2023 62nd IEEE Conference on Decision and Control (CDC)
December 13-15, 2023. Marina Bay Sands, Singapore

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

Licio Romao, Ashish R. Hota, and Alessandro Abate

Abstract— **We present a distributionally robust framework for dynamic programming that uses kernel methods to design control policies satisfying both safety and optimality specifications. Specifically, we leverage kernel mean embedding to map the transition probabilities governing state evolution into an associated reproducing kernel Hilbert space. Our key idea lies in combining conditional mean embedding estimated from past data of system trajectories with the maximum mean discrepancy distance to construct an ambiguity set, and then design a robust control policy using techniques from distributionally robust optimization. The main theoretical contribution of this paper is to leverage functional analytical tools to prove that optimal policies for this infinite-dimensional min-max problem are Markovian. Additionally, we discuss approximation schemes based on discretization of inputs to make the approach computationally tractable. We validate the main theoretical findings of the paper in a benchmark control problem involving safe control of thermostatically controlled loads.**

I. Introduction

We focus on discrete-time stochastic control problems, where states evolve according to a stochastic transition kernel that is unknown. We introduce a novel design approach based on dynamic programming, leveraging data on available trajectories of the system dynamics. To address sampling errors resulting from finite data, we employ techniques from "distributionally robust" optimization and control [1]–[3]. The core of distributionally robust techniques is to compute a feedback control policy that either minimizes an uncertain cost function or maximizes the probability of satisfying safety specifications [4], [5], subject to worst-case realization of the transition kernel over a set of probability distributions or *ambiguity set*. These ideas build upon the class of min-max control problems investigated in the seminal work [6].

In most of the past work on distributionally robust optimal and safe control, such as [2], [3], the ambiguity set is defined in an exogenous manner independent of the current state and action. 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 additive disturbance). However, more generally, when the state evolution is given by a stochastic transition kernel, it becomes necessary to define the ambiguity associated with it as a function of the current state and chosen action, giving rise to *decision-dependent ambiguity sets* [7].

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.

In this paper, we leverage the framework of Hilbert space embedding of conditional distributions [8] to define the ambiguity set associated with the transition kernel. Conditional mean embedding and its empirical estimate have been applied in the context of dynamical systems [9], reachability analysis [10], [11], and more recently for control synthesis in stochastic systems [12]–[15]. Similarly, distributionally robust optimization (DRO) subject to ambiguity sets defined via kernel mean embedding have been studied recently [16] where the authors established the strong duality result for this class of problems. However, kernel DRO problems where the ambiguity set is defined via the conditional mean embedding has not received much attention; [17] being an 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 [8], 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 [10], [11]. 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], [3], [6], we show that there exists a non-randomized Markovian 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 [16]. We then formulate the problem of control synthesis subject to safety specifications within the proposed framework. Numerical results on a benchmark problem provide valuable insights into the performance of 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 $\mathcal{X}$. A mea-

979-8-3503-0123-6/23/$31.00 ©2023 IEEE

surable function $k : \mathcal{X} \times \mathcal{X} \mapsto \mathbb{R}$, is called a *positive definite kernel*, if it satisfies three properties: (i) *boundedness*: $\sup_{x \in \mathcal{X}} |k(x,x)| < \infty$, (ii) *symmetry*: for any $x, x' \in \mathcal{X}$, we have $k(x,x') = k(x',x)$, and (iii) *positive definiteness*: for any finite collection of points $(x_i)_{i=1}^m$, where $x_i \in \mathcal{X}$ for all $i = \{1, \ldots, m\}$, the Gram matrix $K \in \mathbb{R}^{m \times m}$ whose (i,j)-th entry is given by $k(x_i, x_j)$ is a positive definite matrix.

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)

$$\mathcal{H}_{\mathcal{X}} := \overline{\bigcup_{\substack{I \subset \mathcal{X} \\ I \text{ finite}}} \text{span}\{k(x,\cdot) : \mathcal{X} \mapsto \mathbb{R}, x \in I\}}, \tag{1}$$

which is the closure of all possible finite dimensional subspaces induced by the kernel k. The RKHS $\mathcal{H}_{\mathcal{X}}$ is equipped with an inner product given by $\langle k(\cdot,x_1), k(\cdot,x_2)\rangle_{\mathcal{H}_{\mathcal{X}}} = k(x_1,x_2)$. By definition, for any function $f \in \mathcal{H}_{\mathcal{X}}$, there exist a sequence of integers $(m_n)_{n \in \mathbb{N}}$ and a sequence of functions $f_n(x) = \sum_{i=1}^{m_n} \beta_i^n k(x_i^n, x)$ such that $f(x) = \lim_{n \to \infty} f_n(x)$, which then implies that

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

justifying the reproducing property of the Hilbert space $\mathcal{H}_{\mathcal{X}}$.

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 $\phi(x)(\cdot) := k(x,\cdot)$, and we use the notation $\phi(x)(x') = k(x,x')$. The inner product defined earlier induces a norm over the RKHS defined as $||f||_{\mathcal{H}_{\mathcal{X}}} := \sqrt{\langle f, f\rangle_{\mathcal{H}_{\mathcal{X}}}}$ for all $f \in \mathcal{H}_{\mathcal{X}}$.

We now introduce the notion of *kernel mean embedding* of probability measures [8], [18]. 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

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

We have the following result from [8], [18] on the reproducing property of the expectation operator in the RKHS.

Lemma 1 (Lemma 3.1 [8])**.** *If $\mathbb{E}_{\mathbb{P}}[\sqrt{k(X,X)}] < \infty$, then $\Psi(\mathbb{P}) \in \mathcal{H}_{\mathcal{X}}$ and $\mathbb{E}_{\mathbb{P}}[f(X)] = \langle f, \Psi(\mathbb{P})\rangle_{\mathcal{H}_{\mathcal{X}}}$.*

Lemma 1 implies that the expectation of any function of the random variable X can be computed by means of an inner product between the corresponding function and the kernel mean embedding. Let $\{\widehat{x}_{(1)}, \ldots, \widehat{x}_{(M)}\}$ be a collection of M independent samples from the distribution $\mathbb{P}$. Then, the empirical estimate of the kernel mean embedding is

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

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 leverage the kernel mean embedding to define a metric or distance between two distributions $\mathbb{P}$ and $\mathbb{Q}$, called the *Maximum mean discrepancy (MMD)*, which is defined as

$$\begin{aligned} \texttt{MMD}(\mathbb{P}, \mathbb{Q}) &= \sup_{\|f\|_{\mathcal{H}_{\mathcal{X}}} \leq 1} \langle f, \Psi(\mathbb{P})\rangle_{\mathcal{H}_{\mathcal{X}}} - \langle f, \Psi(\mathbb{Q})\rangle_{\mathcal{H}_{\mathcal{X}}} \\ &= ||\Psi(\mathbb{P}) - \Psi(\mathbb{Q})||_{\mathcal{H}_{\mathcal{X}}}. \end{aligned} \tag{4}$$

In this work, we consider data-driven MMD ambiguity sets induced by observed samples $\{\widehat{x}_{(1)}, \ldots, \widehat{x}_{(M)}\}$ defined as

$$\widehat{\mathcal{M}}_M^{\epsilon} := \{\mathbb{P} \in \mathcal{P}(\mathcal{X}) \mid \texttt{MMD}(\mathbb{P}, \widehat{\mathbb{P}}_M) \leq \epsilon\}, \tag{5}$$

where $\widehat{\mathbb{P}}_M$ is the empirical distribution defined earlier. Thus, $\widehat{\mathcal{M}}_M^{\epsilon}$ contains all distributions whose kernel mean embedding is within distance $\epsilon \geq 0$ of the kernel mean embedding of the empirical distribution. The above ambiguity set also enjoys a sharp uniform convergence guarantees of $\mathcal{O}\left(\frac{1}{\sqrt{M}}\right)$ [19].

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 now introduce the notion of conditional mean embedding.

Definition 1 (Definition 4.1 [8])**.** *Given a stochastic kernel $T : \mathcal{Y} \mapsto \mathcal{P}(\mathcal{X})$, its conditional mean embedding is a mapping $\psi : \mathcal{Y} \mapsto \mathcal{H}_{\mathcal{X}}$ such that, for all $f \in \mathcal{H}_{\mathcal{X}}$,*

$$\langle \psi(y), f\rangle_{\mathcal{H}_{\mathcal{X}}} = \int_{\mathcal{X}} f(x) T(dx \mid y) = \mathbb{E}_{X \sim T(\cdot|y)}[f(X)], \tag{6}$$

that is, the inner product of the conditional mean embedding with a function in $f \in \mathcal{H}_{\mathcal{X}}$ coincides with the conditional expectation of f under the stochastic kernel T.

In other words, ψ 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, ψ 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 $\mathcal{Y}$ and $\mathcal{X}$ are not known, rather we have access to i.i.d. samples $\{(\widehat{y}_{(i)}, \widehat{x}_{(i)})\}_{i=1}^M$ drawn from the joint distribution $\mu \in \mathcal{P}(\mathcal{Y} \times \mathcal{X})$. 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 is now stated below.

Theorem 1 (Theorem 4.2 [8])**.** *An empirical estimate of the conditional mean embedding $\psi : \mathcal{Y} \to \mathcal{H}_{\mathcal{X}}$ is given by*

$$\widehat{\psi}_M(y)(\cdot) = \sum_{i=1}^{M} \beta_i(y) k_{\mathcal{X}}(\widehat{x}_{(i)}, \cdot), \tag{7}$$

where $\beta(y) = (K_{\mathcal{Y}} + M\lambda \mathbf{I}_M)^{-1} k_{\mathbf{y}}(y) \in \mathbb{R}^M$ 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.

The above empirical estimate can also be obtained by solving a regularized regression problem as shown in [20].

III. Kernel Distributionally Robust Optimal Control

We now connect the abstract mathematical framework introduced earlier in the context of optimal control problems. Consider the discrete-time stochastic system given by

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

where $x_k \in \mathcal{X} \subset \mathbb{R}^n$ and $a_k \in \mathcal{A}(x) \subset \mathbb{R}^p$ denote the state and control input at time $k \in \mathbb{N}$, with $\mathcal{A}(x_k)$ being the set of admissible control inputs at time k, and $\bar{x}$ denotes the initial state. We denote the (unknown) system dynamics by the stochastic kernel $T : \mathcal{X} \times \mathcal{A}(\mathcal{X}) \to \mathcal{P}(\mathcal{X})$ which describes a probability distribution over the next state.

We define a sequence of *history-dependant* control policies $\pi_k : (\mathcal{X} \times \mathcal{A}(\mathcal{X}))^k \times \mathcal{X} \to \mathcal{P}(\mathcal{A}(\mathcal{X}))$, which maps a sequence of state-input pair $(x_0, a_0, \ldots, x_{k-1}, a_{k-1}, x_k)$ into a probability measure with support $\mathcal{A}(\mathcal{X})$. The collection of admissible control policies over a horizon of length L is

$$\Pi_L = \{(\pi_0, \pi_1, \ldots, \pi_{L-1}) \mid \pi_k(\mathcal{A}(x_k)|h) = 1, \\ \forall k \in \{0, \ldots, L-1\}, \forall h \in (\mathcal{X} \times \mathcal{A}(\mathcal{X}))^k \times \mathcal{X}\}.$$

When the control policy depends only on the current state, i.e., $\pi_k(h) = \pi_k(h')$ for all $h, h' \in (\mathcal{X} \times \mathcal{A}(\mathcal{X}))^k \times \mathcal{X}$ that agree on the last entry, then we call such a policy *Markovian*.

Our objective is to design an admissible control policy that minimizes a performance index with respect to the generated trajectories of the system dynamics given in (8) by relying on available data set $(\widehat{x}_{(i)}, \widehat{a}_{(i)}, \widehat{x}^+_{(i)})_{i=1}^m$, composed by a sequence of state-input-next-state tuple. To account for the sampling error due to the finite dataset, we leverage techniques from distributionally robust optimization [1], [2] by creating a *state-input-dependant ambiguity set* around the estimate of the system dynamics. Formally, we define the space $\mathcal{Y} := \mathcal{X} \times \mathcal{A}(\mathcal{X})$ and $\mathcal{X}$ with the state space of the dynamics in (8). Let $k_{\mathcal{Y}} : (\mathcal{X} \times \mathcal{A}(\mathcal{X})) \times (\mathcal{X} \times \mathcal{A}(\mathcal{X})) \to \mathbb{R}$ and $k_{\mathcal{X}} : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ be the corresponding kernels that induce, respectively, the RKHS $\mathcal{H}_{\mathcal{Y}}$ and $\mathcal{H}_{\mathcal{X}}$.

Let $T : \mathcal{X} \times \mathcal{A}(\mathcal{X}) \to \mathcal{P}(\mathcal{X})$ be the stochastic kernel associated with the state-transition matrix, and $\widehat{\psi}_M$ be the empirical estimate of conditional mean embedding of T obtained from the dataset $(\widehat{x}_{(i)}, \widehat{a}_{(i)}, \widehat{x}^+_{(i)})_{i=1}^M$ following (7). At a given (x, a), we construct the ambiguity set

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

which is a collection of probability measures over $\mathcal{X}$ whose mean embedding is ϵ close to the empirical estimate of the conditional mean embedding of the system dynamics.

Similar to the control policy definition, we define a collection of *admissible dynamics* for a given time-horizon L. Formally, for a given sequence $(x_0, a_0, \ldots, x_{L-1}, a_{L-1})$ of state-input pairs, we define the set of admissible dynamics as

$$\Gamma_L = \big\{(\mu_0, \mu_1, \ldots, \mu_{L-1}) \mid \mu_k \in \widehat{\mathcal{M}}^{\epsilon}_M(x_k, a_k) \\ \text{for all } k \in \{0, \ldots, L-1\}\big\}. \tag{10}$$

We now state the distributionally robust optimal control problem. Let $L \in \mathbb{N}$ be the time-horizon, and consider the corresponding set of control policies Π_L and admissible dynamics Γ_L. For any $\pi \in \Pi_L$ and $\mu \in \Gamma_L$, and for any initial state $\bar{x} \sim \mu_0$, where μ_0 is the first entry of the admissible dynamics μ, we denote by $\mathbb{P}^{\pi,\mu}$ the induced measure on the space of sequences in $\mathcal{X}$ of length L. For a given stage cost $c : \mathcal{X} \times \mathcal{A}(\mathcal{X}) \to \mathbb{R}$, the finite horizon expected cost

$$V_L(\pi, \mu) := \mathbb{E}^{\pi,\mu}\Big[\sum_{k=0}^{L-1} c(x_k, a_k)\Big], \tag{11}$$

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

$$\inf_{\pi \in \Pi_L} \sup_{\mu \in \Gamma_L} V_L(\pi, \mu). \tag{12}$$

Problem (12) consists of an infinite-dimensional min-max problem. We now show that there exist optimal Markovian policies as the solution of (12). Following [6], we impose the following regularity assumptions on our problem.[1]

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

1) *The stage cost function $c(x, a)$ is lower semicontinuous, and there exists $\bar{c} \geq 0$ 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 $u : \mathcal{X} \to \mathbb{R}$, $\widehat{u}(x, a, \mu) := \int_{\mathcal{X}} u(y)\mu(dy)$ is continuous on $\mathbb{K}$.*
3) *The function $\widehat{w}(x, a, \mu) := \int_{\mathcal{X}} w(y)\mu(dy)$ is continuous on $\mathbb{K}$ and there exists a constant $\beta > 0$ such that $\widehat{w}(x, a, \mu) \leq \beta w(x)$ for all $(x, a, \mu) \in \mathbb{K}$.*
4) *The set $\mathcal{A}(x)$ is compact for each $x \in \mathcal{X}$, and 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 [6], 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., $||u||_w := \sup_{x \in \mathcal{X}} \frac{u(x)}{w(x)} < \infty$.

[1]Due to the fact that we are dealing with infinite-dimensional spaces, we rely on the topological notion of continuity. A function between two topological spaces (please refer to [21] 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 [22, Chapter 4], for more details about these concepts.

For each $u \in \mathcal{B}_w(\mathcal{X})$, and $(x,a,\mu) \in \mathbb{K}$, we define

$$H(u;x,a,\mu) := c(x,a) + \int_{\mathcal{X}} u(y)\mu(dy), \tag{13}$$

$$H^{\#}(u;x,a) := \sup_{\mu \in \widehat{\mathcal{M}}_M^{\epsilon}(x,a)} H(u;x,a,\mu), \tag{14}$$

$$\mathcal{T}(u)(x) := \inf_{a \in \mathcal{A}(x)} H^{\#}(u;x,a)$$

$$= \inf_{a \in \mathcal{A}(x)} \sup_{\mu \in \widehat{\mathcal{M}}_M^{\epsilon}(x,a)} \Big[c(x,a) + \int_{\mathcal{X}} u(y)\mu(dy) \Big]. \tag{15}$$

Specifically, (15) 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

$$v_L(x) := 0,$$
$$v_k(x) := (\mathcal{T} v_{k+1})(x) = (\mathcal{T} \circ \mathcal{T} \ldots \circ \mathcal{T})(v_L)(x), \tag{16}$$

for $0 \leq k \leq L-1$. We now state the following result which shows that the problem (12) admits a non-randomized Markov policy which is optimal.

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 (12).*

Proof. We follow a similar approach as the proof of [6, Theorem 3.1] and [3, Theorem 1]. The primary challenge is to show that the DP operator defined in (16) 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 [6, Lemma 3.3], it can be shown that $H(v_{k+1};x,a,\mu)$ is lower semicontinuous on $\mathbb{K}$.

The next step is to show that $H^{\#}(v_{k+1};x,a)$ is lower semi-continuous over $\mathcal{X} \times \mathcal{A}(\mathcal{X})$. To this end, we need to establish that the mapping $(x,a) \mapsto \widehat{\mathcal{M}}_M^{\epsilon}(x,a)$ is weakly compact and lower semicontinuous (i.e., the condition analogous to [6, Assumption 3.1(g)] holds for the ambiguity set (9)). To show weak compactness of $\widehat{\mathcal{M}}_M^{\epsilon}(x,a)$, let us first study properties of the set

$$\mathcal{C}_{(x,a)} = \{f \in \mathcal{H}_{\mathcal{X}} : \|f - \widehat{\psi}_M(x,a)\|_{\mathcal{H}_{\mathcal{X}}} \leq \epsilon\}, \tag{17}$$

which is a subset of the RKHS $\mathcal{H}_{\mathcal{X}}$. It is clear that this set is convex, hence, by [22, Theorem 3.7], it is also weakly closed. Since Hilbert spaces are reflexive Banach spaces, then by Kakutani's theorem (see [22, Theorem 3.17]) we show that the set $\mathcal{C}_{(x,a)}$ is weakly compact. Now, notice that

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

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[2], we have by the open mapping theorem ([22, Theorem 2.6]) that $\widehat{\mathcal{M}}_M^{\epsilon}(x,a)$ is also weakly compact.

We now show that the mapping $(x,a) \mapsto \widehat{\mathcal{M}}_M^{\epsilon}(x,a)$ is lower semicontinuous. We define the distance function from a distribution μ to a closed and convex subset S of $\widehat{\mathcal{M}}_M^{\epsilon}(x,a)$ as[3]

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

Let $(x,a,\mu) \in \mathbb{K}$, i.e., $a \in \mathcal{A}(x)$ and $\mu \in \widehat{\mathcal{M}}_M^{\epsilon}(x,a)$. Thus,

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

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 [23, Proposition 1.4.7], it follows that the lower semi-continuity of $(x,a) \mapsto \widehat{\mathcal{M}}_M^{\epsilon}(x,a)$ is equivalent to

$$\mu \in \liminf_{n\to\infty} \widehat{\mathcal{M}}_M^{\epsilon}(x_n,a_n) \iff \lim_{n\to\infty} d(\mu, \widehat{\mathcal{M}}_M^{\epsilon}(x_n,a_n)) = 0.$$

To this end, we compute

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

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 $\widehat{\psi}_M(x,a)$ in (7). As a result, $\mu \in \liminf_{n\to\infty} \widehat{\mathcal{M}}_M^{\epsilon}(x_n,a_n)$ and thus, $\widehat{\mathcal{M}}_M^{\epsilon}(x,a) \subseteq \liminf_{n\to\infty} \widehat{\mathcal{M}}_M^{\epsilon}(x_n,a_n)$.

With the above properties of the mapping $(x,a) \mapsto \widehat{\mathcal{M}}_M^{\epsilon}(x,a)$ in hand, it can be shown following identical arguments as the proof of [6, 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 (15) for $u = v_{k+1}$. □

Remark 1. Note that the ambiguity sets considered in related works on distributionally robust control [2], [3] do not depend on the current state and action. 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.

By showing that there exists deterministic and Markovian policies that optimize (12), Theorem 2 allows us to reduce the search of general history-dependant policies to functions

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

[3]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.

of the form $\pi : \mathcal{X} \to \mathcal{A}(\mathcal{X})$. However, the result of Theorem 2 does not directly lead to a computationally tractable formulation towards solving problem (12).

To this end, we exploit the rectangular structure of the admissible dynamics in (10), and observe that the inner supremum in (12) is computed separately for each time instant. Thus, for any $f' \in \mathcal{H}_{\mathcal{X}}$, we have

$$\sup_{\mu_k \in \widehat{\mathcal{M}}_M^{\epsilon}(x_k,a_k)} \mathbb{E}_{X\sim\mu_k}[f'(X)] = \sup_{\Psi(\mu_k)\in\mathcal{C}_{(x_k,a_k)}} \langle f', \Psi(\mu_k)\rangle_{\mathcal{H}_{\mathcal{X}}}, \quad (18)$$

where the last equality is a consequence of the reproducing property, and $\mathcal{C}_{(x_k,a_k)}$ is as defined in (17). Following [16], the support of $\mathcal{C}_{(x_k,a_k)}$ can then be computed as

$$\begin{aligned}\sigma_{\mathcal{C}_{(x_k,a_k)}}(f') &= \sup_{f\in\mathcal{C}_{(x_k,a_k)}} \langle f', f\rangle_{\mathcal{H}_{\mathcal{X}}} \\ &= \sup_{f\in\mathcal{C}_{(x_k,a_k)}} \langle f', f-\widehat{\psi}(x_k,a_k)\rangle_{\mathcal{H}_{\mathcal{X}}} + \langle f', \widehat{\psi}(x_k,a_k)\rangle_{\mathcal{H}_{\mathcal{X}}} \\ &= \epsilon||f'||_{\mathcal{H}_{\mathcal{X}}} + \sum_{i=1}^{M} \beta_i(x_k,a_k) f'(\widehat{x}^+_{(i)}), \end{aligned} \quad (19)$$

where $\beta_i(x_k,a_k)$ denote the coefficients of the empirical estimate of the conditional mean embedding as given in (7) evaluated at the state-action pair at the time instant $k \in \{1,\dots,L\}$, and where $\widehat{x}^+_{(i)}$ represents the collected sample[4] along the observed trajectories of the dynamics. To compute the norm $\|f'\|_{\mathcal{H}_{\mathcal{X}}}$, we solve, similar to Theorem 1, a regression problem given by

$$\min_{\alpha_i, i=1,\dots,m} \left\| \sum_{i=1}^{m} \alpha_i k(\widehat{x}'_{(i)}, \cdot) - f' \right\|_{\mathcal{H}_{\mathcal{X}}} + \lambda\|\alpha\|_2^2, \quad (20)$$

where $\{\widehat{x}'_{(1)},\dots,\widehat{x}'_{(m)}\}$ are arbitrary points in the domain of the function f'. The solution of this regression problem is given by $\|f'\|_{\mathcal{H}_{\mathcal{X}}} = \sqrt{\alpha^\top K'_{\mathcal{X}}\alpha}$, where

$$\alpha = (K'_{\mathcal{X}} + \lambda' I)^{-1} \begin{bmatrix} f'(\widehat{x}'_{(1)}) & \dots & f'(\widehat{x}'_{(m)}) \end{bmatrix}^\top,$$

and $K'_{\mathcal{X}}$ is the $\mathbb{R}^{m\times m}$ Gram matrix whose (i,j)-entry is given by $k_{\mathcal{X}}(\widehat{x}'_{(i)}, \widehat{x}'_{(j)})$. Notice that the value of the regularizer λ' or the collection of points used to estimate $\|f'\|_{\mathcal{H}_{\mathcal{X}}}$ may not necessarily coincide with those points used to estimate the conditional mean embedding in Theorem 1.

We now discuss how to solve for optimal control inputs using a value iteration approach. At given state $x \in \mathcal{X}$, we discretize the input space $\mathcal{A}(x)$ as $\{a^{(1)},\dots,a^{(R)}\}$. To define the value function, we slightly modify the set of admissible dynamics Γ_L in (10) by fixing the initial distribution to a given point $x \in \mathcal{X}$. Then, for a given function $f : \mathcal{X} \to \mathbb{R}$, we recursively define the collection of functions $v_\ell : \mathcal{X} \to \mathbb{R}$, for $\ell \in \{1,\dots,L\}$, with $v_L = f$, and

$$v_\ell(x) = \min_{a^{(j)}: j=1,\dots,R} \Big[c(x,a^{(j)}) + \sup_{\mu\in\widehat{\mathcal{M}}_m^{\epsilon}(x,a^{(j)})} \int_{\mathcal{X}} v_{\ell+1}(\xi)\mu(d\xi) \Big]. \quad (21)$$

[4]That is, for each $i \in \{1,\dots,m\}$, $\widehat{x}^+_{(i)}$ is sample from $T(\cdot \mid \widehat{x}_{(i)}, \widehat{a}_{(i)})$.

For a given $(x, a^{(j)})$, the inner supremum problem is an instance of (18) and can be evaluated by setting $f' = v_{\ell+1}$ in (19). We then find the index j that minimizes the R.H.S., and set the value function at x to be the minimum value.

1) 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. Using the notation of previous sections, let $L \in \mathbb{N}$ be the time-horizon and $S \subset \mathbb{R}^n$ be a measurable safe set. For an admissible control policy $\pi \in \Pi_L$, admissible dynamics $\mu \in \Gamma_L$, and initial state x, the probability of the state trajectory being safe is given by

$$V_S(x;\pi,\mu) = \mathbb{P}^{\pi,\mu}_{x_0}\{x_k \in S, \text{ for all } k \in \{1,\dots,L\}\}, \quad (22)$$

where $(x_1, x_2, \dots, x_L)$ denote the solution of (8) under the dynamics μ and policy π. Our goal is to solve the problem

$$V^\star_S(x) = \sup_{\pi\in\Pi_L} \inf_{\mu\in\Gamma_L} V_S(x;\pi,\mu), \quad (23)$$

using the mathematical framework proposed in this paper. In fact, one can show that under standard assumptions, an analogous result of Theorem 2 holds for (23), namely, there exists an optimal Markovian and deterministic policy. Hence, using similar approximations as in the previous section, function $V^\star_S$ in (23) can be approximated recursively as

$$\begin{aligned} v_L(x) &:= \mathbf{1}_S(x), \\ v_\ell(x) &:= \max_{a^{(j)}: j=1,\dots,R} \inf_{\mu\in\widehat{\mathcal{M}}_m^{\epsilon}(x,a^{(j)})} \mathbf{1}_S(x) \int_{\mathcal{X}} v_{\ell+1}(\xi)\mu(d\xi), \end{aligned} \quad (24)$$

for $\ell \in \{0,\dots,L-1\}$, where $\mathbf{1}_S$ is the indicator function of the set S. 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 (19) and (20).

IV. Numerical Examples

Inspired by past works [3], [4], we apply our methods to study safety probability of a thermostatically controlled load, whose dynamics is given by

$$x_{k+1} = \alpha x_k + (1-\alpha)(\theta - \eta R P u_k) + \omega_k, \quad (25)$$

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 (25) are given by $\alpha = \exp(h/CR)$, where $R = 2°\text{C/kW}$, $C = 2\text{kWh}/°\text{C}$, $\theta = 32°\text{C}$, $h = 5/60$ 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 based on available sampled trajectories for the model (25). Let $(\widehat{x}_{(i)}, \widehat{u}_{(i)}, \widehat{x}^+_{(i)})_{i=1}^M$ be a collection of observed transitions from the model, where the pair $(\widehat{x}_{(i)}, \widehat{u}_{(i)})$ is one of the random chosen points in the set $[19,22] \times \{0,1\}$ and $\widehat{x}^+_{(i)}$ represents the observed future state. We solve the dynamic programming recursion given in (24) by partitioning the state

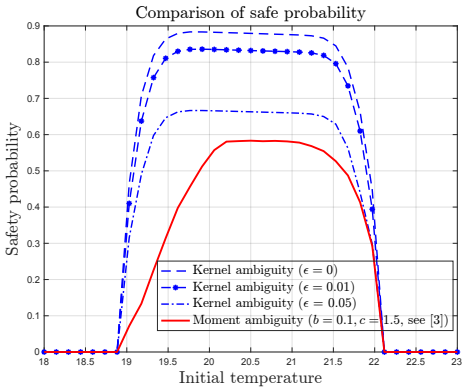

Fig. 1: Solution of the dynamic programming recursion given in (24) for the kernel ambiguity sets with different values of the radius (blue lines). The solid red line is the value function using the methods proposed in [3].

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 (20). We choose $\lambda = 200$ as the regularisation parameter, and use the kernel functions

$$k_{\mathcal{X}}(x, x') = e^{100|x-x'|^2},$$
$$k_{\mathcal{Y}}((x,u),(x',u')) = e^{-100|x-x'|^2} + k_1(u,u'), \qquad (26)$$

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 compared to the Gaussian kernel.

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 [3] (see [3] for the definition of the parameters c and b shown in the legend).

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

[4] 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.

[5] 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.

[6] 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.

[7] 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.

[8] 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.

[9] 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.

[10] 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.

[11] 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.

[12] S. Grünewälder, G. Lever, L. Baldassarre, M. Pontil, and A. Gretton, "Modelling transition dynamics in MDPs with RKHS embeddings," in *Proceedings of the 29th International Coference on International Conference on Machine Learning*, pp. 1603–1610, 2012.

[13] A. J. Thorpe and M. M. Oishi, "Stochastic optimal control via Hilbert space embeddings of distributions," in *2021 60th IEEE Conference on Decision and Control (CDC)*, pp. 904–911, IEEE, 2021.

[14] A. Thorpe and M. Oishi, "SOCKS: A stochastic optimal control and reachability toolbox using kernel methods," in *Proceedings of the 25th ACM International Conference on Hybrid Systems: Computation and Control*, pp. 1–12, 2022.

[15] A. J. Thorpe, J. A. Gonzales, and M. M. Oishi, "Data-driven stochastic optimal control using kernel gradients," in *2023 American Control Conference (ACC)*, pp. 2548–2553, IEEE, 2023.

[16] 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.

[17] 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.

[18] 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.

[19] Y. Nemmour, H. Kremer, B. Schölkopf, and J.-J. Zhu, "Maximum mean discrepancy distributionally robust nonlinear chance-constrained optimization with finite-sample guarantee," in *2022 IEEE 61st Conference on Decision and Control (CDC)*, pp. 5660–5667, IEEE, 2022.

[20] 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.

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

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

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