Summer 7-11-2019

Verification of Stochastic Reach-Avoid Using RKHS Embeddings

Adam J. Thorpe
University of New Mexico - Main Campus

Adam J. Thorpe

Candidate

Electrical Engineering

Department

This dissertation is approved, and it is acceptable in quality and form for publication:

Approved by the Dissertation Committee:

Dr. Meeko Oishi, Chair

Dr. Eirini Tsiropoulou, Co-Chair

Dr. John Richards, Member

Verification of Stochastic Reach-Avoid Using RKHS Embeddings

by

Adam J. Thorpe

B.A., Languages, University of New Mexico, 2010
B.S., Electrical Engineering, University of New Mexico, 2017

THESIS

Submitted in Partial Fulfillment of the
Requirements for the Degree of

Master of Science
Electrical Engineering

The University of New Mexico

Albuquerque, New Mexico

July, 2019

Dedication

To my family, for believing in me.

Acknowledgments

I would like to thank my advisor, Dr. Meeko Oishi, for all her support and encouragement. I could not have done this without her.

Verification of Stochastic Reach-Avoid Using RKHS Embeddings

by

Adam J. Thorpe

B.A., Languages, University of New Mexico, 2010
B.S., Electrical Engineering, University of New Mexico, 2017

M.S., Electrical Engineering, University of New Mexico, 2019

Abstract

A solution to the terminal-hitting and first-hitting stochastic reach-avoid problem for a Markov control process is presented. This solution takes advantage of a non-parametric representation of the stochastic kernel as a conditional distribution embedding within a reproducing kernel Hilbert space (RKHS). Because the disturbance is modeled as a data-driven stochastic process, this representation avoids intractable integrals in the dynamic recursion of the reach-avoid problem since the expectations can be calculated as an inner product within the RKHS. An example using a high-dimensional chain of integrators is presented, as well as for Clohessy-Wiltshire-Hill (CWH) dynamics.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Overview

Verification is an established tool to provide assurances that a system will remain "safe" over some time horizon. Typically, in order to ensure that a system retains certain "safe" properties over time, the desired system properties are given as constraints on the system. For instance, an aircraft may be set to fly within a certain altitude range, or an autonomous vehicle might be required to remain within a lane as it drives on the street. A system is considered safe if there exists a control action that would keep the system within the desired constraints at any given time. However, under real-world conditions which incorporate system uncertainty, dynamical systems lose the ability to provide guaranteed assurances of safety. In a realistic setting, stochastic analysis methods, such as *stochastic reachability*, can be used as a verification tool to provide probabilistic assurances of system safety. In general, stochastic reachability problems are typically posed as: *Does there exist a control action for which the state will stay within a given constraint set, with at least a desired likelihood?*

The solution to this problem is typically described using a dynamic programming [3, 4] based solution [1, 29], and significant progress has been made to solve this problem in a computationally tractable manner. Solutions have been presented using approximate dynamic programming [6], chance constraints [16, 30], sampling methods [22, 33, 35], and convex optimization with Fourier transforms [31, 34]. However, these solutions often suffer from significant limitations in either the computation efficiency, or on assumptions placed on the system dynamics. In many cases, assumptions about the dynamics or the nature of the uncertainty of the system are unrealistic.

In cases such as human-in-the-loop systems, the human has historically been modeled as a disturbance on the system. Systems which incorporate human-in-the-loop elements have high levels of uncertainty, and models of the human as a disturbance are either highly simplistic or overly conservative. For example, without an accurate estimate of a human's driving patterns, an autonomous vehicle has difficulty predicting the actions of other drivers on the road, leading to overly conservative estimates in order to ensure safety. Further, human probability models may not follow a known distribution, and are often data-driven processes that are unable to be analyzed using traditional stochastic verification techniques.

Similarly, the use of autonomous controllers and learning elements in systems is rapidly increasing. These systems are resistant to traditional models for control and formal methods, and verification of these systems may be overly conservative or even simply inaccurate.

In many cases, assumptions of realistic disturbance models or accurate knowledge of the system dynamics are unreasonable. Because of this, there is a common need for a flexible method which allows for model-free representations of dynamical systems and system uncertainty. Verification methods which allow for these complex, real-world elements while remaining computationally efficient are required in order to provide assurances that a system will remain safe with a given likelihood.

We propose a method for stochastic reachability analysis based on conditional distribution embeddings within a reproducing kernel Hilbert space (RKHS). Kernel methods are an established learning technique [2, 23, 24], which have been used for data and functional analysis, as well as for analyzing probability measures and their statistical features. As a nonparametric technique, kernel methods do not suffer from biases or prior assumptions on the system model, and are computationally efficient because they are historically a convergent, non-iterative learning method [23]. Recently, kernel methods have emerged which capture the features of arbitrary statistical distributions in a data-driven fashion [9, 25]. These methods broadly enable nonparametric inference using kernel embeddings of conditional distributions. These techniques have been applied to several problems involving dynamical systems, such as providing a solution to classical dynamic programming problems [14], controller synthesis for partially-observable dynamical system models [19], as well as for estimation of graphical models [27]. The proposed method has several advantages which are useful for analyzing dynamical systems:

1. Conditional distribution embeddings do not suffer from the curse of dimensionality, which is a significant limitation for dynamical systems of any significant complexity. Traditional dynamic programming solutions which leverage discretization approaches, state space limitations, or Monte Carlo methods, quickly become intractable as the complexity of the system representation increases.

2. Kernel methods take advantage of a high-dimensional feature representation of system observations. Because kernel methods are primarily data-driven, they allow for a model-free representation of the system dynamics and the uncertainty on the system. Because the system representation is model-free, it avoids many of the limiting assumptions placed on traditional analysis methods.

3. Conditional distribution embeddings avoid the issues of numerical quadrature which typically arise in dynamic programming-based solutions. The solution to many control problems involves the computation of a high-dimensional integral in order to obtain an expectation. By using conditional distribution embeddings within a reproducing kernel Hilbert space, the evaluation of an expected value can be computed as an inner product within the RKHS.

4. The proposed method also provides convergence results in the infinite-sample case, meaning the approximation improves as more observations of the system are available. The convergence results also include a probabilistic error bound which determines a confidence bound on the quality of the approximation. Because of this, we can provide an estimate of the error produced by the stochastic reachability analysis.

We apply kernel methods to the problem of computing the stochastic reachability probability measure, which is the expected value of the value function typically obtained through dynamic programming. The stochastic kernel that captures the underlying dynamical system (including autonomous or human elements) can be represented as a conditional distribution embedding within a reproducing kernel Hilbert space. Because we can capture the statistical features of the stochastic transition kernel as an arbitrary distribution using observations, the proposed method for stochastic reachability enables a numerically efficient method for analyzing systems with a poorly-characterized or unknown disturbance model. This means we can perform stochastic reachability analysis for systems or environments with human elements, as well as for systems with autonomous or learning elements, such as a neural network or learning controller. Furthermore, one of the key features of this approach is that it is agnostic to system dimensionality, the typical bottleneck for computational feasibility of the stochastic reachability problem.

We consider two problems for stochastic reachability as outlined in [29]. The first is a safety measure known as the terminal-hitting time problem. This presents a probability of retaining certain safe system properties and reaching a particular state. This can, for example, be thought of as a trajectory-planning exercise, where a system reaches a target state while avoiding obstacles in the environment. Second, we consider the first-hitting time problem, which is the probability that a system will reach a target before it violates any safety constraints. An example of the first-hitting time problem could be to intercept a target before it enters an unsafe region. Traditionally, stochastic reachability analysis has focused on the computation of the terminal-time hitting problem, mainly due to the numerical complexity involved with computing the expectation of the value function in the first-hitting time problem. Because the proposed method avoids the typical computational intractability involved with computing the expected values within the value functions for stochastic reachability, we are able to demonstrate the proposed method on both the terminal-hitting and the first-hitting time problem.

The main contribution of this paper is a conditional distribution embedding approach to stochastic reachability that enables model-free verification without invoking a statistical approach. This is particularly relevant for systems with black-box elements, such as autonomous or human-in-the-loop systems, which have previously been hindered by model assumptions or numerical efficiency.

The paper organization is as follows. Section 1.2 formulates the problem by representing the system as a Markov control process [29]. This allows us to represent the system dynamics and underlying system uncertainty as a stochastic kernel. Section 2 outlines the proposed method for conditional distribution embeddings for stochastic reachability. Sections 2.2 and 2.4 describe the use of kernel methods to perform safety verification for the terminal-hitting time problem and first-hitting time problem as outlined in [29]. In section 3, our approach is demonstrated on three examples: a double integrator to enable validation with a "truth" model via dynamic programming, a high-dimensional integrator model up to 10000-dimensions, and a spacecraft example using Clohessy-Wiltshire-Hill dynamics controlled by an open-loop chance-constrained controller as described in [16].

1.2 Problem Formulation

The following notation is used throughout the paper. Set difference is denoted with the backslash operator so that for sets $\mathcal{A}$ and $\mathcal{B}$, the set of all elements of $\mathcal{A}$ which are not in $\mathcal{B}$ is denoted as $\mathcal{A} \backslash \mathcal{B}$. For some nonempty set $\mathcal{A} \subseteq \mathcal{B}$, let $\mathbf{1}_{\mathcal{A}} : \mathcal{B} \rightarrow \{0, 1\}$ denote the indicator function where $\mathbf{1}_{\mathcal{A}}(x) = 1$ if $x \in \mathcal{A}$ and $\mathbf{1}_{\mathcal{A}}(x) = 0$ if $x \notin \mathcal{A}$.

1.2.1 Probability Theory

Let Ω denote a sample space and $\mathcal{F}(\Omega)$ denote the σ-algebra relative to Ω. A probability measure Pr assigned to the measurable space $(\Omega, \mathcal{F}(\Omega))$ is defined as the probability space $(\Omega, \mathcal{F}(\Omega), \mathrm{Pr})$. When $\Omega \equiv \Re$, the σ-algebra of Ω is denoted as $\mathscr{B}(\Omega)$, and is the Borel σ-algebra associated with Ω. A random variable x is a measurable function on the probability space $(\Omega, \mathcal{F}(\Omega), \mathrm{Pr}_x)$. A random vector $\boldsymbol{x} = [x_1, \ldots, x_n]^\top$ of n random variables is defined on the induced probability space $(\Omega^n, \mathcal{F}(\Omega^n), \mathrm{Pr}_{\boldsymbol{x}})$, where $\mathrm{Pr}_{\boldsymbol{x}}$ is the induced probability measure. A stochastic process is defined as a sequence of random vectors $\{\boldsymbol{x}_k : k \in [0, N]\}$, $N \in \mathbb{N}$, where $\boldsymbol{x}_k$ are defined on the probability space $(\Omega^n, \mathcal{F}(\Omega^n), \mathrm{Pr}_{\boldsymbol{x}})$. See [5, 7] for more details.

The expectation operator is denoted as $\mathbb{E}[\,\cdot\,]$, where for some function f, $\mathbb{E}_{\boldsymbol{x} \sim \mathrm{Pr}_{\boldsymbol{x}}\{\cdot\}}[f(\boldsymbol{x})]$ denotes the expectation operator with respect to the probability measure $\mathrm{Pr}_{\boldsymbol{x}}$.

1.2.2 System Model

Consider a Markov control process $\mathcal{H}$, which is defined in [29] as a 3-tuple,

$$\mathcal{H} = \langle \mathcal{X}, \mathcal{U}, Q \rangle \tag{1.1}$$

where $\mathcal{X} \subseteq \Re^n$ is the state space of the system, $\mathcal{U} \subseteq \Re^m$ is the control space, and Q is a stochastic kernel $Q : \mathscr{B}(\mathcal{X}) \times \mathcal{X} \times \mathcal{U} \to [0, 1]$, which is a Borel-measurable function that maps a probability measure $Q(\,\cdot\,|\,x_k, u_k)$ to each $x \in \mathcal{X}$ and $u \in \mathcal{U}$ on the Borel space $(\mathcal{X}, \mathscr{B}(\mathcal{X}))$. Further, let $\mathcal{X}$ and $\mathcal{U}$ be compact Borel spaces. The system evolves over a finite horizon $k \in [0, N]$ with inputs chosen from a Markov policy as defined in [4, 20]. A Markov policy is a sequence $\pi = \{\alpha_0, \alpha_1, \ldots, \alpha_{N-1}\}$ of universally-measurable maps $\alpha_k : \mathcal{X} \to \mathcal{U}$ from the state space $\mathcal{X}$ to the control space $\mathcal{U}$. The set of all Markov control policies π is denoted as $\mathcal{M}$.

We also consider a discrete-time stochastic system that can be described as a Markov control process (1.1), with dynamics given by

$$\boldsymbol{x}_{k+1} = f(\boldsymbol{x}_k, u_k, \boldsymbol{w}_k) \tag{1.2}$$

with state $\boldsymbol{x}_k \in \mathcal{X} \subseteq \Re^n$, input $u_k \in \mathcal{U} \subseteq \Re^m$, and disturbance $\boldsymbol{w}_k \in \mathcal{W} \subseteq \Re^p$. Let $\boldsymbol{w}[\,\cdot\,]$ be an i.i.d. Markov process with elements $\boldsymbol{w}_k$ defined on the probability space $(\mathcal{W}, \mathscr{B}(\mathcal{W}), \mathrm{Pr}_{\boldsymbol{w}})$. Given a policy π and the initial state $x_0 \in \mathcal{X}$, $\boldsymbol{x}_k$ is an n-dimensional random vector defined on the probability space $(\mathcal{X}, \mathscr{B}(\mathcal{X}), \mathrm{Pr}_{\boldsymbol{x}})$ from (1.2). The one-step stochastic transition kernel for (1.2), given by $T(\,\cdot\,|\,x, u)$, is equivalent to the stochastic kernel Q in (1.1).

1.2.3 Terminal-Hitting Time Problem

The goal of the terminal-hitting time problem, extended from [29, Section 4] as a specific case of the safety problem defined in [1, Section 4], is to evaluate the probability that a system controlled by some policy $\pi \in \mathcal{M}$ will stay within a *safe set* $\mathcal{K} \in \mathscr{B}(\mathcal{X})$ and reach a *target set* $\mathcal{T} \in \mathscr{B}(\mathcal{X})$ while avoiding $\mathcal{X} \backslash \mathcal{K}$ over the time horizon $[0, N]$. Given the control policy π such that $u_k = \alpha_k(x)$, and the initial condition $x_0 \in \mathcal{X}$, let the probability that the state x_k will remain within $\mathcal{K}$ for all $k \in [0, N-1]$ and the state $\boldsymbol{x}_N$ will be in $\mathcal{T}$ at $k = N$ be denoted as

$$\begin{aligned} r_{x_0}^{\pi}(\mathcal{K}, \mathcal{T}) \triangleq \Pr_{x_0}^{\pi}\{&\boldsymbol{x}_N \in \mathcal{T} \wedge \\ &\forall i \in [0, N-1], \boldsymbol{x}_i \in \mathcal{K}\} \end{aligned} \tag{1.3}$$

where the probability measure $\Pr_{x_0}^{\pi}$ is uniquely defined by the stochastic kernel Q, the control policy $\pi \in \mathcal{M}$, and the initial condition x_0 [4].

By representing $r_{x_0}^{\pi}(\mathcal{K}, \mathcal{T})$ as a multiplicative cost function, it can be expressed as the expectation [29]

$$r_{x_0}^{\pi}(\mathcal{K}, \mathcal{T}) = \mathbb{E}\left[\left(\prod_{i=0}^{N-1} \mathbf{1}_{\mathcal{K}}(\boldsymbol{x}_i)\right) \mathbf{1}_{\mathcal{T}}(\boldsymbol{x}_N)\right] \tag{1.4}$$

The associated terminal-hitting value functions $V_k^{\pi} : \mathcal{X} \to [0, 1]$ for $k \in [0, N]$ are defined as

$$V_N^{\pi}(x) = \mathbf{1}_{\mathcal{T}}(x) \tag{1.5}$$

$$V_k^{\pi}(x) = \mathbf{1}_{\mathcal{K}}(x) \int_{\mathcal{X}} V_{k+1}^{\pi}(y) Q(\mathrm{d}y \,|\, x, \alpha_k(x)) \tag{1.6}$$

where $x \in \mathcal{X}$. Note that $V_0^{\pi}(x) = r_{x_0}^{\pi}(\mathcal{K}, \mathcal{T})$ when $x = x_0$. As shown in [1, 29], $r_{x_0}^{\pi}(\mathcal{K}, \mathcal{T})$ can be computed via *backward recursion*.

By expressing V_k^{π} in terms of the expected value of V_{k+1}^{π}, (1.6) can alternatively be written [4] as

$$V_k^{\pi}(x) = \mathbf{1}_{\mathcal{K}}(x) \mathbb{E}_{\boldsymbol{y} \sim Q(\cdot \,|\, x, \alpha_k(x))}\left[V_{k+1}^{\pi}(\boldsymbol{y})\right] \tag{1.7}$$

From [29, Definition 10], a policy $\pi^* \in \mathcal{M}$ is denoted as the *maximal reach-avoid policy in the terminal sense* if and only if $r_{x_0}^{*}(\mathcal{K}, \mathcal{T}) = \sup_{\pi \in \mathcal{M}}\{r_{x_0}^{\pi}(\mathcal{K}, \mathcal{T})\}$ for all

$x_0 \in \mathcal{X}$. The maximal reach-avoid policy in the terminal sense is the control policy which maximizes the probability $\Pr_{x_0}^{\pi}$ in (1.3) for all $x_0 \in \mathcal{X}$. As in [29, Theorem 11], we denote the optimal value functions as $V_k^* : \mathcal{X} \to [0,1]$, such that

$$V_k^*(x) = \sup_{u \in \mathcal{U}} \left\{ \mathbf{1}_{\mathcal{K}}(x) \mathbb{E}_{\boldsymbol{y} \sim Q(\,\cdot\,|\,x,u)} \left[V_{k+1}^*(\boldsymbol{y}) \right] \right\} \tag{1.8}$$

initialized with $V_N^*(x) = \mathbf{1}_{\mathcal{T}}(x)$. Then, the maximal reach-avoid policy in the terminal sense is given by $\pi^* = \{\alpha_0^*, \alpha_1^*, \ldots\}$, where

$$\alpha_k^*(x) = \arg\sup_{u \in \mathcal{U}} \left\{ \mathbf{1}_{\mathcal{K}}(x) \mathbb{E}_{\boldsymbol{y} \sim Q(\,\cdot\,|\,x,u)} \left[V_{k+1}^*(\boldsymbol{y}) \right] \right\} \tag{1.9}$$

denotes the optimal maps from $\mathcal{X}$ to $\mathcal{U}$ for $k \in [0, N-1]$.

1.2.4 First-Hitting Time Problem

The goal of the first-hitting time problem [29, Section 3] is to evaluate the probability that a system controlled by some policy $\pi \in \mathcal{M}$ will reach a *target set* $\mathcal{T} \in \mathscr{B}(\mathcal{X})$ before hitting the *unsafe set* $\mathcal{X} \backslash \mathcal{K}$, $\mathcal{K} \in \mathscr{B}(\mathcal{X})$, during the time horizon $[0, N]$. Given the control policy π such that $u_k = \alpha_k(x)$, and the initial condition $x_0 \in \mathcal{X}$, let the probability that the state x_k will hit $\mathcal{T}$ before hitting $\mathcal{X} \backslash \mathcal{K}$ be denoted as

$$\begin{aligned} \bar{r}_{x_0}^{\pi}(\mathcal{K}, \mathcal{T}) \triangleq \Pr_{x_0}^{\pi}\{ & \exists j \in [0, N] : \boldsymbol{x}_j \in \mathcal{T} \wedge \\ & \forall i \in [0, j-1], \boldsymbol{x}_i \in \mathcal{K} \backslash \mathcal{T} \} \end{aligned} \tag{1.10}$$

As shown in [29], by representing (1.10) as a multiplicative cost function, it can be expressed as the expectation

$$\bar{r}_{x_0}^{\pi}(\mathcal{K}, \mathcal{T}) = \mathbb{E}\left[\sum_{j=0}^{N} \left(\prod_{i=0}^{j-1} \mathbf{1}_{\mathcal{K} \backslash \mathcal{T}}(\boldsymbol{x}_i) \right) \mathbf{1}_{\mathcal{T}}(\boldsymbol{x}_j) \right] \tag{1.11}$$

The associated first-hitting value functions $W_k^{\pi} : \mathcal{X} \to [0,1]$ for $k \in [0, N]$ are given by

$$W_N^\pi(x) = \mathbf{1}_{\mathcal{T}}(x) \tag{1.12}$$

$$W_k^\pi(x) = \mathbf{1}_{\mathcal{T}}(x) + \mathbf{1}_{\mathcal{K}\setminus\mathcal{T}}(x) \int_{\mathcal{X}} W_{k+1}^\pi(y) Q(\mathrm{d}y \,|\, x, \alpha_k(x)) \tag{1.13}$$

$$W_0^\pi(x) = \bar{r}_{x_0}^\pi(\mathcal{K}, \mathcal{T}) \tag{1.14}$$

where $x \in \mathcal{X}$. As in (1.7), $W_k^\pi(x)$, $k \in [0, N-1]$ can be expressed in terms of the expected value of $W_{k+1}^\pi(x)$, which is written as [4]

$$W_k^\pi(x) = \mathbf{1}_{\mathcal{T}}(x) + \mathbf{1}_{\mathcal{K}\setminus\mathcal{T}}(x) \mathbb{E}_{\boldsymbol{y} \sim Q(\,\cdot\,|\, x, \alpha_k(x))} \left[W_k^\pi(\boldsymbol{y})\right] \tag{1.15}$$

From [29, Definition 5], a policy $\pi^* \in \mathcal{M}$ is denoted as the *maximal reach-avoid policy in the first sense* if and only if $\bar{r}_{x_0}^*(\mathcal{K}, \mathcal{T}) = \sup_{\pi \in \mathcal{M}} \{\bar{r}_{x_0}^\pi(\mathcal{K}, \mathcal{T})\}$ for all $x_0 \in \mathcal{X}$. The maximal reach-avoid policy in the first sense is the control policy which maximizes the probability $\mathrm{Pr}_{x_0}^\pi$ in (1.10) for all $x_0 \in \mathcal{X}$. As in [29, Theorem 6], we denote the optimal value functions as $W_k^* : \mathcal{X} \to [0, 1]$, such that

$$W_k^*(x) = \sup_{u \in \mathcal{U}} \left\{ \mathbf{1}_{\mathcal{T}}(x) + \mathbf{1}_{\mathcal{K}\setminus\mathcal{T}}(x) \mathbb{E}_{\boldsymbol{y} \sim Q(\,\cdot\,|\, x, u)} \left[W_{k+1}^*(\boldsymbol{y})\right] \right\} \tag{1.16}$$

initialized with $W_N^*(x) = \mathbf{1}_{\mathcal{T}}(x)$. Then, the maximal reach-avoid policy in the first sense is given by $\pi^* = \{\alpha_0^*, \alpha_1^*, \ldots\}$, where

$$\alpha_k^*(x) = \arg\sup_{u \in \mathcal{U}} \left\{ \mathbf{1}_{\mathcal{T}}(x) + \mathbf{1}_{\mathcal{K}\setminus\mathcal{T}}(x) \mathbb{E}_{\boldsymbol{y} \sim Q(\,\cdot\,|\, x, u)} \left[W_{k+1}^*(\boldsymbol{y})\right] \right\} \tag{1.17}$$

denotes the optimal maps from $\mathcal{X}$ to $\mathcal{U}$ for $k \in [0, N-1]$.

1.3 Problem Statement

Consider a set $\mathcal{S}$ of M samples of the form $\mathcal{S} = \{(x_i{}', x_i, u_i)\}_{i=1}^M$ such that $x_i{}'$, $i = 1, \ldots, M$, is drawn i.i.d. from Q according to $x_i{}' \sim Q(\,\cdot\,|\, x_i, u_i)$.

Problem 1 *Without direct knowledge of Q or* (1.2)*, use samples $\mathcal{S}$ to construct an efficient approximation of* (1.7) *that converges in probability.*

Problem 2 *Without direct knowledge of Q or* (1.2)*, use samples $\mathcal{S}$ to construct an efficient approximation of* (1.8) *that converges in probability in order to compute an approximation of the maximal reach-avoid policy in the terminal sense.*

Problem 3 *Without direct knowledge of Q or* (1.2)*, use samples $\mathcal{S}$ to construct an efficient approximation of* (1.15) *that converges in probability.*

Problem 4 *Without direct knowledge of Q or* (1.2)*, use samples $\mathcal{S}$ to construct an efficient approximation of* (1.16) *that converges in probability in order to compute an approximation of the maximal reach-avoid policy in the first sense.*

Conditional distribution embeddings provide a solution to the problem of enabling computation of the stochastic reach-avoid probability for high-dimensional, non-Gaussian systems. The unique computational efficiencies afforded by reproducing kernel Hilbert spaces transforms computation of (1.7), (1.8), (1.15), and (1.16) into simple matrix operations and inner products.

Chapter 2

Conditional Distribution Embeddings for Stochastic Reachabilty

2.1 Kernel Embeddings of Conditional Distributions

For some set $\mathcal{X}$, let $\mathscr{H}_{\mathcal{X}}$ denote a reproducing kernel Hilbert space [23] with the kernel $K_{\mathcal{X}} : \mathcal{X} \times \mathcal{X} \rightarrow \Re$ over the domain $(\mathcal{X}, \mathscr{B}(\mathcal{X}))$, which is a Hilbert space of real-valued functions on $\mathcal{X}$ with inner product $\langle \cdot, \cdot \rangle_{\mathscr{H}_{\mathcal{X}}}$ and norm $\|x\|_{\mathscr{H}_{\mathcal{X}}} = (\langle x, x \rangle_{\mathscr{H}_{\mathcal{X}}})^{1/2}$. A reproducing kernel Hilbert space has two important properties [2]:

1. For any $x \in \mathcal{X}$, $K_{\mathcal{X}}(x, \cdot) : x' \rightarrow K_{\mathcal{X}}(x, x')$ is an element of $\mathscr{H}_{\mathcal{X}}$.

2. An element $K_{\mathcal{X}}(x, x')$ of $\mathscr{H}_{\mathcal{X}}$ satisfies the *reproducing property* such that $\forall h \in$

$\mathscr{H}_{\mathcal{X}}$ and $x \in \mathcal{X}$,

$$h(x) = \langle K_{\mathcal{X}}(x, x'), h(x') \rangle_{\mathscr{H}_{\mathcal{X}}} \tag{2.1}$$

This means that for any $x \in \mathcal{X}$, the evaluation of a function h can be viewed as an inner product. According to [27], an element $K_{\mathcal{X}}(x, \cdot)$ can also be viewed as a nonlinear feature map from $\mathcal{X}$ to $\mathscr{H}_{\mathcal{X}}$, denoted as $\phi_{\mathcal{X}} : \mathcal{X} \to \mathscr{H}_{\mathcal{X}}$, such that $K_{\mathcal{X}}(x, x') = \langle \phi_{\mathcal{X}}(x), \phi_{\mathcal{X}}(x') \rangle_{\mathscr{H}_{\mathcal{X}}}$.

In order to compute an approximation of the expected value in (1.7), we implicitly map the stochastic kernel Q into an infinite dimensional feature space using kernels [14, 25]. We are interested in evaluating the expectation of a function with respect to the stochastic kernel Q as an inner product within the reproducing kernel Hilbert space. Given a probability distribution of $\Pr_{\boldsymbol{x}}$ and positive definite (p.d.) [8, Definition 4.15] kernel $K_{\mathcal{X}} : \mathcal{X} \times \mathcal{X} \to \Re$ over the domain $(\mathcal{X}, \mathscr{B}(\mathcal{X}))$, a *distribution embedding* of $\Pr_{\boldsymbol{x}}$ in $\mathscr{H}_{\mathcal{X}}$ is an element $\mu \in \mathscr{H}_{\mathcal{X}}$ [14, 28] such that for all $h \in \mathscr{H}_{\mathcal{X}}$,

$$\langle \mu, h \rangle_{\mathscr{H}_{\mathcal{X}}} = \mathbb{E}_{\boldsymbol{y} \sim \Pr_{\boldsymbol{x}}\{\cdot\}} [h(\boldsymbol{y})] \tag{2.2}$$

Because constructing the feature mapping $\phi_{\mathcal{X}}(\cdot)$ and computing $K_{\mathcal{X}}(x, x') = \langle \phi_{\mathcal{X}}(x), \phi_{\mathcal{X}}(x') \rangle_{\mathscr{H}_{\mathcal{X}}}$ explicitly can be computationally expensive in a high-dimensional feature space, the inner product can be computed using $K_{\mathcal{X}}(x, x')$ directly for a $K_{\mathcal{X}}$ that is p.d. This is known as the *kernel trick* [24]. One common choice of kernel function is the Gaussian kernel,

$$K_{\mathcal{X}}(x, x') = \exp\left(-\frac{\|x - x'\|_2^2}{2\sigma^2}\right) \tag{2.3}$$

which is a symmetric, p.d. [28] kernel, where σ is a bandwidth parameter. However, many functions can serve as kernels, such as Laplacian and quadratic kernels.

Let $\mathscr{H}_{\mathcal{X}}$ denote the unique reproducing kernel Hilbert space for the state space $\mathcal{X}$ with the p.d. kernel $K_{\mathcal{X}} : \mathcal{X} \times \mathcal{X} \to \Re$. Similarly, let $\mathscr{H}_{\mathcal{X} \times \mathcal{U}}$ denote the reproducing

Table 2.1: Reproducing Kernel Hilbert Space Notation

Domain	$\mathcal{X}$	$\mathcal{U}$	$\mathcal{X} \times \mathcal{U}$
Random Vector	$\boldsymbol{x}$	$\boldsymbol{u}$	$(\boldsymbol{x}, \boldsymbol{u})$
Realization	x	u	(x, u)
RKHS	$\mathscr{H}_{\mathcal{X}}$	$\mathscr{H}_{\mathcal{U}}$	$\mathscr{H}_{\mathcal{X}} \otimes \mathscr{H}_{\mathcal{U}}$
Kernel	$K_{\mathcal{X}}(\cdot, \cdot)$	$K_{\mathcal{U}}(\cdot, \cdot)$	$K_{\mathcal{X}\times\mathcal{U}}(\cdot, \cdot)$
RKHS Element	$K_{\mathcal{X}}(x, \cdot)$	$K_{\mathcal{U}}(u, \cdot)$	$K_{\mathcal{X}\times\mathcal{U}}((x, u), \cdot)$

kernel Hilbert space for $\mathcal{X} \times \mathcal{U}$ with the p.d. kernel $K_{\mathcal{X}\times\mathcal{U}} : (\mathcal{X}, \mathcal{U}) \times (\mathcal{X}, \mathcal{U}) \to \Re$. Using $\mathscr{H}_{\mathcal{X}}$ and $\mathscr{H}_{\mathcal{X}\times\mathcal{U}}$, the goal is to find an element $\mu : \mathscr{H}_{\mathcal{X}\times\mathcal{U}} \to \mathscr{H}_{\mathcal{X}}$ which represents the *conditional distribution embedding* for the stochastic kernel Q. As shown in [13, 14, 18], by representing the conditional probabilities of Q as functions within $\mathscr{H}_{\mathcal{X}}$, it becomes possible to compute the conditional expectation of any function in $\mathscr{H}_{\mathcal{X}}$ as a linear operation, i.e. an inner product with a conditional distribution embedding.

Following [14, 27, 28], the conditional distribution embedding of the stochastic kernel Q is given by the element $\mu_{(x,u)} \in \mathscr{H}_{\mathcal{X}}$ such that $\forall h \in \mathscr{H}_{\mathcal{X}}$,

$$\langle \mu_{(x,u)}, h \rangle_{\mathscr{H}_{\mathcal{X}}} = \mathbb{E}_{\boldsymbol{y} \sim Q(\,\cdot\,|\,x,u)} [h(\boldsymbol{y})] \tag{2.4}$$

We can construct an estimate $\bar{\mu}_{(x,u)}$ of $\mu_{(x,u)}$ [18, 19, 25] from samples $\mathcal{S}$ to approximate (2.4),

$$\langle \bar{\mu}_{(x,u)}, h \rangle_{\mathscr{H}_{\mathcal{X}}} \approx \mathbb{E}_{\boldsymbol{y} \sim Q(\,\cdot\,|\,x,u)} [h(\boldsymbol{y})] \tag{2.5}$$

where the estimate $\bar{\mu}_{(x,u)}$ is given by the weighted linear combination

$$\bar{\mu}_{(x,u)} \triangleq \sum_{i=1}^{M} \hat{\beta}_i(x, u) K_{\mathcal{X}}(x_i{}', \cdot\,) \tag{2.6}$$

To find the weights $\hat{\beta}_i(x, u) \in \Re$ in (2.6), we first define

$$\beta_i(x, u) = \sum_{j=1}^{M} W_{ij} K_{\mathcal{X}\times\mathcal{U}}((x_j, u_j), (x, u)) \tag{2.7}$$

where W_{ij} is the $(i,j)^{\text{th}}$ element of $\boldsymbol{W}$, a regularized weight matrix for samples $\mathcal{S}$ given by

$$\boldsymbol{W} = (\boldsymbol{G} + \lambda M I)^{-1} \tag{2.8}$$

where λ is a regularization parameter to avoid overfitting [13, 17]. The matrix $\boldsymbol{G}$ is the Gram matrix, and is defined such that the $(i,j)^{\text{th}}$ element is given by

$$G_{ij} = K_{\mathcal{X}\times\mathcal{U}}((x_i, u_i), (x_j, u_j)) \tag{2.9}$$

As in [14], we then normalize (2.7) to obtain

$$\hat{\beta}_i(x,u) = \frac{\beta_i(x,u)}{\sum_{j=1}^{M} |\beta_j(x,u)|} \tag{2.10}$$

such that $\hat{\beta}_i(x,u) \in [0,1]$. By the reproducing property of $K_{\mathcal{X}}$ in $\mathscr{H}_{\mathcal{X}}$, $\forall h \in \mathscr{H}_{\mathcal{X}}$, we can rewrite (2.5) as

$$\langle \bar{\mu}_{(x,u)}, h \rangle_{\mathscr{H}_{\mathcal{X}}} = \sum_{i=1}^{M} \hat{\beta}_i(x,u) h(x_i{}') \tag{2.11}$$

This means an approximation of the value function expectation $\mathbb{E}_{\boldsymbol{y}\sim Q(\,\cdot\,|\,x,\alpha_k(x))}\left[V_{k+1}^{\pi}(\boldsymbol{y})\right]$ *in* (1.7) *can be evaluated as a linear operation in* $\mathscr{H}_{\mathcal{X}}$.

2.2 Terminal-Hitting Time Problem

With the conditional distribution embedding $\mu_{(x,u)}$, the value functions in (1.7) can be written as

$$V_k^{\pi}(x) = \mathbf{1}_{\mathcal{K}}(x)\langle \mu_{(x,\alpha_k(x))}, V_{k+1}^{\pi} \rangle_{\mathscr{H}_{\mathcal{X}}} \tag{2.12}$$

With the estimate $\bar{\mu}_{(x,u)}$ (2.6), we obtain the approximation,

$$V_k^{\pi}(x) \approx \mathbf{1}_{\mathcal{K}}(x)\langle \bar{\mu}_{(x,\alpha_k(x))}, V_{k+1}^{\pi} \rangle_{\mathscr{H}_{\mathcal{X}}} \tag{2.13}$$

Algorithm 1: Nonlinear Weights $\hat{\beta}$

Input: samples $\mathcal{S}$ drawn i.i.d. from Q, kernels $K_{\mathcal{X}}$ and $K_{\mathcal{X}\times\mathcal{U}}$, state x, input u

Output: weights $\hat{\beta}_i(x,u)$

Compute Gram matrix $\boldsymbol{G}$ using $\mathcal{S}$

$\boldsymbol{W} \leftarrow (\boldsymbol{G} + \lambda M I)^{-1}$

for $i \leftarrow 1$ **to** M **do** `/* compute` β `*/`

    $\beta_i(x,u) \leftarrow \sum_{j=1}^{M} W_{ij} K_{\mathcal{X}\times\mathcal{U}}((x_j,u_j),(x,u))$

end

for $i \leftarrow 1$ **to** M **do** `/* compute` $\hat{\beta}$ `*/`

    $\hat{\beta}_i(x,u) \leftarrow \beta_i(x,u) / \sum_{j=1}^{M} |\beta_j(x,u)|$

end

return $\hat{\beta}_i(x,u)$

We define the approximate value functions $\overline{V}_k^{\pi} : \mathcal{X} \to [0,1]$, $k \in [0, N-1]$, as

$$\overline{V}_k^{\pi}(x) = \mathbf{1}_{\mathcal{K}}(x)\langle \bar{\mu}_{(x,\alpha_k(x))}, V_{k+1}^{\pi}\rangle_{\mathscr{H}_{\mathcal{X}}} \tag{2.14}$$

An approximation for the reach-avoid probability $r_{x_0}^{\pi}(\mathcal{K},\mathcal{T})$ computed via backward recursion is described in Algorithm 2, such that

$$r_{x_0}^{\pi}(\mathcal{K},\mathcal{T}) \approx \overline{V}_0^{\pi}(x) \tag{2.15}$$

We now seek to characterize the quality of the approximation and the conditions for its convergence. As in [11, 12, 28], we define a pseudometric that characterizes the accuracy of the estimate $\bar{\mu}_{(x,u)}$.

Definition 1 (*Distance Pseudometric*)**.** The distance pseudometric in $\mathscr{H}_{\mathcal{X}}$ between the conditional distribution embedding $\mu_{(x,u)} \in \mathscr{H}_{\mathcal{X}}$ and the estimate $\bar{\mu}_{(x,u)} \in \mathscr{H}_{\mathcal{X}}$ is defined as $\|\mu_{(x,u)} - \bar{\mu}_{(x,u)}\|_{\mathscr{H}_{\mathcal{X}}}$.

Algorithm 2: Sample-Based Terminal-Hitting Value Function Estimate

Input: samples $\mathcal{S}$ drawn i.i.d. from Q, kernels $K_{\mathcal{X}}$ and $K_{\mathcal{X}\times\mathcal{U}}$, policy π, time horizon N

Output: value function estimate $\overline{V}_0^\pi(x)$

Compute Gram matrix $\boldsymbol{G}$ using $\mathcal{S}$

$\boldsymbol{W} \leftarrow (\boldsymbol{G} + \lambda M I)^{-1}$

$\overline{V}_N^\pi(x) \leftarrow \mathbf{1}_{\mathcal{T}}(x)$

for $k \leftarrow N-1$ **to** 0 **do**

  Compute $\hat{\beta}_i(x, \alpha_k(x))$ using Algorithm 1

  $\overline{V}_k^\pi(x) \leftarrow \mathbf{1}_{\mathcal{K}}(x)$

  $$\times \sum_{i=1}^{M} \hat{\beta}_i(x, \alpha_k(x)) \overline{V}_{k+1}^\pi({x_i}')$$

end

return $\overline{V}_0^\pi(x)$

It is shown in [9] that if $K_{\mathcal{X}}$ is a *characteristic*, bounded kernel, then $\|\mu_{(x,u)} - \bar{\mu}_{(x,u)}\|_{\mathscr{H}_{\mathcal{X}}} = 0$ if and only if $\mu_{(x,u)} = \bar{\mu}_{(x,u)}$. A kernel is characteristic if the kernel embedding is injective, meaning the embeddings for any two different conditional distributions are represented by different elements within the reproducing kernel Hilbert space. Thus, as $\|\mu_{(x,u)} - \bar{\mu}_{(x,u)}\|_{\mathscr{H}_{\mathcal{X}}}$ converges [14, 26, 27], the estimate converges in probability to the conditional distribution embedding within $\mathscr{H}_{\mathcal{X}}$.

Lemma 1. *[14, Lemma 2.2] For any $\varepsilon > 0$, if the regularization prameter λ in* (2.8) *is chosen such that $\lambda \to 0$ and $\lambda^3 M \to \infty$, and if $|\mathcal{X}| < \infty$ and $K_{\mathcal{X}}$ is strictly positive definite, then*

$$\Pr_{\mathcal{S}\sim Q}\left\{ \sup_{(x,u)\in\mathcal{X}\times\mathcal{U}} \|\mu_{(x,u)} - \bar{\mu}_{(x,u)}\|_{\mathscr{H}_{\mathcal{X}}} > \varepsilon \right\} \to 0 \tag{2.16}$$

Because the conditional distribution embedding estimate converges in probability as the number of samples increases according to Lemma 1, ε can be seen as a probabilistic error bound on the approximation in (2.13).

Proposition 1 (*Terminal-Hitting Value Function Convergence*)**.** *For any $\varepsilon > 0$, if the regularization prameter λ in* (2.8) *is chosen such that $\lambda \to 0$ and $\lambda^3 M \to \infty$, and if $|\mathcal{X}| < \infty$ and $K_{\mathcal{X}}$ is strictly positive definite, $|V_k^\pi(x) - \overline{V}_k^\pi(x)|$ converges in probability.*

Proof: By subtracting (2.14) from (2.12), we define the absolute value function error $\mathcal{E}_k(x)$ at time k,

$$\mathcal{E}_k(x) \triangleq |V_k^\pi(x) - \overline{V}_k^\pi(x)| \tag{2.17}$$

$$= |\mathbf{1}_{\mathcal{K}}(x)\langle \mu_{(x,\alpha_k(x))}, V_{k+1}^\pi \rangle_{\mathscr{H}_{\mathcal{X}}} - \mathbf{1}_{\mathcal{K}}(x)\langle \bar{\mu}_{(x,\alpha_k(x))}, V_{k+1}^\pi \rangle_{\mathscr{H}_{\mathcal{X}}}| \tag{2.18}$$

We can rewrite (2.18) using the parallelogram law and Cauchy–Schwarz to obtain

$$\mathcal{E}_k(x) = \mathbf{1}_{\mathcal{K}}(x)|\langle \mu_{(x,\alpha_k(x))} - \bar{\mu}_{(x,\alpha_k(x))}, V_{k+1}^\pi \rangle_{\mathscr{H}_{\mathcal{X}}}| \tag{2.19}$$

$$\leq \mathbf{1}_{\mathcal{K}}(x)\|V_{k+1}^\pi\|_{\mathscr{H}_{\mathcal{X}}}\|\mu_{(x,\alpha_k(x))} - \bar{\mu}_{(x,\alpha_k(x))}\|_{\mathscr{H}_{\mathcal{X}}} \tag{2.20}$$

Since $\|\mu_{(x,\alpha_k(x))} - \bar{\mu}_{(x,\alpha_k(x))}\|_{\mathscr{H}_{\mathcal{X}}}$ converges in probability according to Lemma 1, $|V_k^\pi(x) - \overline{V}_k^\pi(x)|$ also converges in probability with the probabilistic error bound ε.
□

Using this, the value function approximation in (2.13) converges in probability for some probabilistic error bound ε as the number of samples increases. By generalizing to the infinite-sample case, it is possible to define ε as the maximum error on the reach-avoid probability computed in (2.15).

Corollary 1. *For any $\varepsilon > 0$, the error in the reach-avoid probability computed* *using Algorithm 2 is given by*

$$|V_0^\pi(x) - \overline{V}_0^\pi(x)| \leq N\varepsilon \tag{2.21}$$

Proof: By subtracting (2.14) from (2.12), we obtain the absolute value function error $\mathcal{E}_{N-1}(x)$ at time $k = N-1$,

$$\mathcal{E}_{N-1}(x) = |V_{N-1}^\pi(x) - \overline{V}_{N-1}^\pi(x)| \tag{2.22}$$

Using Proposition 1, if the error in the approximate value function is at most ε in the infinite-sample case, then the error in (2.22) is at most ε for all $x \in \mathcal{X}$.

$$\mathcal{E}_{N-1}(x) \leq \varepsilon \tag{2.23}$$

Because the error in the approximate value function for $k = N-1$ is at most ε, then by approximating and recursively substituting $\overline{V}_k^\pi(x)$ for $k < N-1$, the error at time k is at most $(N-k)\varepsilon$. Thus by induction the error obtained by the backward recursion in Algorithm 2 is at most $N\varepsilon$,

$$|V_0^\pi(x) - \overline{V}_0^\pi(x)| \leq N\varepsilon \tag{2.24}$$

which concludes the proof. □

Thus in the infinite-sample case, we can choose $\varepsilon > 0$ to be arbitrarily small such that the approximation of the reach-avoid probability $r_{x_0}^\pi(\mathcal{K}, \mathcal{T})$ computed using Algorithm 2 is given by (2.15).

2.3 Maximal Reach-Avoid Policy in the Terminal Sense

As in (2.12), we write the optimal value functions V_k^* from (1.8) using the conditional distribution embedding $\mu_{(x,u)}$.

$$V_k^*(x) = \sup_{u \in \mathcal{U}} \left\{ \mathbf{1}_{\mathcal{K}}(x) \langle \mu_{(x,u)}, V_{k+1}^* \rangle_{\mathscr{H}_{\mathcal{X}}} \right\} \tag{2.25}$$

With the estimate $\bar{\mu}_{(x,u)}$ from (2.6), we obtain an approximation of (2.25) given by

$$V_k^*(x) \approx \sup_{u \in \mathcal{U}} \left\{ \mathbf{1}_{\mathcal{K}}(x) \langle \bar{\mu}_{(x,u)}, V_{k+1}^* \rangle_{\mathscr{H}_{\mathcal{X}}} \right\} \tag{2.26}$$

As in (2.14), we define the approximate optimal value functions $\overline{V}_k^* : \mathcal{X} \to [0, 1]$, $k \in [0, N-1]$, as

$$\overline{V}_k^*(x) = \sup_{u \in \mathcal{U}} \left\{ \mathbf{1}_{\mathcal{K}}(x) \langle \bar{\mu}_{(x,u)}, V_{k+1}^* \rangle_{\mathscr{H}_{\mathcal{X}}} \right\} \tag{2.27}$$

If $\bar{\alpha}_k^* : \mathcal{X} \to \mathcal{U}$, $k \in [0, N-1]$ is such that $\forall x \in \mathcal{X}$

$$\bar{\alpha}_k^*(x) = \arg\sup_{u \in \mathcal{U}} \left\{ \mathbf{1}_{\mathcal{K}}(x) \langle \bar{\mu}_{(x,u)}, V_{k+1}^* \rangle_{\mathscr{H}_{\mathcal{X}}} \right\} \tag{2.28}$$

then $\bar{\pi}^* = \{\bar{\alpha}_0^*, \bar{\alpha}_1^*, \ldots\}$ is the approximate maximal reach-avoid policy in the terminal sense. The approximate optimal reach-avoid probability obtained via back recursion under policy $\bar{\pi}^*$ initialized with $\overline{V}_k^*(x) = \mathbf{1}_{\mathcal{T}}(x)$ is then given by Algorithm 2 as

$$r_{x_0}^*(\mathcal{K}, \mathcal{T}) \approx \overline{V}_0^*(x) \tag{2.29}$$

2.4 First-Hitting Time Problem

With the conditional distribution embedding $\mu_{(x,u)}$, the value functions in (1.15) can be written as

$$W_k^\pi(x) = \mathbf{1}_{\mathcal{T}}(x) + \mathbf{1}_{\mathcal{K} \setminus \mathcal{T}}(x) \langle \mu_{(x, \alpha_k(x))}, W_{k+1}^\pi \rangle_{\mathscr{H}_{\mathcal{X}}} \tag{2.30}$$

Algorithm 3: Sample-Based First-Hitting Value Function Estimate

Input: samples $\mathcal{S}$ drawn i.i.d. from Q, kernels $K_{\mathcal{X}}$ and $K_{\mathcal{X}\times\mathcal{U}}$, policy π, time horizon N

Output: value function estimate $\overline{V}_0^\pi(x)$

Compute Gram matrix $\boldsymbol{G}$ using $\mathcal{S}$

$\boldsymbol{W} \leftarrow (\boldsymbol{G} + \lambda M I)^{-1}$

$\overline{V}_N^\pi(x) \leftarrow \mathbf{1}_{\mathcal{T}}(x)$

for $k \leftarrow N-1$ **to** 0 **do**

  Compute $\hat{\beta}_i(x, \alpha_k(x))$ using Algorithm 1

  $\overline{V}_k^\pi(x) \leftarrow \mathbf{1}_{\mathcal{T}}(x) + \mathbf{1}_{\mathcal{K}\setminus\mathcal{T}}(x) \times \sum_{i=1}^{M} \hat{\beta}_i(x, \alpha_k(x)) \overline{V}_{k+1}^\pi({x_i}')$

end

return $\overline{V}_0^\pi(x)$

With the estimate $\bar{\mu}_{(x,u)}$ from (2.6), we obtain the approximation,

$$W_k^\pi(x) \approx \mathbf{1}_{\mathcal{T}}(x) + \mathbf{1}_{\mathcal{K}\setminus\mathcal{T}}(x) \langle \bar{\mu}_{(x,\alpha_k(x))}, W_{k+1}^\pi \rangle_{\mathscr{H}_{\mathcal{X}}} \tag{2.31}$$

We define the approximate value functions $\overline{W}_k^\pi : \mathcal{X} \to [0,1]$, $k \in [0, N-1]$, as

$$\overline{W}_k^\pi(x) = \mathbf{1}_{\mathcal{T}}(x) + \mathbf{1}_{\mathcal{K}\setminus\mathcal{T}}(x) \langle \bar{\mu}_{(x,\alpha_k(x))}, W_{k+1}^\pi \rangle_{\mathscr{H}_{\mathcal{X}}} \tag{2.32}$$

such that an approximation of the reach-avoid probability $\bar{r}_{x_0}^\pi(\mathcal{K}, \mathcal{T})$ computed via backward recursion is described in Algorithm 3, such that

$$\bar{r}_{x_0}^\pi(\mathcal{K}, \mathcal{T}) \approx \overline{W}_0^\pi(x) \tag{2.33}$$

The approximation converges according to the same logic as in Proposition 2. Since $\|\mu_{(x,\alpha_k(x))} - \bar{\mu}_{(x,\alpha_k(x))}\|_{\mathscr{H}_{\mathcal{X}}}$ converges in probability according to Lemma 1, the approximation of the reach-avoid probability $\bar{r}^{\pi}_{x_0}(\mathcal{K}, \mathcal{T})$ converges as the number of samples increases.

Proposition 2 (*First-Hitting Value Function Convergence*). *For any $\varepsilon > 0$, if the regularization prameter λ in (2.8) is chosen such that $\lambda \to 0$ and $\lambda^3 M \to \infty$, and if $|\mathcal{X}| < \infty$ and $K_{\mathcal{X}}$ is strictly positive definite, $|W^{\pi}_k(x) - \overline{W}^{\pi}_k(x)|$ converges in probability.*

Proof: The proof follows the logic of Proposition 1. □

2.5 Maximal Reach-Avoid Policy in the First Sense

As in (2.30), we write the optimal value functions W^*_k from (1.16) using the conditional distribution embedding $\mu_{(x,u)}$.

$$W^*_k(x) = \sup_{u \in \mathcal{U}} \left\{ \mathbf{1}_{\mathcal{T}}(x) + \mathbf{1}_{\mathcal{K} \setminus \mathcal{T}}(x) \langle \mu_{(x,u)}, W^*_{k+1} \rangle_{\mathscr{H}_{\mathcal{X}}} \right\} \tag{2.34}$$

With the estimate $\bar{\mu}_{(x,u)}$ from (2.6), we obtain an approximation of (2.34) given by

$$W^*_k(x) \approx \sup_{u \in \mathcal{U}} \left\{ \mathbf{1}_{\mathcal{T}}(x) + \mathbf{1}_{\mathcal{K} \setminus \mathcal{T}}(x) \langle \bar{\mu}_{(x,u)}, W^*_{k+1} \rangle_{\mathscr{H}_{\mathcal{X}}} \right\} \tag{2.35}$$

As in (2.32), we define the approximate optimal value functions $\overline{W}^*_k : \mathcal{X} \to [0, 1]$, $k \in [0, N-1]$, as

$$\overline{W}^*_k(x) = \sup_{u \in \mathcal{U}} \left\{ \mathbf{1}_{\mathcal{T}}(x) + \mathbf{1}_{\mathcal{K} \setminus \mathcal{T}}(x) \langle \bar{\mu}_{(x,u)}, W^*_{k+1} \rangle_{\mathscr{H}_{\mathcal{X}}} \right\} \tag{2.36}$$

If $\hat{\alpha}^*_k : \mathcal{X} \to \mathcal{U}$, $k \in [0, N-1]$ is such that $\forall x \in \mathcal{X}$

$$\hat{\alpha}^*_k(x) = \arg\sup_{u \in \mathcal{U}} \left\{ \mathbf{1}_{\mathcal{T}}(x) + \mathbf{1}_{\mathcal{K} \setminus \mathcal{T}}(x) \langle \bar{\mu}_{(x,u)}, W^*_{k+1} \rangle_{\mathscr{H}_{\mathcal{X}}} \right\} \tag{2.37}$$

then $\hat{\pi}^* = \{\hat{\alpha}_0^*, \hat{\alpha}_1^*, \ldots\}$ is the approximate maximal reach-avoid policy in the first sense. The approximate optimal reach-avoid probability obtained via back recursion under policy $\hat{\pi}^*$ initialized with $\overline{W}_k^*(x) = \mathbf{1}_{\mathcal{T}}(x)$ is then given by Algorithm 3 as

$$\bar{r}_{x_0}^*(\mathcal{K}, \mathcal{T}) \approx \overline{W}_0^*(x) \tag{2.38}$$

Chapter 3

Numerical Results

We considered a double integrator with a known Gaussian disturbance and compared the result to the dynamic programming solution for validation. Then, we applied a disturbance with a Beta(α, β) distribution with shape parameters $\alpha = 2$, $\beta = 2$, in order to demonstrate the system for non-Gaussian disturbances. Then, a 10000-D stochastic integrator with a Gaussian disturbance was demonstrated as an example of a high-dimensional system.

We also demonstrated the method using CWH dynamics, which is a linear-affine system with a Gaussian disturbance under a chance-affine open-loop controller as described in [16] in order to demonstrate the method for systems with a fixed policy.

3.1 Implementation

For all problems, we used a Gaussian kernel with $\sigma = 0.1$, and chose $\lambda = 1$ as the default regularization parameter for the evaluation. We considered $M = 1024$ samples drawn i.i.d. from Q to calculate the sample-based estimate of the conditional distribution embedding $\bar{\mu}_{(x,u)}$ using (2.6). For the low-dimensional systems, $T =$

Table 3.1: Computation Time

System	Problem	Dim. [n]	N	M	T	CDE	DP	CC Open
Double Integrator	TH	2	3	1024	10201	530 ms	34.12 s	–
CWH	TH	4	5	1024	10201	2.92 s	–	25.32 s
Stochastic Integrator	TH	10000	3	1024	1	33.74 s	–	–
Double Integrator	FH	2	3	1024	10201	601 ms	–	–
CWH	FH	4	5	1024	10201	3.29 s	–	–
Stochastic Integrator	FH	10000	3	1024	1	34.27 s	–	–

10201 points of the form $\tau = \{(x_i, u_i)\}_{i=1}^{T}$ were used in order to evaluate the reach-avoid probabilities. The sample-based estimates of $\overline{V}_k^{\pi} : \mathcal{X} \rightarrow [0, 1]$, $k \in [0, N-1]$ and $\overline{W}_k^{\pi} : \mathcal{X} \rightarrow [0, 1]$, $k \in [0, N-1]$ were computed using Algorithm 2 for a time horizon of $N = 3$ for the stochastic integrator and $N = 5$ for the CWH system. For the high-dimensional example we restricted ourselves to a single point $T = 1$ to evaluate the reach-avoid probability.

We compared the computation time for the method, where possible, to other existing methods. The results are shown in Table 3.1. The main computational limitation for the conditional distribution embedding (CDE) approach is in computing the Gram matrix $\boldsymbol{G}$, which is on the order of $\mathcal{O}(M^3)$. However, the computational complexity can be reduced to log-linear time by using a more computationally efficient representation of $\boldsymbol{G}$, as shown in [15, 21].

3.2 n-D Stochastic Integrator System

The discrete time dynamics of an n-D stochastic integrator are given by

$$\boldsymbol{x}_{k+1} = A\boldsymbol{x}_k + Bu_k + \boldsymbol{w}_k \tag{3.1}$$

where $A \in \Re^{n \times n}$ and $B \in \Re^{n \times m}$. The disturbance $\boldsymbol{w}[\,\cdot\,]$ in (3.1) is a Markov process with elements $\boldsymbol{w}_k$ defined on the probability space $(\mathcal{W}, \mathscr{B}(\mathcal{W}), \Pr_{\boldsymbol{w}})$. The distribution of $\Pr_{\boldsymbol{w}}$ is first modeled as a Gaussian distribution with variance $\Sigma = 0.01I$ such that $\boldsymbol{w}_k \sim \mathcal{N}(0, \Sigma)$, and then as a Beta distribution such that $\boldsymbol{w}_k \sim \text{Beta}(2, 2)$. For the first-hitting time problem, we also demonstrate the method with an exponential distribution with parameter $\lambda = 0.5$, such that $\boldsymbol{w}_k \sim \text{Exp}(0.5)$. In order to verify the result against a dynamic programming solution for the terminal-hitting time problem, the control policy π is chosen to be a stationary policy such that $\pi = \{\alpha_0, \alpha_1, \ldots\}$ where $\alpha_0(x) = \alpha_1(x) = \ldots = \mathbf{0}$.

For the terminal-hitting time problem, we computed the approximate safety probabilities using Algorithm 2 for a time horizon $N = 3$. The approximate safety probabilities for $k \in [0, 2]$ are shown in Fig. 3.2 (a-c). We compared the terminal-hitting time result with a dynamic-programming based approach using [32], with the error $|V_0^\pi(x) - \overline{V}_0^\pi(x)|$ shown in Fig. 3.2 (d). The approximate safety probabilities for $k = 0$ for the stochastic double integrator with a Beta distribution disturbance are shown in Fig. 3.2 (e).

Fig. 3.1 shows the computation time in seconds for high-dimensional integrator systems. The increase in computation time is roughly linear in the dimensionality of the system because the system dimensionality only appears in the norm of the kernel function.

For the first-hitting time problem, we computed the approximate safety probabilities using Algorithm 3 for a time horizon $N = 3$. The approximate safety probabilities for $k \in [0, 2]$ are shown in Fig. 3.3 (a-c). The approximate safety probabilities for $k = 0$ for the stochastic double integrator with an exponential distribution disturbance are shown in Fig. 3.2 (d). The approximate safety probabilities for $k = 0$ for the stochastic double integrator with a beta distribution disturbance are shown in Fig. 3.2 (e).

3.3 Clohessy-Wiltshire-Hill System

The CWH dynamics are given by

$$\begin{aligned} \ddot{\boldsymbol{x}} - 3\omega \boldsymbol{x} - 2\omega \dot{\boldsymbol{y}} &= F_x/m_d \\ \ddot{\boldsymbol{y}} + 2\omega \dot{\boldsymbol{x}} &= F_y/m_d \end{aligned} \tag{3.2}$$

We define the state vector as $[\boldsymbol{x}, \boldsymbol{y}, \dot{\boldsymbol{x}}, \dot{\boldsymbol{y}}] \in \mathcal{X} \subseteq \Re^4$ and input vector as $[F_x, F_y] \in \mathcal{U} \subseteq \Re^2$, where $\mathcal{U} = [-0.1, 0.1] \times [-0.1, 0.1]$. We discretize the dynamics in time to obtain the discrete-time system dynamics

$$\boldsymbol{z}_{k+1} = A\boldsymbol{z}_k + Bu_k + \boldsymbol{w}_k \tag{3.3}$$

The disturbance $\boldsymbol{w}[\,\cdot\,]$ in (3.3) is a Markov process with elements $\boldsymbol{w}_k$ defined on the probability space $(\mathcal{W}, \mathscr{B}(\mathcal{W}), \Pr_{\boldsymbol{w}})$. The distribution of $\Pr_{\boldsymbol{w}}$ is modeled as a Gaussian distribution with variance $\Sigma = \mathrm{diag}(1 \times 10^{-4}, 1 \times 10^{-4}, 5 \times 10^{-8}, 5 \times 10^{-8})$ such that $\boldsymbol{w}_k \sim \mathcal{N}(0, \Sigma)$, We define the target set $\mathcal{T}$ and safe set $\mathcal{K}$ as in [10],

$$\begin{aligned} \mathcal{T} = \{z \in \Re^4 : {}&|z_1| \leq 0.1, -0.1 < z_2 < 0, \\ &|z_3| \leq 0.01, |z_4| \leq 0.01\} \end{aligned} \tag{3.4}$$

$$\mathcal{K} = \{z \in \Re^4 : |z_1| < z_2, |z_3| \leq 0.5, |z_4| \leq 0.5\} \tag{3.5}$$

and generate samples using [32] with a chance-constrained open loop controller as described in [16]. The approximate terminal-hitting safety probabilities for $k = 0$ are shown in Fig. 3.2 (f). We also computed the approximate first-hitting safety probabilities for $k = 0$, which is shown in Fig. 3.3 (f).

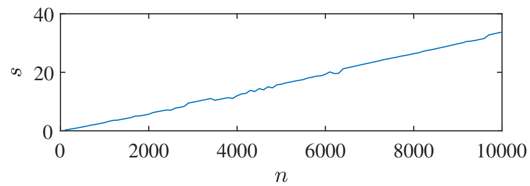

Figure 3.1: System dimensionality $[n]$ vs. average computation time $[s]$ for an n-D stochastic integrator system.

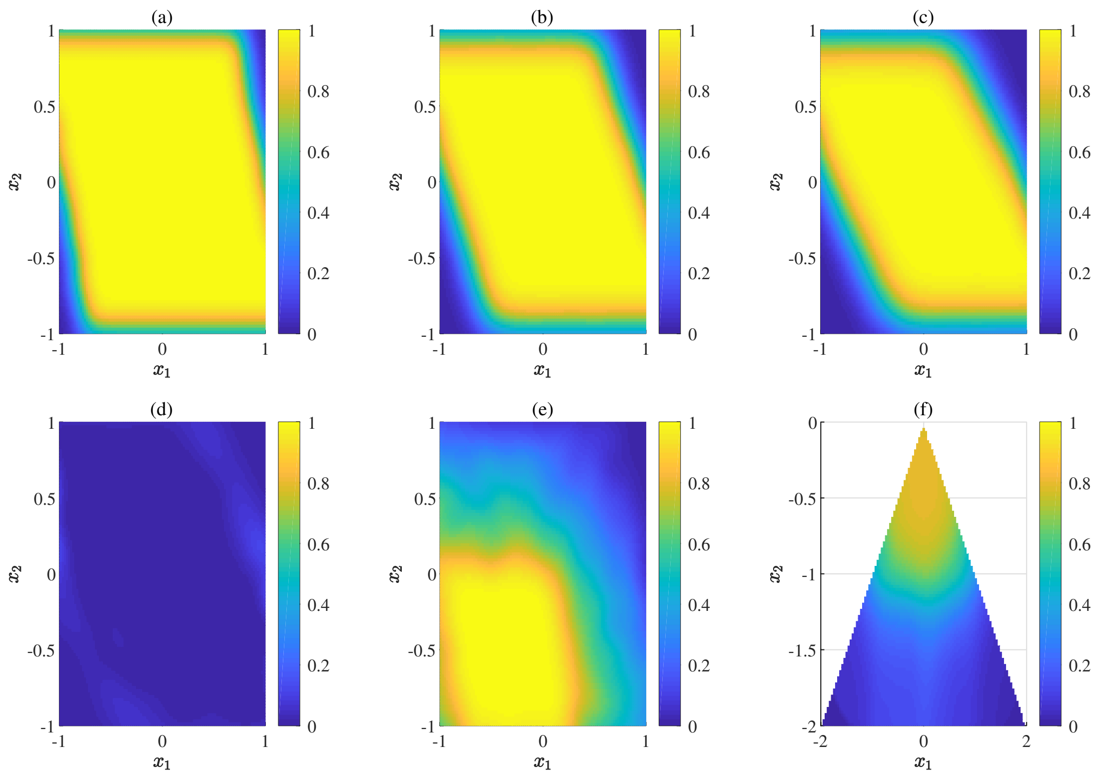

Figure 3.2: (a-c) Approximate terminal-hitting safety probabilities for a double integrator system with a Gaussian disturbance over a time horizon $N = 3$ at (a) $k = 2$, (b) $k = 1$, and (c) $k = 0$. (d) Error $|V_0^\pi(x) - \overline{V}_0^\pi(x)|$ between the dynamic programming solution and the conditional distribution embedding solution for $k = 0$ with $M = 1024$ samples. (e) Terminal-hitting safety probabilities for a double integrator system with a Beta$(2, 2)$ distribution disturbance over a time horizon $N = 3$ at $k = 0$. (f) Terminal-hitting safety Probabilities for a CWH system with a Gaussian disturbance and an open-loop, chance-affine controller over a time horizon $N = 5$ at $k = 0$.

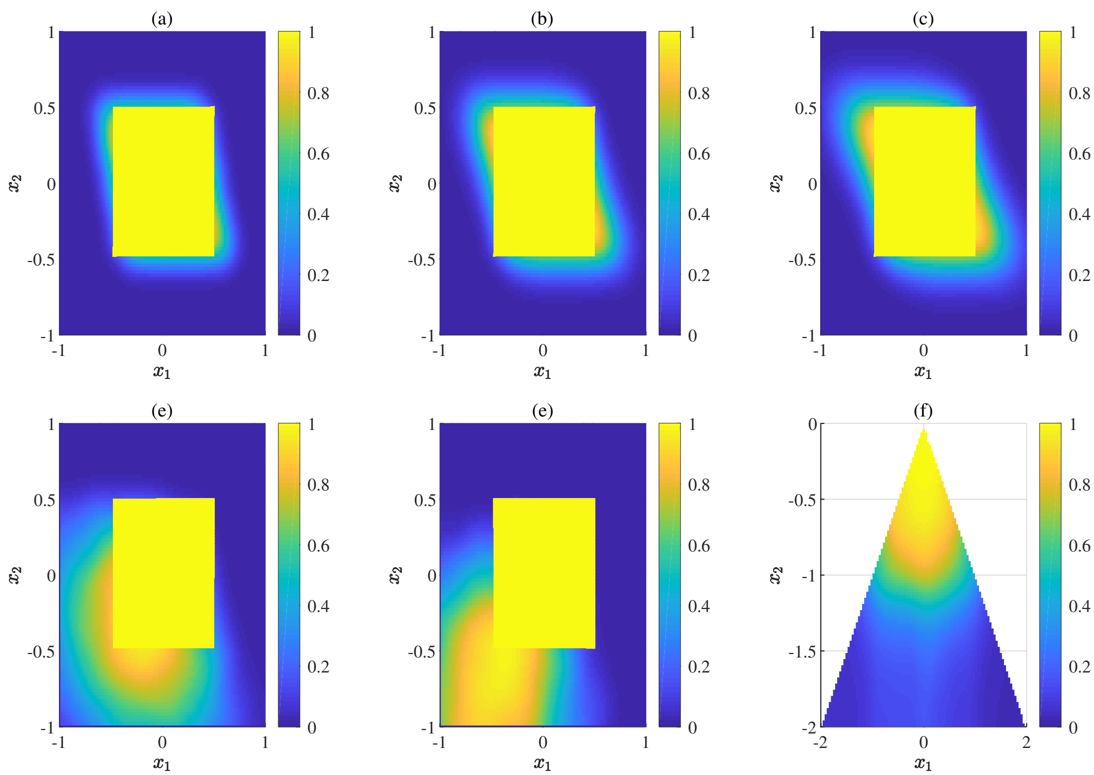

Figure 3.3: (a-c) Approximate first-hitting safety probabilities for a double integrator system with a Gaussian disturbance over a time horizon $N = 3$ at (a) $k = 2$, (b) $k = 1$, and (c) $k = 0$. (d) First-hitting safety probabilities for a double integrator system with a Exp(0.5) distribution disturbance over a time horizon $N = 3$ at $k = 0$. (e) First-hitting safety probabilities for a double integrator system with a Beta$(2, 2)$ distribution disturbance over a time horizon $N = 3$ at $k = 0$. (f) First-hitting safety Probabilities for a CWH system with a Gaussian disturbance and an open-loop, chance-affine controller over a time horizon $N = 5$ at $k = 0$.

Chapter 4

Conclusions & Future Work

In this paper, a sample-based method is outlined to solve the terminal-hitting time and first-hitting time problem for stochastic reach-avoid calculations with arbitrary disturbances. The method is demonstrated on a toy problem using an n-dimensional integrator, and the feasibility of the method is demonstrated for high-dimensional systems up to ten thousand dimensions. State-of-the-art in this area, as far as the authors are aware, is ≈ 40 dimensions [31].

Future work includes performing characteristic function analysis for controller synthesis. Because the characteristic function can be evaluated as an expectation, the conditional distribution embedding method outlined in 2 can be used to evaluate the characteristic function as an inner product within the RKHS. Prior work that leverages the characteristic function for maximally safe controller synthesis [31] could thus be used to compute a maximally safe policy using established methods.

We would also like to demonstrate the method for autonomous systems with a neural network controller, discrete-time stochastic hybrid systems, as well as for a human-in-the-loop system. Until now, systems with autonomous and human elements have yielded overly conservative safety estimates using traditional verification techniques. Because the method outlined above is largely agnostic to the dynamical system model, it could easily be extended to these complex or highly stochastic systems.

References

[1] Alessandro Abate, Maria Prandini, John Lygeros, and Shankar Sastry. Probabilistic reachability and safety for controlled discrete time stochastic hybrid systems. *Automatica*, 44(11):2724–2734, November 2008.

[2] N. Aronszajn. Theory of reproducing kernels. *Transactions of the American Mathematical Society*, 68(3):337–404, 1950.

[3] Richard E Bellman and Stuart E Dreyfus. *Applied dynamic programming*, volume 2050. Princeton university press, 2015.

[4] Dimitri P. Bertsekas and Steven E. Shreve. *Stochastic optimal control: the discrete-time case*. Athena Scientific Publ., 1996.

[5] Patrick Billingsley. *Probability and measure*. John Wiley & Sons, 2012.

[6] Francesco Borghesan, Riccardo Vignali, Luigi Piroddi, Maria Prandini, and Martin Strelec. Approximate dynamic programming-based control of a building cooling system with thermal storage. In *IEEE PES ISGT Europe 2013*, pages 1–5. IEEE, 2013.

[7] Yuan Shih Chow and Henry Teicher. *Probability theory: independence, interchangeability, martingales*. Springer Science & Business Media, 2012.

[8] Andreas Christmann and Ingo Steinwart. Support vector machines. 2008.

[9] Kenji Fukumizu, Arthur Gretton, Xiaohai Sun, and Bernhard Schölkopf. Kernel measures of conditional dependence. In J. C. Platt, D. Koller, Y. Singer, and S. T. Roweis, editors, *Advances in Neural Information Processing Systems 20*, pages 489–496. Curran Associates, Inc., 2008.

[10] J. D. Gleason, A. P. Vinod, and M. M. K. Oishi. Underapproximation of reach-avoid sets for discrete-time stochastic systems via lagrangian methods. In *2017*

IEEE 56th Annual Conference on Decision and Control (CDC), pages 4283–4290, Dec 2017.

[11] A. Gretton, K. Borgwardt, M. Rasch, B. Schölkopf, and A. Smola. A kernel two-sample test. *Journal of Machine Learning Research*, 13:723–773, March 2012.

[12] Arthur Gretton, Karsten Borgwardt, Malte Rasch, Bernhard Schölkopf, and Alex J. Smola. A kernel method for the two-sample-problem. In B. Schölkopf, J. C. Platt, and T. Hoffman, editors, *Advances in Neural Information Processing Systems 19*, pages 513–520. MIT Press, 2007.

[13] S. Grünewälder, G. Lever, L. Baldassarre, S. Patterson, A. Gretton, and M. Pontil. Conditional mean embeddings as regressors. In *Proceedings of the 29th International Conference on Machine Learning*, pages 1823–1830, New York, NY, USA, 2012. Omnipress.

[14] S. Grünewälder, G. Lever, Ll. Baldassarre, M. Pontil, and A. Gretton. Modelling transition dynamics in mdps with rkhs embeddings. In *Proceedings of the 29th International Conference on Machine Learning*, pages 535–542, New York, NY, USA, 2012. Omnipress.

[15] Quoc Le, Tamás Sarlós, and Alex Smola. Fastfood-approximating kernel expansions in loglinear time. In *Proceedings of the international conference on machine learning*, volume 85, 2013.

[16] K. Lesser, M. Oishi, and R. S. Erwin. Stochastic reachability for control of spacecraft relative motion. In *52nd IEEE Conference on Decision and Control*, pages 4705–4712, Dec 2013.

[17] C. A. Micchelli and M. Pontil. On learning vector-valued functions. *Neural Computation*, 17(1):177–204, Jan 2005.

[18] K. Muandet, K. Fukumizu, B. Sriperumbudur, and B. Schölkopf. *Kernel Mean Embedding of Distributions: A Review and Beyond*. now, 2017.

[19] Y. Nishiyama, A. Boularias, A. Gretton, and K. Fukumizu. Hilbert space embeddings of pomdps. 2012.

[20] Martin L Puterman. *Markov decision processes: discrete stochastic dynamic programming*. John Wiley & Sons, 2014.

[21] Ali Rahimi and Benjamin Recht. Random features for large-scale kernel machines. In *Advances in neural information processing systems*, pages 1177–1184, 2008.

[22] Hossein Sartipizadeh, Abraham P Vinod, Behcet Acikmese, and Meeko Oishi. Voronoi partition-based scenario reduction for fast sampling-based stochastic reachability computation of lti systems. *arXiv preprint arXiv:1811.03643*, 2018.

[23] B. Schölkopf and AJ. Smola. *Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond.* Adaptive Computation and Machine Learning. MIT Press, Cambridge, MA, USA, December 2002.

[24] John Shawe-Taylor and Nello Cristianini. *Kernel Methods for Pattern Analysis.* Cambridge University Press, New York, NY, USA, 2004.

[25] L. Song, K. Fukumizu, and A. Gretton. Kernel Embeddings of Conditional Distributions: A Unified Kernel Framework for Nonparametric Inference in Graphical Models. *IEEE Signal Processing Magazine*, 30:98–111, July 2013.

[26] Le Song, Arthur Gretton, and Carlos Guestrin. Nonparametric tree graphical models. In *Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics*, pages 765–772, 2010.

[27] Le Song, Jonathan Huang, Alex Smola, and Kenji Fukumizu. Hilbert space embeddings of conditional distributions with applications to dynamical systems. In *Proceedings of the 26th Annual International Conference on Machine Learning*, ICML '09, pages 961–968, New York, NY, USA, 2009. ACM.

[28] Bharath K. Sriperumbudur, Arthur Gretton, Kenji Fukumizu, Bernhard Schölkopf, and Gert R.G. Lanckriet. Hilbert space embeddings and metrics on probability measures. *J. Mach. Learn. Res.*, 11:1517–1561, August 2010.

[29] Sean Summers and John Lygeros. Verification of discrete time stochastic hybrid systems: A stochastic reach-avoid decision problem. *Automatica*, 46(12):1951–1961, December 2010.

[30] A Vinod, Vignesh Sivaramakrishnan, and M Oishi. Piecewise-affine approximation-based stochastic optimal control with gaussian joint chance constraints. In *Proc. Amer. Ctrl. Conf.(submitted)*, 2019.

[31] A. P. Vinod and M. M. K. Oishi. Scalable underapproximation for the stochastic reach-avoid problem for high-dimensional lti systems using fourier transforms. *IEEE Control Systems Letters*, 1(2):316–321, Oct 2017.

[32] Abraham P. Vinod, Joseph D. Gleason, and Meeko M. K. Oishi. SReachTools: A MATLAB Stochastic Reachability Toolbox, April 16–18 2019.

[33] Abraham P Vinod, Baisravan HomChaudhuri, Christoph Hintz, Anup Parikh, Stephen P Buerger, Meeko MK Oishi, Greg Brunson, Shakeeb Ahmad, and Rafael Fierro. Multiple pursuer-based intercept via forward stochastic reachability. In *2018 Annual American Control Conference (ACC)*, pages 1559–1566. IEEE, 2018.

[34] Abraham P. Vinod and Meeko M. K. Oishi. Scalable underapproximative verification of stochastic lti systems using convexity and compactness. In *Proceedings of the 21st International Conference on Hybrid Systems: Computation and Control (Part of CPS Week)*, HSCC '18, pages 1–10, New York, NY, USA, 2018. ACM.

[35] Abraham P Vinod, Sean Rice, Yuanqi Mao, Meeko MK Oishi, and Behçet Açıkmeşe. Stochastic motion planning using successive convexification and probabilistic occupancy functions. In *2018 IEEE Conference on Decision and Control (CDC)*, pages 4425–4432. IEEE, 2018.