A Dynamic Game Approach to Distributionally Robust Safety Specifications for Stochastic Systems

Insoon Yang*

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. This control method can be used to minimize another cost function while ensuring safety in a probabilistic way. However, an associated Bellman equation involves infinite-dimensional minimax optimization problems since the disturbance distribution may have a continuous density. To resolve 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 specification tool is not.

Key words. Distributionally robust optimization, safety specifications, reachability analysis, distributional ambiguity, stochastic systems, dynamic games, duality.

1 Introduction

Various critical decision-making and control problems associated with engineering and socio-technical systems are subject to uncertainties. Large-scale data collected from the Internet of Things and cyber-physical systems can provide information about the probability distribution of these uncertainties. Statistical learning and filtering methods also support the construction of data-driven distribution models of uncertainty based on the observed data. Such distributional information can be used to dramatically improve the performance of closed-loop systems if they adopt appropriate controllers, that reduce the conservativeness of classical techniques, such as robust control. Several concerns have been raised about how best to incorporate the collected data into critical control and decision-making problems. These concerns center on safety, risk, robustness and reliability because the data and statistical models extracted from the data often result in inaccurate distributional information. Among them, we focus on the safety issue and develop a safety specification and management tool with ambiguous information about the probability distribution of disturbances.

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

*Department of Electrical Engineering, University of Southern California, Los Angeles, CA 90089, USA (insoonya@usc.edu).

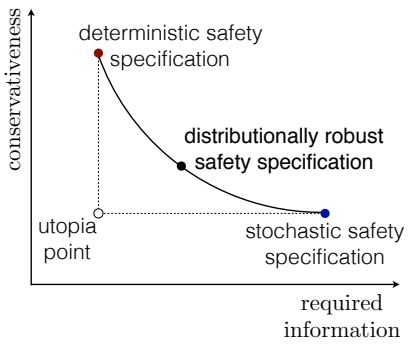

Figure 1: Tradeoff between required information and conservativeness.

verify that a system is evolving within a safe range of operation and to synthesize controllers to satisfy safety constraints (e.g., [5], [47], [28], [7], [23], [36], [17], [40], [2], [29], [16], [8]). 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 the probability distribution of 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 not available 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 in various systems can now provide data to help estimate the probability distribution of disturbances. Stochastic reachability analysis tools are based on the assumption that the full probability distribution of disturbances is available and can often be used to reduce the conservativeness of deterministic safe set computations. 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 with computational methods is often unreliable as it is subject to the quality of the observations, statistical learning or filtering methods, and prior knowledge about the disturbances. Thus, probabilistic safety specification tools can be misleading and lead to the design of an unreliable controller that may violate safety constraints when distributional information about disturbances is inaccurate.

Fig. 1 illustrates the tradeoff between required information and conservativeness in deterministic and stochastic safety specification methods. This study 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. If, for example, only the support, mean, and variance of an uncertain variable are reliably estimated, the ambiguity set can be chosen so as to contain distributions consistent with the empirical estimates. The proposed *distrbutionally 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 the ambiguity set.

This paper proposes a dynamic game formulation of constructing distributionally robust safe policies and safe sets. 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 adversarially 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, the Bellman equation for this dynamic game problem involves infinite-dimensional optimization problems that are computationally challenging. Furthermore, the existence of a distributionally robust safe policy is not guaranteed.

The contributions of this work are threefold. First, we characterize conditions under which a non-randomized Markov policy is optimal for Player I (controller). 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. Furthermore, the existence of a non-randomized Markov policy guarantees that the outer maximization problem (for Player I) of an associated Bellman equation is solvable—that is, it is feasible and has an optimal solution. Second, we develop a design approach to a safety-oriented stochastic 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, regardless of how the disturbance distribution is chosen in an ambiguity set. Third, we propose a duality-based reformulation method for the Bellman equation in cases with moment uncertainty. We show that there is no duality gap in the inner minimization problem (for Player II) of the Bellman equation, which is an infinite-dimensional optimization problem. Using the strong duality result, we reformulate each infinite-dimensional minimax problem in 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 through the use of existing convergent algorithms for semi-infinite programs.

1.1 Related Work

A probabilistic reachability tool for stochastic differential equations with jumps has been proposed; it uses a Markov chain approximation to propagate the transition probabilities of the Markov chain backward in time starting from a target set [21], [38], [39]. In [37], barrier certificates are employed to calculate an upper bound of the probability that a system will reach a target set. Additionally, [31] proposes a toolbox that supports expectation-based reachability problems associated with a class of continuous-time stochastic (hybrid) systems by extending the celebrated Hamilton–Jacobi–Isaacs reachability analysis [48], [30]. A partial differential equation characterization of continuous-time stochastic reach-avoid problems is studied in [32] based on the theory of discontinuous viscosity solutions. For discrete-time stochastic hybrid systems, an elegant dynamic programming approach has been proposed to compute the maximal probability of safety [1]. This method has been extended to stochastic reach–avoid problems [45], stochastic hybrid games [10], and partially observable stochastic hybrid systems [25], [26]. However, all the aforementioned methods are based on the possibly restrictive assumption that the probability distribution of disturbances is completely known.

This work also closely relates to *distributionally robust control*, which is an emerging stochastic control method. This method is based on single-stage distributionally robust stochastic optimization that minimizes the worst-case cost, assuming that the probability distribution of uncertain variables lies within an ambiguity set of distributions (e.g., [43], [12], [13], [6], [9], [50]). For multistage problems, a distributionally robust Markov decision process (MDP) formulation has recently been developed while focusing on finite-state, finite-action MDPs [51], [53]. For cases with moment uncertainty, [49] investigates linear feedback strategies in linear-quadratic settings with risk constraints and proposes a semidefinite programming approach. We extend the theory of distributionally robust control to the case of continuous state spaces and apply it to reachability analysis and safety specifications.

1.2 Organization

The remainder of this paper is organized as follows. In Section 2, we define distributionally robust safe sets and policies and introduce a dynamic game formulation to construct them. Section 3 contains a dynamic programming solution to the problem of computing optimal policies. In particular, we characterize conditions under which a Markov control policy is optimal, and propose a safety-oriented controller design method. In Section 4, we consider ambiguity sets with moment uncertainty and show that an associated Bellman equation can be formulated as a semi-infinite program. An application of the proposed safety specification tool is illustrated through examples in Section 5.

1.3 Notation

Given a Borel space X, $\mathcal{B}(X)$ represents its Borel σ-algebra. Given Borel spaces X and Y, $Q(A|y)$ denotes a transition probability (or a stochastic kernel) from Y to X, where $Q(\cdot|y)$ is a probability measure on $\mathcal{B}(X)$ for each $y \in Y$ and $Q(A|\cdot)$ is a measurable function on Y for each $A \in \mathcal{B}(X)$. Also, $M(X)$ denotes the Banach space of finite signed measures on $\mathcal{B}(X)$ and $M_+(X)$ represents its positive cone. For simplicity, we use the following notation of time indexes: $\mathcal{T} := \{0, 1, \cdots, T-1\}$ and $\bar{\mathcal{T}} := \{0, 1, \cdots, T\}$.

2 Distributionally Robust Safe Sets and Policies

2.1 Stochastic Systems with Distributional Ambiguity

Consider a discrete-time stochastic system of the form

$$x_{t+1} = f(x_t, u_t, w_t) \quad t \in \mathcal{T}, \quad x_0 = \boldsymbol{x}, \tag{2.1}$$

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. In many cases, only estimates of its support, mean, and variance are available. Furthermore, such estimates are rarely accurate. This issue of imperfect disturbance distributions degrades the practicality of classical stochastic reachability tools. To resolve this problem, we allow errors in such distributional information and quantify the probability of safety with the worst-case disturbance distribution. To mathematically model distributional ambiguity, we assume that the true probability distribution μ_t of w_t is contained in a so-called *ambiguity set* of distributions, denoted by $\mathbb{D}_t$. Note that $\mathbb{D}_t$ is a subset of the space $M_+(\mathbb{R}^l)$ of signed measures and thus is infinite dimensional. An example of such ambiguity sets can be found in Section 4.1. We assume that the set $\mathbb{D}_t$ of distributions is not empty for each t.

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 $h_t = (x_0, u_0, \mu_0, \cdots, x_{t-1}, u_{t-1}, \mu_{t-1}, x_t)$. The set of admissible control strategies is chosen as $\Pi := \{\pi = (\pi_0, \cdots, \pi_{T-1}) \mid \pi_t(\mathbb{U}(x_t)|h_t) = 1 \ \forall h_t \in H_t\}$, 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. We assume that there exists a measurable function $\pi : \mathbb{R}^n \to \mathbb{R}^m$ such that $\pi(\boldsymbol{x}) \in \mathbb{U}(\boldsymbol{x})$ for all $\boldsymbol{x} \in \mathbb{R}^n$. By viewing the disturbance as an adversarial player who chooses the disturbance's probability distribution based on the available information, the set of admissible

disturbance distribution strategies is similarly defined as $\Gamma := \{\gamma = (\gamma_0, \cdots, \gamma_{T-1}) | \gamma_t(\mathbb{D}_t | h_t^e) = 1 \ \forall h_t^e \in H_t^e\}$, where H_t^e is an extended set of histories up to stage t, whose element is of the form $h_t^e = (x_0, u_0, \mu_0, \cdots, x_{t-1}, u_{t-1}, \mu_{t-1}, x_t, u_t)$. Note that the distributional constraints in the ambiguity set $\mathbb{D}_t$ is encoded in this strategy space. The disturbance (or the adversarial player) can use slightly more information than the controller; the disturbance is aware of the controller's action at stage t, u_t, in addition to the history h_t. However, the controller cannot be aware of the disturbance distribution's realization μ_t when making a decision at stage t.

2.2 Distributionally Robust Safety Specifications

Our goal is to develop a probabilistic safety specification tool that is robust against errors in the probability distribution of the disturbance $\{w_t\}$. Specifically, we 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, $\mathbb{D}_t$. The proposed *distributionally robust* safety specification tool will be used to design a controller for safety-critical stochastic systems under imperfect information about disturbance distributions.

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 definition of a probabilistic safe set:

Definition 1 (Probabilistic Safe Set)**.** *We define the probability that the system* (2.1) *is safe for all $t \in \bar{\mathcal{T}}$ given the strategy pair (π, γ) and the initial value $\boldsymbol{x}$ as*

$$P_{\boldsymbol{x}}^{\text{safe}}(\pi, \gamma; A) := \mathbb{P}^{\pi,\gamma}\{x_t \in A \ \ \forall t \in \bar{\mathcal{T}}, \ \ x_0 = \boldsymbol{x}\}, \tag{2.2}$$

which we call the probability of safety *for the set A. We also define the* probabilistic safe set *with probability α under (π, γ) as the set*

$$S_\alpha(\pi, \gamma; A) := \{\boldsymbol{x} \in \mathbb{R}^n | P_{\boldsymbol{x}}^{\text{safe}}(\pi, \gamma; A) \geq \alpha\}.$$

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. [1] to the case 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*

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

In other words, 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 distributionally robust safe policy $\pi^\star$. Once we obtain $\pi^\star$ and $P_{\boldsymbol{x}}^{\text{safe}}$ for each $\boldsymbol{x}$, we can calculate the distributionally robust safe set $S_\alpha^\star(A)$ through simple thresholding. To this end, in the next subsection, we consider a game theoretic formulation to construct a distributionally robust safe policy.

2.3 A Dynamic Game Formulation

Let $\mathbf{1}_A : \mathbb{R}^n \to \{0,1\}$ be the indicator function of the set $A \subseteq \mathbb{R}^n$ such that $\mathbf{1}_A(\boldsymbol{x}) = 1$ if $\boldsymbol{x} \in A$ and $\mathbf{1}_A(\boldsymbol{x}) = 0$ otherwise. Then, given $x_0 = \boldsymbol{x}$, we have

$$P_{\boldsymbol{x}}^{\text{safe}}(\pi, \gamma; A) = \mathbb{E}^{\pi,\gamma}\left[\prod_{t=0}^{T} \mathbf{1}_A(x_t)\right], \tag{2.3}$$

where $\mathbb{E}^{\pi,\gamma}$ 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 zero-sum dynamic game problem:

$$\sup_{\pi \in \Pi} \inf_{\gamma \in \Gamma} P_{\boldsymbol{x}}^{\text{safe}}(\pi, \gamma; A). \tag{2.4}$$

In this two-player 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. Recall that information about the ambiguity set $\mathbb{D}_t$ of probability distributions is encoded in Player II's strategy space Γ. In general, the action space of Player II at each stage is infinite-dimensional because the disturbance may have a continuous probability density. Therefore, the Bellman equation associated with this dynamic game problem involves infinite-dimensional optimization problems, which are computationally challenging. To alleviate this computational issue, we propose a duality-based approach to reformulate the Bellman equation as a semi-infinite program that can be solved by convergent algorithms. In the next section, we first establish some analytical results about the dynamic game problem. In particular, we show that under mild conditions an associated value function is upper semi-continuous and a non-randomized Markov control policy is optimal.

3 Dynamic Programming

3.1 A Semi-Continuous Model

We let $\mathbb{K}_t \in \mathcal{B}(\mathbb{R}^n \times \mathbb{R}^m \times M_+(\mathbb{R}^l))$ 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) $\boldsymbol{\mu} \in \mathbb{D}_t$. The stochastic kernel $Q_t(\boldsymbol{\xi}|\boldsymbol{x}, \boldsymbol{u}, \boldsymbol{\mu})$ represents the probability that the state of the system (2.1) at stage $t+1$ is equal to $\boldsymbol{\xi} \in \mathbb{R}^n$ given $(x_t, u_t, \mu_t) = (\boldsymbol{x}, \boldsymbol{u}, \boldsymbol{\mu})$.

Assumption 1. *The problem* (2.4) *of constructing a distributionally robust safe policy satisfies the following conditions:*

(*i*) *For each bounded continuous function $g_t : \mathbb{R}^n \to \mathbb{R}$, the function*

$$\hat{g}_t(\boldsymbol{x}, \boldsymbol{u}, \boldsymbol{\mu}) := \int_{\mathbb{R}^n} g_t(\xi) Q_t(d\xi|\boldsymbol{x}, \boldsymbol{u}, \boldsymbol{\mu})$$

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 ambiguity set $\mathbb{D}_t$ is σ-compact for each $t \in \mathcal{T}$.*

These conditions are standard *measurable selection conditions* for semi-continuous stochastic control models (e.g., [11], [19], [18]) and will be used to ensure the existence of a distributionally robust safe policy. We will further show that a non-randomized Markov policy is optimal. For each measurable function $\mathbf{v}$ on $\mathbb{R}^n$, we define a dynamic programming operator, denoted by $\mathbf{T}_t$, $t \in \mathcal{T}$, as

$$\mathbf{T}_t\mathbf{v}(\boldsymbol{x}, \boldsymbol{y}) := \sup_{\boldsymbol{u}\in\mathbb{U}(\boldsymbol{x})} \inf_{\boldsymbol{\mu}\in\mathbb{D}_t} \mathbf{1}_A(\boldsymbol{x}) \int_{\mathbb{R}^n} \mathbf{v}(\xi) Q_t(d\xi|\boldsymbol{x}, \boldsymbol{u}, \boldsymbol{\mu}).$$

We can first show the following properties of the dynamic programming operator. In particular, the outer "sup" problem has an optimal solution under a mild condition on $\mathbf{v}$.

Lemma 1. *Let $\mathbf{v} : \mathbb{R}^n \to \mathbb{R}$ be a measurable upper semi-continuous function with $\|\mathbf{v}\|_\infty < \infty$ and $\mathbf{v} \geq 0$. Then,*

(i) $\mathbf{T}_t\mathbf{v}$ is upper semi-continuous.

(ii) 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

$$\mathbf{T}_t\mathbf{v}(\boldsymbol{x}) = \inf_{\boldsymbol{\mu}\in\mathbb{D}_t} \left[\mathbf{1}_A(\boldsymbol{x}) \int_{\mathbb{R}^n} \mathbf{v}(\xi) Q_t(d\xi|\boldsymbol{x}, \kappa(\boldsymbol{x}), \boldsymbol{\mu}) \right].$$

Proof. Define a function $\hat{\mathbf{v}} : \mathbb{R}^n \times \mathbb{R}^m \times M_+(\mathbb{R}^l) \to \mathbb{R}$ as

$$\hat{\mathbf{v}}(\boldsymbol{x}, \boldsymbol{u}, \boldsymbol{\mu}) := \int_{\mathbb{R}^n} \mathbf{v}(\xi) Q_t(d\xi|\boldsymbol{x}, \boldsymbol{u}, \boldsymbol{\mu}).$$

Since $\mathbf{v}$ is a measurable, upper semi-continuous and nonnegative function with finite L^∞-norm, 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 all k, $\int_{\mathbb{R}^n} \mathbf{v}(\xi) Q_t(d\xi|\boldsymbol{x}, \boldsymbol{u}, \boldsymbol{\mu}) \leq \int_{\mathbb{R}^n} \mathbf{v}_k(\xi) Q_t(d\xi|\boldsymbol{x}, \boldsymbol{u}, \boldsymbol{\mu})$, and for any $(\boldsymbol{x}_j, \boldsymbol{u}_j, \boldsymbol{\mu}_j) \to (\boldsymbol{x}, \boldsymbol{u}, \boldsymbol{\mu})$ we have that for all k

$$\limsup_{j\to\infty} \int_{\mathbb{R}^n} \mathbf{v}(\xi) Q_t(d\xi|\boldsymbol{x}_j, \boldsymbol{u}_j, \boldsymbol{\mu}_j) \leq \int_{\mathbb{R}^n} \mathbf{v}_k(\xi) Q_t(d\xi|\boldsymbol{x}, \boldsymbol{u}, \boldsymbol{\mu})$$

due to Assumption 1 (ii). Letting $k \to \infty$, we conclude that $\hat{\mathbf{v}}$ is upper semi-continuous. 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 as well. Furthermore, for all $(\boldsymbol{x}, \boldsymbol{u}, \boldsymbol{\mu}) \in \mathbb{K}_t$,

$$\begin{aligned} |\mathbf{1}_A(\boldsymbol{x})\hat{\mathbf{v}}(\boldsymbol{x}, \boldsymbol{u}, \boldsymbol{\mu})| &= \mathbf{1}_A(\boldsymbol{x}) \int_{\mathbb{R}^n} |\mathbf{v}(\xi)| Q_t(d\xi|\boldsymbol{x}, \boldsymbol{u}, \boldsymbol{\mu}) \\ &\leq \|\mathbf{v}\|_\infty < \infty. \end{aligned}$$

Thus, by Lemma 3.2. (b) in [18] have that $\mathbf{T}_t\mathbf{v}$ is upper semi-continuous and that there exists a measurable function $h : \mathbb{R}^n \to \mathbb{R}^m$ such that $\mathbf{T}_t\mathbf{v}(\boldsymbol{x}) = \inf_{\boldsymbol{\mu}\in\mathbb{D}_t} \left[\mathbf{1}_A(\boldsymbol{x}) \int_{\mathbb{R}^n} \mathbf{v}(\xi) Q_t(d\xi|\boldsymbol{x}, \kappa(\boldsymbol{x}), \boldsymbol{\mu}) \right]$ and $\kappa(\boldsymbol{x}) \in \mathbb{U}(\boldsymbol{x})$ for all $\boldsymbol{x} \in \mathbb{R}^n$ under Assumption 1 (iii) and (iv). □

For each $t \in \mathcal{T}$, we define the value function of the distributionally robust safe control problem (2.4) 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 $v_T(\boldsymbol{x}) = \mathbf{1}_A(\boldsymbol{x})$. By definition, $v_0(x_0) = \sup_{\pi\in\Pi} \inf_{\gamma\in\Gamma} P_{x_0}^{\text{safe}}(\pi, \gamma; A)$ for some initial state x_0. Setting $\mathbf{v} = v_{t+1}$ in Lemma 1, we can show that the distributionally robust safe control problem (2.4) admits a non-randomized Markov policy, which is optimal.

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

$$v_t(\boldsymbol{x}) = \inf_{\boldsymbol{\mu} \in \mathbb{D}_t} \left[\mathbf{1}_A(\boldsymbol{x}) \int_{\mathbb{R}^n} v_{t+1}(\xi) Q_t(d\xi | \boldsymbol{x}, \phi_t(\boldsymbol{x}), \boldsymbol{\mu}) \right]$$

for all $\boldsymbol{x} \in \mathbb{R}^n$. The non-randomized Markov policy $\pi^\star := (\phi_0, \cdots, \phi_{T-1}) \in \Pi$ is a distributionally robust safe policy, i.e.,

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

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

Proof. We first show that v_t is a measurable upper semi-continuous function with $\|v_t\|_\infty < \infty$ for each t by mathematical induction. For $t = T$, $v_T = \mathbf{1}_A$, which is upper semi-continuous because A is closed. Furthermore, it is measurable and has a finite L^∞-norm. Suppose now that v_{t+1} is a measurable upper semi-continuous function with $\|v_{t+1}\|_\infty < \infty$ Then, by Lemma 1 (i), $v_t = \mathbf{T}_t v_{t+1}$ is upper semi-continuous. Furthermore, it is clear that v_t is measurable and $\|v_t\|_\infty \leq \|v_{t+1}\|_\infty < \infty$. This completes our inductive argument.

We now use Lemma 1 (ii) to conclude that there exists a measurable function $\phi_t : \mathbb{R}^n \to \mathbb{R}^m$ such that $v_t(\boldsymbol{x}) = \sup_{\boldsymbol{\mu} \in \mathbb{D}_t} [\mathbf{1}_A(\boldsymbol{x}) \int_{\mathbb{R}^n} v_{t+1}(\xi) Q_t(d\xi | \boldsymbol{x}, \phi_t(\boldsymbol{x}), \boldsymbol{\mu})]$ and $\phi_t(\boldsymbol{x}) \in \mathbb{U}(\boldsymbol{x})$ 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 $x_0 = \boldsymbol{x}$. □

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 $\mathbb{D}_t$ using the history of information. 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: for $t = T$*

$$v_T(\boldsymbol{x}) = \mathbf{1}_A(\boldsymbol{x}),$$

and for $t \in \mathcal{T}$

$$v_t(\boldsymbol{x}) = \max_{\boldsymbol{u} \in \mathbb{U}(\boldsymbol{x})} \inf_{\boldsymbol{\mu} \in \mathbb{D}_t} \mathbf{1}_A(\boldsymbol{x}) \int_{\mathbb{R}^l} v_{t+1}(f(\boldsymbol{x}, \boldsymbol{u}, \boldsymbol{w})) d\boldsymbol{\mu}(\boldsymbol{w}). \tag{3.1}$$

Note that "max" is used instead of "sup" in the outer problem because its optimal solution exists due to Theorem 1. This Bellman equation is computationally challenging because it involves infinite-dimensional minimax optimization problems over the ambiguity set $\mathbb{D}_t$ of probability distributions. In Section 4, we propose a duality-based approach to alleviate the computational issue that arises from the infinite-dimensional minimax programs. Before introducing this method, we discuss how to construct distributionally robust safe sets and controllers through a stochastic reachability analysis in the following subsection.

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 backward in time, i.e., from $t = T$ to $t = 0$. A distribu-

tionally robust safe policy can then be constructed as

$$\phi_t^{\text{safe}}(x_t) \in \arg\max_{\boldsymbol{u}\in\mathbb{U}(x_t)} \inf_{\boldsymbol{\mu}\in\mathbb{D}_t} \mathbf{1}_A(x_t) \int_{\mathbb{R}^l} v_{t+1}(f(x_t,\boldsymbol{u},\boldsymbol{w}))d\boldsymbol{\mu}(\boldsymbol{w}).$$

The distributionally robust safe set with probability α can 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 with probability α as

$$S_{\alpha,t}^\star(A) := \{\boldsymbol{x}\in\mathbb{R}^n \mid \forall\gamma\in\Gamma\ \exists\pi\in\Pi \text{ s.t. } \mathbb{P}^{\pi,\gamma}(x_s\in A\ \forall s=t,\cdots,T, x_t=\boldsymbol{x}) \geq \alpha\}.$$

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 equal to α. Note that $S_{\alpha,0}^\star(A) = S_\alpha^\star(A)$ under Assumption 1 which guarantees the existence of a distributionally robust safe policy. Using the value function, we can compute the set as

$$S_{\alpha,t}^\star(A) = \{\boldsymbol{x}\in\mathbb{R}^n \mid v_t(\boldsymbol{x}) \geq \alpha\}$$

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 W_t of μ_t for each t. Consider the following controller: given x_t at stage t, the control action is determined as

$$u_t^{\text{safe}} \begin{cases} \in \mathbb{U}(x_t) & \text{if } x_t \in \{\boldsymbol{x} \mid f(\boldsymbol{x},\boldsymbol{u},\boldsymbol{w}) \in S_{\alpha,t+1}^\star(A)\ \forall\boldsymbol{u}\in\mathbb{U}(\boldsymbol{x})\ \forall\boldsymbol{w}\in W_t\} \\ = \phi_t^{\text{safe}}(x_t) & \text{otherwise.} \end{cases} \tag{3.2}$$

This controller chooses an arbitrary admissible control action if it can drive the system into $S_{\alpha,t+1}^\star(A)$ at stage $t+1$ for any realization of the disturbance. Otherwise, it uses a distributionally robust safe policy. This procedure is motivated by the deterministic safe controller synthesis method of Lygeros et al. [28]. We can show that at each t this controller ensures that the system will remain safe for stages $t+1,\cdots,T$ with probability greater than or equal to α under any disturbance distribution strategy $\gamma\in\Gamma$.

Proposition 2. *Suppose that Assumption 1 holds. If the initial state is chosen so that $x_0\in S_\alpha^\star(A)$ and the Markov control policy* (3.2) *is used, then*

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

i.e., 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 the ambiguity set.

Proof. 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 $x_t\in\{\boldsymbol{x}\mid f(\boldsymbol{x},\boldsymbol{u},\boldsymbol{w})\in S_{\alpha,t+1}^\star(A)\ \forall\boldsymbol{u}\in\mathbb{U}(\boldsymbol{x})\ \forall\boldsymbol{w}\in W_t\}$. Fix an arbitrary $u_t\in\mathbb{U}(x_t)$. The controller 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 $\mu_t(W_t)=\int_{W_t}d\mu_t(\boldsymbol{w})=1$. On the other hand, when $x_t\notin\{\boldsymbol{x}\mid f(\boldsymbol{x},\boldsymbol{u},\boldsymbol{w})\in S_{\alpha,t+1}^\star(A)\ \forall\boldsymbol{u}\in\mathbb{U}(\boldsymbol{x})\ \forall\boldsymbol{w}\in W_t\}$, by using a distributionally robust safe policy ϕ_t^{safe}, we can ensure that $x_{t+1}\in S_{\alpha,t+1}^\star(A)$ since

$$\begin{aligned}\mathbb{P}^{\pi^{\text{safe}},\gamma}(x_s\in A\ \forall s\in\mathcal{T}_{t+1}) &\geq \mathbb{P}^{\pi^{\text{safe}},\gamma}(x_s\in A\ \forall s\in\mathcal{T}_t)\\ &\geq \alpha \quad \forall\gamma\in\Gamma,\end{aligned}$$

where $\pi^{\text{safe}} := \{\phi_0^{\text{safe}},\cdots,\phi_{T-1}^{\text{safe}}\}$ and $\mathcal{T}_t := \{t,t+1,\cdots,T\}$. This completes our inductive argument. □

If another objective function needs to be minimized in a distributionally robust way while ensuring safety, one may solve

$$\inf_{\pi\in\Pi}\sup_{\gamma\in\Gamma}\mathbb{E}^{\pi,\gamma}\left[\sum_{t=0}^{T-1} r(x_t,u_t)+q(x_T)\right]$$

to obtain an optimal distributionally robust policy π^{opt}, where $r:\mathbb{R}^n\times\mathbb{R}^m\to\mathbb{R}$ is a running cost function and $q:\mathbb{R}^n\to\mathbb{R}$ is a terminal cost function of interest. Then, one can employ the proposed controller (3.2) that chooses $\pi_t^{opt}(x_t)$ whenever $x_t\in\{\boldsymbol{x}\mid f(\boldsymbol{x},\boldsymbol{u},\boldsymbol{w})\in S^\star_{\alpha,t+1}(A)\ \forall\boldsymbol{u}\in\mathbb{U}(\boldsymbol{x})\ \forall\boldsymbol{w}\in W_t\}$. 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 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 Dual Bellman Equations

4.1 Ambiguity Sets with Moment Uncertainty

Recall that the proposed distributionally robust safe policies maximize the worst-case probability for a system to be safe, assuming that the probability distribution of the disturbance lies within an ambiguity set, $\mathbb{D}_t$, of distributions. Therefore, modeling the ambiguity set may critically affect the resulting safe policies. Several ambiguity set modeling approaches have been developed in the context of single-stage optimization problems. The approaches can be categorized as *moment-based* and *statistical distance-based* methods. A moment-based approach employs an ambiguity set of distributions whose moments (e.g., mean and covariance) satisfy certain constraints [43], [9], [35], [55], [50]. 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 [3], [4], [22], [46], Prokhorov metric [14] and Wasserstein distance [33], [54], [15].

In this work, we take a moment-based approach. 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 $\mathbf{\Sigma}_t\in\mathbb{R}^{l\times l}$ be the estimate of the mean and covariance matrix, respectively. The set of all the probability distributions (i.e., distribution measures) that are consistent with these estimates can be modeled as

$$\begin{aligned}\mathbb{D}_t := \{&\mu_t\in M_+(W_t)\mid \mu_t(W_t)=1,\\ &|\mathbb{E}_{\mu_t}[w_t]-\mathbf{m}_t|\le b_t,\\ &\mathbb{E}_{\mu_t}[(w_t-\mathbf{m}_t)(w_t-\mathbf{m}_t)^\top]\preceq c_t\mathbf{\Sigma}_t\},\end{aligned}\tag{4.1}$$

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 $c_t\ge 1$ are given constants that depend on one's confidence in the estimates $\mathbf{m}_t$ and $\mathbf{\Sigma}_t$. Any probability distribution in this set satisfies the following properties: (i) the support of w_t is W_t; (ii) the mean of $w_{t,i}$ lies in a circle of size $b_{t,i}$; and (iii) the centered second moment matrix of w_t lies in a positive semidefinite cone. It models how likely w_t is to be close to $\mathbf{m}_t$ in terms of the correlation matrix $c_t\mathbf{\Sigma}_t$ with $c_t\ge 1$.

4.2 Zero Duality Gap

In general, solving the Bellman equation (3.1) to evaluate $v_t(\boldsymbol{x})$ with the ambiguity set (4.1) is challenging as it involves infinite-dimensional minimax optimization problems. To resolve this

difficulty, we propose a dual formulation method.[1] 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 (4.1), the inner minimization problem in the Bellman equation (3.1) can be written as the following infinite-dimensional conic linear program:

$$\begin{aligned}
\mathbf{P}: \inf_{\boldsymbol{\mu}} \quad & \int_{W_t} v_{t+1}(f(\boldsymbol{x}, \boldsymbol{u}, \boldsymbol{w}))d\boldsymbol{\mu}(\boldsymbol{w}) && \text{(4.2a)}\\
\text{s.t.} \quad & \mathbf{m}_t - b_t \leq \int_{W_t} \boldsymbol{w} d\boldsymbol{\mu}(\boldsymbol{w}) \leq \mathbf{m}_t + b_t && \text{(4.2b)}\\
& \int_{W_t} (\boldsymbol{w} - \mathbf{m}_t)(\boldsymbol{w} - \mathbf{m}_t)^\top d\boldsymbol{\mu}(\boldsymbol{w}) \preceq c_t \boldsymbol{\Sigma}_t && \text{(4.2c)}\\
& \int_{W_t} d\boldsymbol{\mu}(\boldsymbol{w}) = 1 && \text{(4.2d)}\\
& \boldsymbol{\mu} \in M_+(W_t). && \text{(4.2e)}
\end{aligned}$$

Here, (4.2b) and (4.2c) represent the first- and second-moment constraints encoded in the ambiguity set $\mathbb{D}_t$, respectively. The constraints (4.2d) and (4.2e) ensure that $\boldsymbol{\mu}$ is a probability distribution measure whose support is $\mathbb{D}_t$. Let $\underline{\boldsymbol{\lambda}}, \overline{\boldsymbol{\lambda}} \in \mathbb{R}^l$ be the Lagrange multipliers associated with the inequality constraints (4.2b). We also let $\boldsymbol{\Lambda} \in \mathbb{S}^l$ and $\boldsymbol{\nu} \in \mathbb{R}$ be the Lagrange multipliers associated with (4.2c) and (4.2d), respectively. Its dual problem can then be derived as

$$\begin{aligned}
\mathbf{P}^*: \sup_{\underline{\boldsymbol{\lambda}}, \overline{\boldsymbol{\lambda}}, \boldsymbol{\Lambda}, \boldsymbol{\nu}} \quad & -\underline{b}_t^\top \underline{\boldsymbol{\lambda}} - \overline{b}_t^\top \overline{\boldsymbol{\lambda}} - c_t \mathrm{Tr}(\boldsymbol{\Sigma}_t \boldsymbol{\Lambda}) - \boldsymbol{\nu}\\
\text{s.t.} \quad & \boldsymbol{w}^\top(\overline{\boldsymbol{\lambda}} - \underline{\boldsymbol{\lambda}}) + (\boldsymbol{w} - \mathbf{m}_t)^\top \boldsymbol{\Lambda}(\boldsymbol{w} - \mathbf{m}_t)\\
& \quad + \boldsymbol{\nu} + v_{t+1}(f(\boldsymbol{x}, \boldsymbol{u}, \boldsymbol{w})) \geq 0 \; \forall \boldsymbol{w} \in W_t\\
& \underline{\boldsymbol{\lambda}}, \overline{\boldsymbol{\lambda}} \geq 0, \; \boldsymbol{\Lambda} \succeq 0\\
& \underline{\boldsymbol{\lambda}}, \overline{\boldsymbol{\lambda}} \in \mathbb{R}^l, \; \boldsymbol{\Lambda} \in \mathbb{S}^l, \; \boldsymbol{\nu} \in \mathbb{R},
\end{aligned} \tag{4.3}$$

where $\underline{b}_t := b_t - \mathbf{m}_t$ and $\overline{b}_t := b_t + \mathbf{m}_t$. This is a semi-infinite program because the first constraint must be satisfied for all $\boldsymbol{w} \in W_t \subseteq \mathbb{R}^l$. Let $\inf \mathbf{P}$ and $\sup \mathbf{P}^*$ denote the optimal values of the primal and dual problems, respectively. By weak duality, we have $\sup \mathbf{P}^* \leq \inf \mathbf{P}$. However, we can further show that strong duality holds, i.e., the dual program is exact in the sense that the duality gap is zero.

Proposition 3 (Zero duality gap)**.** *Suppose that Assumption 1 holds,* $\mathbf{P}$ *has a feasible solution, and* W_t *is compact. Then,* $\mathbf{P}$ *has an optimal solution and there is no duality gap, i.e.,*

$$\sup \mathbf{P}^* = \min \mathbf{P}.$$

Proof. Note that $\mathbf{P}$ has a feasible solution with finite value since the objective function value lies in $[0, 1]$. We introduce the following convex cone:

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

Fix an arbitrary $\epsilon > 0$. Choose $(\underline{\boldsymbol{\lambda}}^{\text{feas}}, \overline{\boldsymbol{\lambda}}^{\text{feas}}, \Lambda^{\text{feas}}, \boldsymbol{\nu}^{\text{feas}}, 1) \in P(W_t)$ such that $\underline{\boldsymbol{\lambda}}^{\text{feas}} = \overline{\boldsymbol{\lambda}}^{\text{feas}} = 0$, $\boldsymbol{\Lambda}^{\text{feas}} = \epsilon I$ and

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

[1] 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 system state $\boldsymbol{x}$ even if the proposed approach is employed.

where

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

Note that $\rho \in \mathbb{R}$ because W_t is compact, and that $\boldsymbol{\nu}^{\text{feas}} \in \mathbb{R}$ because the value of v_{t+1} is in $[0,1]$. Therefore, any $(\underline{\boldsymbol{\lambda}}, \overline{\boldsymbol{\lambda}}, \Lambda, \boldsymbol{\nu}, \lambda_0)$ that is contained in the ϵ-ball centered at $(\underline{\boldsymbol{\lambda}}^{\text{feas}}, \overline{\boldsymbol{\lambda}}^{\text{feas}}, \Lambda^{\text{feas}}, \boldsymbol{\nu}^{\text{feas}}, 1)$ lies in the cone $P(W_t)$. This implies that $(\underline{\boldsymbol{\lambda}}^{\text{feas}}, \overline{\boldsymbol{\lambda}}^{\text{feas}}, \Lambda^{\text{feas}}, \boldsymbol{\nu}^{\text{feas}}, 1)$ is an interior point of $P(W_t)$. Thus, due to Theorem 1.2 in [24], there is no duality gap and the inf in $\mathbf{P}$ is attained.

□

This proof ensures that a Slater type condition holds, and the rest follows from results of conic duality in infinite-dimensional convex optimization (see also [44]).

4.3 Dual Bellman Equation as a Semi-Infinite Program

Using the zero duality gap result, we can substitute the inner minimization problem ($\inf \mathbf{P}$) in the Bellman equation as its dual problem ($\sup \mathbf{P}^*$) without sacrificing optimality. This substitution allows us to evaluate the value function $v_t(\boldsymbol{x})$ via a dual version of the Bellman equation.

Theorem 2 (Dual Bellman equation)**.** *Suppose that Assumption 1 holds,* $\mathbf{P}$ *has a feasible solution, and* W_t *is compact. For all* $(t, \boldsymbol{x}) \in \mathcal{T} \times \mathbb{R}^n$*, the Bellman equation* (3.1) *is equivalent to the following semi-infinite program:*

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

with the terminal condition $v_T(\boldsymbol{x}) = \mathbf{1}_A(\boldsymbol{x})$*.*

Remark 1. *For a compact representation, we merged* "$\max_{\boldsymbol{u}}$" *and* "$\sup_{\underline{\boldsymbol{\lambda}},\overline{\boldsymbol{\lambda}},\boldsymbol{\Lambda},\boldsymbol{\nu}}$"*. However, it should be noted that this semi-infinite program has an optimal feasible* $\boldsymbol{u}$*.*

Since the dual problem ($\sup \mathbf{P}^*$) is a semi-infinite program, the dual Bellman equation also involves semi-infinite optimization problems. Each semi-infinite optimization program can be solved by using several convergent methods, such as discretization methods, exchange methods, homotopy methods, and primal-dual methods (e.g., [20], [42], [27] and the reference therein). In Section 5, we employ the discretization method proposed by Reemtsen [41]. This algorithm adaptively generates a grid on W_t and converges to a locally optimal value of the semi-infinite program in the dual Bellman equation (see [41] for a proof). When the semi-infinite program is concave, it converges to the globally optimal value. We can show that each semi-infinite program is concave under the next assumption.

Assumption 2. *The following properties hold:*

(*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;*

(*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* $\lambda \boldsymbol{u}_1 + (1-\lambda)\boldsymbol{u}_2 \in \mathbb{U}(\lambda \boldsymbol{x}_1 + (1-\lambda)\boldsymbol{x}_2)$.

When $\mathbb{U}$ is independent of $\boldsymbol{x}$, Assumption 2 (*iii*) is equivalent to the concavity of $\mathbb{U}$. Note that this assumption is independent of the ambiguity set $\mathbb{D}_t$. Thus, the following concavity result holds with general ambiguity sets:

Proposition 4. *Suppose that Assumptions 1 and 2 hold. Then, the value function* $v_t(\boldsymbol{x})$ *is concave with respect to* $\boldsymbol{x} \in A$ *for each* $t \in \mathcal{T}$.

Its proof is contained in the Appendix. Since $\boldsymbol{u} \mapsto v_t(f(\boldsymbol{x}, \boldsymbol{u}, \boldsymbol{w}))$ is concave for each $(t, \boldsymbol{x}, \boldsymbol{w}) \in \bar{\mathcal{T}} \times A \times W_t$ under Assumption 2, the semi-infinite program in the dual Bellman equation is concave for each $(t, \boldsymbol{x}) \in \mathcal{T} \times A$.

Remark 2. *When control actions are chosen from a discrete set, i.e.,* $\mathbb{U}(\boldsymbol{x})$ *is a discrete set, the semi-infinite program in Theorem 2 can be formulated as a mixed-integer program, where "*$\max_{\boldsymbol{u}}$*" is a discrete-optimization problem and "*$\sup_{\boldsymbol{\lambda}, \bar{\boldsymbol{\lambda}}, \boldsymbol{\Lambda}, \boldsymbol{\nu}}$*" is a continuous optimization problem. In particular, under Assumption 2, we can employ a linear approximation-based method to obtain an approximate solution with a provable suboptimality bound [52].*

5 Numerical Examples: Thermostatically Controlled Loads

Thermostatically controlled loads (TCLs)—such as air conditioners, refrigerators, water heaters, and battery pack cooling systems—are used to guarantee human comfort, and food provision and battery safety, etc. Therefore, ensuring safe TCL operation is critical in a wide range of applications. We consider the following model of the temperature being controlled through a TCL:

$$x_{t+1} = \alpha x_t + (1-\alpha)(\theta - \eta R P u_t) + w_t,$$

which was originally developed by Mortensen and Haggerty [34]. 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 environmental and/or human behavioral uncertainty. Note also that $\alpha = \exp(-h/CR)$, where C is the thermal capacitance (kWh/°C), R is the thermal resistance (°C/kW) and h is the time interval between stages t and $t+1$. In addition, the parameter P represents the range of energy transfer to or from the thermal mass (kW) and the coefficient η is the control efficiency. In our numerical experiments, the following parameters are used: $R = 2$ °C/kW, $C = 2$ kWh/°C, $\theta = 32$ °C, $h = 5/60$ hour, $P = 14$ kW, and $\eta = 0.7$. We compute the distributionally robust safe sets and policies for 90 minutes with 5-minute time intervals between two consecutive stages. The desired safe set of temperature values is chosen as $A = [19, 22]$ (°C).

We now assume that an empirical estimate of the first- and second-moments is the only available information about w_t except its support. Let $\mathbf{m}$ and $\boldsymbol{\Sigma}$ be the empirical mean and variance of w_t. As discussed in Section 4.1, it is reasonable to consider the following constraints: $|\mathbb{E}_{\mu_t}[w_t] - \mathbf{m}| \leq b$, and $\mathbb{E}_{\mu_t}[(w_t - \mathbf{m})^2] \leq c\boldsymbol{\Sigma}$, where $b \geq 0$ and $c \geq 1$ are adjustable parameters depending on one's confidence in the estimate. We call them *confidence parameters*. We set $\mathbf{m} = 0$, $\boldsymbol{\Sigma} = 0.25^2$, and the disturbance's support $\mathcal{K} = [-\frac{1}{2}\sqrt{\boldsymbol{\Sigma}/12}, \frac{1}{2}\sqrt{\boldsymbol{\Sigma}/12}]$.

5.1 Effect of the Confidence Parameters

Fig. 2 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

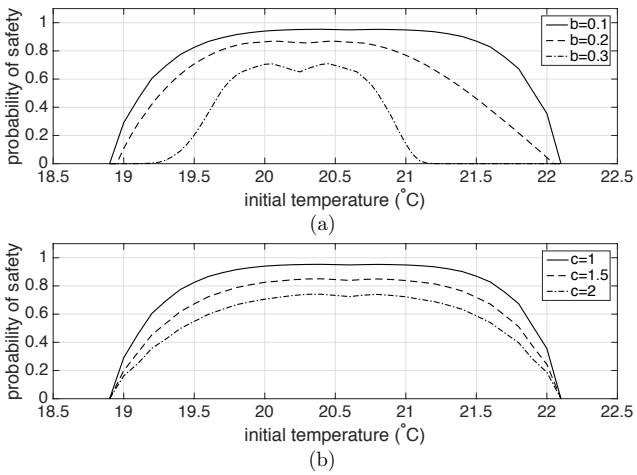

Figure 2: Effect of (a) the confidence parameter b for the mean $\mathbf{m}$ and (b) the confidence parameter c for the variance $\mathbf{\Sigma}$ on the probability of safety.

control inputs: it can be considered as the point-wise maximum of the probability function with OFF control and that with ON control. At initial states between the two peaks, ON or OFF control input at time 0 does not maximize the probability of safety. Aditionally, as the initial state approaches the boundary of the set $A = [19, 22]$, the probability of safety decreases.

As shown in Fig. 2 (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. 2 (b) illustrates that the probability of safety decreases as the uncertainty of the variance estimate $\mathbf{\Sigma}$ increases when $b = 0.1$. The probability $P_{\boldsymbol{x}}^{\text{safe}}$ of safety scales down with the variance confidence parameter c without any change in its support. On the other hand, an increase in the mean confidence parameter b reduces the support of $P_{\boldsymbol{x}}^{\text{safe}}$. This is because a change in b may shift the worst-case disturbance distribution while a change in c can only scale the worst-case distribution.

5.2 Safety-Oriented Distributionally Robust Control

To demonstrate the performance of the proposed safety-oriented distributionally robust controller, we compare it to a safety-oriented controller synthesized with standard probabilistic safe sets. We use the safety-oriented controller design approach proposed in Section 3.2. When the control action is allowed to be arbitrarily chosen in (3.2), we choose OFF control input to minimize the energy cost. In other words, it is an energy cost-minimizing safety-oriented controller. Suppose that the true disturbance distribution is uniformly distributed over the support $\mathcal{K}$ with mean $\mathbf{m}$, and variance $\mathbf{\Sigma}$. We consider the situation in which we misestimate the distribution as a truncated normal distribution with the same support $\mathcal{K}$, mean $\mathbf{m}$ and variance $\mathbf{\Sigma}/2$.[2] We set the probability threshold as $\alpha = 0.95$ and construct the probabilistic safe sets using the method proposed by Abate et al. [1] with the inaccurately estimated distribution. As shown in Fig. 3 (a), the safety-oriented controller obtained using the probabilistic safe sets fails to guarantee that the probability of safety will be greater than or equal to the threshold $\alpha = 0.95$. Specifically, in our numerical experiment with the initial state $\boldsymbol{x} = 21$, 1,365 of 10,000 sample trajectories violated the safety

[2]The variance of the truncated normal distribution must be greater than the variance of the uniform distribution with the same support and mean.

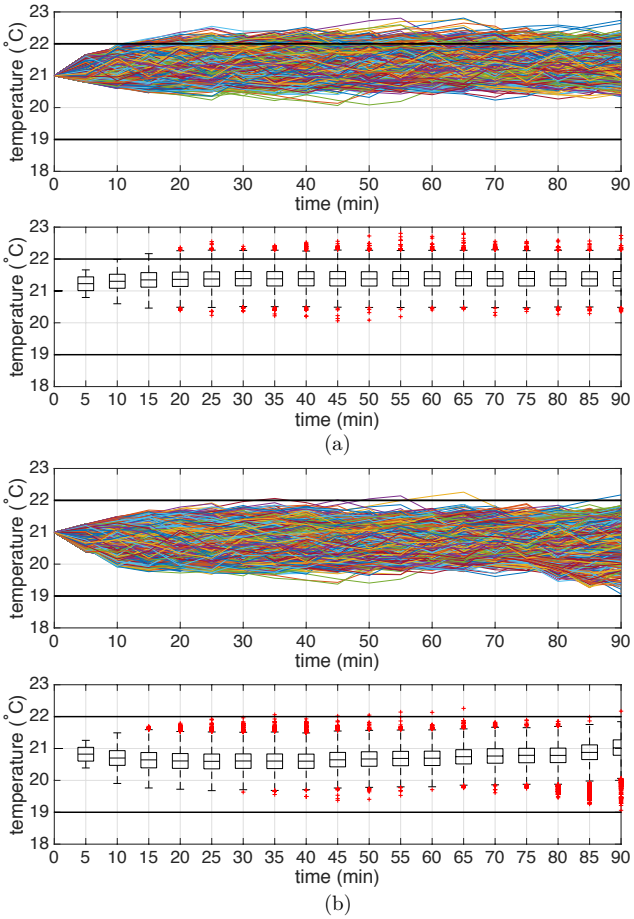

Figure 3: State trajectories and their Tukey box plot generated by the safety-oriented controller obtained using (a) standard probabilistic safe sets and (b) distributionally robust safe sets.

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 sample trajectories move out from the set $A = [19, 22]$. In other words, the probability of safety is 0.995.[3]

6 Conclusions

We proposed a dynamic game approach to computing distributionally robust safe sets and policies concerning ambiguous information about the probability distribution of disturbances. We identified conditions under which a Markov policy is an optimal distributionally robust safe policy. Such a policy leads to a practical design method for safety-oriented stochastic controllers that guarantee that the probability of the system being safe for all remaining stages exceeds a pre-specified threshold, regardless of how the disturbance distribution is chosen in an ambiguity set. We also proved that there is no duality gap in the inner minimization problem of the Bellman equation

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

when ambiguity sets with moment uncertainty are considered. This strong duality result allows us to reformulate the infinite-dimensional minimax problem in the Bellman equation as a semi-infinite program. Since the reformulated dual Bellman equation can be solved using existing convergent algorithms, the proposed dual formulation method alleviates computational issues that otherwise occur when solving the distributionally robust safety specification problem. 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 probability distribution of disturbances is inaccurately estimated; meanwhile, the same controller based on standard probabilistic safe sets cannot.

The following future directions are of interests to generalize and improve the proposed approach. First, this method is readily applicable to backward reachability analysis. Given a target set B, the probability that reaches B for some $t \in \bar{\mathcal{T}}$ given the strategy pair (π, γ) and the initial value $\boldsymbol{x}$ can be expressed as $P_{\boldsymbol{x}}^{\text{reach}}(\pi, \gamma; B) = \mathbb{E}^{\pi,\gamma}[\max_{t\in\bar{\mathcal{T}}} \mathbf{1}_B(x_t)]$. We can then use dynamic programming and the same dual formulation to obtain a control policy that maximizes this probability. Similarly, the proposed distributionally robust approach can be extended to reach-avoid problems. Second, the proposed method is compatible with other types of ambiguity sets. The existence result and safety-oriented controller design approach remain valid when different ambiguity sets are considered. However, different reformulations of associated Bellman equations are required to handle other ambiguity sets such as statistical distance-based ones by using different existing strong duality results. Third, this method presents a scalability issue arising from both dynamic programming and semi-infinite programming. Applying and developing advanced computational techniques including approximate dynamic programming, and sampling- and moment-based approximation methods is of great interests to alleviate the scalability issue.

References

[1] A. Abate, M. Prandini, J. Lygeros, and S. Sastry. Probabilistic reachability and safety for controlled discrete time stochastic hybrid systems. *Automatica*, 44:2724–2734, 2008.

[2] M. Althoff, C. Le Guernic, and B. H. Krogh. Reachable set computation for uncertain time-varying linear systems. In *Hybrid Systems: Computation and Control*, pages 93–102. Springer, 2011.

[3] G. Bayraksan and D. K. Love. Data-driven stochastic programming using phi-divergences. *Tutorials in Operations Research*, pages 1–19, 2015.

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

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

[6] G. C. Calafiore and L. El Ghaoui. On distributionally robust chance-constrained linear programs. *Journal of Optimization Theory and Applications*, 130(1):1–22, 2006.

[7] P. Cardaliaguet, M. Quincampoix, and P. Saint-Pierre. Set-valued numerical analysis for optimal control and differential games. In *Stochastic and Differential Games*, pages 177–247. Birkhäuser, 1999.

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

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

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

[11] L. E. Dubins and L. J. Savage. *Inequalities for Stochastic Processes: How to Gamble If You Must.* McGraw-Hill, 1965.

[12] J. Dupačová. The minimax approach to stochastic programming and an lllustrative application. *Stochastics*, 20:73–88, 1987.

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

[14] E. Erdoğan and G. Iyengar. Ambiguous chance constrained problems and robust optimization. *Mathematical Programming, Ser. B*, 107:37–61, 2006.

[15] R. Gao and A. J. Kleywegt. Distributionally robust stochastic optimization with Wasserstein distance. *arXiv preprint arXiv:1604.02199*, 2016.

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

[17] A. Girard. Reachability of uncertain linear systems using zonotopes. In *International Workshop on Hybrid Systems: Computation and Control*, pages 291–305. Springer, 2005.

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

[19] O. Hernández-Lerma and J. B. Lasserre. *Discrete-Time Markov Control Processes: Basic Optimality Criteria*. Springer, 2012.

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

[21] J. Hu, M. Prandini, and S. Sastry. Aircraft conflict prediction in the presence of a spatially correlated wind field. *IEEE Transactions on Intelligent Transportation Systems*, 6(3):326–340, 2005.

[22] R. Jiang and Y. Guan. Data-driven chance constrained stochastic program. *Mathematical Programming, Ser. A*, 158:291–327, 2016.

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

[24] J. B. Lasserre. *Moments, Positive Polynomials and Their Applications*. World Scientific, 2009.

[25] K. Lesser and M. Oishi. Reachability for partially observable discrete time stochastic hybrid systems. *Automatica*, 50:1989–1998, 2014.

[26] K. Lesser and M. Oishi. Approximate safety verification and control of partially observable stochastic hybrid systems. *IEEE Transactions on Automatic Control*, to appear.

[27] M. López and G. Still. Semi-infinite programming. *European Journal of Operational Research*, 180:491–518, 2007.

[28] J. Lygeros, C. Tomlin, and S. Sastry. Controllers for reachability specifications for hybrid systems. *Automatica*, 35:349–370, 1999.

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

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

[31] I. M. Mitchell and J. A. Templeton. A toolbox of Hamilton-Jacobi solvers for analysis of non-deterministic continuous and hybrid systems. In *International Workshop on Hybrid Systems: Computation and Control*, pages 480–494. Springer, 2005.

[32] P. Mohajerin Esfahani, D. Chatterjee, and J. Lygeros. The stochastic reach-avoid problem and set characterization for diffusions. *Automatica*, 70:43–56, 2016.

[33] P. Mohajerin Esfahani and D. Kuhn. Data-driven distributionally robust optimization using the Wasserstein metric: Performance guarantees and tractable reformulations. *arXiv preprint arXiv:1505.05116.*, 2015.

[34] R. E. Mortensen and K. P. Haggerty. A stochastic computer model for heating and cooling loads. *IEEE Transactions on Power Systems*, 3(3):1213–1219, 1988.

[35] I. Popescu. Robust mean-covariance solutions for stochastic optimization. *Operations Research*, 55(1):98–112, 2007.

[36] S. Prajna and A. Jadbabaie. Safety verification of hybrid systems using barrier certificates. In *International Workshop on Hybrid Systems: Computation and Control*, pages 477–492. Springer, 2004.

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

[38] M. Prandini and J. Hu. A stochastic approximation method for reachability computations. In *Stochastic Hybrid Systems*, pages 107–139. Springer, 2006.

[39] M. Prandini and J. Hu. Stochastic reachability: Theory and numerical approximation. *Stochastic Hybrid Systems, Automation and Control Engineering Series*, 24:107–138, 2006.

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

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

[42] R. Reemtsen and S. Görner. Numerical methods for semi-infinite programming: a survey. In *Semi-Infinite Programming*, pages 195–275. Springer, 1998.

[43] H. Scarf, K. J. Arrow, and S. Karlin. A min-max solution of an inventory problem. *Studies in the Mathematical Theory of Inventory and Production*, pages 201–209, 1958.

[44] A. Shapiro. On duality theory of conic linear problems. In *Semi-Infinite Programming*, pages 135–165. Springer, 2001.

[45] S. Summers and J. Lygeros. Verification of discrete time stochastic hybrid systems: A stochastic reach-avoid decision problem. *Automatica*, 46:1951–1961, 2010.

[46] H. Sun and H. Xu. Convergence analysis for distributionally robust optimization and equilibrium problems. *Mathematics of Operations Research*, 41(2):377–401, 2016.

[47] C. Tomlin, G. J. Pappas, and S. Sastry. Conflict resolution for air traffic management: A study in multiagent hybrid systems. *IEEE Transactions on Automatic Control*, 43(4):509–521, 1998.

[48] C. J. Tomlin, J. Lygeros, and S. S. Sastry. A game theoretic approach to controller design for hybrid systems. *Proceedings of the IEEE*, 88(7):949–970, 2000.

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

[50] W. Wiesemann, D. Kuhn, and M. Sim. Distributionally robust convex optimization. *Operations Research*, 62(6):1358–1376, 2014.

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

[52] I. Yang, S. A. Burden, R. Rajagopal, S. S. Sastry, and C. J. Tomlin. Approximation algorithms for optimization of combinatorial dynamical systems. *IEEE Transactions on Automatic Control*, 61(9):2644–2649, 2016.

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

[54] C. Zhao and Y. Guan. Data-driven risk-averse stochastic optimization with Wasserstein metric. *Available on Optimization Online*, 2015.

[55] S. Zymler, D. Kuhn, and B. Rustem. Distributionally robust joint chance constraints with second-order moment information. *Mathematical Programming, Ser. A*, 137:167–198, 2013.

A Proof of Proposition 4

We use mathematical induction to prove this proposition. For $t = T$, $v_T(\boldsymbol{x}) = \mathbf{1}_A(\boldsymbol{x})$ is concave with respect to $\boldsymbol{x} \in A$. Suppose that $\boldsymbol{x} \mapsto v_s(\boldsymbol{x})$ is concave for $\boldsymbol{x} \in A$ for $s = T-1, \cdots, t+1$. Fix arbitrary $\boldsymbol{x}^1, \boldsymbol{x}^2 \in \mathbb{R}^n$ and $\lambda \in (0,1)$. Let $\boldsymbol{u}^i \in \mathbb{U}(\boldsymbol{x}^i)$ be a solution to the outer maximization

problem in (3.1) at $(t, \boldsymbol{x}^i)$ for $i = 1, 2$. We also let $\boldsymbol{x}^\lambda := \lambda \boldsymbol{x}^1 + (1-\lambda)\boldsymbol{x}^2$ and $\boldsymbol{u}^\lambda := \lambda \boldsymbol{u}^1 + (1-\lambda)\boldsymbol{u}^2$. Since A is a convex set, $\boldsymbol{x}^\lambda \in A$. In addition, Assumption 2 (iii) implies that $\boldsymbol{u}^\lambda \in \mathbb{U}(\boldsymbol{x}^\lambda)$. Therefore, we have that

$$v_t(\boldsymbol{x}^\lambda) \geq \inf_{\boldsymbol{\mu} \in \mathbb{D}_t} \int_{\mathbb{R}^l} v_{t+1}(f(\boldsymbol{x}^\lambda, \boldsymbol{u}^\lambda, \boldsymbol{w}))d\boldsymbol{\mu}(\boldsymbol{w}).$$

Since f is an affine function, for each $\boldsymbol{w} \in \mathbb{R}^l$, $f(\boldsymbol{x}^\lambda, \boldsymbol{u}^\lambda, \boldsymbol{w}) = \lambda f(\boldsymbol{x}^1, \boldsymbol{u}^1, \boldsymbol{w}) + (1-\lambda) f(\boldsymbol{x}^2, \boldsymbol{u}^2, \boldsymbol{w})$. Combining this with the concavity of v_{t+1}, we further have that

$$\begin{aligned}
v_t(\boldsymbol{x}^\lambda) &\geq \inf_{\boldsymbol{\mu} \in \mathbb{D}_t} \int_{\mathbb{R}^l} \lambda \big[v_{t+1}(f(\boldsymbol{x}^1, \boldsymbol{u}^1, \boldsymbol{w})) \\
&\qquad + (1-\lambda) v_{t+1}(f(\boldsymbol{x}^2, \boldsymbol{u}^2, \boldsymbol{w})) \big] d\boldsymbol{\mu}(\boldsymbol{w}) \\
&\geq \lambda \inf_{\boldsymbol{\mu} \in \mathbb{D}_t} \int_{\mathbb{R}^l} v_{t+1}(f(\boldsymbol{x}^1, \boldsymbol{u}^1, \boldsymbol{w}))d\boldsymbol{\mu}(\boldsymbol{w}) \\
&\quad + (1-\lambda) \inf_{\boldsymbol{\mu} \in \mathbb{D}_t} \int_{\mathbb{R}^l} v_{t+1}(f(\boldsymbol{x}^2, \boldsymbol{u}^2, \boldsymbol{w}))d\boldsymbol{\mu}(\boldsymbol{w}) \\
&= \lambda v_{t+1}(\boldsymbol{x}^1) + (1-\lambda) v_{t+1}(\boldsymbol{x}^2),
\end{aligned}$$

which implies that v_t is concave. This completes our inductive argument. □