Distributionally Robust Optimal and Safe Control of Stochastic Systems via Kernel Conditional Mean Embedding

Licio Romao, Ashish R. Hota, and Alessandro Abate

***Abstract*— We develop a distributionally robust framework to perform dynamic programming using kernel methods and apply our framework to design feedback control policies that satisfy safety and optimality specifications. To this end, we leverage kernels to map the transition probabilities associated with the state evolution to functions of the associated reproducing kernel Hilbert space (RKHS) using available data to construct an empirical estimate of the conditional kernel mean embedding. Hence, the proposed method is model-free, as we only require trajectories of the associated dynamical system to design the underlying control policy. We construct an ambiguity set on the space of probability measures and show that backward propagation using dynamic programming is well-defined and that optimal policies are Markovian. We verify the main theoretical findings of the paper in benchmark control problems, such as safe control of thermostatically controlled loads (TCL).**

I. Introduction

In recent years, study of stochastic optimization and control problems where the distribution of the uncertain parameters or the transition probabilities are not known with certainty has received a lot of attention. This class of problems, termed as "distributionally robust" optimization or control problems [1]–[4], assume that the distribution of unknown parameters belongs to a family of distributions that is often constructed from past data or samples collected from the underlying environment. The goal is to compute an optimal solution or control policy to minimize an uncertain cost function subject to worst case realization of the uncertainty distribution from the ambiguity set. These ideas are related to the class of min-max control problems investigated in the seminal work [5], and later explored in [6]. In addition to optimality, there is a growing interest in the problem of control synthesis for stochastic systems subject to safety specifications [7], [8]. Recently, [9] studied the distributionally robust version of this problem where the ambiguity set is defined as the set of all distributions that satisfy certain first and second order moment constraints.

In most of the past work on distributionally robust optimal and safe control, such as [2]–[4], [9], the ambiguity set is defined in an exogenous manner independent of the current state and action. While this is a reasonable assumption when the state evolution is uncertain in a parametric manner (e.g., the state transition being governed by known dynamics affected by a parametric uncertainty or additive disturbance), the more general case of state evolution given by a stochastic transition kernel necessitates defining the stochastic state evolution and the ambiguity associated with it as a function of the current state and chosen action. This class of ambiguity sets are referred to as *decision dependent ambiguity sets* and are often not amenable for tractable reformulations [10], [11].

In this paper, we leverage the framework of Hilbert space embedding of conditional distributions [12] to define the ambiguity set associated with the transition kernel. While conditional mean embedding and its empirical estimate have been applied in the context of Bayesian inference [13], dynamical systems [14] and more recently for reachability analysis [15], [16], we are not aware of any work that leverages this framework for control synthesis (of safety-critical systems). In fact, distributionally robust optimization (DRO) subject to ambiguity sets defined via kernel mean embedding have only been studied recently in [17] where the authors established the strong duality result for this class of problems. Kernel DRO problems where the ambiguity set is defined via the conditional mean embedding has not received much attention; with [18] being the only exception.

In this paper, we build upon the above line of work and treat the transition probability associated with state evolution as a conditional distribution that depends on the chosen state and action. Following [12], the expectation of any function of the subsequent state can be viewed as a linear function evaluation of the function and the conditional mean embedding in the underlying Hilbert space. When the transition probability is not known, rather we have access to state-input trajectories, the empirical estimate of the conditional mean embedding has been used to evaluate the expectation operator in [15], [16]. However, when the number of samples is not sufficiently large, the empirical estimate may not be rich enough to approximate the (true) conditional mean embedding sufficiently well, thus undermining its use in safety-critical control applications.

In order to robustify this approach, we consider a distributionally robust or min-max control problem where the transition probabilities are assumed to reside in an ambiguity set that contains all distributions whose kernel mean embedding are within a certain distance from the empirical estimate of the conditional mean embedding. Following a similar approach as [2], [5], [9], we show that there exists a non-randomized Markov policy which is optimal and then discuss how to compute an optimal control input via value iteration by leveraging duality results associated with Kernel DRO problems [17]. We then formulate the problem of control synthesis subject to safety specifications within the proposed framework. Numerical results on benchmark problems provide valuable insights into the performance of

L. Romao and A. Abate are with the Department of Computer Science, Oxford University, UK. Email adresses: {licio.romao,aabate}@cs.ox.ac.uk. A. R. Hota is with the Department of Electrical Engineering, Indian Institute of Technology, Kharagpur, India. Email: ahota@ee.iitkgp.ac.in.

Manuscript 976 submitted to 2023 62nd IEEE Conference on Decision and Control (CDC) . Received March 31, 2023.

the proposed formulation.

II. Preliminaries

A. Reproducing kernel Hilbert spaces (RKHS) and kernel mean embeddings

Let $(\mathcal{X}, \mathcal{F}_X)$ be a measurable space, where $\mathcal{X}$ is an abstract set and $\mathcal{F}_X$ represents a σ-algebra on $\mathcal{X}$. Consider a measurable function $k : \mathcal{X} \times \mathcal{X} \mapsto \mathbb{R}$, called a *kernel*, that satisfies the following properties.

- *Boundedness*: For any $x \in \mathcal{X}$, we have that $\sup_x |k(x,x)| < \infty$.
- *Symmetry*: For any $x, x' \in \mathcal{X}$, we have $k(x,x') = k(x',x)$;
- *Positive semidefinite (PSD)*: For any finite collection of points $(x_i)_{i=1}^m$, where $x_i \in \mathcal{X}$ for all $i = 1, \ldots, m$, and for any vector $\alpha \in \mathbb{R}^m$, we have that

$$\sum_{i,j=1}^{m} \alpha_i \alpha_j k(x_i, x_j) \geq 0.$$

  In other words, the Gram matrix K whose (i,j)-th entry is given by $k(x_i, x_j)$ is a positive semidefinite matrix for any choice of points $(x_i)_{i=1}^m$.

A kernel function k satisfying the above three properties is called a *positive definite* kernel.

Two consequences are in place with the presence of a positive definite kernel. First, every positive definite kernel is associated with a *reproducing kernel Hilbert space* (RKHS) $\mathcal{H}_\mathcal{X}$, which is defined as[1]

$$\mathcal{H}_\mathcal{X} = \overline{\{k(x,\cdot) : \mathcal{X} \mapsto \mathbb{R}, \text{ for all } x \in \mathcal{X}\}}, \tag{1}$$

with an inner product given by $\langle k(\cdot, x_1), k(\cdot, x_2)\rangle_{\mathcal{H}_\mathcal{X}} = k(x_1, x_2)$. We may notice that the boundedness, symmetry and PSD properties above ensure that such an inner product is well-defined; hence, $\mathcal{H}_\mathcal{X}$ equipped with this inner is a Hilbert space (see [19] for more details). By definition, for any function $f \in \mathcal{H}_\mathcal{X}$ there exists a sequence of functions $f_n(x) = \sum_{i=1}^m \beta_i^n k(x_i^n, x)$ such that $f(x) = \lim_{n\to\infty} f_n(x)$, which then implies that

$$\langle f, k(\cdot, x)\rangle_{\mathcal{H}_\mathcal{X}} = f(x),$$

thus justifying the reproducing kernel property of the Hilbert space $\mathcal{H}_\mathcal{X}$. Alternatively, RKHS can be characterized by the reproducing kernel property and Riesz representation theorem (Theorem 4.11 in [19]). Details are omitted for brevity, but the interested reader is referred to [20] and references therein for more details.

Second, there exists a feature map $\phi : \mathcal{X} \mapsto \mathcal{H}_\mathcal{X}$ with the property $k(x_1, x_2) = \langle \phi(x_1), \phi(x_2)\rangle_{\mathcal{H}_\mathcal{X}}$. The canonical feature map is given by $\phi(x) := k(x, \cdot)$. The inner product defined earlier induces a norm over the RKHS defined as $||f(\cdot)||_{\mathcal{H}_\mathcal{X}} := \sqrt{\langle f(\cdot), f(\cdot)\rangle_{\mathcal{H}_\mathcal{X}}}$ for all $f \in \mathcal{H}_\mathcal{X}$.

[1]For separable spaces $\mathcal{X}$, which are topological spaces with a countable dense subset, the Hilbert space in (1) could be defined using any (countable) dense subset of $\mathcal{X}$. Details are omitted for brevity.

We now introduce the notion of *kernel mean embedding* of probability measures [12], [20]. Let $\mathcal{P}(\mathcal{X})$ be the set of probability measures on $\mathcal{X}$. Let X be a random variable defined on $\mathcal{X}$ with distribution $\mathbb{P}$. The kernel mean embedding is a mapping $\Psi : \mathcal{P}(\mathcal{X}) \mapsto \mathcal{H}_\mathcal{X}$ defined as

$$\Psi(\mathbb{P})(\cdot) := \mathbb{E}_\mathbb{P}[\phi(X)] = \int_\mathcal{X} k(x,\cdot) d\mathbb{P}(x). \tag{2}$$

We have the following result from [12], [20] on the reproducing property of the expectation operator in the RKHS.

Lemma 1 (Lemma 3.1 [12]). *If $\mathbb{E}_\mathbb{P}[\sqrt{k(X,X)}] < \infty$, then $\Psi(\mathbb{P}) \in \mathcal{H}_\mathcal{X}$ and $\mathbb{E}_\mathbb{P}[f(X)] = \langle f, \Psi(\mathbb{P})\rangle_{\mathcal{H}_\mathcal{X}}$.*

In other words, $\Psi(\mathbb{P})$ is indeed an element of the RKHS $\mathcal{H}_\mathcal{X}$, and the expectation of any function of the random variable X can be computed by means of an inner product of the function with the kernel mean embedding. If a collection of i.i.d. samples $\{\widehat{x}_i\}_{i=1}^M$ drawn from the distribution $\mathbb{P}$ of the random variable X are given, then the empirical estimate of the kernel mean embedding can be defined as

$$\widehat{\Psi}(\mathbb{P})(\cdot) := \frac{1}{M}\sum_{i=1}^M \phi(\widehat{x}_i) = \frac{1}{M}\sum_{i=1}^M k(\widehat{x}_i, \cdot). \tag{3}$$

In other words, $\widehat{\Psi}(\mathbb{P})$ is the mean embedding of the empirical distribution $\widehat{\mathbb{P}}_M := \frac{1}{M}\sum_{i=1}^M \delta_{\widehat{x}_i}$ induced by the samples.

B. Kernel-based ambiguity sets

We can leverage the kernel mean embedding to define a metric or distance between two distributions $\mathbb{P}$ and $\mathbb{Q}$ as $||\Psi(\mathbb{P}) - \Psi(\mathbb{Q})||_{\mathcal{H}_\mathcal{X}}$ which is called the *Maximum Mean Discrepancy* (MMD) distance. Due to the reproducing property, it can be shown that

$$\begin{aligned}\texttt{MMD}(\mathbb{P}, \mathbb{Q}) &= ||\Psi(\mathbb{P}) - \Psi(\mathbb{Q})||_{\mathcal{H}_\mathcal{X}} \\ &= \mathbb{E}_{\xi,\xi'\sim\mathbb{P}}[k(\xi,\xi')] + \mathbb{E}_{\omega,\omega'\sim\mathbb{Q}}[k(\omega,\omega')] \\ &\quad - 2\mathbb{E}_{\xi\sim\mathbb{P},\omega\sim\mathbb{Q}}[k(\xi,\omega)].\end{aligned} \tag{4}$$

In this work, we consider data-driven MMD ambiguity sets induced by observed samples $\{\widehat{x}_i\}_{i\in[N]}$ defined as

$$\texttt{MMD}_N^\theta := \{\mathbb{P} \in \mathcal{P}(\mathcal{X}) | \|\Psi(\mathbb{P}) - \Psi(\widehat{\mathbb{P}}_N)\|_{\mathcal{H}_\mathcal{X}} \leq \theta\}, \tag{5}$$

where $\widehat{\mathbb{P}}_M := \frac{1}{M}\sum_{i=1}^M \delta_{\widehat{x}_i}$ as defined earlier. Thus, $\texttt{MMD}_N^\theta$ contains all distributions whose kernel mean embedding is within distance $\theta \geq 0$ of the kernel mean embedding of the empirical distribution. The above ambiguity set also enjoys a sharp finite sample guarantee as shown in [21]. If the observed samples are drawn i.i.d. from an underlying distribution $\mathbb{P}_0$, then with probability $1-\delta$, we have

$$\texttt{MMD}(\mathbb{P}_0, \widehat{\mathbb{P}}) \leq \sqrt{\frac{C}{N}} + \sqrt{\frac{2C\log(1/\delta)}{N}}$$

where C is a constant such that $\sup_x k(x,x) \leq C \leq \infty$.

Manuscript 976 submitted to 2023 62nd IEEE Conference on Decision and Control (CDC) . Received March 31, 2023.

C. RKHS embedding of conditional distributions

We now consider random variables of the form (Y, X) taking values over the space $\mathcal{Y} \times \mathcal{X}$. Let $\mathcal{H}_\mathcal{Y}$ be the RKHS of real valued functions defined on $\mathcal{Y}$ with positive definite kernel $k_\mathcal{Y} : \mathcal{Y} \times \mathcal{Y} \mapsto \mathbb{R}$ and feature map $\phi_\mathcal{Y} : \mathcal{Y} \mapsto \mathcal{H}_\mathcal{Y}$. We define the cross-covariance operator $C_{XY} : \mathcal{H}_\mathcal{Y} \mapsto \mathcal{H}_\mathcal{X}$ as

$$C_{XY} := \mathbb{E}_{\mathbb{P}_{XY}}[\phi_\mathcal{Y}(Y) \otimes \phi_\mathcal{X}(X)], \tag{6}$$

where $\otimes$ denotes the tensor product[2] and $\mathbb{P}_{XY}$ is the joint distribution of (Y, X). For any $f \in \mathcal{H}_\mathcal{X}, g \in \mathcal{H}_\mathcal{Y}$, the cross-covariance operator satisfies

$$\langle f, C_{XY} g\rangle_{\mathcal{H}_\mathcal{X}} = \mathrm{cov}[g(Y), f(X)].$$

When $Y = X$, the analogous operator C_{YY} is called the covariance operator. We now formally define the conditional mean embedding of the conditional distribution $\mathbb{P}(X|Y)$.

Definition 1 (Definition 4.1 [12])**.** *The conditional mean embedding of the conditional distribution $\mathbb{P}(X|Y = y)$, denoted $\mathcal{U}_{X|y} \in \mathcal{H}_\mathcal{X}$ is defined as*

$$\mathcal{U}_{X|y} := \mathbb{E}_{X|y}[\phi_\mathcal{X}(X)|Y = y] = C_{XY} C_{YY}^{-1} k_\mathcal{Y}(y, \cdot),$$

and for any $f \in \mathcal{H}_\mathcal{X}$, we have

$$\mathbb{E}_{X|y}[f(X)|Y = y] = \langle f, \mathcal{U}_{X|y}\rangle_{\mathcal{H}_\mathcal{X}}.$$

More generally, the conditional mean embedding for the distribution $\mathbb{P}(X|Y)$, denoted $\mathcal{U}_{X|Y} : \mathcal{H}_\mathcal{Y} \mapsto \mathcal{H}_\mathcal{X}$ is defined as

$$\mathcal{U}_{X|Y} := C_{XY} C_{YY}^{-1},$$

and satisfies $\mathcal{U}_{X|y} = \mathcal{U}_{X|Y} k_\mathcal{Y}(y, \cdot)$.

In other words, $\mathcal{U}_{X|Y}$ is a mapping from the RKHS associated with the conditioned variable to the RKHS associated with the observed variable. When the conditioned variable $Y = y$ is specified, $\mathcal{U}_{X|y}$ gives a specific element within the RKHS $\mathcal{H}_\mathcal{X}$ which satisfies the reproducing property of the conditional expectation operator.

In many applications, the joint or conditional distributions involving X and Y are not known, rather we have access to i.i.d. samples $\{(\widehat{x}_i, \widehat{y}_i)\}_{i=1}^M$ drawn from the joint distribution $\mathbb{P}_{XY}$. Let $K_\mathcal{Y} \in \mathbb{R}^{M \times M}$ be the gram matrix associated with $\{\widehat{y}_i\}_{i=1}^M$ with its (i, j)-th entry given by $[K_\mathcal{Y}]_{ij} = k_\mathcal{Y}(\widehat{y}_i, \widehat{y}_j)$. The empirical estimate of the conditional mean embedding $\mathcal{U}_{X|y}$ was derived in [22] and is stated below.

Theorem 1 (Theorem 4.2 [12])**.** *An empirical estimate of the conditional mean embedding $\mathcal{U}_{X|y}$ is given by*

$$\widehat{\mu}_{X|y} = \sum_{i=1}^{M} \beta_i(y) k_\mathcal{X}(x_i, \cdot), \tag{7}$$

where $\beta = (K_\mathcal{Y} + M\lambda \mathbf{I}_M)^{-1} k_\mathbf{y}(y) \in \mathbb{R}^M$ with $k_\mathbf{y}(y) = [k_\mathcal{Y}(\widehat{y}_1, y) \quad k_\mathcal{Y}(\widehat{y}_2, y) \quad \ldots k_\mathcal{Y}(\widehat{y}_M, y)]^\top \in \mathbb{R}^M$, $\mathbf{I}_M$ being the identity matrix of dimension M and $\lambda > 0$ being the regularization parameter.

[2]Formally, this is the product between two functions that are member of the tensor product between $\mathcal{H}_\mathcal{Y}$ and $\mathcal{H}_\mathcal{Y}$.

The above empirical estimate can also be obtained by solving a regularized regression problem as established in [23], [24].

III. Kernel Distributionally Robust Optimal Control

We now introduce the distributionally robust optimal control problem. Let $x \in \mathcal{X} \subset \mathbb{R}^n$ and $a \in \mathcal{A}(x) \subset \mathbb{R}^p$ denote the state and control input of a discrete-time stochastic dynamical system with $\mathcal{A}(x)$ being the set of admissible control inputs at state x. We are interested in the paths generated by the discrete-time stochastic process

$$x_{k+1} \sim T_t(\cdot|x_k, a_k), \quad a_k \in \mathcal{A}(x_k), \quad x_0 = \bar{x}, \tag{8}$$

where the next state is sampled from the stochastic kernel or probability measure defined by the pair (x_k, a_k), denoted by $T_k(\cdot|x_k, a_k)$. Let H_k denote the set of histories up to time k with elements of the form $h_k = (x_0, u_0, T_0, \ldots, x_{k-1}, u_{k-1}, T_{k-1}, x_k)$.

We consider a finite horizon optimal control problem in this work. The set of admissible control policies over a horizon of length L is given by $\Pi = \{(\pi_0, \pi_1, \ldots, \pi_{L-1}) | \pi_k(\mathcal{A}(x_k)|h_k) = 1, \forall h_k \in H_k\}$ where $\mathcal{A}(x_k)$ is the set of admissible control inputs at state x_k. We assume that the transition probability T_k is not known with certainty, rather it belongs to an *ambiguity set* possibly dependant on the current state and input chosen by the controller. Formally, we define $\Gamma := \{(T_0, T_1, \ldots, T_{L-1}) | T_k(\cdot|x_k, a_k) \in \mathcal{M}_{(x_k, a_k)}\}$, where $\mathcal{M}_{(x_k, a_k)}$ denotes a set of distributions or ambiguity set.

In this work, we define the ambiguity set at (x, a) to be the set of all distributions whose Hilbert space embedding is within MMD distance ϵ from the empirical estimate of the conditional mean embedding, i.e.,

$$\mathcal{M}^\epsilon_{(x,a)} := \{\mathbb{P} \in \mathcal{P}(\mathcal{X}) | \|\Psi(\mathbb{P}) - \widehat{\mu}_{X|(x,a)}\|_{\mathcal{H}_\mathcal{X}} \leq \epsilon\}. \tag{9}$$

In the terminology of the previous subsection, the input space $\mathcal{Y}$ is the Cartesian product between state and control input $\mathcal{X} \times \mathcal{A}(\mathcal{X})$ and the output space $\mathcal{X}$ is the state space $\mathcal{X}$. Each of these spaces are associated with a suitable continuous positive definite kernels denoted by $k_\mathcal{Y} : (\mathcal{X}, \mathcal{A}(\mathcal{X})) \times (\mathcal{X}, \mathcal{A}(\mathcal{X})) \to \mathbb{R}$ and $k_\mathcal{X} : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$, respectively. We omit the dependence on ϵ in the following analysis for brevity of notation.

For a given finite-horizon policy $\pi \in \Pi$, sequence of transitions kernels $T \in \Gamma$ and initial state x_0, we denote the distribution over the trajectories generated by (8) by $\mathbb{P}^{\pi,T}_{x_0}$. For a given stage cost $c : \mathbb{R}^n \times \mathbb{R}^p \to \mathbb{R}$, we define the finite horizon expected cost as

$$V_T(x_0, \pi, T) := \mathbb{E}^{\pi,T}_{x_0}\left[\sum_{k=0}^{L-1} c(x_k, a_k)\right], \tag{10}$$

where $\mathbb{E}^{\pi,T}_{x_0}$ denotes the expectation operator with respect to $\mathbb{P}^{\pi,T}_{x_0}$. Our goal is to find a policy $\pi^\star \in \Pi$ that solves the distributionally robust control problem given by

$$\inf_{\pi \in \Pi} \sup_{T \in \Gamma} V_T(x_0, \pi, T). \tag{11}$$

Following [5], we impose the following regularity assumptions on our problem[3].

Assumption 1. *Let $\mathbb{K} \in \mathcal{B}(\mathbb{R}^n \times \mathbb{R}^n \times \mathcal{P}(\mathbb{R}^n))$ be the set containing elements (x, a, T) satisfying $x \in \mathcal{X}, a \in \mathcal{A}(x)$ and $T \in \mathcal{M}_{(x,a)}$. The following conditions hold.*

1) *The stage cost function $c(x, a)$ is lower semicontinuous, and there exists a constant $\bar{c} \geq 0$ and a continuous function w defined on $\mathcal{X}$ satisfying $w(x) \geq 1, \forall x \in \mathcal{X}$ such that*

$$|c(x, a)| \leq \bar{c} w(x), \quad \forall a \in \mathcal{A}(x),\ x \in \mathcal{X}.$$

2) *The transition kernels are weakly continuous, i.e., for every bounded, continuous function $u : \mathcal{X} \to \mathbb{R}$, the function $\widehat{u}(x, a, T) := \int_{\mathcal{X}} u(y)T(dy)$ is continuous on $\mathbb{K}$.*
3) *The function $\widehat{w}(x, a, T) := \int_{\mathcal{X}} w(y)T(dy)$ is continuous on $\mathbb{K}$ and there exists a constant $\beta > 0$ such that $\widehat{w}(x, a, T) \leq \beta w(x)$ for all $(x, a, T) \in \mathbb{K}$.*
4) *The set $\mathcal{A}(x)$ is compact for each $x \in \mathcal{X}$. Furthermore, the set-valued mapping $x \mapsto \mathcal{A}(x)$ is upper semi-continuous.*
5) *The ambiguity set is as defined in (9).*

Before stating the main result, which is inspired by the analysis in [5], we introduce the relevant terminology. Let $\mathcal{B}_w(\mathcal{X})$ denote the Banach space of measurable functions u defined on $\mathcal{X}$ with finite w-norm, i.e., $||u||_w := \sup_{x \in \mathcal{X}} \frac{u(x)}{w(x)} < \infty$. For each $u \in \mathcal{B}_w(\mathcal{X})$, and $(x, a, T) \in \mathbb{K}$, we define

$$H(u; x, a, T) := c(x, a) + \int_X u(y)T(dy), \tag{12}$$

$$H^{\#}(u; x, a) := \sup_{T \in \mathcal{M}_{(x,a)}} H(u; x, a, T), \tag{13}$$

$$\begin{aligned}\mathcal{T}(u)(x) &:= \inf_{a \in \mathcal{A}(x)} H^{\#}(u; x, a) \\ &= \inf_{a \in \mathcal{A}(x)} \sup_{T \in \mathcal{M}_{(x,a)}} \left[c(x, a) + \int_X u(y)T(dy) \right].\end{aligned} \tag{14}$$

Specifically, (14) defines the distributionally robust dynamic programming (DP) operator under MMD ambiguity set centered at the empirical conditional mean embedding. The value function of the distributionally robust control problem can be defined iteratively as

$$\begin{aligned} v_L(x) &:= 0, \\ v_k(x) &:= (\mathcal{T} v_{k+1})(x) = (\mathcal{T} \circ \mathcal{T} \ldots \circ \mathcal{T})(v_L)(x), \end{aligned} \tag{15}$$

for $0 \leq t \leq L - 1$. We now state the following result which shows that the problem (11) admits a non-randomized Markov policy which is optimal.

[3]Due to the fact that we are dealing with some infinite-dimensional spaces, we rely on the topological notion of continuity. A function between two topological spaces (please refer to [25] for an introduction to these concepts) $(\mathcal{Y}, \tau_{\mathcal{Y}})$ and $(\mathcal{X}, \tau_{\mathcal{X}})$, $f : \mathcal{Y} \mapsto \mathcal{X}$ is continuous if for all $U \in \tau_{\mathcal{X}}$ we have that $f^{-1}(U) \in \mathcal{X}$. The notions of weakly continuous, weakly compact, etc, are used due to the fact that we equip the infinite-dimensional spaces $\mathcal{P}(\mathcal{X})$ and $\mathcal{H}_{\mathcal{X}}$ with the weak* topology. We refer the reader to [19], Chapter 4, for more details about these concepts.

Theorem 2. *Suppose Assumption 1 holds. Then, v_k is lower semi-continuous for $k \in \{0, 1, \ldots, L-1\}$. Further, there exists a function f_k on $\mathcal{X}$ such that $v_k(x) = H^{\#}(v_{k+1}; x, f_k(x))$ and the Markov policy $(f_0, f_1, \ldots, f_{L-1})$ is the optimal solution to the distributionally robust control problem (11).*

Proof. We follow a similar approach as the proof of [5, Theorem 3.1] and [9, Theorem 1]. The primary challenge is to show that the DP operator defined in (15) preserves the lower semi-continuity of the value function. We show this via induction. Let v_{k+1} be lower semicontinuous. Following identical arguments as [5, Lemma 3.3], it can be shown that $H(v_{k+1}; x, a, T)$ is lower semicontinuous on $\mathbb{K}$.

The next step is to show that $H^{\#}(v_{k+1}; x, a)$ is lower semicontinuous over $\mathcal{X} \times \mathcal{A}(\mathcal{X})$. To this end, we need to establish that the mapping $(x, a) \mapsto \mathcal{M}_{(x,a)}$ is weakly compact and lower semicontinuous (i.e., the condition analogous to [5, Assumption 3.1(g)] holds for the ambiguity set (9)).

To show weak compactness of $\mathcal{M}_{(x,a)}$, let us first study properties of the set

$$\mathcal{C}_{(x,a)} = \{f \in \mathcal{H}_{\mathcal{X}} : \|f - \widehat{\mu}_{X|(x,a)}\| \leq \epsilon\},$$

which is a subset of the RKHS $\mathcal{H}_{\mathcal{X}}$. It is clear that this set is convex, hence, by [19, Theorem 3.7], it is also weakly closed. Since Hilbert spaces are reflexive Banach spaces, then by Kakutani's theorem (see [19, Theorem 3.17]) we show that the set $\mathcal{C}_{(x,a)}$ is weakly compact. Now, notice that

$$\mathcal{M}_{(x,a)} = \Psi^{-1}(\mathcal{C}_{(x,a)}),$$

where we recall that the mapping $\Psi : \mathcal{P}(\mathcal{X}) \mapsto \mathcal{H}_{\mathcal{X}}$ is continuous, thus weakly continuous. Since Ψ is also surjective[4], we have by the open mapping theorem ([19, Theorem 2.6]) that $\mathcal{M}_{(x,a)}$ is also weakly compact.

We now show that the mapping $(x, a) \mapsto \mathcal{M}_{(x,a)}$ is lower semicontinuous. We define the distance function from a distribution T to a subset S of $\mathcal{M}_{(x,a)}$ as[5]

$$d(T, S) := \inf_{\xi \in S} ||\Psi(T) - \Psi(\xi)||_{\mathcal{H}_{\mathcal{X}}}.$$

Let $(x, a, T) \in \mathbb{K}$, i.e., $a \in \mathcal{A}(x)$ and $T \in \mathcal{M}_{(x,a)}$. Thus,

$$||\Psi(T) - \widehat{\mu}_{X|(x,a)}||_{\mathcal{H}_{\mathcal{X}}} \leq \epsilon.$$

Consider a sequence $(x_n, a_n)_{n \geq 0}$ with $a_n \in \mathcal{A}(x_n), \forall n \geq 0$ and $\lim_{n \to \infty}(x_n, a_n) = (x, a)$. From [26, Proposition 1.4.7], it follows that the lower semi-continuity of $(x, a) \mapsto \mathcal{M}_{(x,a)}$ is equivalent to

$$T \in \liminf_{n \to \infty} \mathcal{M}_{(x_n, a_n)} \iff \lim_{n \to \infty} d(T, \mathcal{M}_{(x_n, a_n)}) = 0.$$

[4]For any function $f \in \mathcal{H}_{\mathcal{X}}$ there exists a sequence $f_n(x) = \sum_{i=1}^m \beta_i(n) k(x_i, x)$ such that $\lim_{n \to \infty} f_n(x) = f(x)$. Let $\mathbb{P}_n = \sum_{i=1}^m \beta_i(n) \delta_{x_i}$ and notice that $\Psi(\mathbb{P}) = \Psi(\lim_{n \to \infty} \mathbb{P}_n) = f$, thus showing that Ψ is a surjective mapping.

[5]This distance is only well-defined due to the weak compactness result shown above, as it would allow us to take convergent subsequences for any sequence achieving the infimum in this definition.

To this end, we compute

$$\begin{aligned}
&||\Psi(T) - \widehat{\mu}_{X|(x_n,a_n)}||_{\mathcal{H}_\mathcal{X}} \leq ||\Psi(T) - \widehat{\mu}_{X|(x,a)}||_{\mathcal{H}_\mathcal{X}} \\
&\qquad + ||\widehat{\mu}_{X|(x,a)} - \widehat{\mu}_{X|(x_n,a_n)}||_{\mathcal{H}_\mathcal{X}} \\
\implies &\lim_{n\to\infty} ||\Psi(T) - \widehat{\mu}_{X|(x_n,a_n)}||_{\mathcal{H}_\mathcal{X}} \leq \epsilon \\
&\qquad + \lim_{n\to\infty} ||\widehat{\mu}_{X|(x,a)} - \widehat{\mu}_{X|(x_n,a_n)}||_{\mathcal{H}_\mathcal{X}} \\
\implies &\lim_{n\to\infty} ||\Psi(T) - \widehat{\mu}_{X|(x_n,a_n)}||_{\mathcal{H}_\mathcal{X}} \leq \epsilon \\
\implies &\lim_{n\to\infty} d(T, \mathcal{M}_{(x_n,a_n)}) = 0,
\end{aligned}$$

from the definition of the ambiguity set. In the second last step, the second term goes to 0 as $\lim_{n\to\infty}(x_n, a_n) = (x, a)$ due to the continuity of the kernel function and the definition of $\widehat{\mu}_{X|(x,a)}$ in (7). As a result, $T \in \liminf_{n\to\infty} \mathcal{M}_{(x_n,a_n)}$ and thus, $\mathcal{M}_{(x,a)} \subseteq \liminf_{n\to\infty} \mathcal{M}_{(x_n,a_n)}$.

With the above properties of the mapping $(x, a) \mapsto \mathcal{M}_{(x,a)}$ in hand, it can be shown following identical arguments as the proof of [5, Theorem 3.1] that both $H^{\#}(v_{k+1}; x, a)$ and $\mathcal{T}(v_{k+1})$ are lower semicontinuous and there exists a function $f_k : \mathcal{X} \to \mathcal{A}(x)$ such that $a_k = f_k(x_k)$ is the minimizer of (14) for $u = v_{k+1}$. This concludes the proof. □

Remark 1. Note that the ambiguity sets considered in related works on distributionally robust control [2], [9] do not depend on the current state and action, i.e., the mapping $(x, a) \mapsto \mathcal{M}_{(x,a)}$ is independent of (x, a), and as a result, properties such as compactness and lower semi-continuity are easily shown. In contrast, the ambiguity set considered here is more general and directly captures the dependence of the transition kernel on the current state-action pair. In the DRO literature, such ambiguity sets are referred to as *decision-dependent ambiguity sets* [10], [11], [27], which are relatively challenging to handle, and less explored in the past.

The above theorem establishes that a non-randomized Markovian policy solves the min-max optimal control problem. We now discuss how to solve for optimal control inputs. We adopt a value iteration approach where starting from the value function v_{k+1}, we compute v_k by solving (14). To this end, at a given state x, we discretize the input space $\mathcal{A}(x)$ as $\{a^{(1)}, a^{(2)}, \ldots, a^{(N)}\}$ and compute

$$v_k(x) = \min_{a^{(j)}: j=1,\ldots,N} c(x, a^{(j)}) + \sup_{\gamma \in \mathcal{M}^{\epsilon}_{(x,a^{(j)})}} \int_X v_{k+1}(y)\gamma(dy).$$

For a given $(x, a^{(j)})$, the inner supremum problem is an instance of a Kernel DRO problem (subject to MMD ambiguity sets) which we reformulate below building upon the results in [17].

In the following discussion, we suppress the dependence of the ambiguity set on the radius ϵ and on (x, a) for better readability. Note that the inner supremum problem can be stated as

$$\sup_{T\in\mathcal{M}} \mathbb{E}_{X\sim T}[u(X)] = \sup_{T\in\mathcal{M}} \langle T, u\rangle = \sup_{\Psi(T)\in\mathcal{C}} \langle u, \Psi(T)\rangle_{\mathcal{H}_\mathcal{X}}.$$

following the reproducing property of the underlying RKHS with the set $\mathcal{C} := \{f \in \mathcal{H}_\mathcal{X} | ||f - \widehat{\mu}_{X|(x,a)}||_{\mathcal{H}_\mathcal{X}} \leq \epsilon\} \subseteq \mathcal{H}_\mathcal{X}$. The right-hand side is essentially the support function $\sigma_\mathcal{C}(u)$.

Following [17], the support of $\mathcal{C}$ can then be computed as

$$\begin{aligned}
\sigma_\mathcal{C}(u) &= \sup_{f\in\mathcal{C}} \langle u, f\rangle_{\mathcal{H}_\mathcal{X}} \\
&= \sup_{f: ||f-\widehat{\mu}_{X|(x,a)}||_{\mathcal{H}_\mathcal{X}} \leq \epsilon} \langle u, f - \widehat{\mu}_{X|(x,a)}\rangle_{\mathcal{H}_\mathcal{X}} + \langle u, \widehat{\mu}_{X|(x,a)}\rangle_{\mathcal{H}_\mathcal{X}} \\
&= \epsilon||u||_{\mathcal{H}_\mathcal{X}} + \sum_{i=1}^{M} \beta_i(x, a)u(x_i),
\end{aligned}$$

where $\beta_i(x, a)$ denote the coefficients of the empirical estimate of the conditional mean embedding as given in (7). To compute the norm $\|u\|_{\mathcal{H}_\mathcal{X}}$ we solve, similar to Theorem 1, a regression problem given by

$$\underset{\alpha_i, i=1,\ldots,m}{\text{minimise}} \quad \|\sum_{i=1}^{m} \alpha_i k(x_i, \cdot) - u\|_{\mathcal{H}_\mathcal{X}} + \lambda\|\alpha\|_2^2, \tag{16}$$

where $\{x_1, \ldots, x_m\}$ are arbitrary points in the domain of the function u. The solution of this regression problem is given by $\|u\|_{\mathcal{H}_\mathcal{X}} = \sqrt{\alpha' K_\mathcal{X} \alpha}$, where

$$\alpha = (K_\mathcal{X} + \lambda I)^{-1} \begin{bmatrix} u(x_1) & \ldots & u(x_m) \end{bmatrix}^\top.$$

Variable $K_\mathcal{X}$ is the Gram matrix associated with the available points, namely, it is the positive semi-definite matrix whose (i, j)-entry is given by $k_\mathcal{X}(x_i, x_j)$. Notice that the collection of points used to estimate $\|u\|_{\mathcal{H}_\mathcal{X}}$ may not necessarily coincide with those points used to estimate the conditional mean embedding in Theorem 1. Same comment applies for the regularizer λ appearing in (16) and in the expression of β in Theorem 1.

A. Distributionally Robust Safe Control

While the discussion thus far has focused on the general problem of optimal control, this approach can also be leveraged for synthesizing control inputs meeting safety specifications. Let $S \subset \mathbb{R}^n$ be a measurable safe set. For an admissible policy $\pi = (\pi_0, \pi_1, \ldots, \pi_{L-1})$, and transition kernels $(T_0, T_1, \ldots, T_{L-1})$ and initial state x_0, the probability of the state trajectory being safe is given by

$$V(S; \pi, T, x_0) = \mathbb{P}^{\pi,T}_{x_0}\{x_k \in S, \text{ for all } k \in \{1, \ldots, L\} | x_0\}, \tag{17}$$

$$= \mathbb{E}^{\pi,T}_{x_0}[\Pi_{k=1}^{L} \mathbf{1}_S(x_k) | x_0], \tag{18}$$

where $(x_1, x_2, \ldots, x_L)$ denote the solution of (8) and $\mathbf{1}_S$ is the indicator function of the safe set S.

We aim to choose a policy that maximizes the safety probability (17) in the worst came among all distributions in the ambiguity set, that is, to solve the problem

$$V^\star(S; x_0) = \sup_{\pi\in\Pi} \inf_{T\in\Gamma} V(S; \pi, T, x_0), \tag{19}$$

and find a distributionally robust safe policy $\pi^* \in \Pi$ which satisfies

$$\inf_{T\in\Gamma} V(S; \pi^*, T, x_0) \geq \inf_{T'\in\Gamma} V(S; \pi, T', x_0), \forall \pi \in \Pi. \tag{20}$$

We define the dynamic programming operator as

$$\mathcal{T}(v)(x) = \sup_{a\in\mathcal{A}(x)} \inf_{T\in\mathcal{M}^{\epsilon}_{(x,a)}} \mathbf{1}_S(x) \int_X v(y)T(dy). \tag{21}$$

The value function of the distributionally robust safety problem can be defined recursively as

$$v_L(x) := \mathbf{1}_S(x),$$
$$v_k(x) := (\mathcal{T} \circ \mathcal{T} \circ \ldots \circ \mathcal{T})(v_L)(x), \tag{22}$$

for $0 \leq k \leq L-1$. We now state the following result, which is analogous to the earlier theorem, for the safe control synthesis problem.

Theorem 3. *Suppose Assumption 1 holds. Then, there exists a Markov policy $(f_0, f_1, \ldots, f_{L-1})$ which is an optimal solution to the distributionally robust safe control problem (19).*

The proof is analogous to the proof of Theorem 2 and is omitted in the interest of space. The value function in this case is bounded in the range $[0, 1]$. In the numerical example reported in the following section, we compute the safe control inputs by solving the inner problem in an identical manner as discussed in the previous subsection.

IV. Numerical examples

Inspired by papers [7], [9], we apply our methods to study safety probability of a thermostatically controlled load given by the dynamics

$$x(k+1) = \alpha x(k) + (1-\alpha)(\theta - \eta R P u(k)) + \omega(k), \tag{23}$$

where the state $x_k \in \mathbb{R}$ is the temperature, $u_k \in \{0, 1\}$ is a binary control input, representing whether the load is on or off, and ω_k is a stochastic disturbance taking values in the uncertainty space $(\Omega, \mathcal{F}, \mathbb{P})$. The parameters of (23) are given by $\alpha = \exp(h/CR)$, where $R = 2°\text{C/kW}$, $C = 2\text{kWh}/°C$, $\theta = 32°\text{C}$, $h = 5/60$ hour, $P = 14\text{kW}$, and $\eta = 0.7$. Our goal is keep the temperature within the range $\mathcal{S} = [19°\text{C}, 22°C]$ for 90 minutes.

Our goal is to compute a control policy purely on available sampled trajectories for the model (23) and without solving expensive optimisation problems at each iteration. To this end, we let $S = (x_i, u_i, x_{i,+})_{i=1}^M$ be a collection of observed transitions from the model, where the pair (x_i, u_i) is the set of random chosen points in the set $[19, 22] \times \{0, 1\}$ and $x_{i,+}$ represents the observed transitions from such a state-input pair[6]. We then solve the dynamic programming recursion given in (21) and (22) by partitioning the state space uniformly from 18°C to 23°C with 35 points, and using 7000 data points to estimate the conditional kernel mean embedding map (see Theorem 1) and to compute the norm of the value function, as shown in (16). We choose $\lambda = 200$ as the regularisation parameter, and use the kernel function $k : (\mathcal{X} \times \mathcal{U})^2 \mapsto \mathbb{R}$, defined as

$$k((x,u),(x',u')) = e^{-\gamma|x-x'|^2} + k_1(u,u'),$$

where $k_1(u,u') = 1 + uu' + uu' \min(u,u') - \frac{u+u'}{2}\min(u,u')^2 + \frac{1}{3}\min(u,u')^3$, for the numerical examples. The choice of the kernel k_1 has shown better results for this problem when compared with a Gaussian kernel.

[6]We report the dynamical model in (23) for the sake of reproducibility. Notice that our data-driven method relies only on the data set S to come up with the feedback control policy and does not require the solution of expensive convex optimisation problems to obtain the value function.

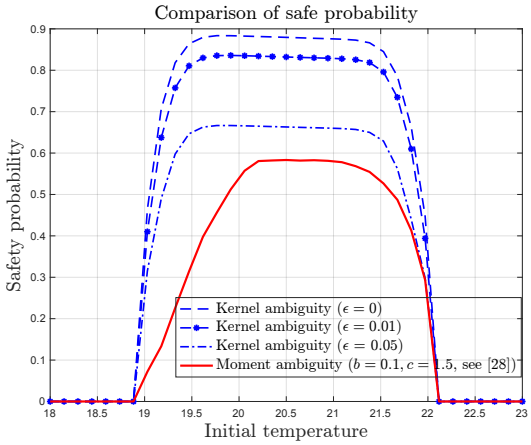

Fig. 1: Solution of the dynamic programming recursion given in (22) for the kernel ambiguity sets with different values of the radius (blue lines) with the kernel parameter $\gamma = 100$. We have used 7000 state-action pairs as samples to estimate the conditional mean embedding and the norm of vectors in the RKHS. The regularisation parameter λ is equal to 200. The solid red line is the value function using the methods proposed in [9] for $b = 0.1$ and $c = 1.5$ (as defined in [9]).

Figure 1 shows the obtained value function for different values of the radius ϵ, where we notice a decrease in the returned value function with the increase in the size of the ambiguity set. The y-axis represents the safety probability and the x-axis is the temperature; notice that the value function is zero outside the safe set $[19°\text{C}, 22°\text{C}]$. We also compare the returned value function with the one obtained using the method proposed in [9] (we refer to the reader to this paper for the definition of the parameters c and b shown in the legend).

A key feature of our approach is the fact that we are able to recover the shape of value function and perform controller design using available data only. We also do not need to solve an optimisation problem at each discretisation step as in [9], thus mitigating the computation burden when compared to existing approaches in the literature.

V. Conclusion

We analyzed the problem of distributionally robust (safe) control of stochastic systems where the ambiguity set is defined as the set of distributions whose kernel mean embedding is within a certain distance from the empirical estimate of the conditional kernel mean embedding derived from data. We showed that there exists a non-randomized Markovian policy that is optimal and discussed how to compute the value iteration by leveraging strong duality associated with kernel DRO problems. Numerical results illustrate the performance of the proposed formulations and the impact of the radius of the ambiguity set. There are several possible directions for future research, including deriving efficient algorithms to

compute the value iteration without resorting to discretization, representing multistage state evolution using composition of conditional mean embedding operators, and a thorough empirical investigation on the impact of dataset size on the performance and computational complexity of the problem.

REFERENCES

[1] P. Mohajerin Esfahani and D. Kuhn, "Data-driven distributionally robust optimization using the wasserstein metric: Performance guarantees and tractable reformulations," *Mathematical Programming*, vol. 171, no. 1, pp. 115–166, 2018.

[2] I. Yang, "Wasserstein distributionally robust stochastic control: A data-driven approach," *IEEE Transactions on Automatic Control*, vol. 66, no. 8, pp. 3863–3870, 2020.

[3] M. Schuurmans and P. Patrinos, "A general framework for learning-based distributionally robust mpc of markov jump systems," *IEEE Transactions on Automatic Control*, 2023.

[4] M. Fochesato and J. Lygeros, "Data-driven distributionally robust bounds for stochastic model predictive control," in *2022 IEEE 61st Conference on Decision and Control (CDC)*, pp. 3611–3616, IEEE, 2022.

[5] J. 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, 2002.

[6] 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*, vol. 49, no. 9, pp. 2665–2674, 2013.

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

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

[9] I. Yang, "A dynamic game approach to distributionally robust safety specifications for stochastic systems," *Automatica*, vol. 94, pp. 94–101, 2018.

[10] F. Luo and S. Mehrotra, "Distributionally robust optimization with decision dependent ambiguity sets," *Optimization Letters*, vol. 14, pp. 2565–2594, 2020.

[11] N. Noyan, G. Rudolf, and M. Lejeune, "Distributionally robust optimization under a decision-dependent ambiguity set with applications to machine scheduling and humanitarian logistics," *INFORMS Journal on Computing*, vol. 34, no. 2, pp. 729–751, 2022.

[12] K. Muandet, K. Fukumizu, B. Sriperumbudur, and B. Schölkopf, "Kernel mean embedding of distributions: A review and beyond," *Foundations and Trends® in Machine Learning*, vol. 10, no. 1-2, pp. 1–141, 2017.

[13] K. Fukumizu, L. Song, and A. Gretton, "Kernel bayes' rule: Bayesian inference with positive definite kernels," *The Journal of Machine Learning Research*, vol. 14, no. 1, pp. 3753–3783, 2013.

[14] B. Boots, A. Gretton, and G. J. Gordon, "Hilbert space embeddings of predictive state representations," in *Proceedings of the Twenty-Ninth Conference on Uncertainty in Artificial Intelligence*, pp. 92–101, 2013.

[15] A. J. Thorpe and M. M. Oishi, "Model-free stochastic reachability using kernel distribution embeddings," *IEEE Control Systems Letters*, vol. 4, no. 2, pp. 512–517, 2019.

[16] A. J. Thorpe, K. R. Ortiz, and M. M. Oishi, "State-based confidence bounds for data-driven stochastic reachability using hilbert space embeddings," *Automatica*, vol. 138, p. 110146, 2022.

[17] J.-J. Zhu, W. Jitkrittum, M. Diehl, and B. Schölkopf, "Kernel distributionally robust optimization: Generalized duality theorem and stochastic approximation," in *International Conference on Artificial Intelligence and Statistics*, pp. 280–288, PMLR, 2021.

[18] Y. Chen, J. Kim, and J. Anderson, "Distributionally robust decision making leveraging conditional distributions," in *2022 IEEE 61st Conference on Decision and Control (CDC)*, pp. 5652–5659, IEEE, 2022.

[19] H. Brezis, *Functional Analysis, Sobolev Spaces and Partial Differential Equations*. Springer, 2011.

[20] A. Smola, A. Gretton, L. Song, and B. Schölkopf, "A hilbert space embedding for distributions," in *International Conference on Algorithmic Learning Theory*, pp. 13–31, Springer, 2007.

[21] Y. Nemmour, H. Kremer, B. Schölkopf, and J.-J. Zhu, "Maximum mean discrepancy distributionally robust nonlinear chance-constrained optimization with finite-sample guarantee," *arXiv preprint arXiv:2204.11564*, 2022.

[22] L. Song, J. Huang, A. Smola, and K. Fukumizu, "Hilbert space embeddings of conditional distributions with applications to dynamical systems," in *Proceedings of the 26th Annual International Conference on Machine Learning*, pp. 961–968, 2009.

[23] S. Grünewälder, G. Lever, L. Baldassarre, S. Patterson, A. Gretton, and M. Pontil, "Conditional mean embeddings as regressors," in *Proceedings of the 29th International Coference on International Conference on Machine Learning*, pp. 1803–1810, 2012.

[24] C. A. Micchelli and M. Pontil, "On learning vector-valued functions," *Neural computation*, vol. 17, no. 1, pp. 177–204, 2005.

[25] J. Munkres, *Topology: A First Course*. Prentice Hall, 1974.

[26] J.-P. Aubin and H. Frankowska, *Set-valued analysis*. Springer Science & Business Media, 2009.

[27] K. Wood, G. Bianchin, and E. Dall'Anese, "Online projected gradient descent for stochastic optimization with decision-dependent distributions," *IEEE Control Systems Letters*, vol. 6, pp. 1646–1651, 2021.

[28] I. Yang, "A dynamic game approach to distributionally robust safety specifications for stochastic systems," *Automatica*, vol. 94, pp. 94–101, 2018.