Distributionally Robust Stochastic Control with Conic Confidence Sets

Insoon Yang

***Abstract*— The theory of (standard) stochastic optimal control is based on the assumption that the probability distribution of uncertain variables is fully known. In practice, however, obtaining an accurate distribution is often challenging. To resolve this issue, we study a *distributionally robust stochastic control* problem that minimizes a cost function of interest given that the distribution of uncertain variables is not known but lies in a so-called *ambiguity set*. We first investigate a dynamic programming approach and identify conditions for the existence and optimality of non-randomized Markov policies. We then propose a duality-based reformulation method for an associated Bellman equation in cases with conic confidence sets. This reformulation alleviates the computational issues inherent in the infinite-dimensional minimax optimization problem in the Bellman equation without sacrificing optimality. The effectiveness of the proposed method is demonstrated through an application to a stochastic inventory control problem.**

I. Introduction

Standard stochastic control methods assume that the probability distribution of uncertain variables (e.g., disturbances) is available. 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. Situations in which uncertain variables are not directly observed could be much more challenging; computational methods, such as filtering or statistical learning techniques, are often used to obtain the (posterior) distribution of the uncertain variables given limited observations. The accuracy of the obtained distribution is often poor, as it is subject to the quality of the observations, computational methods, and prior knowledge about the variables. If poor distributional information is employed in constructing a stochastic optimal controller, it does not guarantee optimality and can even cause catastrophic system behaviors [1], [2].

To overcome this issue of limited distribution information in stochastic control, we investigate a *distributionally robust control* approach. This emerging stochastic control method minimizes a cost function of interest, assuming that the distribution of uncertain variables is not completely known but contained in a pre-specified *ambiguity set* of probability distributions. In finite-state Markov decision process settings, several types of ambiguity sets have been considered: [3], [4], [5] employ ambiguity sets with moment constraints, confidence intervals and Wasserstein distance, respectively. However, for continuous-state control problems, only a few cases with moment constraint-based and total variation distance ambiguity are well studied [6], [7], [8].

In this paper, we consider an important class of ambiguity sets that are characterized with confidence sets and probability intervals in a continuous-state stochastic control setting. The contributions of this work are twofold. First, we investigate a dynamic programming solution to discrete-time distributionally robust control problems in general finite-horizon cases and provide conditions for the existence and optimality of non-randomized Markov policies. This existence result is based on the lower semi-continuity of an associated dynamic programming operator. Our dynamic programming approach identifies an important structural property: the system state is a sufficient statistic. Second, we propose a dual formulation of the Bellman equation in cases with conic confidence sets. This approach converts the computationally challenging infinite-dimensional minimax problem into a semi-infinite program, which can be solved by existing convergent algorithms. We also show that the proposed reformulation is exact using strong duality and the *nesting condition* for confidence sets proposed by Wiesemann *et al.* [9]. The utility of our approach is demonstrated through an application to a stochastic inventory control problem.

The remainder of this paper is organized as follows. In Section II, we introduce the problem setup, including a dynamic game formulation of distributionally robust control problems. In Section III, we provide conditions for the existence and optimality of non-randomized Markov policies and another set of conditions for the convexity of the value function. Based on these analytical results, we develop a duality-based reformulation method for the Bellman equation in Section IV. A stochastic inventory control problem is considered in Section V as an application of the proposed method.

We use the following notation throughout the paper. Given a Borel space X, $\mathcal{P}(X)$ denotes the set of Borel probability measures on X. Given a cone K, K^* represents its dual cone. We also let $\mathcal{T} := \{0, 1, \cdots, T-1\}$ and $\bar{\mathcal{T}} := \{0, 1, \cdots, T\}$.

II. Problem Setup

A. Ambiguity in the Distribution of Disturbances

Consider the following stochastic system subject to the disturbance $\{w_t\}_{t=0}^{T-1}$, $w_t \in \mathbb{R}^l$, defined on a standard probability space $(\Omega, \mathcal{F}, \mathbb{P})$:

$$x_{t+1} = f(x_t, u_t, w_t), \tag{1}$$

where $x_t \in \mathbb{R}^n$ is the state, $u_t \in \mathbb{R}^m$ is the control input at stage t and $f : \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^l \to \mathbb{R}^n$ is a measurable function. We assume that w_s and w_t are independent for $s \neq t$. In

This work is supported in part by NSF under ECCS-1708906 and CNS-1657100. I. Yang is with the Electrical Engineering Department, University of Southern California, Los Angeles, CA 90089, USA insoonya@usc.edu

many practical situations, the full disturbance distribution may not be available. To overcome this challenge, we will investigate a distributionally robust control approach that minimizes the worst-case cost associated with the system evolution under information constraints characterized as a set of probability distributions. We assume that the true probability distribution μ_t of w_t is not completely known but contained in a so-called *ambiguity set*, $\mathbb{D}_t$.

Example 1 (Ambiguity with Confidence Sets)**.** *Consider the following set of probability measures:*

$$\mathbb{D}_t := \{\boldsymbol{\mu} \in \mathcal{P}(\mathbb{R}^l) \mid \boldsymbol{\mu}(\mathcal{C}_t^i) \in [\underline{\mathbf{p}}_t^i, \overline{\mathbf{p}}_t^i],\ i \in \mathcal{I}_t\}, \qquad (2)$$

where $\underline{\mathbf{p}}_t^i, \overline{\mathbf{p}}_t^i \in [0,1]$ are model parameters, $\mathcal{C}_t^i$'s are confidence sets, and $\mathcal{I}_t := \{1, \cdots, N_t\}$ is a given index set. With any measure in this set, the probability that w_t is contained in the set $\mathcal{C}_t^i$ is between $\underline{\mathbf{p}}_t^i$ and $\overline{\mathbf{p}}_t^i$.

Note that the user can specify the distributional information, such as $\mathcal{C}_t^i$ and $(\underline{\mathbf{p}}_t^i, \overline{\mathbf{p}}_t^i)$, that the controller trusts given data and statistical information about the disturbance. Another popular type of ambiguity sets in single-stage optimization problems is based on moment constraints and confidence sets [10], [11], [12], [13], [14], [9]. Recently, statistical distance-based ambiguity sets also received a great deal of attention since they can be easily designed from an empirical distribution without requiring a large enough data samples to estimate moments [15], [16], [17], [18], [19], [20]. In Section IV, we consider ambiguity sets of the form (2) that are characterized with conic confidence sets. This class of ambiguity sets can be intuitively constructed from empirical distributions, as will be illustrated in Section V.

B. Distributionally Robust Control as a Dynamic Game

Let H_t be the set of *histories* up to stage t, whose element is of the form[1] $h_t = (x_0, u_0, w_0, \cdots, x_{t-1}, u_{t-1}, w_{t-1}, x_t)$. The set of admissible control strategies is chosen as

$$\Pi := \{\pi := (\pi_0, \cdots, \pi_{T-1}) \,|\, \pi_t(\mathbb{U}(x_t) \,|\, h_t) = 1 \ \ \forall h_t \in H_t\},$$

where $\mathbb{U}(x_t)$ is the set of admissible actions given state x_t, and π_t is a stochastic kernel from H_t to $\mathbb{R}^m$. Note that the strategy space is broad enough to contain randomized non-Markov policies.[2] Similarly, we let H_t^e be the set of extended histories up to stage t, whose element takes the form $h_t^e := (x_0, u_0, w_0, \mu_0, \cdots, x_{t-1}, u_{t-1}, w_{t-1}, \mu_{t-1}, x_t, u_t)$. By viewing the disturbance as an adversarial player who chooses the disturbance distribution given available information h_t^e, we define the set of admissible distribution strategies as

$$\Gamma := \{\gamma = (\gamma_0, \cdots, \gamma_{T-1}) \mid \gamma_t(\mathbb{D}_t \mid h_t^e) = 1 \ \forall h_t^e \in H_t^e\}.$$

[1]All the results in this paper are valid with histories of the form $\tilde{h}_t := (x_0, u_0, w_0, \mu_0, \cdots, x_{t-1}, u_{t-1}, w_{t-1}, \mu_{t-1}, x_t)$ that also contains Player II's actions $(\mu_0, \cdots, \mu_{t-1})$. However, we use the reduced version of histories because the realized distributions may not be observable in practice.

[2]Suppose that for each t, $\pi_t(\cdot \mid h_t)$ is concentrated at a measurable function $\phi_t : \mathbb{R}^n \to \mathbb{R}^m$ such that $\phi_t(x) \in \mathbb{U}(x)$ for all $x \in \mathbb{R}^n$ and for all $h_t \in H_t$. Then, π is a (non-randomized) Markov policy and by a slight abuse of notation π_t is considered to be identical with ϕ_t.

Consider the following cost function associated with the system evolution starting from $x_0 = \boldsymbol{x}$:

$$J_{\boldsymbol{x}}[\pi, \gamma] := \mathbb{E}^{\pi,\gamma}\left[\sum_{t=0}^{T-1} r(x_t, u_t) + q(x_T)\right],$$

where $r : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$ and $q : \mathbb{R}^n \to \mathbb{R}$ are stage-wise and terminal cost functions of interest. 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).

Our goal is to choose a control policy that minimizes the worst-case cost given ambiguous information about the probability distribution $\{\mu_t\}_{t=0}^{T-1}$ of the disturbance $\{w_t\}_{t=0}^{T-1}$. To be more precise, we define the optimal distributionally robust control policies as follows:

Definition 1. *A control strategy $\pi^\star \in \Pi$ is said to be an* optimal distributionally robust policy *if it satisfies*

$$\sup_{\gamma \in \Gamma} J_{\boldsymbol{x}}[\pi^\star, \gamma] \leq \sup_{\gamma' \in \Gamma} J_{\boldsymbol{x}}[\pi, \gamma'] \quad \forall \pi \in \Pi.$$

A desired control strategy can be obtained by solving the following minimax control problem:

$$\inf_{\pi \in \Pi} \sup_{\gamma \in \Gamma} J_{\boldsymbol{x}}[\pi, \gamma]. \qquad (3)$$

The most important part of this problem formulation is the inner maximization of the cost function over all probability distribution policies in the strategy space Γ, which encodes distributional ambiguity through $\mathbb{D}_t$. One can view this problem as a two-player zero-sum dynamic game, in which Player I chooses a control policy to minimize the cost and Player II selects the disturbance's distribution strategy that maximizes the cost. The proposed minimax formulation implies the following property:

Proposition 1. *Suppose that an optimal solution to the distributionally robust control problem* (3) *exists and is denoted by $(\pi^\star, \gamma^\star)$. Then, the following inequalities hold:*

$$J_{\boldsymbol{x}}[\pi^\star, \gamma] \leq J_{\boldsymbol{x}}[\pi^\star, \gamma^\star] \leq \sup_{\gamma' \in \Gamma} J_{\boldsymbol{x}}[\pi, \gamma'] \quad \forall (\pi, \gamma) \in \Pi \times \Gamma.$$

The first and second equalities hold with $\gamma = \gamma^\star$ and $\pi = \pi^\star$, respectively. In addition, $\pi^\star$ is an optimal distributionally robust policy.

The first inequality implies that when the optimal policy $\pi^\star$ is employed, the worst-case cost value is equal to $J_{\boldsymbol{x}}[\pi^\star, \gamma^\star]$ for any distributional error consistent with the constraints in the ambiguity set $\mathbb{D}_t$ for each t. Thus, this approach provides a performance guarantee in the form of an upper-bound, $J_{\boldsymbol{x}}[\pi^\star, \gamma^\star]$, of the cost value, which is tight. Note that this performance guarantee may not be valid when a different control policy is employed as shown in the second inequality. Furthermore, the second inequality confirms that an optimal solution to the dynamic game problem (3) provides an optimal distributionally robust policy.

III. Dynamic Programming Solution

A. Existence of Optimal Distributionally Robust Policies

We begin by introducing the following dynamic programming operator $\mathbf{T}_t$, $t \in \mathcal{T}$:

$$\mathbf{T}_t\mathbf{v}(\boldsymbol{x}) := \inf_{\boldsymbol{u}\in\mathbb{U}(\boldsymbol{x})} \sup_{\boldsymbol{\mu}\in\mathbb{D}_t} \left[r(\boldsymbol{x},\boldsymbol{u}) + \int_{\mathbb{R}^l} \mathbf{v}(f(\boldsymbol{x},\boldsymbol{u},\boldsymbol{w}))\mathrm{d}\boldsymbol{\mu}(\boldsymbol{w}) \right],$$

where $\mathbf{v}$ is a measurable function on $\mathbb{R}^n$. We then define the value function of the distributionally robust control problem (3) as

$$v_t(\boldsymbol{x}) := \mathbf{T}_t \circ \mathbf{T}_{t+1} \circ \cdots \circ \mathbf{T}_{T-1} q(\boldsymbol{x})$$

for each $t \in \mathcal{T}$ and $v_T(\boldsymbol{x}) := q(\boldsymbol{x})$. By definition, $v_t(\boldsymbol{x})$ represents the minimal worst-case expected cost value from stage t to T given $x_t = \boldsymbol{x}$. Under the following assumption for the *measurable selection condition*, the value function is lower semi-continuous and thus the distributionally robust control problem (3) admits a non-randomized Markov policy, which is optimal.

Assumption 1. *The following properties hold:*

- (i) *$r(\boldsymbol{x},\boldsymbol{u})$ and $q(\boldsymbol{x})$ are lower semi-continuous and bounded below for all $(\boldsymbol{x},\boldsymbol{u}) \in \mathbb{R}^n \times \mathbb{R}^m$ such that $\boldsymbol{u} \in \mathbb{U}(\boldsymbol{x})$;*
- (ii) *For each 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 for all $(\boldsymbol{x},\boldsymbol{u},\boldsymbol{\mu}) \in \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{D}_t$ such that $\boldsymbol{u} \in \mathbb{U}(\boldsymbol{x})$;
- (iii) *The set $\mathbb{U}(\boldsymbol{x})$ is compact for each $\boldsymbol{x} \in \mathbb{R}^n$. In addition, the set-valued mapping $\boldsymbol{x} \mapsto \mathbb{U}(\boldsymbol{x})$ is upper semi-continuous.*

Theorem 1. *Suppose that Assumption 1 holds. Then, the value function v_t is lower semi-continuous for each $t \in \mathcal{T}$. Furthermore, there exists a measurable function $\phi_t : \mathbb{R}^n \to \mathbb{R}^m$ for each $t \in \mathcal{T}$ such that $\phi_t(\boldsymbol{x}) \in \mathbb{U}(\boldsymbol{x})$ and*

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

for all $\boldsymbol{x} \in \mathbb{R}^n$. The non-randomized Markov policy $\pi^\star := (\phi_0, \cdots, \phi_{T-1}) \in \Pi$ is an optimal solution to the distributionally robust control problem (3)*, i.e.,*

$$v_0(\boldsymbol{x}) = \sup_{\gamma\in\Gamma} J_{\boldsymbol{x}}[\pi^\star, \gamma].$$

This theorem can be shown by extending the proof of Theorem 1 in [7] and Theorem 3.1 in [21]. The key idea is to show that the lower semi-continuity of v_t is preserved through the dynamic programming operator. We can then use mathematical induction to show that v_t is lower semi-continuous and thus, the outer minimization problem in the definition of value functions admits an optimal solution. Note that Theorem 1 allows us to identify an important structural property of the distributionally robust control problem: the system state is a sufficient statistic for Player I (controller) under Assumption 1. Theorem 1 yields the practical advantage that it suffices to focus on non-randomized Markov polices when designing an optimal distributionally robust controller.

B. Bellman Equation

Applying the dynamic programming principle [22], [23], we can evaluate the value function backward in time as follows:

Proposition 2. *Suppose that Assumption 1 holds. Then, the value function v_t satisfies the following Bellman equation:*

$$v_t(\boldsymbol{x}) = \min_{\boldsymbol{u}\in\mathbb{U}(\boldsymbol{x})} \left[r(\boldsymbol{x},\boldsymbol{u}) + \sup_{\boldsymbol{\mu}\in\mathbb{D}_t} \int_{\mathbb{R}^l} v_{t+1}(f(\boldsymbol{x},\boldsymbol{u},\boldsymbol{w}))\mathrm{d}\boldsymbol{\mu}(\boldsymbol{w}) \right] \quad (4)$$

with $v_T(\boldsymbol{x}) = q(\boldsymbol{x})$.

Note that due to Theorem 1 the outer minimization problem in the Bellman equation admits an optimal solution. From the numerical perspective, however, solving the Bellman equation is challenging. In addition to the scalability issue inherent in dynamic programming, this Bellman equation involves an infinite-dimensional minimax optimization problem because the disturbance may have a continuous density. In the next section, we will resolve this computational issue for an important class of distributionally robust stochastic control problems in which confidence sets are specified. With such an ambiguity set, we will show that an optimal distributionally robust policy can be obtained by solving computationally tractable semi-infinite programs when the value function is convex.

C. Convexity of the Value Function

We now show that the value function is convex under the following conditions:

Assumption 2. *The distributionally robust control problem* (3) *satisfies the followings:*

- (i) *$r : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$ and $q : \mathbb{R}^n \to \mathbb{R}$ are convex functions;*
- (ii) *$f : \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^l \to \mathbb{R}^n$ is an affine function;*
- (iii) *For all $\lambda \in (0,1)$ and for all $\boldsymbol{x}^1, \boldsymbol{x}^2 \in \mathbb{R}^n$, if $\boldsymbol{u}^i \in \mathbb{U}(\boldsymbol{x}^i)$, $i = 1, 2$, then $\lambda\boldsymbol{u}^1 + (1-\lambda)\boldsymbol{u}^2 \in \mathbb{U}(\lambda\boldsymbol{x}^1 + (1-\lambda)\boldsymbol{x}^2)$.*

Proposition 3. *Suppose that Assumption 2 holds. Then, the value function $v_t : \mathbb{R}^n \to \mathbb{R}$ is convex for each $t \in \bar{\mathcal{T}}$.*

Proof. We use mathematical induction. For $t = T$, $v_T : \mathbb{R}^n \to \mathbb{R}$ is convex since $v_T = q$. Suppose that $v_\tau : \mathbb{R}^n \to \mathbb{R}$ is convex for $\tau = T-1, \cdots, t+1$. We now the consider the value function at stage t. Fix $\lambda \in (0,1)$ and $\boldsymbol{x}^1, \boldsymbol{x}^2 \in \mathbb{R}^n$. For any $\epsilon > 0$, there exists an ϵ-optimal solution $\boldsymbol{u}^i \in \mathbb{U}(\boldsymbol{x}^i)$ to the outer minimization problem in the Bellman equation (4) for $(t, \boldsymbol{x}^i)$, i.e.,

$$v_t(\boldsymbol{x}^i) + \epsilon > r(\boldsymbol{x}^i,\boldsymbol{u}^i) + \sup_{\boldsymbol{\mu}\in\mathbb{D}_t} \int_{\mathbb{R}^l} v_{t+1}(f(\boldsymbol{x}^i,\boldsymbol{u}^i,\boldsymbol{w}))\mathrm{d}\boldsymbol{\mu}(\boldsymbol{w}).$$

Let $\boldsymbol{x}^\lambda := \lambda \boldsymbol{x}^1 + (1-\lambda)\boldsymbol{x}^2$ and $\boldsymbol{u}^\lambda := \lambda \boldsymbol{u}^1 + (1-\lambda)\boldsymbol{u}^2$. Due to Assumption 2 (iii), $\boldsymbol{u}^\lambda \in \mathbb{U}(\boldsymbol{x}^\lambda)$. Thus,

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

We now notice that $f(\boldsymbol{x}^\lambda, \boldsymbol{u}^\lambda, \boldsymbol{w}) = \lambda f(\boldsymbol{x}^1, \boldsymbol{u}^1, \boldsymbol{w}) + (1-\lambda) f(\boldsymbol{x}^2, \boldsymbol{u}^2, \boldsymbol{w})$ due to Assumption 2 (ii). Since v_{t+1} and r are convex, we obtain that

$$\begin{aligned} v_t(\boldsymbol{x}^\lambda) &\leq \lambda r(\boldsymbol{x}^1, \boldsymbol{u}^1) + (1-\lambda) r(\boldsymbol{x}^2, \boldsymbol{u}^2) \\ &\quad + \sup_{\boldsymbol{\mu}\in\mathbb{D}_t} \int_{\mathbb{R}^l} \lambda v_{t+1}(f(\boldsymbol{x}^1, \boldsymbol{u}^1, \boldsymbol{w}))\mathrm{d}\boldsymbol{\mu}(\boldsymbol{w}) \\ &\quad + \sup_{\boldsymbol{\mu}\in\mathbb{D}_t} \int_{\mathbb{R}^l} (1-\lambda) v_{t+1}(f(\boldsymbol{x}^2, \boldsymbol{u}^2, \boldsymbol{w}))\mathrm{d}\boldsymbol{\mu}(\boldsymbol{w}) \\ &< \lambda v_t(\boldsymbol{x}^1) + (1-\lambda) v_t(\boldsymbol{x}^2) + \epsilon. \end{aligned}$$

Letting $\epsilon \to 0$, we conclude that v_t is convex. □

Note that this proposition does not require the existence of optimal distributionally robust policies. Furthermore, we do not impose any specific structure on the ambiguity set $\mathbb{D}_t$ for the convexity of the value function.

IV. Strong Duality-Based Reformulation

A. Distributional Ambiguity with Conic Confidence Sets

We now focus on the distributionally robust control problem with a particular ambiguity set of the form (2) in Example 1:

$$\mathbb{D}_t := \{\boldsymbol{\mu} \in \mathcal{P}(\mathbb{R}^l) \mid \boldsymbol{\mu}(\mathcal{C}_t^i) \in [\underline{\mathbf{p}}_t^i, \overline{\mathbf{p}}_t^i],\ i \in \mathcal{I}_t\},$$

where $\mathcal{I}_t := \{1, \cdots, N_t\}$ and the confidence set $\mathcal{C}_t^i$ has the following conic representation:

$$\mathcal{C}_t^i = \{\boldsymbol{w} \in \mathbb{R}^l \mid \mathbf{C}_t^i \boldsymbol{w} \preceq_{K_t^i} \mathbf{d}_t^i\},$$

where $\mathbf{C}_t^i \in \mathbb{R}^{L_t^i \times l}$ and $\mathbf{d}_t^i \in \mathbb{R}^{L_t^i}$ are model parameters, and K_t^i is a proper cone. We impose the following assumption:

Assumption 3. *The ambiguity set satisfies the following conditions:*

- (i) *The confidence set $\mathcal{C}_t^{N_t}$ is compact and $\underline{\mathbf{p}}_t^{N_t} = \overline{\mathbf{p}}_t^{N_t} = 1$.*
- (ii) *There exists a distribution measure $\boldsymbol{\mu} \in \mathbb{D}_t$ such that $\boldsymbol{\mu}(\mathcal{C}_t^i) \in (\underline{\mathbf{p}}_t^i, \overline{\mathbf{p}}_t^i)$ whenever $\underline{\mathbf{p}}_t^i < \overline{\mathbf{p}}_t^i$, $i \in \mathcal{I}_t$.*
- (iii) (Nesting Condition) *For each $t \in \mathcal{T}$ and all $i, i' \in \mathcal{I}_t$ such that $i \neq i'$, we have either $\mathcal{C}_t^i \subset^s \mathcal{C}_t^{i'}$, $\mathcal{C}_t^{i'} \subset^s \mathcal{C}_t^i$ or $\mathcal{C}_t^i \cap \mathcal{C}_t^{i'} = \emptyset$, where $X \subset^s Y$ represents that X is a strict subset of Y.*

The first condition represents that $\mathcal{C}_t^{N_t}$ is the support of μ_t. The second condition ensures that there exists a probability distribution that satisfies the probabilistic constraints in $\mathbb{D}_t$ as strict inequalities whenever $\underline{\mathbf{p}}_t^i < \overline{\mathbf{p}}_t^i$. These two regularity conditions will guarantee that strong duality holds based on the generalized Slater-type results from Shapiro [24] when the Bellman equation is reformulated in the next subsection. The third condition is called the *nesting condition* [9], which implies that there exists a strict partial order on the confidence sets regarding the set inclusion and that any incomparable sets are disjoint. This nesting condition, together with the two regularity conditions, provides a tractable dual formulation of the Bellman equation without loss of optimality.

B. Dual Bellman Equation

We now reformulate the infinite dimensional minimax optimization problem in the Bellman equation (4) as a semi-infinite program, which can be numerically solved by existing convergent algorithms. Furthermore, this reformulation based on strong duality is exact as shown in the following theorem.

Theorem 2. *Suppose that Assumptions 1, 2 and 3 hold. Then, the following equality holds for all $(t, \boldsymbol{x}) \in \mathcal{T} \times \mathbb{R}^n$:*[3]

$$\begin{aligned} v_t(\boldsymbol{x}) = \inf_{\boldsymbol{u},\kappa,\lambda,\nu}\ & r(\boldsymbol{x}, \boldsymbol{u}) + \sum_{i\in\mathcal{I}_t} (\overline{\mathbf{p}}_t^i \kappa^i - \underline{\mathbf{p}}_t^i \lambda^i) \\ \text{s.t.}\ & (\mathbf{C}_t^i \boldsymbol{w} - \mathbf{d}_t^i)^\top \nu^i + \sum_{i'\in\mathcal{A}_t(i)} (\kappa^{i'} - \lambda^{i'}) \\ & \geq v_{t+1}(f(\boldsymbol{x}, \boldsymbol{u}, \boldsymbol{w}))\ \ \forall \boldsymbol{w} \in \mathcal{C}_t^{N_t}\ \ \forall i \in \mathcal{I}_t \\ & \boldsymbol{u} \in \mathbb{U}(\boldsymbol{x}),\ \lambda, \kappa \in \mathbb{R}_+^{N_t},\ \nu^i \in K_t^{i*}, \end{aligned}$$

where $\mathcal{A}_t(i) := \{i\} \cup \{i' \in \mathcal{I}_t \,|\, \mathcal{C}_t^i \subset^s \mathcal{C}_t^{i'}\}$, with the terminal condition $v_T(\boldsymbol{x}) = q(\boldsymbol{x})$.

Proof. By introducing a slack variable $z \in \mathbb{R}$, we can rewrite the Bellman equation (4) in the following equivalent form:

$$\begin{aligned} v_t(\boldsymbol{x}) = \inf_{\boldsymbol{u}\in\mathbb{U}(\boldsymbol{x}), z\in\mathbb{R}}\ & z \\ \text{s.t.}\ & \sup_{\boldsymbol{\mu}\in\mathbb{D}_t} \left\{ r(\boldsymbol{x}, \boldsymbol{u}) \right. \\ & \left. + \int_{\mathbb{R}^l} v_{t+1}(f(\boldsymbol{x}, \boldsymbol{u}, \boldsymbol{w}))\mathrm{d}\boldsymbol{\mu}(\boldsymbol{w}) \right\} \leq z \end{aligned}$$

for each $(t, \boldsymbol{x}) \in \mathcal{T} \times \mathbb{R}^n$. We first focus on the maximization problem in the inequality constraint. It can be rewritten as the following infinite-dimensional linear program:

$$\begin{aligned} \sup_{\boldsymbol{\mu}\in\mathcal{P}(\mathbb{R}^l)}\ & r(\boldsymbol{x}, \boldsymbol{u}) + \int_{\mathbb{R}^l} v_{t+1}(f(\boldsymbol{x}, \boldsymbol{u}, \boldsymbol{w}))\mathrm{d}\boldsymbol{\mu}(\boldsymbol{w}) \\ \text{s.t.}\ & \int_{\mathcal{C}_t^{N_t}} \mathbf{1}_{\{\boldsymbol{w}\in\mathcal{C}_t^i\}} \mathrm{d}\boldsymbol{\mu}(\boldsymbol{w}) \geq \underline{\mathbf{p}}_t^i\ \ \forall i \in \mathcal{I}_t \\ & \int_{\mathcal{C}_t^{N_t}} \mathbf{1}_{\{\boldsymbol{w}\in\mathcal{C}_t^i\}} \mathrm{d}\boldsymbol{\mu}(\boldsymbol{w}) \leq \overline{\mathbf{p}}_t^i\ \ \forall i \in \mathcal{I}_t. \end{aligned}$$

Under Assumption 3, the generalized Slater condition holds [24]. Thus, there is no duality gap and we have the following dual formulation of the problem above without loss of optimality:

$$\begin{aligned} \inf_{\kappa,\lambda\in\mathbb{R}_+^{N_t}}\ & r(\boldsymbol{x}, \boldsymbol{u}) + \sum_{i\in\mathcal{I}_t} (\overline{\mathbf{p}}_t^i \kappa^i - \underline{\mathbf{p}}_t^i \lambda^i) \\ \text{s.t.}\ & v_{t+1}(f(\boldsymbol{x}, \boldsymbol{u}, \boldsymbol{w})) + \sum_{i\in\mathcal{I}_t} \mathbf{1}_{\{\boldsymbol{w}\in\mathcal{C}_t^i\}} (\lambda^i - \kappa^i) \leq 0 \\ & \forall \boldsymbol{w} \in \mathcal{C}_t^{N_t}. \end{aligned}$$

[3] In the reformulated Bellman equation, $\min_{\boldsymbol{u}}$ is merged with $\inf_{\kappa,\lambda,\nu}$ for a compact representation. The minimization problem admits an optimal solution $\boldsymbol{u}$.

Let $\mathcal{B}_t(i)$ be the index sets of all the strict subsets of $\mathcal{C}_t^i$, i.e., $\mathcal{B}_t(i) := \{i' \in \mathcal{I}_t \mid \mathcal{C}_t^{i'} \subset^s \mathcal{C}_t^i\}$. We also let

$$\bar{\mathcal{C}}_t^i := \mathcal{C}_t^i \setminus \bigcup_{i' \in \mathcal{B}_t(i)} \mathcal{C}_t^{i'}.$$

Due to the nesting condition, $\{\bar{\mathcal{C}}_t^1, \cdots, \bar{\mathcal{C}}_t^{N_t}\}$ is a disjoint partition of the support $\mathcal{C}_t^{N_t}$. Therefore, the inequality constraint of the dual problem can be rewritten as

$$v_{t+1}(f(\boldsymbol{x}, \boldsymbol{u}, \boldsymbol{w})) + \sum_{i' \in \mathcal{A}_t(i)} (\lambda^{i'} - \kappa^{i'}) \leq 0 \ \forall \boldsymbol{w} \in \bar{\mathcal{C}}_t^i \ \forall i \in \mathcal{I}_t.$$

The inequality constraint associated with the index $i \in \mathcal{I}_t$ is equivalent to

$$\sup_{\boldsymbol{w} \in \bar{\mathcal{C}}_t^i} \quad v_{t+1}(f(\boldsymbol{x}, \boldsymbol{u}, \boldsymbol{w})) + \sum_{i' \in \mathcal{A}_t(i)} (\lambda^{i'} - \kappa^{i'}) \leq 0.$$

Since $\boldsymbol{w} \mapsto v_{t+1}(f(\boldsymbol{x}, \boldsymbol{u}, \boldsymbol{w}))$ is convex for each $(\boldsymbol{x}, \boldsymbol{u})$, the objective function is convex with respect to $\boldsymbol{w}$. Therefore, the maximum is attained at the outer boundary of $\bar{\mathcal{C}}_t^i$. We now observe that the outer boundary of $\bar{\mathcal{C}}_t^i$ corresponds to the outer boundary of $\mathcal{C}_t^i$ due to the nesting condition [9]. Thus, we can rewrite the ith constraint as

$$\begin{aligned} \sup_{\boldsymbol{w} \in \mathcal{C}_t^{N_t}} \quad & v_{t+1}(f(\boldsymbol{x}, \boldsymbol{u}, \boldsymbol{w})) + \sum_{i' \in \mathcal{A}_t(i)} (\lambda^{i'} - \kappa^{i'}) \\ \text{s.t.} \quad & \mathbf{C}_t^i \boldsymbol{w} \preceq_{K_t^i} \mathbf{d}_t^i. \end{aligned}$$

Its dual is given by the following semi-infinite program:

$$\begin{aligned} \inf_{\nu^i \in K_t^{i*}, \theta^i \in \mathbb{R}} \quad & {\mathbf{d}_t^i}^\top \nu^i + \theta^i + \sum_{i' \in \mathcal{A}_t(i)} (\lambda^{i'} - \kappa^{i'}) \\ \text{s.t.} \quad & \sup_{\boldsymbol{w} \in \mathcal{C}_t^{N_t}} v_{t+1}(f(\boldsymbol{x}, \boldsymbol{u}, \boldsymbol{w})) - (\mathbf{C}_t^i \boldsymbol{w})^\top \nu^i \leq \theta^i. \end{aligned}$$

Putting the reformulation results all together, we have that

$$\begin{aligned} & v_t(\boldsymbol{x}) = \\ & \inf \ z \\ & \text{s.t. } r(\boldsymbol{x}, \boldsymbol{u}) + \sum_{i \in \mathcal{I}_t} (\overline{\mathbf{p}}_t^i \kappa^i - \underline{\mathbf{p}}_t^i \lambda^i) \leq z \\ & \quad {\mathbf{d}_t^i}^\top \nu^i + \theta^i + \sum_{i' \in \mathcal{A}_t(i)} (\lambda^{i'} - \kappa^{i'}) \leq 0 \ \ \forall i \in \mathcal{I}_t \\ & \quad v_{t+1}(f(\boldsymbol{x}, \boldsymbol{u}, \boldsymbol{w})) - (\mathbf{C}_t^i \boldsymbol{w})^\top \nu^i \leq \theta^i \ \forall \boldsymbol{w} \in \mathcal{C}_t^{N_t} \ \forall i \in \mathcal{I}_t \\ & \quad \boldsymbol{u} \in \mathcal{U}(\boldsymbol{x}), \ \kappa, \lambda \in \mathbb{R}_+^{N_t}, \ \nu^i \in K_t^{i*}, \ z \in \mathbb{R}, \ \theta \in \mathbb{R}^{N_t}. \end{aligned}$$

Viewing z and θ as slack variables, we conclude that the statement in the theorem holds. □

Note that the convexity of $\boldsymbol{w} \mapsto v_{t+1}(f(\boldsymbol{x}, \boldsymbol{u}, \boldsymbol{w}))$ and the nesting condition play a critical role in preserving optimality in the proposed reformulation as originally observed by Wiesemann *et al.* in the context of single-stage optimization [9]. When $(\boldsymbol{u}, \boldsymbol{w}) \mapsto v_{t+1}(f(\boldsymbol{x}, \boldsymbol{u}, \boldsymbol{w}))$ is also piecewise affine, our result is consistent with Theorem 1 in [9]. Theorem 2 allows us to avoid solving the computationally challenging infinite-dimensional minimax optimization problems in the original Bellman equation. Instead, we can evaluate the value function backward in time by solving a semi-infinite program at each (discretized) state. This semi-infinite program can be solved by several convergent methods such as primal-dual methods, discretization methods, homotopy methods, exchange methods, and constraint sampling methods (see [25], [26], [27] and the references therein). Among them, we used the convergent discretization method developed by Reemtsen [28] in the next section.

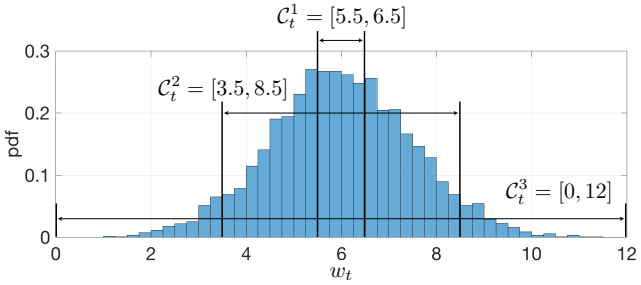


Fig. 1: The empirical distribution $\bar{\mu}_t$ of w_t and the confidence sets $\mathcal{C}_t^i$, $i = 1, 2, 3$.

V. Application to Inventory Control

We consider a stochastic inventory control problem to demonstrate the performance of our distributionally robust control method. We use the standard setting of stochastic Newsvendor problems (e.g., [29]). Let $x_t \in \mathbb{R}$ be an inventory level of interest at stage t. Given the quantity $u_t \in \mathbb{U} := [0, 10]$ ordered and the stochastic demand w_t at stage t, the inventory level evolves as

$$x_{t+1} = x_t + u_t - w_t,$$

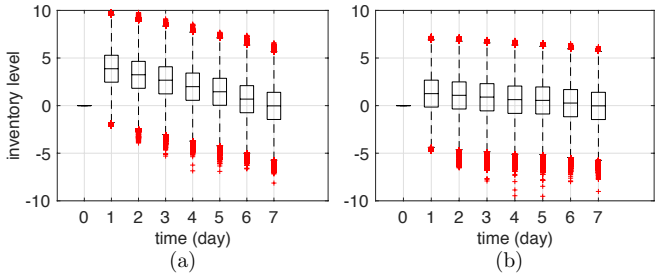
for $t \in \mathcal{T} := \{0, 1, \cdots, 6\}$. We assume that any unsatisfied demand is backlogged for the next stage and thus allow negative state values. The stage-wise cost function is given by

$$r(x_t, u_t, w_t) = c_o(x_t + u_t - w_t)_+ + c_u(w_t - x_t - u_t)_+,$$

where $c_o = 1$ is the overage (or storage) cost and $c_u = 1$ is the underage cost (or the cost of lost sales). Fig. 1 shows the empirical distribution $\bar{\mu}_t$ of w_t for all $t \in \mathcal{T}$, and the confidence sets used in our simulations. We choose $\underline{\mathbf{p}}_t^i$ and $\overline{\mathbf{p}}_t^i$ as 90% and 110% of $\bar{\mu}_t(\mathcal{C}_t^i)$. Thus, $\bar{\mu}_t$ is contained in the constructed ambiguity set $\mathbb{D}_t$.

We compare our distributionally robust controller designed using $\mathbb{D}_t$ and the standard stochastic optimal controller constructed with the empirical distribution $\bar{\mu}_t$ when $x_0 = 10$. Suppose that the actual distribution μ_t^{true} of w_t is uniform in each confidence set and $\mu_t^{true}(\mathcal{C}_t^i) = \underline{\mathbf{p}}_t^i$. Then, $\mu_t^{true} \in \mathbb{D}_t$ but μ_t^{true} is different from the empirical distribution $\bar{\mu}_t$. In our simulation with 10^5 trajectories of $\{w_t\}$ sampled from $\{\mu_t^{true}\}$, the distributionally robust control method reduces the total expected cost incurred by the standard controller by 29.6%. This result confirms that our controller is robust against errors in disturbance distributions while the standard controller is not.

To investigate why the proposed controller performs better than the standard controller under distributional ambiguity,



(a) (b)

Fig. 2: Tukey box plots of state trajectories controlled by (a) the standard stochastic optimal controller and (b) the distributionally robust controller.

we now compare their controlled state trajectories. As shown in Fig. 2, the distributionally robust controller drives the median (and the first and third quantiles) of state trajectories closer to the origin than the standard controller. Since our controller considers the worst-case distribution in the ambiguity set, it can control the system in a desirable manner (maintaining the inventory level close to zero) even when μ_t^{true} deviates from the empirical distribution $\bar{\mu}_t$. On the other hand, the standard controller optimizes the system performance only when $\mu_t^{true} = \bar{\mu}_t$; otherwise, there is no performance guarantee. In particular, the standard controller initially increases the inventory level by using approximately 97% of the maximum allowable control value. This aggressive control action is intended to satisfy demand at later stages with the limited control range $\mathbb{U} := [0, 10]$. However, as μ_t^{true} deviates from $\bar{\mu}_t$, this standard control strategy generates higher overage costs than expected. On the other hand, the distributionally robust controller is designed to take into account such possibilities and is capable of balancing the overage and underage costs when $\mu_t^{true} \neq \bar{\mu}_t$.

VI. Conclusion and Future Work

We have proposed a duality-based dynamic programming approach to distributionally robust control problems with conic confidence sets. The structural property we identified allows us to focus on non-randomized Markov policies with state feedback. Our exact duality-based reformulation method also alleviates the computational issues in the original Bellman equation that involves infinite-dimensional minimax optimization problems. As a future research, it is of great interest to develop a scalable numerical method for the reformulated Bellman equation. Furthermore, adding risk constraints may help in systematically discouraging undesirable system behaviors.

References

[1] A. Nilim and L. El Ghaoui, "Robust control of Markov decision processes with uncertain transition matrices," *Operations Research*, vol. 53, no. 5, pp. 780–798, 2005.

[2] S. Samuelson and I. Yang, "Data-driven distributionally robust control of energy storage to manage wind power fluctuations," in *Proceedings of the 1st IEEE Conference on Control Technology and Applications*, 2017.

[3] H. Xu and S. Mannor, "Distributionally robust Markov decision processes," *Mathematics of Operations Research*, vol. 37, no. 2, pp. 288–300, 2012.

[4] P. Yu and H. Xu, "Distributionally robust counterpart in Markov decision processes," *IEEE Transactions on Automatic Control*, vol. 61, no. 9, pp. 2538–2543, 2016.

[5] I. Yang, "A convex optimization approach to distributionally robust Markov decision processes with Wasserstein distance," *IEEE Control Systems Letters*, vol. 1, no. 1, pp. 164–169, 2017.

[6] 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*, vol. 61, no. 2, pp. 430–442, 2016.

[7] I. Yang, "A dynamic game approach to distributionally robust safety specifications for stochastic systems," *arXiv:1701.06260*, 2017.

[8] I. Tzortzis, C. D. Charalambous, and T. Charalambous, "Dynamic programming subject to total variation distance ambiguity," *SIAM Journal on Control and Optimization*, vol. 53, no. 4, pp. 2040–2075, 2015.

[9] W. Wiesemann, D. Kuhn, and M. Sim, "Distributionally robust convex optimization," *Operations Research*, vol. 62, no. 6, pp. 1358–1376, 2014.

[10] 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*, pp. 201–209, 1958.

[11] J. Dupačová, "The minimax approach to stochastic programming and an lllustrative application," *Stochastics*, vol. 20, pp. 73–88, 1987.

[12] E. Delage and Y. Ye, "Distributionally robust optimization under moment uncertainty with application to data-driven problems," *Operations Research*, vol. 58, no. 3, pp. 595–612, 2010.

[13] I. Popescu, "Robust mean-covariance solutions for stochastic optimization," *Operations Research*, vol. 55, no. 1, pp. 98–112, 2007.

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

[15] 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*, vol. 59, no. 2, pp. 341–357, 2013.

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

[17] H. Sun and H. Xu, "Convergence analysis for distributionally robust optimization and equilibrium problems," *Mathematics of Operations Research*, vol. 41, no. 2, pp. 377–401, 2016.

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

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

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

[21] 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*, vol. 41, no. 5, pp. 1626–1659, 2003.

[22] R. Bellman, "Dynamic programming and Lagrange multipliers," *Proceedings of the National Academy of Sciences*, vol. 42, no. 10, pp. 767–769, 1956.

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

[24] A. Shapiro, "On duality theory of conic linear problems," in *Semi-Infinite Programming*. Springer, 2001, pp. 135–165.

[25] R. Hettich and K. O. Kortanek, "Semi-infinite programming: Theory, methods, and applications," *SIAM Review*, vol. 35, no. 3, pp. 380–429, 1993.

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

[27] G. Calafiore and M. C. Campi, "Uncertain convex programs: randomized solutions and confidence levels," *Mathematical Programming, Ser. A*, vol. 102, pp. 25–46, 2005.

[28] R. Reemtsen, "Discretization methods for the solution of semi-infinite programming problems," *Journal of Optimization Theory and Applications*, vol. 71, no. 1, pp. 85–103, 1991.

[29] R. Levi, R. O. Roundy, and D. B. Shmoys, "Provably near-optimal sampling-based policies for stochastic inventory control models," *Mathematics of Operations Research*, vol. 32, no. 4, pp. 821–839, 2007.