolve the scalability issue inherent in dynamic programming: the complexity of computing $v_t(\boldsymbol{x})$ is still exponential with the dimension of $\boldsymbol{x}$.

[4] Note that "$\max_{\boldsymbol{u}}$" and "$\sup_{\underline{\lambda},\overline{\lambda},\Lambda,\nu}$" are merged for a compact representation. However, it should be noted that this semi-infinite program has an optimal solution $\boldsymbol{u}$.

[5] It is worth noting that the number of grid points or constraints normally grows exponentially with the dimension of $\mathbb{W}_t$. This issue can be alleviated by using the random constraint sampling method (Calafiore & Campi, 2005) in which the required number of constraints is independent of $\mathbb{W}_t$'s dimension.

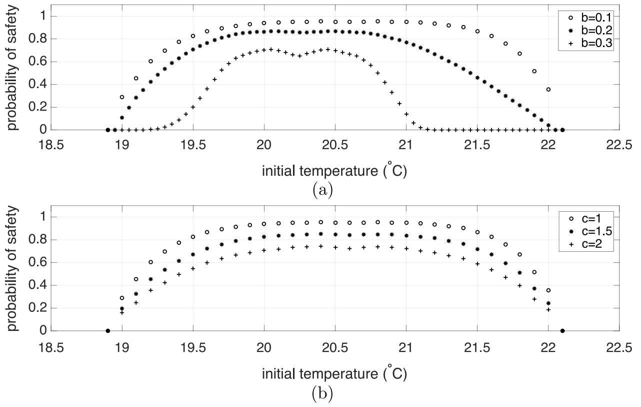

Fig. 1. Effect of (a) the parameter b and (b) the parameter c on the probability of safety.

Proposition 3. *Suppose that (i) $f : \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^l \to \mathbb{R}^n$ is an affine function, (ii) A is a convex set, and (iii) for all $\boldsymbol{x}_1, \boldsymbol{x}_2 \in \mathbb{R}^n$ and for all $\lambda \in (0, 1)$, if $\boldsymbol{u}_i \in \mathbb{U}(\boldsymbol{x}_i)$, $i = 1, 2$, then $\lambda \boldsymbol{u}_1 + (1 - \lambda)\boldsymbol{u}_2 \in \mathbb{U}(\lambda \boldsymbol{x}_1 + (1 - \lambda)\boldsymbol{x}_2)$. Then, the value function $v_t(\boldsymbol{x})$ is concave with respect to $\boldsymbol{x} \in A$ for each $t \in \mathcal{T}$.*

This proposition can be shown using the proof of Proposition 3 in Yang (2017b). Since $\boldsymbol{u} \mapsto v_t(f(\boldsymbol{x}, \boldsymbol{u}, \boldsymbol{w}))$ is concave for each $(t, \boldsymbol{x}, \boldsymbol{w}) \in \bar{\mathcal{T}} \times A \times W_t$, the semi-infinite program in the dual Bellman equation is concave for each $(t, \boldsymbol{x}) \in \mathcal{T} \times A$. Note that the dual Bellman equation in Theorem 2 is exact even when the value function is not convex. If only local optima can be found due to the nonconvexity of $\boldsymbol{u} \mapsto v_{t+1}(f(\boldsymbol{x}, \boldsymbol{u}, \boldsymbol{w}))$, $v_t(\boldsymbol{x})$ is bounded below by the local optima. As a result, the maximum probability of safety and the distributionally robust safe set are underestimated. Thus, the locally optimal distributionally robust policies can still be used for a safety certification but they produce conservative results. To reduce the conservativeness, one may use techniques for nonconvex optimization (e.g., Horst, Pardalos, and Van Thoai, 2000; Lasserre, 2001).

5. Numerical examples

Thermostatically controlled loads (TCLs) – such as air conditioners, refrigerators, and water heaters – are used to guarantee human comfort, and food provision, etc. Therefore, ensuring safe TCL operation is critical in a wide range of applications, and thus has been studied in safety verification literature (Abate et al., 2008). Consider the following model of the temperature being controlled through a TCL: $x_{t+1} = \alpha x_t + (1-\alpha)(\theta - \eta RPu_t) + w_t$. Here, $x_t \in \mathbb{R}$ is the temperature of interest (e.g., indoor temperature, food temperature), $u_t \in \{0, 1\}$ is an ON/OFF control input, and $w_t \in \mathbb{R}$ is a disturbance variable that takes into account uncertainty. For a detailed explanation of our simulation setting, refer to the extended version (Yang, 2017c).

5.1. Effect of the confidence parameters

Fig. 1 shows the probability $P_{\boldsymbol{x}}^{\text{safe}}$ of safety as a function of the initial state $\boldsymbol{x} \in [18, 23]$ for multiple confidence parameters b and c. The function has a bimodal structure due to the discrete ON/OFF control inputs: it can be considered as the point-wise maximum of the probability function with OFF control and that with ON control. As the initial state approaches the boundary of the set $A = [19, 22]$, the probability of safety decreases. As shown in Fig. 1(a), given $c = 1$, the probability of safety decreases with the confidence parameter b for the mean. In other words, an inaccurate mean estimate $\mathbf{m}$ makes it difficult for the system to remain safe for all stages. Fig. 1(b) illustrates that the probability of safety decreases as the uncertainty of the variance estimate Σ increases.

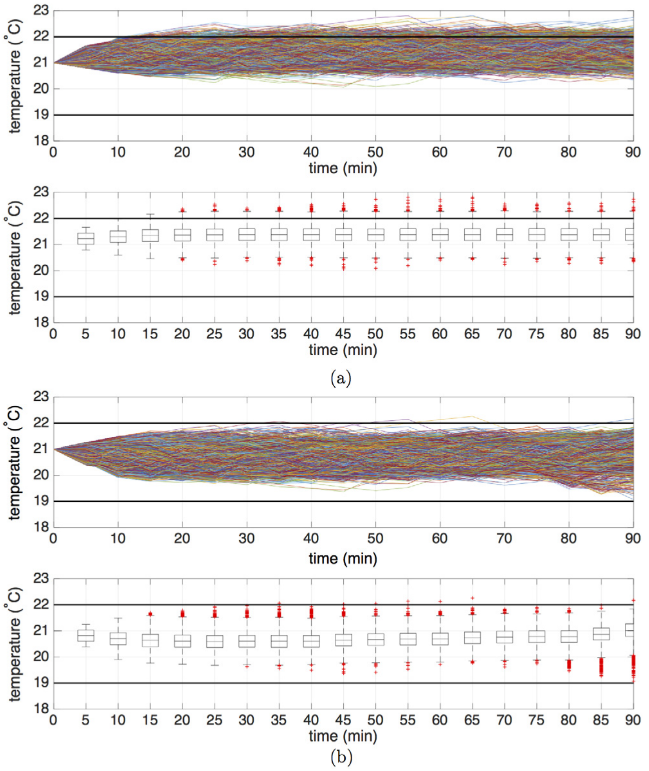

Fig. 2. State trajectories and their Tukey box plot generated by the safety-oriented controller obtained using (a) standard safe sets and (b) distributionally robust safe sets.

5.2. Safety-oriented distributionally robust control

To demonstrate the performance of the safety-oriented distributionally robust controller proposed in Section 3.2, we compare it to a safety-oriented controller synthesized with standard probabilistic safe sets.[6] When the control action is allowed to be arbitrarily chosen in (4), we choose OFF control input to minimize the energy cost. Suppose that the true disturbance distribution is uniformly distributed over a compact support with mean $\mathbf{m}$, and variance Σ. We consider the situation in which we misestimate the distribution as a truncated normal distribution with the same support, mean $\mathbf{m}$ and variance $\Sigma/2$. Setting $\alpha = 0.95$, we construct the probabilistic safe sets using the method proposed by Abate et al. (2008) with the inaccurately estimated distribution. As shown in Fig. 2(a), the safety-oriented controller obtained using the probabilistic safe sets fails to guarantee the desired probability of safety $\alpha = 0.95$. Specifically, in our numerical experiment with the initial state $\boldsymbol{x} = 21$, 1365 of 10,000 sample trajectories violated the safety constraints. Thus, the probability of the system being safe for all stages is only 0.8635 even though the proposed safety-oriented controller is constructed in a conservative way for safety. However, when the distributionally robust safe sets are employed to construct the safety-oriented controllers, only 5

[6] Both methods have the same computational complexity with respect to the dimension of $\boldsymbol{x}$. However, when solving an associated Bellman equation for each $(t, \boldsymbol{x})$, the optimization variable of the standard method is m-dimensional, while the distributionally robust method has the $(m + 2l + l(l + 1)/2 + 1)$-dimensional optimization variable and involves (discretized or sampled) semi-infinite constraints.

sample trajectories move out from the set $A = [19, 22]$. In other words, the probability of safety is 0.995.[7]

6. Conclusions

We proposed a dynamic game approach to computing distributionally robust safe sets and policies concerning inaccurate probability distribution of disturbances. We identified conditions for the existence and optimality of non-randomized Markov policies. Such a policy leads to a practical design method for distributionally robust safety-oriented controllers. To develop a computationally tractable method, we established a strong duality result that allows us to reformulate the infinite-dimensional minimax problem in the Bellman equation as a semi-infinite program. Through numerical simulations, we demonstrated that the safety-oriented controller constructed using the distributionally robust safe sets and policies can guarantee the desired probability of safety even when the disturbance distribution is inaccurately estimated; meanwhile, the same controller based on standard probabilistic safe sets cannot.

Appendix A. Proof of Theorem 1

We first show the following properties of the dynamic programming operator $\mathbf{T}_t$.

Lemma 1. *Let $\mathbf{v} : \mathbb{R}^n \to [0, 1]$ be an upper semi-continuous function. Then, $\mathbf{T}_t\mathbf{v}$ is upper semi-continuous. Furthermore, there exists a measurable function $\kappa : \mathbb{R}^n \to \mathbb{R}^m$ such that, for all $\boldsymbol{x} \in \mathbb{R}^n$, $\kappa(\boldsymbol{x}) \in \mathbb{U}(\boldsymbol{x})$ and*

$$\mathbf{T}_t\mathbf{v}(\boldsymbol{x}) = \inf_{\boldsymbol{\mu}\in\mathbb{D}_t} \left[\mathbf{1}_A(\boldsymbol{x}) \int_{\mathbb{R}^l} \mathbf{v}(f(\boldsymbol{x}, \kappa(\boldsymbol{x}), \boldsymbol{w}))\mathrm{d}\boldsymbol{\mu}(\boldsymbol{w}) \right].$$

Proof. Define a function $\hat{\mathbf{v}} : \mathbb{R}^n \times \mathbb{R}^m \times \mathcal{P}(\mathbb{R}^l) \to \mathbb{R}$ as $\hat{\mathbf{v}}(\boldsymbol{x}, \boldsymbol{u}, \boldsymbol{\mu}) := \int_{\mathbb{R}^l} \mathbf{v}(f(\boldsymbol{x}, \boldsymbol{u}, \boldsymbol{w}))\mathrm{d}\boldsymbol{\mu}(\boldsymbol{w})$. Since $\mathbf{v}$ is a measurable, upper semi-continuous function with values in [0, 1], there exists a sequence $\{\mathbf{v}^{(k)}\}$ such that $\mathbf{v}^{(k)} \downarrow \mathbf{v}$ pointwise and each $\mathbf{v}^{(k)}$ is a bounded continuous function. Thus, for each k, $\int_{\mathbb{R}^l} \mathbf{v}(f(\boldsymbol{x}, \boldsymbol{u}, \boldsymbol{w}))\mathrm{d}\boldsymbol{\mu}(\boldsymbol{w}) \leq \int_{\mathbb{R}^l} \mathbf{v}^{(k)}(f(\boldsymbol{x}, \boldsymbol{u}, \boldsymbol{w}))\mathrm{d}\boldsymbol{\mu}(\boldsymbol{w})$, and for any $(\boldsymbol{x}^{(j)}, \boldsymbol{u}^{(j)}, \boldsymbol{\mu}^{(j)}) \to (\boldsymbol{x}, \boldsymbol{u}, \boldsymbol{\mu})$,

$$\limsup_{j\to\infty} \int_{\mathbb{R}^l} \mathbf{v}(f(\boldsymbol{x}^{(j)}, \boldsymbol{u}^{(j)}, \boldsymbol{w}))\mathrm{d}\boldsymbol{\mu}^{(j)}(\boldsymbol{w})$$
$$\leq \int_{\mathbb{R}^l} \mathbf{v}_k(f(\boldsymbol{x}, \boldsymbol{u}, \boldsymbol{w}))\mathrm{d}\boldsymbol{\mu}(\boldsymbol{w})$$

due to Assumption 1(i). Letting $k \to \infty$, we conclude that $\hat{\mathbf{v}}$ is upper semi-continuous on $\mathbb{K}_t$. Since $\mathbf{1}_A : \mathbb{R}^n \to \mathbb{R}$ is upper semi-continuous with a compact set A, $(\boldsymbol{x}, \boldsymbol{u}, \boldsymbol{\mu}) \mapsto \mathbf{1}_A(\boldsymbol{x})\hat{\mathbf{v}}(\boldsymbol{x}, \boldsymbol{u}, \boldsymbol{\mu})$ is upper semi-continuous on $\mathbb{K}_t$ as well. Let

$$\varphi(\boldsymbol{x}, \boldsymbol{u}) := \inf_{\boldsymbol{\mu}\in\mathbb{D}_t} \mathbf{1}_A(\boldsymbol{x})\hat{\mathbf{v}}(\boldsymbol{x}, \boldsymbol{u}, \boldsymbol{\mu}).$$

Since $\hat{\mathbf{v}}(\boldsymbol{x}, \boldsymbol{u}, \boldsymbol{\mu})$ has values in [0, 1] for any $(\boldsymbol{x}, \boldsymbol{u}, \boldsymbol{\mu}) \in \mathbb{K}_t$, φ is bounded on $\hat{\mathbb{K}}_t := \{(\boldsymbol{x}, \boldsymbol{u}) \in \mathbb{R}^n \times \mathbb{R}^m : \boldsymbol{u} \in \mathbb{U}(\boldsymbol{x})\}$. We now show that φ is upper semi-continuous on $\mathbb{K}_t$. Choose a sequence $\{(\boldsymbol{x}^{(k)}, \boldsymbol{u}^{(k)})\}_{k=0}^\infty$ on $\hat{\mathbb{K}}_t$ converging to $(\boldsymbol{x}, \boldsymbol{u}) \in \hat{\mathbb{K}}_t$. Also, for any $\boldsymbol{\mu} \in \mathbb{D}_t$, select a sequence $\{\boldsymbol{\mu}^{(k)}\}_{k=0}^\infty$ on $\mathbb{D}_t$ such that $\boldsymbol{\mu}^{(k)} \to \boldsymbol{\mu}$. We notice that

$$\varphi(\boldsymbol{x}^{(k)}, \boldsymbol{u}^{(k)}) \leq \mathbf{1}_A(\boldsymbol{x}^{(k)})\hat{\mathbf{v}}(\boldsymbol{x}^{(k)}, \boldsymbol{u}^{(k)}, \boldsymbol{\mu}^{(k)}),$$

which implies that

$$\limsup_{k\to\infty} \varphi(\boldsymbol{x}^{(k)}, \boldsymbol{u}^{(k)}) \leq \mathbf{1}_A(\boldsymbol{x})\hat{\mathbf{v}}(\boldsymbol{x}, \boldsymbol{u}, \boldsymbol{\mu})$$

for any $\boldsymbol{\mu} \in \mathbb{D}_t$ due to the upper semi-continuity of $(\boldsymbol{x}, \boldsymbol{u}, \boldsymbol{\mu}) \mapsto \mathbf{1}_A(\boldsymbol{x})\hat{\mathbf{v}}(\boldsymbol{x}, \boldsymbol{u}, \boldsymbol{\mu})$ on $\mathbb{K}_t$. Thus,

$$\limsup_{k\to\infty} \varphi(\boldsymbol{x}^{(k)}, \boldsymbol{u}^{(k)}) \leq \varphi(\boldsymbol{x}, \boldsymbol{u})$$

and the upper semi-continuity of φ follows. Recall that

$$\mathbf{T}_t\mathbf{v}(\boldsymbol{x}) = \sup_{\boldsymbol{u}\in\mathbb{U}(\boldsymbol{x})} \varphi(\boldsymbol{x}, \boldsymbol{u}).$$

Since (i) $\mathbb{U}(\boldsymbol{x})$ is compact and upper semi-continuous on $\mathbb{X}$, and (ii) φ is upper semi-continuous and bounded on $\hat{\mathbb{K}}_t$, $\mathbf{T}_t\mathbf{v}$ is upper semi-continuous and bounded on $\mathbb{R}^n$, and there exists a measurable function $\kappa : \mathbb{R}^n \to \mathbb{R}^m$ such that $\mathbf{T}_t\mathbf{v}(\boldsymbol{x}) = \varphi(\boldsymbol{x}, \kappa(\boldsymbol{x}))$ and $\kappa(\boldsymbol{x}) \in \mathbb{U}(\boldsymbol{x})$ for each $\boldsymbol{x} \in \mathbb{R}^n$ (Schäl, 1975). □

To prove Theorem 1, we first show that v_t is an upper semi-continuous function with $v_t(\boldsymbol{x}) \in [0, 1]$ $\forall \boldsymbol{x} \in \mathbb{R}^n$ for each t by mathematical induction. For $t = T$, $v_T = \mathbf{1}_A$, which is upper semi-continuous because A is closed. Furthermore, its co-domain is [0, 1]. Suppose now that v_{t+1} is an upper semi-continuous function with $v_{t+1}(\boldsymbol{x}) \in [0, 1]$ for all $\boldsymbol{x} \in \mathbb{R}^n$. Then, by Lemma 1, $v_t = \mathbf{T}_t v_{t+1}$ is upper semi-continuous. Furthermore, $0 \leq v_t(\boldsymbol{x}) \leq v_{t+1}(\boldsymbol{x}) \leq 1$ $\forall \boldsymbol{x} \in \mathbb{R}^n$. This completes our inductive argument.

We now use Lemma 1 to conclude that there exists a measurable function $\pi_t^\star : \mathbb{R}^n \to \mathbb{R}^m$ such that $v_t(\boldsymbol{x}) = \inf_{\boldsymbol{\mu}\in\mathbb{D}_t}[\mathbf{1}_A(\boldsymbol{x}) \int_{\mathbb{R}^l} v_{t+1}(f(\boldsymbol{x}, \pi_t^\star(\boldsymbol{x}), \boldsymbol{w}))\mathrm{d}\boldsymbol{\mu}(\boldsymbol{w})]$ and $\pi_t^\star(\boldsymbol{x}) \in \mathbb{U}(\boldsymbol{x})$ for all $(t, \boldsymbol{x}) \in \mathcal{T} \times \mathbb{R}^n$. Lastly, by the dynamic programming principle, we obtain that $v_0(\boldsymbol{x}) = \inf_{\gamma\in\Gamma}\mathbb{E}^{\pi^\star,\gamma}[\prod_{t=0}^T \mathbf{1}_A(x_t)]$ with $x_0 = \boldsymbol{x}$. □

Appendix B. Proof of Proposition 2

We use mathematical induction. For stage $t = 0$, the statement is true since x_0 is assumed to be contained in $S_\alpha^\star(A) = S_{\alpha,0}^\star(A)$. Suppose that $x_t \in S_{\alpha,t}^\star(A)$ for some $t \in \mathcal{T}$. Fix an arbitrary $t \in \mathcal{T}$. Assume first that $x_t \in \{\boldsymbol{x} \mid f(\boldsymbol{x}, \boldsymbol{u}, \boldsymbol{w}) \in S_{\alpha,t+1}^\star(A)\ \forall(\boldsymbol{u}, \boldsymbol{w}) \in \mathbb{U}(\boldsymbol{x}) \times \mathbb{W}_t\}$. Fix an arbitrary $u_t^\star \in \mathbb{U}(x_t)$. The controller (4) guarantees that $x_{t+1} = f(x_t, u_t, w_t) \in S_{\alpha,t+1}^\star(A)$ with probability 1 for any $\mu_t \in \mathbb{D}_t$ because $\mu_t(\mathbb{W}_t) = \int_{\mathbb{W}_t} \mathrm{d}\mu_t(\boldsymbol{w}) = 1$. On the other hand, when $x_t \notin \{\boldsymbol{x} \mid f(\boldsymbol{x}, \boldsymbol{u}, \boldsymbol{w}) \in S_{\alpha,t+1}^\star(A)\ \forall(\boldsymbol{u}, \boldsymbol{w}) \in \mathbb{U}(\boldsymbol{x}) \times \mathbb{W}_t\}$, by using a distributionally robust safe policy $\pi_t^\star$, we can ensure that $x_{t+1} \in S_{\alpha,t+1}^\star(A)$ since for all $\gamma \in \Gamma$

$$\mathbb{P}^{\pi^\star,\gamma}(x_s \in A\ \forall s \in \bar{\mathcal{T}}_{t+1}) \geq \mathbb{P}^{\pi^\star,\gamma}(x_s \in A\ \forall s \in \bar{\mathcal{T}}_t) \geq \alpha,$$

where $\bar{\mathcal{T}}_t := \{t, t+1, \ldots, T\}$. This completes our inductive argument. □

Appendix C. Proof of Theorem 2

Fix an arbitrary $(t, \boldsymbol{x}, \boldsymbol{u}) \in \mathcal{T} \times \mathbb{R}^n \times \mathbb{R}^m$ such that $\boldsymbol{u} \in \mathbb{U}(\boldsymbol{x})$. With the ambiguity set (5), the inner minimization problem in the Bellman equation (3) can be written as the following infinite-dimensional conic linear program:

$$\begin{aligned}
\mathbf{P}: \quad & \inf_{\boldsymbol{\mu}\in M_+(\mathbb{W}_t)} \int_{\mathbb{W}_t} v_{t+1}(f(\boldsymbol{x}, \boldsymbol{u}, \boldsymbol{w}))\mathrm{d}\boldsymbol{\mu}(\boldsymbol{w}) \\
& \text{s.t. } \mathbf{m}_t - b_t \leq \int_{\mathbb{W}_t} \boldsymbol{w}\mathrm{d}\boldsymbol{\mu}(\boldsymbol{w}) \leq \mathbf{m}_t + b_t \\
& \int_{\mathbb{W}_t} (\boldsymbol{w} - \mathbf{m}_t)(\boldsymbol{w} - \mathbf{m}_t)^\top \mathrm{d}\boldsymbol{\mu}(\boldsymbol{w}) \preceq c_t \Sigma_t \\
& \int_{\mathbb{W}_t} \mathrm{d}\boldsymbol{\mu}(\boldsymbol{w}) = 1,
\end{aligned}$$

where $M_+(\mathbb{W}_t)$ is the set of finite measures on $\mathbb{W}_t$ with nonnegative values. Let $\underline{\lambda}, \overline{\lambda} \in \mathbb{R}_+^l$, $\Lambda \in \mathbb{S}_+^l$ and $\nu \in \mathbb{R}$ be the Lagrange multipliers associated with the constraints. Its dual problem can

[7] The resulting probability of safety is significantly greater than the threshold 0.95 because the uniform distribution is not the worst-possible distribution.

then be derived as

$$\begin{aligned} \mathbf{P}^* : \sup_{\underline{\lambda},\overline{\lambda},\Lambda,\nu} \quad & -\underline{b}_t^\top \underline{\lambda} - \overline{b}_t^\top \overline{\lambda} - c_t \mathrm{Tr}(\Sigma_t \Lambda) - \nu \\ \text{s.t. } & \boldsymbol{w}^\top(\overline{\lambda} - \underline{\lambda}) + (\boldsymbol{w} - \mathbf{m}_t)^\top \Lambda (\boldsymbol{w} - \mathbf{m}_t) + \nu \\ & + v_{t+1}(f(\boldsymbol{x}, \boldsymbol{u}, \boldsymbol{w})) \geq 0 \quad \forall \boldsymbol{w} \in \mathbb{W}_t \\ & \underline{\lambda}, \overline{\lambda} \in \mathbb{R}_+^l, \; \Lambda \in \mathbb{S}_+^l, \; \nu \in \mathbb{R}. \end{aligned}$$

To show that there is no duality gap, we use results of conic duality in infinite-dimensional convex optimization (e.g., Lasserre, 2009; Shapiro, 2001). Specifically, we claim that a generalized Slater condition holds. We first introduce the following convex cone:

$$\begin{aligned} P(\mathbb{W}_t) := \Big\{ & (\underline{\lambda}, \overline{\lambda}, \Lambda, \nu, \lambda_0) \in \mathbb{R}_+^l \times \mathbb{R}_+^l \times \mathbb{S}_+^l \times \mathbb{R} \times \mathbb{R} \mid \\ & \boldsymbol{w}^\top(\overline{\lambda} - \underline{\lambda}) + (\boldsymbol{w} - \mathbf{m}_t)^\top \Lambda (\boldsymbol{w} - \mathbf{m}_t) + \nu \\ & + \lambda_0 v_{t+1}(f(\boldsymbol{x}, \boldsymbol{u}, \boldsymbol{w})) \geq 0 \; \forall \boldsymbol{w} \in \mathbb{W}_t \Big\}. \end{aligned}$$

Due to Theorem 1.2 in Lasserre (2009) or Proposition 3.4 in Shapiro (2001), it suffices to show the generalized Slater condition that there exists $(\underline{\lambda}, \overline{\lambda}, \Lambda, \nu) \in \mathbb{R}^l \times \mathbb{R}^l \times \mathbb{S}^l \times \mathbb{R} \times \mathbb{R}$ such that $(\underline{\lambda}, \overline{\lambda}, \Lambda, \nu, 1) \in \mathrm{Int}\, P(\mathbb{W}_t)$. Fix an arbitrary $\epsilon > 0$. We set $\underline{\lambda}^{\text{feas}} = \overline{\lambda}^{\text{feas}} = \epsilon \mathbf{1}$, $\Lambda^{\text{feas}} = \epsilon I$ and

$$\nu^{\text{feas}} = \rho - \inf_{\boldsymbol{w} \in \mathbb{W}_t, \delta \in [-\epsilon, \epsilon]} (1 + \delta) v_{t+1}(f(\boldsymbol{x}, \boldsymbol{u}, \boldsymbol{w})),$$

where $\mathbf{1}$ is the l-dimensional vector whose entries are all 1's and

$$\rho := \epsilon - \inf_{\boldsymbol{w} \in \mathbb{W}_t, \delta' \in [-2\epsilon, 2\epsilon]^l} \boldsymbol{w}^\top \delta' \geq \epsilon.$$

Note that $\rho < +\infty$ because $\mathbb{W}_t$ is compact, and that $\nu^{\text{feas}} < +\infty$ because the value of v_{t+1} lies in $[0, 1]$. We can check that $(\underline{\lambda}^{\text{feas}}, \overline{\lambda}^{\text{feas}}, \Lambda^{\text{feas}}, \nu^{\text{feas}}, 1) \in P(\mathbb{W}_t)$. Furthermore, any $(\underline{\lambda}, \overline{\lambda}, \Lambda, \nu, \lambda_0)$ that is contained in the ϵ-ball centered at $(\underline{\lambda}^{\text{feas}}, \overline{\lambda}^{\text{feas}}, \Lambda^{\text{feas}}, \nu^{\text{feas}}, 1)$ lies in the cone $P(\mathbb{W}_t)$, because for any $\boldsymbol{w} \in \mathbb{W}_t$

$$\begin{aligned} & \inf_{\underline{\delta}', \overline{\delta}' \in [-\epsilon, \epsilon]^l} \boldsymbol{w}^\top((\overline{\lambda}^{\text{feas}} + \overline{\delta}') - (\underline{\lambda}^{\text{feas}} + \underline{\delta}')) + \\ & \inf_{\delta \in [-\epsilon, \epsilon]} (\nu^{\text{feas}} + \delta) + \inf_{\delta \in [-\epsilon, \epsilon]} (1 + \delta) v_{t+1}(f(\boldsymbol{x}, \boldsymbol{u}, \boldsymbol{w})) \geq 0. \end{aligned}$$

This implies that $(\underline{\lambda}^{\text{feas}}, \overline{\lambda}^{\text{feas}}, \Lambda^{\text{feas}}, \nu^{\text{feas}}, 1)$ is an interior point of $P(\mathbb{W}_t)$. Thus, the generalized Slater condition is satisfied, and thus there is no duality gap. Furthermore, $\mathbf{P}$ has an optimal solution because v_{t+1} in the objective of $\mathbf{P}$ is bounded below by zero. □

References

Abate, A., Prandini, M., Lygeros, J., & Sastry, S. (2008). Probabilistic reachability and safety for controlled discrete time stochastic hybrid systems. *Automatica*, *44*(11), 2724–2734.

Althoff, M., Le Guernic, C., & Krogh, B. H. (2011). Reachable set computation for uncertain time-varying linear systems. In *Hybrid systems: computation and control* (pp. 93–102). Springer.

Ben-Tal, A., Den Hertog, D., De Waegenaere, A., Melenberg, B., & Rennen, G. (2013). Robust solutions of optimization problems affected by uncertain probabilities. *Management Science*, *59*(2), 341–357.

Bertsekas, D. P., & Rhodes, I. B. (1971). On the minmax reachability of target sets and target tubes. *Automatica*, *7*(2), 233–247.

Calafiore, G., & Campi, M. C. (2005). Uncertain convex programs: randomized solutions and confidence levels. *Mathematical Programming, Ser. A*, *102*, 25–46.

Cardaliaguet, P., Quincampoix, M., & Saint-Pierre, P. (1999). Set-valued numerical analysis for optimal control and differential games. In *Stochastic and differential games* (pp. 177–247). Birkhäuser.

Chen, M., Herbert, S. L., Vashishtha, M. S., Bansal, S., & Tomlin, C. J. (2016). A general system decomposition method for computing reachable sets and tubes. arXiv preprint arXiv:1611.00122.

Delage, E., & Ye, Y. (2010). Distributionally robust optimization under moment uncertainty with application to data-driven problems. *Operations Research*, *58*(3), 595–612.

Ding, J., Kamgarpour, M., Summers, S., Abate, A., Lygeros, J., & Tomlin, C. (2013). A stochastic games framework for verification and control of discrete time stochastic hybrid systems. *Automatica*, *49*, 2665–2674.

Dubins, L. E., & Savage, L. J. (1965). *Inequalities for stochastic processes: how to gamble if you must*. McGraw-Hill.

El Ghaoui, L., Oks, M., & Oustry, F. (2003). Worst-case value-at-risk and robust portfolio optimization: a conic programming approach. *Operations Research*, *51*(4), 543–556.

Erdoğan, E., & Iyengar, G. (2006). Ambiguous chance constrained problems and robust optimization. *Mathematical Programming, Ser. B*, *107*, 37–61.

Gao, R., & Kleywegt, A. J. (2016). Distributionally robust stochastic optimization with wasserstein distance. arXiv preprint arXiv:1604.02199.

Ghaemi, R., & Del Vecchio, D. (2014). Control for safety specifications of systems with imperfect information on a partial order. *IEEE Transactions on Automatic Control*, *59*(4), 982–995.

Girard, A. (2005). Reachability of uncertain linear systems using zonotopes. In *International workshop on hybrid systems: computation and control* (pp. 291–305). Springer.

González-Trejo, J. I., Hernández-Lerma, O., & Hoyos-Reyes, L. F. (2003). Minimax control of discrete-time stochastic systems. *SIAM Journal on Control and Optimization*, *41*(5), 1626–1659.

Gordon, G. J. (1995). Stable function approximation in dynamic programming. In: *Proceedings of the 12th international conference on machine learning*.

Hernández-Lerma, O., & Lasserre, J. B. (2012). *Discrete-time Markov control processes: basic optimality criteria*. Springer.

Hettich, R., & Kortanek, K. O. (1993). Semi-infinite programming: theory, methods, and applications. *SIAM Review*, *35*(3), 380–429.

Horst, R., Pardalos, P., & Van Thoai, N. (2000). *Introduction to global optimization*. Springer Science & Business Media.

Hu, J., Prandini, M., & Sastry, S. (2005). Aircraft conflict prediction in the presence of a spatially correlated wind field. *IEEE Trans. Intell. Transp. Syst.*, *6*(3), 326–340.

Jiang, R., & Guan, Y. (2016). Data-driven chance constrained stochastic program. *Mathematical Programming, Ser. A*, *158*, 291–327.

Kurzhanski, A. B., & Varaiya, P. (2002). Reachability analysis for uncertain systems–the ellipsoidal technique. *Dynamics of Continuous Discrete and Impulsive Systems Series B*, *9*(3), 347–367.

Lasserre, J. B. (2001). Global optimization with polynomials and the problem of moments. *SIAM Journal on Optimization*, *11*(3), 796–817.

Lasserre, J. B. (2009). *Moments, positive polynomials and their applications*. World Scientific.

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

López, M., & Still, G. (2007). Semi-infinite programming. *European Journal of Operational Research*, *180*, 491–518.

Lygeros, J., Tomlin, C., & Sastry, S. (1999). Controllers for reachability specifications for hybrid systems. *Automatica*, *35*(3), 349–370.

Margellos, K., & Lygeros, J. (2011). Hamilton–Jacobi formulation for reach–avoid differential games. *IEEE Transactions on Automatic Control*, *56*(8), 1849–1861.

Mitchell, I. M., Bayen, A. M., & Tomlin, C. J. (2005). A time-dependent Hamilton-Jacobi formulation of reachable sets for continuous dynamic games. *IEEE Transactions on Automatic Control*, *50*(7), 947–957.

Mitchell, I. M., & Templeton, J. A. (2005). A toolbox of Hamilton-Jacobi solvers for analysis of nondeterministic continuous and hybrid systems. In *International workshop on hybrid systems: computation and control* (pp. 480–494). Springer.

Mohajerin Esfahani, P., & Kuhn, D. (2017). Data-driven distributionally robust optimization using the wasserstein metric: Performance guarantees and tractable reformulations. *Mathematical Programming, Ser. A*.

Prajna, S., Jadbabaie, A., & Pappas, G. J. (2007). A framework for worst-case and stochastic safety verification using barrier certificates. *IEEE Transactions on Automatic Control*, *52*(8), 1415–1429.

Prandini, M., & Hu, J. (2006). A stochastic approximation method for reachability computations. In *Stochastic hybrid systems* (pp. 107–139). Springer.

Rakovic, S., Kerrigan, E. C., Mayne, D. Q., & Lygeros, J. (2006). Reachability analysis of discrete-time systems with disturbances. *IEEE Transactions on Automatic Control*, *51*(4), 546–561.

Reemtsen, R. (1991). Discretization methods for the solution of semi-infinite programming problems. *Journal of Optimization Theory and Applications*, *71*(1), 85–103.

Schäl, M. (1975). Conditions for optimality in dynamic programming and for the limit of n-stage optimal policies to be optimal. *Probability Theory and Related Fields*, *32*(3), 179–196.

Shapiro, A. (2001). On duality theory of conic linear problems. In *Semi-infinite programming* (pp. 135–165). Springer.

Summers, S., & Lygeros, J. (2010). Verification of discrete time stochastic hybrid systems: a stochastic reach-avoid decision problem. *Automatica*, *46*(12), 1951–1961.

Van Parys, B. P. G., Kuhn, D., Goulart, P. J., & Morari, M. (2016). Distributionally robust control of constrained stochastic systems. *IEEE Transactions on Automatic Control*, *61*(2), 430–442.

Wiesemann, W., Kuhn, D., & Sim, M. (2014). Distributionally robust convex optimization. *Operations Research*, *62*(6), 1358–1376.

Xu, H., & Mannor, S. (2012). Distributionally robust Markov decision processes. *Mathematics of Operations Research*, *37*(2), 288–300.

Yang, I. (2017a). A convex optimization approach to distributionally robust Markov decision processes with Wasserstein distance. *IEEE Control Systems Letters*, *1*(1), 164–169.

Yang, I. (2017b). Distributionally robust stochastic control with conic confidence sets. In *Proceedings of the 56th IEEE conference on decision and control*.

Yang, I. (2017c). A dynamic game approach to distributionally robust safety specifications for stochastic systems. arXiv:1701.06260.

Yu, P., & Xu, H. (2016). Distributionally robust counterpart in Markov decision processes. *IEEE Transactions on Automatic Control*, *61*(9), 2538–2543.

Insoon Yang is an Assistant Professor in the ECE department at Seoul National University (SNU) and in the EE department at USC. He received B.S. degrees in Mathematics and in Mechanical Engineering (summa cum laude) from SNU in 2009; and an M.S. in EECS, an M.A. in Mathematics and a Ph.D. in EECS from UC Berkeley in 2012, 2013 and 2015, respectively. He was a Postdoctoral Associate at the Laboratory for Information and Decision Systems in MIT from 2015 to 2016. Insoon's research interests are in stochastic control and optimization with application to cyber–physical systems and safe autonomy. He is a recipient of the 2015 Eli Jury Award and a finalist for the Best Student Paper Award at the 55th IEEE Conference on Decision and Control 2016.