Data-Driven Stochastic Reachability Using Hilbert Space Embeddings

Adam J. Thorpe, Kendric R. Ortiz, Meeko M. K. Oishi

Department of Electrical and Computer Engineering, University of New Mexico, Albuquerque, USA

Abstract

We compute finite sample bounds for approximations of the solution to stochastic reachability problems computed using kernel distribution embeddings, a non-parametric machine learning technique. Our approach enables assurances of safety from observed data, through construction of probabilistic violation bounds on the computed stochastic reachability probability. By embedding the stochastic kernel of a Markov control process in a reproducing kernel Hilbert space, we can compute the safety probabilities for systems with arbitrary disturbances as simple matrix operations and inner products. We present finite sample bounds for the approximation using elements from statistical learning theory. We numerically evaluate the approach, and demonstrate its efficacy on neural net-controlled pendulum system.

Key words: Machine Learning, Stochastic Systems, Reachability, Safety, Optimal Control

1 Introduction

In expensive, high risk, or safety-critical systems, tools for verification are important for ensuring correctness before testing, implementation, or deployment. As autonomy grows in prevalence, and systems continue to grow in size and complexity, there is a need to extend such tools to accommodate learning-enabled components in autonomous systems that resist traditional modeling. Stochastic reachability is an established tool for model-based verification and probabilistic safety [1, 39], that has been applied to safety-critical problems in a variety of domains, including spacecraft rendezvous and docking [20], and robotics and path planning [23], and vehicle control [16]. Safety refers to the ability of trajectories of the system to respect known constraints on the state space, with at least a desired likelihood, despite bounded control authority. In this paper, we characterize confidence bounds on a data-driven approach for stochastic reachability, enabling rigorous assurances of safety in a model-free manner.

The solution to stochastic reachability problems is framed as the solution to a dynamic program [39], which scales exponentially with the dimensionality of the system, and hence suffers from the curse of dimensionality [4]. Model-based methods have been developed that leverage approximate dynamic programming [17], particle filtering [20, 24], and abstractions [37], but are limited to systems of moderate dimensionality. Optimization-based solutions have garnered modest computational tractability via chance constraints [20, 47], sampling methods [31, 43, 46], and convex optimization with Fourier transforms [44, 45], but these approaches are generally limited to linear dynamical systems and Gaussian or log-concave disturbances.

We propose a method to compute the stochastic reachability safety probabilities with confidence bounds, us-
ing a machine learning technique known as kernel methods. Kernel methods have long been used in probability and statistics, [5, 28], and more recently applied to Markov models [15], partially observable systems [27, 33], and policy synthesis [21]. Finite sample bounds for conditional distribution embeddings in these contexts are presented in [14, 15, 33, 35, 36], which show that the estimated expectation converges in probability to the true expected value with known, asymptotic convergence rate. We have applied this approach to the stochastic reachability problem [41], and obtained an asymptotic bound obtained in a similar manner as [13, 36]. However, the confidence interval that results from this approach is often so large as to be impractical.

In this paper, we construct finite sample bounds for kernel embedding based computation of the stochastic reachability probability, that are specific to the stochastic reachability problem. Our main contribution is the construction of state- and input-based upper and lower bounds, obtained via concentration inequalities [25] and tools from statistical learning theory [42]. Our proposed approach can provide much tighter bounds than those available via [41] because of its state dependence, and because it accommodates the fact that the gathered data may not be available uniformly through the state and input space. That is, the bound is dependent upon the sample set from which the kernel embedding is inferred. Further, in contrast to statistical verification [29, 48, 50], this approach is amenable to controller synthesis as well as construction of stochastic reachable sets.

The paper is organized as follows. In Section 2, we formulate the problem and provide relevant background information. In Section 3, we derive the finite sample bounds. We discuss implications of the proposed bounds in Section 4, and the problem of parameter selection. In Section 5, we numerically validate the bounds on stochastic reachability problem for a nonlinear pendulum system and a nonlinear cart-pole system with black box neural network controllers. Concluding remarks are provided in Section 6.

2 Preliminaries

We employ the following notational conventions. Let $E$ be an arbitrary nonempty space. The indicator function $1_A : E \to \{0, 1\}$ of $A \subseteq E$ is defined such that $1_A(x) = 1$ if $x \in A$, and $1_A(x) = 0$ if $x \notin A$. Let $E$ denote the stochastic process, $X \in \mathbb{R}$, and $F$ denote the Borel-$\sigma$-algebra on $E$. If $E$ is a topological space [9], the $\sigma$-algebra generated by the set of all open subsets of $E$ is called the Borel $\sigma$-algebra, denoted by $\mathcal{B}(E)$. Let $\mathcal{P} : F \to [0, 1]$ be a probability measure on the measurable space $(\Omega, F)$. A measurable function $X : \Omega \to E$ is called a random variable taking values in $(E, F)$. The image of $\mathcal{P}$ under $X$, $\mathcal{P}(X^{-1}A), A \in \mathcal{E}$ is called the distribution of $X$. Let $T$ be an arbitrary set, and for each $t \in T$, let $X_t$ be a random variable. The collection of random variables $\{X_t : t \in T\}$ on $(\Omega, \mathcal{F})$ is a stochastic process.

2.1 System Model & Stochastic Reachability Problem

Consider a stochastic dynamical system described by a Markov control process as defined in [39].

Definition 1 (Markov Control Process). A Markov control process $\mathcal{H}$ is defined as a 3-tuple, $\mathcal{H} = (\mathcal{X}, \mathcal{U}, Q)$, consisting of a Borel space $\mathcal{X} \subseteq \mathbb{R}^n$ representing the state space, a compact Borel space $\mathcal{U} \subseteq \mathbb{R}^m$ representing the control space, and $Q : \mathcal{B}(\mathcal{X}) \times \mathcal{U} \times [0, 1]$, a Borel-measurable stochastic kernel which assigns a probability measure $Q(\cdot | x, u)$ to every $x \in \mathcal{X}$ and $u \in \mathcal{U}$ on the Borel space $(\mathcal{X}, \mathcal{B}(\mathcal{X}))$.

The system evolves over a finite time horizon $k \in [0, N]$ from an initial condition $x_0 \in \mathcal{X}$. The control inputs are chosen according to a Markov control policy $\pi = \{\pi_0, \pi_1, \ldots, \pi_{N-1}\}$ [6], which is a sequence of universally-measurable maps $\pi_i : \mathcal{X} \to \mathcal{U}, i = 0, 1, \ldots, N - 1$.

Let $K, T \in \mathcal{B}(\mathcal{X})$, denote the safe set and target set, respectively. From [39], the terminal hitting time safety probability is defined as the probability that a system $\mathcal{H}$ following a Markov policy $\pi$ from the initial condition $x_0$ will reach the target set $T$ at time $N$ while remaining within the safe set $K$ for all $k = 0, 1, \ldots, N - 1$.

$$r_{x_0}^\pi(K, T) = \mathbb{E}_{x_0}^\pi \left[ \prod_{i=0}^{N-1} 1_K(x_i) 1_T(x_N) \right] \tag{1}$$

For a fixed Markov policy $\pi$, we define the value functions $V_k^\pi : \mathcal{X} \to [0, 1], k = 0, \ldots, N$, via backward recursion:

$$V_0^\pi(x) = 1_T(x) \tag{2}$$

$$V_k^\pi(x) = 1_K(x) \int_{\mathcal{X}} V_{k+1}^\pi(y) Q(dy | x, \pi_k(x)) \tag{3}$$

Then, according to [39], we have that $V_0^\pi(x) = r_{x_0}^\pi(K, T)$ for every $x_0 \in \mathcal{X}$.

We presume that a finite sample $S = \{(y_i, x_i, u_i)\}_{i=1}^M$ of $M$ observations are available, taken i.i.d. from the stochastic kernel $Q$, where $y_i \sim Q(\cdot | x_i, u_i)$ and $u_i = \pi(x_i)$. We assume that the stochastic kernel $Q$ is unknown, which means the integral in (3) becomes intractable. In order to numerically compute the expectation without prior knowledge of the stochastic kernel, we utilize a technique known as conditional distribution embeddings [15, 34, 36] to model the probability measures $Q(\cdot | x, u)$ in Hilbert space and approximate the integral operator of the stochastic kernel in (3).
2.2 Conditional Distribution Embeddings

**Definition 2 (RKHS).** For any arbitrary space $E$, let $\mathcal{H}_E$ denote a Hilbert space of functions $f : E \to \mathbb{R}$ equipped with the inner product $\langle \cdot, \cdot \rangle_{\mathcal{H}_E}$ and the induced norm $\|\cdot\|_{\mathcal{H}_E}$. A Hilbert space $\mathcal{H}_E$ is a reproducing kernel Hilbert space (RKHS) if there exists a positive definite kernel $k_E : E \times E \to \mathbb{R}$ that satisfies the following properties [2]:

\[ k_E(x, \cdot) \in \mathcal{H}_E \quad \forall x \in E \]  
\[ f(x) = \langle f, k_E(x, \cdot) \rangle_{\mathcal{H}_E} \quad \forall f \in \mathcal{H}_E, \forall x \in E \]  

where (4b) is called the reproducing property, and for any $x, x' \in E$, we denote $k_E(x, \cdot)$ in the RKHS $\mathcal{H}_E$ as a function on $E$ such that $x' \mapsto k_E(x, x')$.

Conversely, by the Moore-Aronszajn theorem [2], for any positive definite kernel function $k_E$, there exists a unique RKHS $\mathcal{H}_E$ with $k_E$ as its reproducing kernel, where $\mathcal{H}_E$ is the closure of the linear span of functions $k_E(x, \cdot)$.

For the measurable space $X$, let $\mathcal{P}$ denote the set of probability measures on $X$ which are densities of $Y \in \mathcal{X}$ conditioned on $(X, U) \in X \times U$. For any measure $Q \in \mathcal{P}$, if the sufficient condition $E_{Y \sim Q}[k_X(X, Y)] < \infty$ is satisfied, there exists an element $m_{Y|x,u}$ in the RKHS $\mathcal{H}_X$ called the conditional distribution embedding [36], which is defined as:

\[ m : \mathcal{P} \to \mathcal{H}_X \]  
\[ Q \mapsto m_{Y|x,u} := \int_X k_X(y, \cdot)Q(dy | x, u) \]  

Using the reproducing property (4b), we can compute the integral in (3) as an inner product in Hilbert space with the conditional distribution embedding,

\[ \langle m_{Y|x,u}, V_{k+1}^\pi \rangle_{\mathcal{H}_X} = \int_X V_{k+1}^\pi(y)Q(dy | x, u) \]  

However, because $Q$ is unknown, we do not have access to $m_{Y|x,u}$ directly. Instead, the sample $S$ drawn i.i.d. from $Q$ can be used to compute an estimate $\hat{m}_{Y|x,u} \in \mathcal{H}_X$ which is found by minimizing the following regularized least-squares optimization problem [14]:

\[ \sum_{i=1}^M \|k_X(y_i, \cdot) - \hat{m}_{Y|x,u}\|_{\mathcal{H}_X}^2 + \lambda\|m_{Y|x,u}\|_{\mathcal{H}_X}^2 \]  

where $\Gamma$ is a vector-valued RKHS [26] and $\lambda > 0$ is the regularization parameter. According to [14], the solution to (7) is unique and has the following form:

\[ \hat{m}_{Y|x,u} = \Phi^T(\Psi\Psi^T + \lambda\Gamma)^{-1}\Psi k_{X \times U}((x, u), \cdot) \]  

Where $\Phi$ and $\Psi$ are known as feature vectors, given by:

\[ \Phi = [k_X(y_1, \cdot), \ldots, k_X(y_M, \cdot)]^T \]  
\[ \Psi = [k_{X \times U}(x_1, u_1), \ldots, k_{X \times U}(x_M, u_M), \cdot]^T \]

We can then approximate the integrals in (3) as an inner product with the estimate $\hat{m}_{Y|x,u}$.

\[ \langle \hat{m}_{Y|x,u}, V_{k+1}^\pi \rangle_{\mathcal{H}_X} \approx \int_X V_{k+1}^\pi(y)Q(dy | x, \pi_k(x)) \]  

For simplicity, we can write the inner product as:

\[ \langle \hat{m}_{Y|x,u}, V_{k+1}^\pi \rangle_{\mathcal{H}_X} = V_{k+1}^\pi \beta(x, u) \]  

where $V_{k+1}^\pi = [V_{k+1}^\pi(y_1), \ldots, V_{k+1}^\pi(y_M)]^T$ and $\beta(x, u) \in \mathbb{R}^M$ is a vector of coefficients that depends on the value of the conditioning variables $(x, u) \in X \times U$.

\[ \beta(x, u) = (\Psi\Psi^T + \lambda\Gamma)^{-1}\Psi k_{X \times U}((x, u), \cdot) \]

As shown in [41], by approximating the integrals in (3) using (12), we obtain an approximation of the safety probabilities in (1).

**Lemma 3 (Approximate Backward Recursion [41]).** Let $\pi$ be a fixed Markov policy. Define the approximate value functions $\tilde{V}_k^\pi : X \to [0, 1]$, $k \in [0, N - 1]$ by the backward recursion:

\[ \bar{V}_N^\pi(x) = 1 \]
\[ \bar{V}_k^\pi(x) = \mathbf{1}_K(x)(\hat{m}_{Y|x,\pi(x)}, \bar{V}_{k+1}^\pi)_{\mathcal{H}_X} \]  

Then $\bar{V}_{x_0}^\pi(K, \mathcal{T}) \approx \tilde{V}_{x_0}^\pi(x)$. 

Lemma 3 provides a model-free approach to approximate of the stochastic reachability probability, and can easily be extended to solve related problems, including the first-hitting time problem [39] and the multiplicative and maximal cost stochastic reachability problems in [1].
3 Finite Sample Bounds

The difficulty in finding bounds on the stochastic reachability probability stems from the underlying structure of the conditional distribution embedding estimate. Unlike the embedding for a marginal distribution [32], which has uniform coefficients $1/M$, the conditional distribution embedding has non-uniform coefficients $\beta(x, u)$ (12) which depend upon the value of the conditioning variables. This complicates the application of existing mathematical techniques from statistical learning theory.

In order to determine a bound on the quality of the approximation obtained using Lemma 3, we seek a bound on the difference between the expectation of the value functions and its empirical counterpart.

3.1 Worst-Case Difference Between the True and the Empirical Expectation

Assume that $f \in [0, 1]$ and $\|f\|_{\mathcal{H}_X} \leq 1$ for all $f \in \mathcal{H}_X$. First, we uniformly bound the difference between the value function expectation and the empirical expectation computed using the estimate $\hat{\mu}_Y|_{x,u}$,

$$\left| E[V_{k+1}^\dagger] - E_S[V_{k+1}^\dagger] \right| \leq \sup_{\|h\|_{\mathcal{H}_X} \leq 1} \left| E[h] - E_S[h] \right| \tag{16}$$

where $E_S[V_{k+1}^\dagger]$ is defined as in (12), $E_S[h] = h^\top \beta(x, u)$, $h = [h(y_1), \ldots, h(y_M)]^\top$, and $\beta(x, u)$ is computed using (13). We then bound the right hand side of (16) using McDiarmid’s inequality.

Theorem 4 (McDiarmid’s inequality [25]). Let $X = \{X_1, \ldots, X_M\}$ be independent random variables taking values in a set $E$, and assume that the function $f : E^M \to \mathbb{R}$ satisfies the bounded differences condition (Definition 5). Then for every $\varepsilon > 0$,

$$\Pr(|f(X) - E[f(X)]| \geq \varepsilon) \leq \exp\left(-\frac{2\varepsilon^2}{\sum_{i=1}^{M} c_i^2}\right) \tag{17}$$

Definition 5 (Bounded Differences Condition). Given coefficients $c_i \geq 0$, $i \in [1, M]$, a function $f : E^M \to \mathbb{R}$ satisfies the bounded differences condition if

$$\sup_{x_1, \ldots, x_M, x'_i \in E} |f(x_1, \ldots, x_M) - f(x_1, \ldots, x'_i, \ldots, x_M)| \leq c_i$$

for every $i \in [1, M]$.

In order to determine the bound on (16), we seek some $C \geq c_i$ for all $i \in [1, M]$ that satisfies the bounded differences condition. However, the effect of changing an individual observation in the empirical expectation term in (16) is non-trivial, since changing a single observation affects all elements of the coefficient vector $\beta(x, u)$. Therefore, in order to determine $C$, we decompose the empirical expectation $\mathbf{h}^\top (\mathbf{Ψ}_1 + \lambda M I)^{-1} \mathbf{Ψ}_{kX \times U}((x, u), \cdot)$ from (12) to determine its dependence on any given observation and bound the individual terms.

Note that when a single observation in $S$ is changed, the row and column corresponding to the changed observation in the positive semi-definite matrix $\mathbf{Ψ}_1 + \lambda M I$, which can be found for example using the Frobenius theorem [12] or using the trace of the matrix [49]. Let $W = (\mathbf{Ψ}_1 + \lambda M I)^{-1}$. Using basic properties of positive semi-definite matrices and Sylvester’s criterion, which gives conditions for positive semi-definite Hermitian matrices, the elements in $W$ are bounded by $|w_{ij}| \leq 1/\ell$. Furthermore, the off-diagonal elements obey the following property: $|w_{ij}| \leq \sqrt{w_{ii}w_{jj}}$. We now prove the following proposition:

Proposition 6. The deviation of $\langle \hat{\mu}_Y|x,u,h \rangle_{\mathcal{H}_X}$ by changing the $n$th observation in $S$ is less than or equal to $(2M - 1)\ell^2$.

Proof. Let $h = [h(y_1), \ldots, h(y_M)]^\top$ as defined in (16), $v = \mathbf{Ψ}_{kX \times U}((x, u), \cdot)$, and $W = (w_{ij}) \in \mathbb{R}^{M \times M}$, where $W = (\mathbf{Ψ}_1 + \lambda M I)^{-1}$. We can write the inner product $\langle \hat{\mu}_Y|x,u,h \rangle_{\mathcal{H}_X}$ using (13) as:

$$\langle \hat{\mu}_Y|x,u,h \rangle_{\mathcal{H}_X} = \sum_{i=1}^{M} \sum_{j=1}^{M} h_i w_{ij} v_j \tag{19}$$

Then, using the form in (19), we collect the terms corresponding to the $n$th observation.

$$\sum_{i=1}^{M} \sum_{j=1}^{M} h_i w_{ij} v_j = \sum_{i=1}^{M} \sum_{j=1}^{M} h_i w_{ij} v_j + \sum_{i=1}^{M} h_i w_{in} v_n + \sum_{j=1}^{M} h_n w_{nj} v_j + h_n w_{nn} v_n \tag{20}$$

Since $h_i \in [0, 1]$ by assumption, $v_j \in [0, \rho]$ by definition, and $|w_{ij}| \leq 1/\ell$, we have from (20) that the deviation for changing a single observation is less than or equal to $(2M - 1)\ell^2$, which proves the result.

Thus, from Proposition 6, we have that $C = (2M - 1)\ell^2$ satisfies the condition in Definition 5. Continuing from
(16), using McDiarmid’s inequality (17), we have that given \( \delta / 2 \in (0, 1) \), with probability \( 1 - \delta / 2 \),

\[
\sup_{\|h\|_{\mathcal{F}_X} \leq 1} |E_{Y \sim Q}[h(Y)] - h^T \beta(x, u)| \\
\leq E_{\mathcal{S}} \left[ \sup_{\|h\|_{\mathcal{F}_X} \leq 1} |E_{Y \sim Q}[h(Y)] - h^T \beta(x, u)| \right] \\
+ \frac{MC^2 \log(2/\delta)}{2}
\]

We now have an expression for the worst-case difference between the true and empirical stochastic reachability probabilities. However, because the expectation on the right hand side of (21) relies upon the true expectation of \( h \), it is not directly computable.

### 3.2 Removing Reliance Upon the True Expectation for Computation

We now bound the first term on the right hand side of (21) via symmetrization [42]. Let \( \mathcal{S} \) be a ghost sample, that is, an independent copy of \( S \) [42]. We replace the expectation in the first term on the right hand side of (21) with a second empirical estimate computed using \( \mathcal{S} \), to obtain

\[
E_{\mathcal{S}} \left[ \sup_{\|h\|_{\mathcal{F}_X} \leq 1} |E_{y \sim Q}[h(y)] - h^T \beta(x, u)| \right] \\
= E_{\mathcal{S}} \left[ \sup_{\|h\|_{\mathcal{F}_X} \leq 1} |E_{\mathcal{S}}[h(\tilde{y})] - h^T \beta(x, u)| \right] \\
\leq E_{\mathcal{S} \tilde{\mathcal{S}}} \left[ \sup_{\|h\|_{\mathcal{F}_X} \leq 1} |\tilde{h}^T \beta(x, u) - h^T \beta(x, u)| \right]
\]

where \( \beta(x, u) \) is computed using \( \mathcal{S} \) as in (13) and \( \tilde{h} = [h(\tilde{y}_1), \ldots, h(\tilde{y}_N)]^T \). Equation (22) follows almost surely from the properties of conditional expectations and (23) follows by the convexity of the supremum.

We next exploit the symmetry of the empirical distributions to upper bound (23).

**Definition 7 (Rademacher Variables [3]).** A random variable \( \sigma \) is called a Rademacher variable if it is independent and identically distributed and \( \Pr(\sigma = 1) = \Pr(\sigma = -1) = 1/2 \).

Let \( \sigma \) be Rademacher variables, with \( \sigma_i \in \{-1, 1\} \) [3], and let \( \beta = \text{diag}(\beta(x, u)) \). Since the distribution of the difference in empirical expectations \( \tilde{h}^T \tilde{\beta}(x, u) - h^T \beta(x, u) \) is symmetric around 0, which follows since \( f^T \beta(x, u) \in [0, 1] \) for all \( f \in \mathcal{F}_X \), we see that \( \tilde{h}^T \tilde{\beta} - h^T \beta \) has the same distribution. In effect, the Rademacher variables randomly exchange observations in \( S \) and \( \tilde{S} \) with probability 1/2. When we take the expectation over \( \sigma \), the expectations of the empirical estimates computed using \( S \) and \( \tilde{S} \) are the same. Using this fact, we obtain the following:

\[
E_{S \tilde{S}} \left[ \sup_{\|h\|_{\mathcal{F}_X} \leq 1} |\tilde{h}^T \beta(x, u) - h^T \beta(x, u)| \right] \\
= E_{S \tilde{S} \sigma} \left[ \sup_{\|h\|_{\mathcal{F}_X} \leq 1} |\tilde{h}^T \beta \sigma - h^T \beta \sigma| \right] \\
\leq 2E_{S \tilde{S} \sigma} \left[ \sup_{\|h\|_{\mathcal{F}_X} \leq 1} |h^T \beta \sigma| \right]
\]

Then by applying the reproducing property (4b) and the definition of the dual norm for Hilbert spaces [30], we remove the dependence on \( h \).

\[
2E_{S \tilde{S} \sigma} \left[ \sup_{\|h\|_{\mathcal{F}_X} \leq 1} |h^T \beta \sigma| \right] \\
= 2E_{S \tilde{S} \sigma} \left[ \sup_{\|h\|_{\mathcal{F}_X} \leq 1} |\langle h, \Phi^T \beta \sigma \rangle_{\mathcal{F}_X}| \right] \\
= 2E_{S \tilde{S} \sigma} \left[ \|\Phi^T \beta \sigma\|_{\mathcal{F}_X} \right]
\]

We then utilize the definition of the Hilbert-Schmidt norm and the concavity of the square root and note that the expectation of the Rademacher variables \( E_{\sigma}[\sigma_i \sigma_j] \) vanishes except when \( i = j \).

\[
2E_{S \tilde{S} \sigma} \left[ \|\Phi^T \beta \sigma\|_{\mathcal{F}_X} \right] \\
\leq 2E_{S \tilde{S}} \left[ \left( E_{\sigma} \left[ \sigma^T \beta^T \Phi^T \beta \sigma \right] \right)^{1/2} \right] \\
= 2E_{S \tilde{S}} \left[ \left( \text{tr} \left( \beta^T \Phi^T \beta \right) \right)^{1/2} \right]
\]

By bounding the expectation in (29) by McDiarmid’s inequality again, we obtain

\[
2E_{S} \left[ \left( \text{tr} \left( \beta^T \Phi^T \beta \right) \right)^{1/2} \right] \\
\leq 2 \sqrt{\text{tr}(\beta^T \Phi^T \beta)} + 2 \sqrt{MC^2 \log(2/\delta) / 2}
\]

which is independent of the true expected value, and hence computable. Continuing from (21), using McDiarmid’s inequality in (30) with \( C = (2M - 1)\Phi \) via Proposition 6, we have that given \( \delta / 2 \in (0, 1) \), with
probability $1 - \delta/2$, 

$$
\sup_{\|h\|_{\mathcal{H}_X} \leq 1} |E_{Y \sim Q}[h(Y)] - h^\top \beta(x, u)| 
\leq 2 \sqrt{\text{tr}(\beta^\top \Phi \Phi^\top \beta)} + 3 \sqrt{MC^2 \log(2/\delta)}
$$

(31)

Thus, we have a computable bound for the worst-case difference between the true and empirical expectation of a function $h \in \mathcal{H}_X$.

### 3.3 Finite Sample Bound on the Stochastic Reachability Probability

We now can state the main result. We present the main result as Theorem 8.

**Theorem 8.** For any value function $V_{k+1}^\pi$, given $\delta/2 \in (0, 1)$, with probability $1 - \delta/2$, the difference between the true and empirical expectation of the value functions is bounded by

$$
|E_{Y \sim Q}[V_{k+1}^\pi(Y)] - V_{k+1}^\pi(x, u)| 
\leq 2 \sqrt{\text{tr}(\beta^\top \Phi \Phi^\top \beta)} + 3 \sqrt{MC^2 \log(2/\delta)}
$$

(32)

**PROOF.** The proof follows from the arguments presented for (31). In (16), we uniformly bound the difference between the true and empirical expectation of the value functions by the worst-case function $h \in \mathcal{H}_X$. We then use McDiarmid’s inequality (Theorem 4) with the bound $C = (2M - 1)/\delta$ satisfying the bounded differences condition (Proposition 6) to obtain (21). Using a ghost sample, the symmetrization argument, and the definition of the dual norm for Hilbert spaces, we then bound the first term on the right hand side of (21) in order to remove the dependence on $h$.

$$
E_S \left[ \sup_{\|h\|_{\mathcal{H}_X} \leq 1} |E_{Y \sim Q}[h(Y)] - h^\top \beta(x, u)| \right] 
\leq 2 \sqrt{\text{tr}(\beta^\top \Phi \Phi^\top \beta)} + 3 \sqrt{MC^2 \log(2/\delta)}
$$

(33)

We then substitute the bound in (33) into (21) and obtain the result,

$$
|E_{Y \sim Q}[V_{k+1}^\pi(Y)] - V_{k+1}^\pi(x, u)| 
\leq 2 \sqrt{\text{tr}(\beta^\top \Phi \Phi^\top \beta)} + 3 \sqrt{MC^2 \log(2/\delta)}
$$

(34)

which proves (32).

Let $B(x, u)$ be the bound on the difference between the expected value of the value functions and its empirical counterpart in (32), given by

$$
B(x, u) = 2 \sqrt{\text{tr}(\beta^\top \Phi \Phi^\top \beta)} + 3 \sqrt{MC^2 \log(2/\delta)}
$$

(35)

This means that for any value function $V_{k+1}^\pi$, given $\delta/2 \in (0, 1)$, with probability $1 - \delta/2$, that the absolute difference between the actual expectation and the empirical expectation computed using $\hat{m}_{Y|x,u}$ is bounded by

$$
-B(x, u) \leq E[V_{k+1}^\pi] - E_S[V_{k+1}^\pi] \leq B(x, u)
$$

(36)

We summarize this result, which provides a bound on the approximation of the value functions used in the backward recursion in Lemma 3, in the following Corollary. Because of the uniform bound on the difference of expectations, this bound can be applied to the expectation of any function $f \in \mathcal{H}_X$.

**Corollary 9.** Assume that $f \in [0, 1]$ for all $f \in \mathcal{H}_X$ with $\|f\|_{\mathcal{H}_X} \leq 1$. For any $\delta/2 \in (0, 1)$ with probability $1 - \delta/2$, the difference between the expectation of a function $f$ and its empirical counterpart is bounded by:

$$
|E_{Y \sim Q}[f(y)] - f^\top \beta(x, u)| 
\leq 2 \sqrt{\text{tr}(\beta^\top \Phi \Phi^\top \beta)} + 3 \sqrt{MC^2 \log(2/\delta)}
$$

(37)

**PROOF.** The proof follows directly from the proof of Theorem 8. Since for any function $f \in \mathcal{H}_X$, the difference between the expectation and its empirical counterpart is uniformly bounded by the difference in expectations of the worst-case function $h \in \mathcal{H}_X$, it follows that

$$
|E[f] - E_S[f]| \leq \sup_{\|h\|_{\mathcal{H}_X} \leq 1} |E[h] - E_S[h]|
$$

(38)

and thus has the bound presented in (31), which concludes the proof.

Thus, by applying this bound to the expectations in the backward recursion, we obtain an overall bound on the approximation of the safety probabilities obtained using Lemma 3. Furthermore, the bound in (37) depends on the value of the conditioning variables $(x, u)$, which means the bound can serve as an indication of the quality of the approximation at a particular point.

### 4 Kernel and Parameter Selection

The quality of the approximation of the stochastic reachability probability is governed by a number of factors,
including the choice of kernel function, parameters associated with the kernel function, the regularization parameter of the least-squares problem (7), and the sample, \( S \). In a realistic setting, we typically do not have explicit control over the sample \( S \) or the number of observations in the sample. Thus, the choice of kernel function and the model parameters plays an important role in the quality of the kernel-based approximation.

4.1 Kernel Selection

The performance of kernel based learning algorithms is closely tied to the choice of kernel function and the structure of the RKHS. In essence, the Hilbert space needs to be rich enough to model the set of probability measures underlying the observed data without overfitting. Thus, we frame the problem of kernel selection as a problem of limiting the complexity of the RKHS \([3, 42]\), where complexity in statistical learning literature refers to the ability of a function class to fit random noise. Intuitively, by choosing a function class that lowers the complexity term, we reduce the possibility that our function class \( \mathcal{H}_X \) will overfit the observed data.

We propose a complexity term that is closely related to the Rademacher complexity \([3]\) from statistical learning theory, but instead based on the bound presented in Corollary 9 that accommodates non-uniform coefficients \( \beta(x, u) \). Define the random variable

\[
\hat{C}(\mathcal{H}_X) = E_{\sigma} \left[ \sup_{\|h\|_{\mathcal{H}_X} \leq 1} |h^\top \beta\sigma| \right]
\]  

(39)

Then the conditional complexity of \( \mathcal{H}_X \) is defined as \( \mathcal{C}(\mathcal{H}_X) = E_S[\hat{C}(\mathcal{H}_X)] \). Using the bound in (29), we have

\[
\mathcal{C}(\mathcal{H}_X) \leq E_S \left[ \left( \text{tr} \left( \beta^\top \Phi \beta \right) \right)^{1/2} \right]
\]  

(40)

This choice of complexity term is equivalent to the first term on the right hand side of (37). Since the conditional complexity appears in the bounds presented in Corollary 9, minimizing the complexity term also minimizes the finite sample bounds on the difference in expectations. Thus, we can choose the kernel functions which minimize the complexity term for all \( (x, u) \), effectively minimizing the finite sample bound on the difference in expectations of the value functions.

The kernel should also be chosen to satisfy universality \([38]\) and boundedness properties. One common choice of kernel that satisfies these properties is the Gaussian RBF kernel \( k_{\gamma}(x, x') = \exp(-\|x - x'\|^2/2\sigma^2) \), \( \sigma > 0 \). Universal \([32, 38]\) kernels capture all statistical information and moments of the underlying distributions. Because the mapping from the set of all probability measures \( \mathcal{P} \) into the RKHS (5) is injective for universal kernels, this means there is a unique element in the RKHS \( \mathcal{H}_X \) for any \( P, Q \in \mathcal{P} \), such that \( \|m_P - m_Q\|_{\mathcal{H}_X} = 0 \) if and only if \( P = Q \). This ensures that the conditional distribution embedding admits a unique solution \([38]\), and that we can distinguish between distributions in Hilbert space. By choosing a bounded kernel function, we ensure that \( \rho < \infty \), and we can achieve tighter bounds by selecting a kernel function with small \( \rho \). The Gaussian kernel function, for example, has \( \rho = 1 \).

Typically, a parameterized kernel which is known to satisfy these properties is chosen, and then kernel parameters are tuned via cross-validation techniques. However, nascent work has posed kernel synthesis for a given sample \( S \) as an optimization problem. This approach has been demonstrated via convex optimization \([18]\) and semi-definite programming \([19]\) for marginal distributions with scalar-valued regression. While this idea is promising, the connection between conditional distribution embeddings and the underlying regression problem has only recently begun to be explored \([14]\). Further, the extension from scalar-valued regression to vector-valued regression \([26]\) that is required by the objective function in (7) is not straightforward. Hence additional work will be needed to evaluate the feasibility of kernel synthesis methods for this problem.

4.2 Parameter Selection

Tighter bounds may be possible by identifying a strict upper bound on the elements of \( \beta(x, u) \), which in turn are influenced by the parameters \( \rho \), the upper bound on the kernel function, \( \lambda \), the least-squares regularization coefficient in (7), and \( \ell \), a positive lower bound on the eigenvalues of \( \Psi \Psi^\top + \lambda M I \).

The value of \( \rho \) is determined primarily by the choice of kernel and its parameters, and can be tuned by minimizing the complexity term \( \mathcal{C}(\mathcal{H}_X) \). The value of \( \lambda \) affects the convergence rate associated with the uniform finite sample bound. The convergence guarantees in \([36]\) and \([15]\) typically depend upon \( \lambda \) going to zero as the number of observations increases. See \([7, 10]\) for a discussion of optimal values of \( \lambda \). Lastly, \( \ell \) is determined in large part based on \( \lambda \), with lower values associated with tighter bounds.

5 Examples

We implemented Lemma 3 on a stochastic chain of integrators for the purpose of validation, and on a nonlinear cart-pole benchmark system \([22]\) with black-box neural network controllers to demonstrate the capabilities of the proposed approach. For each problem, we generated a sample \( S \) of observations via simulation, and
In this section, we consider a 2-D stochastic chain of integrators \[ [44], \] in which the input appears at the second derivative or the structure of the disturbance for the purposes of computing the stochastic reachability probability. We then computed finite sample bounds via Theorem 8.

All computations were done in Matlab, and code to reproduce the analysis and figures is available at: github.com/unm-hscl/ajthor-Automatica2020a.

5.1 Stochastic Chain of Integrators

We first consider a 2-D stochastic chain of integrators \[ [44], \] in which the input appears at the second derivative or the structure of the disturbance for the purposes of computing the stochastic reachability probability. We then computed finite sample bounds via Theorem 8.

For all problems, we chose a Gaussian kernel \[ k(x,x') = \exp(-||x-x'||^2/2\sigma^2) \] with \( \sigma = 0.1 \), and chose \( \lambda \) using the optimal rate computed in \[ [7]. \]

All computations were done in Matlab, and code to reproduce the analysis and figures is available at: github.com/unm-hscl/ajthor-Automatica2020a.

\[
x_{k+1} = \begin{bmatrix} 1 & T \\ 0 & 1 \end{bmatrix} x_k + \begin{bmatrix} \frac{x^2}{2} \\ T \end{bmatrix} u_k + w_k
\]

where \( w_k \) is an i.i.d. disturbance defined on the probability space \( (W, \mathcal{F}(W), \Pr_w) \). We presume a Gaussian disturbance \( w_k \sim \mathcal{N}(0, \Sigma) \) with \( \Sigma = 0.01I \), a control policy \( \pi(x) = 0 \), and target and safe sets \( T = [-1,1]^2 \) and \( K = [-1,1]^2 \).

We consider a sample \( S \) of \( M = 2500 \) observations drawn i.i.d. from \( Q \), a representation of (41) as a Markov control process (Definition 1). The initial conditions \( x \in \mathcal{X} \) in the sample were chosen uniformly in the interval \([-1,1] \times [-1,1] \) in order to ensure that a subset of the initial conditions violate the safety constraints, \( K \) and \( T \). We do this to ensure the “learned” model does not map all initial conditions to a safe set. Otherwise, in the

then assumed no knowledge of the system dynamics or the structure of the disturbance for the purposes of computing the stochastic reachability probability. We then computed finite sample bounds via Theorem 8. For all problems, we chose a Gaussian kernel \( k(x,x') = \exp(-||x-x'||^2/2\sigma^2) \) with \( \sigma = 0.1 \), and chose \( \lambda \) using the optimal rate computed in \[ [7]. \]

All computations were done in Matlab, and code to reproduce the analysis and figures is available at: github.com/unm-hscl/ajthor-Automatica2020a.

\[
x_{k+1} = \begin{bmatrix} 1 & T \\ 0 & 1 \end{bmatrix} x_k + \begin{bmatrix} \frac{x^2}{2} \\ T \end{bmatrix} u_k + w_k
\]

where \( w_k \) is an i.i.d. disturbance defined on the probability space \( (W, \mathcal{F}(W), \Pr_w) \). We presume a Gaussian disturbance \( w_k \sim \mathcal{N}(0, \Sigma) \) with \( \Sigma = 0.01I \), a control policy \( \pi(x) = 0 \), and target and safe sets \( T = [-1,1]^2 \) and \( K = [-1,1]^2 \).

We consider a sample \( S \) of \( M = 2500 \) observations drawn i.i.d. from \( Q \), a representation of (41) as a Markov control process (Definition 1). The initial conditions \( x \in \mathcal{X} \) in the sample were chosen uniformly in the interval \([-1,1] \times [-1,1] \) in order to ensure that a subset of the initial conditions violate the safety constraints, \( K \) and \( T \). We do this to ensure the “learned” model does not map all initial conditions to a safe set. Otherwise, in the

Fig. 1. (a) Safety probabilities at \( k = 0, N = 5 \), for a double integrator computed using dynamic programming. (b) Safety probabilities at \( k = 0 \) for a double integrator computed using Lemma 3 for \( N = 5 \). (c) Absolute error \( |V_0^*(x) - V_0^*(x)| \) between the dynamic programming solution and Lemma 3. (d, e) Upper and lower finite sample bounds, respectively, of the safety probabilities of a double integrator system computed using Theorem 8 with \( \delta = 0.1 \), where (d) is \( V_0^*(x) + B(x, \pi(x)) \) and (e) is \( V_0^*(x) - B(x, \pi(x)) \).

Fig. 2. Figure showing the mean of the finite sample bounds \( B(x, u) \) in the region \([-1,1] \times [-1,1] \) for a double integrator system as a function of \( \delta \).

regression, the value function estimate maps all values to 1. The resulting state \( y \in \mathcal{X} \) is drawn from \( Q(\cdot | x, u) \) using the dynamics in (41). We then presumed no knowledge of the system dynamics or \( Q \) and computed the estimate \( \hat{m}_{y|x,u} \) according to (12), with \( \beta(x, u) \) computed as in (13). Using \( \hat{m}_{y|x,u} \), we then computed the stochastic reachability using Lemma 3 for a time horizon of \( N = 5 \). We compare the stochastic reachability probability computed according to Lemma 3 against the solution via dynamic programming, that presumers the stochastic kernel \( Q \) is known. The absolute error \( |V_0^*(x) - V_0^*(x)| \) between the results obtained from Lemma 3 and the dynamic programming solution is shown in Figure 1(c). As expected, the stochastic reachability probabilities computed using Lemma 3, show low absolute error as compared with the dynamic programming solution, with a maximum absolute error of 0.1158 and a mean absolute error of 0.0122.

We then evaluated the finite sample bounds of the ap-
We can see that the difference between the upper and lower bounds is small, which indicates that the quality of the approximation obtained via Lemma 3 is close to a reasonable probabilistic upper bound.

We then consider the effect of varying the parameters $M$ and $\delta$ in (35) on the finite sample bound computed using Theorem 8. In order to demonstrate the effect of the parameter $M$, we drew 6 new samples from the stochastic kernel $K$, where the number of observations in each sample was chosen to be of size $M \in [100, 3600]$. An estimate was then computed for each sample and the finite sample bounds were computed for each estimate. For each sample of length $M \in [100, 3600]$, we computed the mean of the finite sample bounds in the region $[-1,1] \times [-1,1]$ for different values of $\delta \in [0.1, 1.9]$. Figure 2 shows the mean of $B(x, u)$ (35), the finite sample bounds computed using Theorem 8 for the six samples of length $M \in [100, 3600]$ as a function of $\delta$.

As expected, we can see that as the size of the samples increases, we obtain tighter probabilistic bounds $B(x, u)$ in the region $[-1,1] \times [-1,1]$ via Theorem 8. Effectively, this means that as the number of observations from the stochastic kernel increases, we obtain a better estimate of the conditional distribution embedding $m_{Y|x,u}$, and thus a better estimate of the safety probabilities via Lemma 3. Also as expected, as the violation threshold decreases (i.e., $\delta$ increases), the mean of the probabilistic bound decreases. Figure 2 also shows that for low values of $\delta$, we obtain higher values of the probabilistic bound $B(x, u)$. This corresponds to a high desired confidence. Further, we can see in Figure 2 that the finite sample bounds do not improve appreciably as the number of samples increases beyond a certain point.

5.2 Linearized Cart-Pole System

We then considered a benchmark cart-pole system [22] with a black-box neural network feedback controller. The dynamics for the linearized cart-pole system [22] are given by:

$$\begin{align*}
\ddot{x} &= 0.0043\dot{\theta} - 2.75\theta + 1.94u - 10.95\dot{x} \\
\dot{\theta} &= 28.58\theta - 0.044\dot{\theta} - 4.44u + 24.92\dot{x}
\end{align*}$$

with state $x = [x, \dot{x}, \theta, \dot{\theta}]^\top \in \mathbb{R}^4$ and control input $u \in \mathbb{R}$. We add an additional Gaussian disturbance $w_k \sim \mathcal{N}(0, \Sigma)$ with $\Sigma = 0.01I$ to the dynamical state equations, which can simulate dynamical uncertainty or minor system perturbations. The control input is computed via a neural network controller [22], which takes the current state and outputs a real number $u \in \mathbb{R}$, which can be interpreted as the input torque.

The benchmark is defined [22] such that the neural network controller must keep the lateral position of the cart $x$ within $[-0.7, 0.7]$, maintain a low cart velocity $\dot{x} \in [-1, 1]$, and keep the pendulum angle $\theta$ within $[-\pi/6, \pi/6]$ while the angular velocity $\dot{\theta}$ is unconstrained. We define the safe set $\mathcal{K}$ according to the above constraints, and define the target set $\mathcal{T}$ such that the pendulum angle $\theta$ must be within $[-0.05, 0.05]$.

$$\mathcal{K} = \{x \in \mathbb{R}^4 | |x_1| \leq 0.7, |x_2| \leq 1, |x_3| \leq \pi/6\}$$

$$\mathcal{T} = \{x \in \mathbb{R}^4 | |x_3| \leq 0.05\}$$

---

---
We simulated 10 trajectories from initial conditions taken uniformly from the ranges specified above, and extracted a sample $S$ of $M = 12234$ observations taken i.i.d. from the stochastic kernel $Q$, a representation of the dynamics (42) as a Markov control process (Definition 1).

We then computed the safety probabilities for the system over a time horizon $N = 3$ using Lemma 3 to demonstrate the feasibility of the approach. Figure 3(a) shows the approximate value function $V^k_N$ for the system in (42) in three dimensions. We can see that the closed loop system has a high probability of stabilizing the pendulum from an initial condition within the range $\theta \in [-\pi/6, \pi/6]$ and the results also show an underlying symmetry around $\theta = 0$, as expected. The safety probabilities computed using Lemma 3 at $k = 0$ for $N = 3$ are shown in Figure 3(b). We then computed the finite sample bounds on the approximation using Theorem 8 with $\delta = 0.1$ to obtain probabilistic upper and lower bounds on the safety probabilities. Figure 3(c) shows the probabilistic upper bound on the safety probabilities, while Figure 3(d) shows the lower bound. As expected, because we use a high number of samples $M = 12234$, the bounds computed in Theorem 8 show that with high probability, the solution is close to the true solution.

This means that using the proposed data-driven approach, we can utilize stochastic reachability to analyze the safety properties of a dynamical system with a black-box neural network controller. Similarly, we can expose the underlying structure of the closed-loop system to reveal useful knowledge of the system properties, such as symmetry.

As before, we added a Gaussian disturbance $w_k \sim \mathcal{N}(0, \Sigma)$ with $\Sigma = 0.01I$ to the dynamical equations and represent the system as a Markov control process. The benchmark is defined such that the pole angle $\theta$ will remain within $[-\pi/6, \pi/6]$, while the other variables are unconstrained. As such, we define the target set $\mathcal{T}$ such that $\theta \in [-0.05, 0.05]$ as with the linearized cart-pole system, but define the safe set $K$ to be the entire state space, meaning there is no unsafe region.

$$\mathcal{T} = \{x \in \mathbb{R}^4 \mid |x_3| \leq 0.05\} \quad (46)$$

**5.3 Nonlinear Cart-Pole System**

We then analyzed a nonlinear cart-pole system with a neural network controller [22], with dynamics given by:

$$\ddot{x} = \frac{u + ml\omega^2 \sin(\theta)}{mt} - \frac{ml(g \sin(\theta) - \cos(\theta))(\frac{u + ml\omega^2 \sin(\theta)}{mt} \cos(\theta))}{l(\frac{4}{3} - m\cos^2(\theta))/mt} \quad (45)$$

where $g = 9.8$ is the gravitational constant, the pole mass is $m = 0.1$, half the pole’s length is $l = 0.5$, and $m_t = 1.1$ is the total mass. The control input, $u \in \{-10, 10\}$, which affects the lateral position of the cart, is chosen by the neural network controller [22]. This means the controller is less “smooth” than the neural network controller for the linearized cart-pole system, because when the pendulum is near vertical, the controller rapidly switches between a high positive and negative control input value. Thus, the velocity components of the system state, $\dot{x}$ and $\dot{\theta}$, will not stabilize to zero, as shown in Figure 4.

![Fig. 5. (a) Safety probabilities of the nonlinear cart-pole system computed using Lemma 3 for a time horizon of $N = 3$. (b) Probabilistic lower bound on the safety probabilities computed using Theorem 8 with $\delta = 0.1$.](image-url)
This allows us to analyze the behavior of the controller to reach a pre-specified objective without enforcing constraints on the system before the terminal time. We then simulated 10 trajectories from initial conditions sampled uniformly from the ranges specified above, and collected a sample $S$ of $M = 20020$ observations. Then, we computed the safety probabilities using Lemma 3 with $N = 3$. The results are shown in Figure 5(a). We then computed the finite sample bounds using Theorem 8 for $\delta = 0.1$, and plotted the lower bound on the safety probabilities in Figure 5(b).

As expected, we see that the safety probabilities show the switching behavior of the neural network controller when theta is close to zero, since the angular velocity $\dot{\theta}$ switches between a positive and negative value due to the range of the control input. We also see that the controller does not exhibit a high safety probability for negative values of $\theta$. In simulation, it was observed that the controller did not learn symmetric stabilizing behavior. Instead, the controller learned to quickly force $\theta$ to be positive, regardless of the initial condition, before stabilizing the pendulum. We can see this behavior in the plot above. Fig. 5(b) shows the probabilistic lower bound on safety probabilities computed using Theorem 8 with $\delta = 0.1$.

### 5.4 Nonlinear Pendulum System

We then consider a nonlinear pendulum system with a black box neural network controller. The dynamics of the system are given by:

$$
\begin{align*}
\dot{x}_1 &= x_2 \\
\dot{x}_2 &= -x_1 + 0.1 \sin(x_3) \\
\dot{x}_3 &= x_4 \\
\dot{x}_4 &= u \\
\end{align*}
$$

(47)

with state $x = [x_1, x_2, x_3, x_4]^T \in \mathbb{R}^4$ and control input $u \in \mathbb{R}$. We formulate a discrete time system using the dynamics (47) with a sampling time of $T = 0.1$ and by implementing a zero-order hold on the control input. The control input $u$ is chosen by a feedforward neural network controller [11], which is a function $F : \mathbb{R}^4 \rightarrow \mathbb{R}$, defined by $F = F_L \circ \cdots \circ F_0$, where $F_i(z) = \sigma_i(W_i z + b_i)$, $i = 1, \ldots, L-1, \sigma_i$ is called the activation function, $W_i \in \mathbb{R}^{d \times d}$, and $b_i \in \mathbb{R}^d$, with $d$ being the dimensionality of the input data for each layer. $F_0$ is the function mapping the states into the first layer, and $F_L$ is the function mapping the output of the last layer into the control space. The neural network used for this benchmark is described in [11], where $L = 3$, $D = 300$, and $\sigma_i$ is the ReLU function, i.e. $\sigma_i(x) = \max\{0, x\}$. The neural network is trained using an MPC framework. We refer the reader to [11] for more details.

We seek to compare our approach against the existing result presented in [11]. To do so, we compare the forward reachability analysis in [11] to a comparable backward reachability problem in reverse time. We formulate the backward reachability problem as a first time hitting problem [39], which evaluates the likelihood that the system will reach a target set $T$ at some time $k \leq N$. We define the target set as

$$
T = \{ x \in \mathbb{R}^4 \mid 0.6 \leq x_1 \leq 0.7, -0.7 \leq x_2 \leq -0.6, \\
-0.4 \leq x_3 \leq -0.3, 0.5 \leq x_4 \leq 0.6 \} 
$$

(48)

and seek to determine the set of initial conditions which lead to this set through a time-reversed version of (47).

We reformulate (47) in reverse time as a Markov control process (Definition 1) with stochastic kernel $Q$. We consider a sample $S$ of $M = 10000$ observations drawn i.i.d. from $Q$. First, we generate 20 trajectories evolving over 200 time steps in forward time starting with initial conditions chosen uniformly in $T$, ensuring that the trajectories are viable in reverse time. These trajectories are comparable to the realizations depicted in red in Figure 6. Then, we generate an additional 20 trajectories with initial conditions taken from outside $T$, in order to ensure we have examples of infeasible trajectories for the backward reachability problem. With a sampling time
of $T = 0.1$, we sample in reverse time from these forward time trajectories to obtain 8000 observations. We then generate an additional 2000 observations taken uniformly in the range $[-2.1, 2.1]^4$ using the reverse-time dynamics in order to more fully characterize the state space.

Figure 6 (left) shows the first hitting time safety probabilities computed using the sample $S$. Because our method is primarily point-based, we form a coarse projection by computing the safety probabilities for points distributed uniformly in all 4 state space dimensions and then projecting onto a 2-dimensional subspace. We then computed the finite sample bounds for the safety probabilities using Corollary 9. Figure 6 (right) shows the lower bounds on the safety probabilities.

We can see that the backward reachability analysis identifies a region that loosely encompasses the flowpipes, with a high probability region towards the center, as expected. However, the backward reachability analysis does not identify the entire flowpipe region with high probability values, and does not show with high fidelity the precise path of the trajectories that are evident from [11] in Figure 6 (top).

We anticipate this is due to multiple contributing factors. First, the coarseness of the projection makes it difficult to capture the safety probabilities across a wide range due to the highly nonlinear closed-loop system dynamics. In order to overcome this limitation, we need to sample and evaluate many more points across the state space, which leads to computational limitations due to the matrix inversion in (13) as well as in the dynamic program for Lemma 3. Second, the conditional distribution embedding approach implicitly computes an average in order to determine the safety probabilities. Because of this, we assume that the chosen parameters may be contributing to a poor approximation due to an overly-smooth estimate. Lastly, there is no guarantee that a particular set of observations $S$ is assured to satisfy limits of the finite sample bound. This is because the finite sample bounds are derived in the limit, over the expectation of all possible sets of observations $S$. With a high number of observations, the finite sample bound is low even though our sample $S$ does not fully capture the nonlinear dynamics of the system, which leads to low safety probabilities. Ideally, our result should completely identify the region covered by the flowpipes determined by [11]. We anticipate that with a more uniform set of observations, as well as a higher fidelity sampling scheme, we would be able to generate a better approximation of the flowpipes. Extension of these methods to overcome some of these computational challenges is an area of current work [40].

6 Conclusion

We provided state- and input-based finite sample bounds for the stochastic reachability probability constructed via conditional distribution embeddings. Our approach is based on an application of statistical learning theory, that relates the observed data to the quality of the approximation of the stochastic reachability probability at a given state and input. This approach enables rigorous bounds on model-free stochastic reachability. We validated our approach on a nonlinear dynamical system with a neural net controller, and numerically characterized our approach on the stochastic double integrator.

Acknowledgements

We would like to thank Le Song for helpful discussion about conditional embeddings for Markov processes, as well as Souradeep Dutta and Sriram Sankaranarayanan for helpful assistance with Sherlock and Flow*.

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, 2008.

[2] Nachman Aronszajn. Theory of reproducing kernels. *Transactions of the American mathematical society*, 68(3):337–404, 1950.

[3] Peter L Bartlett and Shahar Mendelson. Rademacher and Gaussian complexities: Risk bounds and structural results. *Journal of Machine Learning Research*, 3(Nov):463–482, 2002.

[4] Richard Ernest Bellman and Stuart E Dreyfus. Applied dynamic programming. 1962.

[5] Alain Berlinet and Christine Thomas-Agnan. Reproducing kernel Hilbert spaces in probability and statistics. 2004.

[6] Dimitri P Bertsekas and Steven E Shreve. Stochastic optimal control: the discrete time case. 1978.

[7] Andrea Caponnetto and Ernesto De Vito. Optimal rates for the regularized least-squares algorithm. *Foundations of Computational Mathematics*, 7(3):331–368, 2007.

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

[9] Erhan Çinlar. Probability and stochastic processes. 2011.

[10] Ernesto De Vito, Andrea Caponnetto, and Lorenzo Rosasco. Model selection for regularized least-squares algorithm in learning theory. *Foundations of Computational Mathematics*, 5(1):59–85, 2005.

[11] Souradeep Dutta, Xin Chen, and Sriram Sankaranarayanan. Reachability analysis for neural feedback systems using regressive polynomial rule inference. In *Proceedings of the 22nd ACM International Conference on Hybrid Systems: Computation and Control*, pages 157–168, 2019.

[12] Kenneth R Garren. *Bounds for the Eigenvalues of a Matrix*. National Aeronautics and Space Administration, 1968.
[45] Abraham P Vinod and Meeko MK 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), pages 1–10, 2018.

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

[47] Abraham P Vinod, Vignesh Sivaramakrishnan, and Meeko MK Oishi. Piecewise-affine approximation-based stochastic optimal control with gaussian joint chance constraints. In 2019 American Control Conference (ACC), pages 2942–2949. IEEE, 2019.

[48] Yu Wang, Nima Roohi, Matthew West, Mahesh Viswanathan, and Geir E. Dullerud. Statistical verification of pctl using antithetic and stratified samples. Formal Methods in System Design, 54:145–163, 2019.

[49] Henry Wolkowicz and George PH Styan. Bounds for eigenvalues using traces. Linear algebra and its applications, 29:471–506, 1980.

[50] Mojtaba Zarei, Yu Wang, and Miroslav Pajic. Statistical verification of learning-based cyber-physical systems. In Proceedings of Hybrid Systems: Computation and Control, New York, NY, 2020.
This figure "benchmark9.png" is available in "png" format from:

http://arxiv.org/ps/2010.08036v1