A Simple and Efficient Sampling-based Algorithm for General Reachability Analysis

Thomas Lew[1] THOMAS.LEW@STANFORD.EDU

Lucas Janson[2] LJANSON@FAS.HARVARD.EDU

Riccardo Bonalli[3] RICCARDO.BONALLI@L2S.CENTRALESUPELEC.FR

Marco Pavone[1] PAVONE@STANFORD.EDU

[1]*Department of Aeronautics and Astronautics, Stanford University*
[2]*Department of Statistics, Harvard University*
[3]*Laboratory of Signals and Systems, University of Paris-Saclay, CNRS, CentraleSupélec*

Editors: R. Firoozi, N. Mehr, E. Yel, R. Antonova, J. Bohg, M. Schwager, M. Kochenderfer

Abstract

In this work, we analyze an efficient sampling-based algorithm for general-purpose reachability analysis, which remains a notoriously challenging problem with applications ranging from neural network verification to safety analysis of dynamical systems. By sampling inputs, evaluating their images in the true reachable set, and taking their ϵ-padded convex hull as a set estimator, this algorithm applies to general problem settings and is simple to implement. Our main contribution is the derivation of asymptotic and finite-sample accuracy guarantees using random set theory. This analysis informs algorithmic design to obtain an ϵ-close reachable set approximation with high probability, provides insights into which reachability problems are most challenging, and motivates safety-critical applications of the technique. On a neural network verification task, we show that this approach is more accurate and significantly faster than prior work. Informed by our analysis, we also design a robust model predictive controller that we demonstrate in hardware experiments.

Keywords: reachability analysis, random set theory, robust control, neural network verification.
Appendix: https://arxiv.org/abs/2112.05745

1. Introduction

Forward reachability analysis entails characterizing the reachable set of outputs of a given function corresponding to a set of inputs. This type of analysis underpins a plethora of applications in model predictive control, neural network verification, and safety analysis of dynamical systems. Sampling-based reachability analysis techniques are a particularly simple class of methods to implement; however, conventional wisdom suggests that if insufficient representative samples are considered, these methods may not be robust in that they cannot rule out edge cases missed by the sampling procedure. Alternatively, by leveraging structure in specific problem formulations or computational methods designed for exhaustivity (e.g., branch and bound), a large range of algorithms with deterministic accuracy and performance

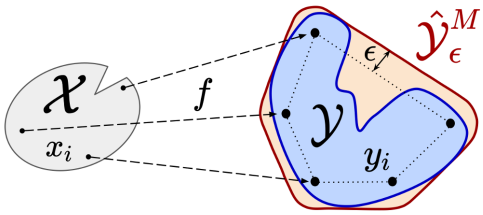

Figure 1: ϵ-RANDUP consists of three simple steps: 1) sampling M inputs x_i in $\mathcal{X}$, 2) propagating these inputs through the reachability map f, and 3) taking the ϵ-padded convex hull $\hat{\mathcal{Y}}_\epsilon^M$ to approximate the reachable set $\mathcal{Y}$.

guarantees have been developed. However, these methods often sacrifice simplicity and generality for their power, motivating the development of algorithms that avoid such restrictions.

In this work, we analyze a simple yet efficient sampling-based algorithm for general-purpose reachability analysis. As depicted in Figure 1, it consists of 1) sampling inputs, 2) propagating these inputs, and 3) taking the padded convex hull of these output samples. We refer to this RANDomized Uncertainty Propagation algorithm as ϵ-RANDUP: it is simple to implement, benefits from statistical accuracy guarantees, and applies to a wide range of problems including reachability analysis of uncertain dynamical systems with neural network controllers. Importantly, ϵ-RANDUP fulfills key desiderata that a general-purpose reachability analysis algorithm should satisfy:

- it works with any choice of possibly nonlinear reachability maps and non-convex input sets,
- its estimate of the reachable set is conservative with high probability and tighter than prior work,
- it is efficient and does not require precomputations, which is a key advantage for learning-based control applications where uncertainty bounds and models are updated in real-time.

Our main contribution is a thorough analysis of the statistical properties of ϵ-RANDUP. Specifically:

1. We prove that the set estimator converges to the ϵ-padded convex hull of the true reachable set as the number of samples increases. Our assumption about the sampling distribution is weaker than in related work and implies that sampling the boundary of the input set is sufficient. This asymptotic result justifies using ϵ-RANDUP as a thrustworthy baseline for offline validation whenever the reachability map and the input set are complex and no tractable algorithm exists.
2. We derive a finite-sample bound for the Hausdorff distance between the output of ϵ-RANDUP and the convex hull of the true reachable set, assuming that the reachability map is Lipschitz continuous. This result informs algorithmic design (e.g., how to choose the number of samples to obtain an ϵ-accurate approximation with high probability), sheds insights into which problems are most challenging, and motivates using this simple algorithm in safety-critical applications.

We demonstrate ϵ-RANDUP on a neural network controller verification task and show that it is highly competitive with prior work. We also embed this algorithm within a robust model predictive controller and present hardware results demonstrating the reliability of the approach.

2. Related work

Reachability analysis has found a wide range of applications ranging from model predictive control (Schürmann et al., 2018), robotics (Shao et al., 2021; Lew et al., 2022), neural network verification (Tran et al., 2019; Hu et al., 2020), to orbital mechanics (Wittig et al., 2015). Reachability analysis is particularly relevant in safety-critical applications which require the strict satisfaction of specifications. For instance, a drone transporting a package should never collide with obstacles and respect velocity bounds for any payload mass in a bounded input set. In contrast to stochastic problem formulations which typically consider the inputs as random variables with known probability distributions (Webb et al., 2019; Sinha et al., 2020; Devonport and Arcak, 2020), we consider robust formulations which are of interest whenever minimal information about the inputs is available.

Deterministic algorithms are often tailored to the particular parameterization of the reachability map and to the shape of the input set. For instance, one finds methods that are particularly designed for neural networks (Tran et al., 2019; Ivanov et al., 2019; Hu et al., 2020), nonlinear hybrid systems (Chen et al., 2013; Kong et al., 2015), linear dynamical systems with zonotopic (Girard, 2005) and

ellipsoidal (Kurzhanski and Varaiya, 2000) parameter sets, etc. We refer to (Liu et al., 2021) and (Althoff et al., 2021) for recent comprehensive surveys. Such algorithms have deterministic accuracy guarantees but require problem-specific structure that restricts the class of systems they apply to. Given the wide range of applications of reachability analysis, there is a pressing need for the development and analysis of simple algorithms that can be applied to general problem formulations.

On the other hand, sampling-based algorithms reconstruct the reachable set from sampled outputs. The stochasticity is typically controlled by the engineer, who selects the number of samples and their distribution. A key strength of this methodology is the possible use of black-box models with arbitrary input sets, which allows using complex simulators of the system. For instance, kernel-based methods (De Vito et al., 2014; Rudi et al., 2017; Thorpe et al., 2021) have been proposed as a strong approach for data-driven reachability analysis. Kernel-based methods are highly expressive, as selecting a completely separating kernel (De Vito et al., 2014) enables reconstructing any closed set to arbitrary precision given enough samples. Their main drawback is the potentially expensive evaluation of the estimator for a large number of samples. Its implicit representation as a level set is also not particularly convenient for downstream applications.

Sampling-based reachable set estimators with pre-specified shapes have been proposed to simplify computations and downstream applications. Recently, (Lew and Pavone, 2020) proposed to approximate reachable sets with the convex hull of the samples, but this approach is not guaranteed to return a conservative approximation. Ellipsoidal and rectangular sets are computed in (Devonport and Arcak, 2020) using the scenario approach, but this work tackles a different problem formulation with inputs that are random variables with known distribution. To tackle the robust reachability analyis problem setting, (Gruenbacher et al., 2022) use a ball estimator that bounds the samples. The statistical analysis is restricted to ball-parameterized input sets, uniform sampling distributions, and smooth diffeomorphic reachability maps that represent the solution of a neural ordinary differential equation (Chen et al., 2018) from the input set. In practice, using an outer-bounding ball is more conservative than taking the convex hull of the samples, see Section 6.

In this work, we slightly modify RANDUP (Lew and Pavone, 2020) with an additional ϵ-padding step to yield finite-sample outer-approximation guarantees, Our analysis leverages random set theory (Matheron, 1975; Molchanov, 2017), which provides a natural mathematical framework to analyze the reachable set estimator. We characterize its accuracy using the Hausdorff distance to the convex hull of the true reachable set, which provides an intuitive error measure that can be directly used for downstream control applications. Our analysis draws inspiration from the vast literature on statistical geometric inference, which proposes different set estimators including union of balls (Devroye and Wise, 1980; Baillo and Cuevas, 2001), convex hulls (Ripley and Rasson, 1977; Schneider, 1988; Dumbgen and Walther, 1996), r-convex hulls (Rodriguez-Casal and Saavedra-Nieves, 2016, 2019; Arias-Castro et al., 2019), Delaunay complexes (Boissonnat and Ghosh, 2013; Aamari, 2017; Aamari and Levrard, 2018), and kernel-based estimators (De Vito et al., 2014; Rudi et al., 2017). This research typically makes assumptions about the set to be reconstructed (e.g., it is convex (Dumbgen and Walther, 1996) or has bounded reach (Cuevas, 2009)) and considers points that are directly sampled from this set. In this work, we derive similar results for reachable sets given known properties of the input set, reachability map, and chosen input sampling distribution.

3. Problem definition

In this section, we introduce our notations and problem formulation. Due to space constraints, we leave measure-theoretic details to Appendix A. We denote $\lambda(\cdot)$ for the Lebesgue measure over $\mathbb{R}^p$,

$\Gamma(\cdot)$ for the gamma function, $\mathrm{H}(A)$ for the convex hull of a subset $A \subset \mathbb{R}^n$, $A^{\mathsf{c}} = \mathbb{R}^n \setminus A$ for its complement, ∂A for its boundary, $\oplus$ for the Minkowski sum, $B(x,r) := \{y \in \mathbb{R}^n : \|y - x\| \leq r\}$ for the closed ball of center $x \in \mathbb{R}^n$ and radius $r \geq 0$, and $\mathring{B}(x,r)$ for the open ball. The family of nonempty compact subsets of $\mathbb{R}^n$ is denoted as $\mathcal{K}$. For any $A \in \mathcal{K}$ and $d > 0$, $D(A,d) := \min\{n \in \mathbb{N} : \exists\{a_1,\ldots,a_n\} \subset \mathbb{R}^n,\ A \subset B(a_1,d) \cup \ldots \cup B(a_n,d)\}$ denotes the d-covering number of A.

Let $\mathcal{X} \subset \mathbb{R}^p$ be a compact nonempty set of inputs and $f : \mathbb{R}^p \to \mathbb{R}^n$ be a continuous function. In this work, we tackle the general problem of reachability analysis, i.e., characterizing the set of reachable outputs $y = f(x)$ for all possible inputs $x \in \mathcal{X}$. This problem is also often referred to as uncertainty propagation. Mathematically, the objective consists of efficiently computing an accurate approximation of the reachable set $\mathcal{Y} \subset \mathbb{R}^n$, which is defined as

$$\mathcal{Y} = f(\mathcal{X}) = \{f(x) \,:\, x \in \mathcal{X}\}. \tag{1}$$

To tackle this problem, ϵ-RANDUP relies on the choice of three parameters: a number of samples $M \in \mathbb{N}$, a padding constant $\epsilon > 0$, and a sampling distribution $\mathbb{P}_{\mathcal{X}}$ on measurable subsets of $\mathbb{R}^p$. As depicted in Figure 1, ϵ-RANDUP consists of sampling M independent identically-distributed inputs x_i in $\mathcal{X}$ according to $\mathbb{P}_{\mathcal{X}}$, of evaluating each output $y_i = f(x_i)$, and of computing the ϵ-padded convex hull

$$\hat{\mathcal{Y}}_\epsilon^M := \mathrm{H}\left(\{y_i\}_{i=1}^M\right) \oplus B(0,\epsilon). \tag{2}$$

Our analysis hinges on the observation that the reachable set estimator $\hat{\mathcal{Y}}_\epsilon^M$ is a *random compact set*, i.e., $\hat{\mathcal{Y}}_\epsilon^M$ is a random variable taking values in the family of nonempty compact sets $\mathcal{K}$. We refer to Appendix A for rigorous definitions using random set theory. Intuitively, different input samples x_i in $\mathcal{X}$ induce different output samples y_i in $\mathcal{Y}$, resulting in different approximated reachable sets $\hat{\mathcal{Y}}_\epsilon^M$. To characterize the accuracy of the estimator, we use the *Hausdorff metric*, which is defined as

$$d_{\mathrm{H}}(A,B) := \max\left(\sup_{x\in B}\inf_{y\in A}\|x-y\|,\ \sup_{x\in A}\inf_{y\in B}\|x-y\|\right) \quad \text{for any } A,B \in \mathcal{K}. \tag{3}$$

This metric induces a topology and an associated σ-algebra, which enables rigorously defining random compact sets as random variables and describing their convergence; see Appendix A. Interestingly, the distribution of a random compact set is characterized by the probability that it intersects any given compact set. We use this fact in Sections 4 and 5, where we characterize the probability that the set estimator $\hat{\mathcal{Y}}_\epsilon^M$ intersects well-chosen sets along the boundary of the true reachable set. By analyzing the distribution of $\hat{\mathcal{Y}}_\epsilon^M$, this approach allows bounding the Hausdorff distance between $\hat{\mathcal{Y}}_\epsilon^M$ and the convex hull of the true reachable set $\mathrm{H}(\mathcal{Y})$ with high probability.

4. Asymptotic analysis

In this section, we provide an asymptotic analysis under minimal assumptions about the input set and the reachability map (namely, that $\mathcal{X}$ is compact and f is continuous). To enable the reconstruction of the true convex hull $\mathrm{H}(\mathcal{Y})$ using the sampling-based set estimator $\hat{\mathcal{Y}}_\epsilon^M$, we make one assumption about the sampling distribution $\mathbb{P}_{\mathcal{X}}$ for the inputs x_i. Note that by definition, $\mathbb{P}_{\mathcal{X}}(\mathcal{X}) = 1$.

Assumption 1 $\mathbb{P}_{\mathcal{X}}(\{x \in \mathcal{X} : f(x) \in \mathring{B}(y,r)\}) > 0$ *for all* $y \in \partial\mathcal{Y}$ *and all* $r > 0$.

This assumption states that the probability of sampling an output arbitrarily close to any point on the boundary of the true reachable set is strictly positive. In other words, the boundary of the reachable

set should be contained in the support of the distribution of the output samples y_i. Assumption 1 is weaker than the associated assumption in (Lew and Pavone, 2020, Theorem 2), which can be restated as "$\mathbb{P}_{\mathcal{X}}(f^{-1}(A)) > 0$ *for any open set* $A \subset \mathbb{R}^n$ *such that* $\mathcal{Y} \cap A \neq \emptyset$". Indeed, Assumption 1 only considers open neighborhoods of the boundary $\partial\mathcal{Y}$, as opposed to all open sets intersecting $\mathcal{Y}$. Selecting a sampling distribution $\mathbb{P}_{\mathcal{X}}$ that satisfies Assumption 1 is easy. For instance, if $\mathcal{X}$ has a smooth boundary (see Assumption 4), then the uniform distribution over $\mathcal{X}$ satisfies Assumption 1.

Assumption 1 is sufficient to prove that the random set estimator $\hat{\mathcal{Y}}_{\epsilon}^{M}$ converges to the ϵ-padded convex hull of $\mathcal{Y}$ as the number of samples M increases. Below, we prove a more general result which allows for variations of the padding radius ϵ as the number of samples increases.

Theorem 1 (Asymptotic Convergence) *Let* $\bar{\epsilon} \geq 0$ *and* $(\epsilon_M)_{M\in\mathbb{N}}$ *be a sequence of padding radii such that* $\epsilon_M \geq 0$ *for all* $M \in \mathbb{N}$ *and* $\epsilon_M \to \bar{\epsilon}$ *as* $M \to \infty$. *For any* $\epsilon \geq 0$, *define the estimator* $\hat{\mathcal{Y}}_{\epsilon}^{M} = H\left(\{y_i\}_{i=1}^{M}\right) \oplus B(0,\epsilon)$. *Then, under Assumption 1, almost surely, as* $M \to \infty$,

$$d_H(\hat{\mathcal{Y}}_{\epsilon_M}^{M}, H(\mathcal{Y}) \oplus B(0,\bar{\epsilon})) \longrightarrow 0.$$

Proof We refer to Appendix B.1. We leverage (Molchanov, 2017, Proposition 1.7.23) which states sufficient conditions for the convergence of random compact sets and use properties of the convex hull to relax the corresponding assumption in (Lew and Pavone, 2020) with Assumption 1. ■

Practically, Theorem 1 justifies using ϵ-RANDUP for general continuous maps f and compact sets $\mathcal{X}$. This consistency result implies that choosing any converging sequence of padding radii (e.g., $\epsilon_M = 1/M$) guarantees the convergence of the random set estimator $\hat{\mathcal{Y}}_{\epsilon_M}^{M}$ to the $\bar{\epsilon}$-padded convex hull of the true reachable set. As a particular case, selecting a constant padding radius ϵ (which yields ϵ-RANDUP) guarantees that $\mathcal{Y}_{\epsilon}^{M}$ converges to the ϵ-padded convex hull $\mathrm{H}(\mathcal{Y}) \oplus B(0,\epsilon)$.

Compared to (Lew and Pavone, 2020, Theorem 2), which only treats the case with constant zero padding radii $\epsilon_M = \bar{\epsilon} = 0$ (i.e., without ϵ-padding the convex hull of the output samples), Theorem 1 allows for variations of the padding radii ϵ_M and is proved under weaker assumptions. Instead of relying on ϵ-covering arguments (e.g., see Corollary 1 in (Dumbgen and Walther, 1996) which assumes that $\mathcal{Y}$ is convex), we use (Molchanov, 2017, Proposition 1.7.23) to conclude asymptotic convergence. This proof technique allows deriving a general result that does not depend on the exact sampling density along the boundary $\partial\mathcal{Y}$ and uses a sequence of padding radii ϵ_M converging arbitrarily slowly to some constant $\bar{\epsilon} \geq 0$.

5. Finite-sample analysis

Theorem 1 provides asymptotic convergence guarantees that support the application of ϵ-RANDUP in general scenarios (e.g., as a baseline for offline validation in complex problem settings), but does not provide finite-sample guarantees which are of practical interest in safety-critical applications. Deriving stronger statistical guarantees requires leveraging more information about the structure of the problem. We derive finite-sample rates under general assumptions in Section 5.1 and analyze a particular case in Section 5.2. We discuss practical implications of our results in Section 5.3.

5.1. General finite-sample statistical guarantees

To derive convergence rates and outer-approximation guarantees given a finite number of samples M, we first make an assumption about the smoothness of the reachability map f.

Assumption 2 *The reachability map $f : \mathbb{R}^p \to \mathbb{R}^n$ is L-Lipschitz: for some constant $L \geq 0$, $\|f(x_1) - f(x_2)\| \leq L \|x_1 - x_2\|$ for all $x_1, x_2 \in \mathcal{X}$.*

Next, we make an assumption about the sampling distribution $\mathbb{P}_{\mathcal{X}}$ along the input set boundary $\partial\mathcal{X}$.

Assumption 3 *Given $\epsilon, L > 0$, there exists $\Lambda_\epsilon^L > 0$ such that $\mathbb{P}_{\mathcal{X}}\left(B\left(x, \frac{\epsilon}{2L}\right)\right) \geq \Lambda_\epsilon^L$ for all $x \in \partial\mathcal{X}$.*

Given any boundary input $x \in \partial\mathcal{X}$, the constant Λ_ϵ^L characterizes the probability of sampling an input x_i that is $\epsilon/(2L)$-close to x. Selecting a sampling distribution that satisfies Assumption 3 is simple; we provide examples in Sections 5.2 and 6. As we show next, these two assumptions are sufficient to derive finite-sample convergence rates for ϵ-RANDUP. Recall that $D(\partial\mathcal{X}, d)$ denotes the d-packing number of $\partial\mathcal{X}$, which is necessarily finite by the compactness of $\mathcal{X}$.

Theorem 2 (Finite-Sample Bound) *Define the estimator $\hat{\mathcal{Y}}^M = H\left(\{y_i\}_{i=1}^M\right)$ and the probability threshold $\delta_M = D(\partial\mathcal{X}, \epsilon/(2L))(1 - \Lambda_\epsilon^L)^M$. Then, under Assumptions 2 and 3 and assuming that $\partial\mathcal{Y} \subseteq f(\partial\mathcal{X})$, with probability at least $1 - \delta_M$,*

$$d_H(\hat{\mathcal{Y}}^M, H(\mathcal{Y})) \leq \epsilon \quad \textit{and} \quad \mathcal{Y} \subseteq \hat{\mathcal{Y}}_\epsilon^M.$$

Proof We refer to Appendix B.2 for a complete proof. ■

Using a similar analysis, one could derive convergence rates for the ϵ-padded union of balls estimator (Devroye and Wise, 1980; Baillo and Cuevas, 2001) that would depend on the ϵ-covering number of the entire input set $D(\mathcal{X}, \epsilon)$. In the general case, $D(\partial\mathcal{X}, \epsilon) \leq D(\mathcal{X}, \epsilon)$: Theorem 2 indicates that using a convex hull is more sample-efficient than a union of balls, assuming that $\partial\mathcal{Y} \subseteq f(\partial\mathcal{X})$ (see Appendix B.2 for further details). It is better suited if $\mathcal{Y}$ is convex or if an approximation of $\mathrm{H}(\mathcal{Y})$ is sufficient for the downstream application, as is usual in control applications which typically use convex reachable set approximations, see (Lew and Pavone, 2020).

5.2. Analysis of a particular setting: smooth input set and continuous distribution

In many applications, the boundary of the input set is smooth (e.g., $\mathcal{X}$ is a 2-norm ball). In this setting, we can apply Theorem 2 to derive finite-sample guarantees for general continuous sampling distributions. We state this smoothness assumption below.

Assumption 4 *$\mathcal{X}^{\mathsf{c}}$ is r-convex for some $r > 0$. Equivalently, for any $x \in \partial\mathcal{X}$, there exists $\tilde{x} \in \mathcal{X}$ such that $x \in B(\tilde{x}, r) \subseteq \mathcal{X}$.*

Assumption 4 guarantees that for any parameter x on the boundary $\partial\mathcal{X}$, one can find a ball of radius r contained in $\mathcal{X}$ that also contains x, see Figure 2. This assumption corresponds to a general inwards-curvature condition of the boundary $\partial\mathcal{X}$. It is a common assumption in the literature (Walther, 1997; Rodriguez-Casal and Saavedra-Nieves, 2016, 2019; Arias-Castro et al., 2019) and is related to the notion of reach (Federer, 1959; Cuevas, 2009; Aamari, 2017) that bounds the curvature of the boundary $\partial\mathcal{X}$. To guarantee its satisfaction, one can replace $\mathcal{X}$ with $\mathcal{X} \oplus B(0, r)$ (Walther, 1997) before performing reachability analysis, which would yield a more conservative estimate of $\mathcal{Y}$. Next, we state an assumption about the sampling distribution $\mathbb{P}_{\mathcal{X}}$.

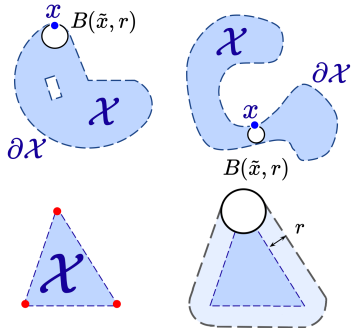

*Figure 2: **Top**: sets $\mathcal{X}$ satisfying Assumption 4 can be non-convex, have holes, and be disconnected. **Bottom**: if $\mathcal{X}^{\mathsf{c}}$ is not r-convex, it is still possible to find a conservative approximation that is r-convex.*

Assumption 5 *$\mathbb{P}_{\mathcal{X}}(A) \geq p_0 \lambda(A)$ for all measurable sets $A \subset \mathcal{X}$ for some constant $p_0 > 0$.*

This assumption states that the sampling distribution admits a lower-bounded continuous density. Specifically, there exists a density function $p_{\mathcal{X}} : \mathbb{R}^p \to \mathbb{R}_+$ such that $\mathbb{P}_{\mathcal{X}}(A) = \int_A p_{\mathcal{X}}(x)\mathrm{d}x \geq p_0 \int_A \mathrm{d}x = p_0 \lambda(A)$ for any measurable subset $A \subset \mathcal{X}$. For instance, the uniform distribution over $\mathcal{X}$ satisfies this assumption. Similarly to Assumption 3, this density assumption can be relaxed to neighborhoods of $\partial\mathcal{X}$; we leave this extension for future work. We obtain the following corollary.

Corollary 1 *Define the estimator $\hat{\mathcal{Y}}^M = H\left(\{y_i\}_{i=1}^M\right)$, the offset vector $\boldsymbol{r} = (r, 0, \ldots, 0) \in \mathbb{R}^p$, the volume $\Lambda_\epsilon^{r,L} = \lambda\big(B(0, \epsilon/(2L)) \cap B(\boldsymbol{r}, r)\big)$, and the threshold $\delta_M = D(\partial\mathcal{X}, \epsilon/(2L))(1 - p_0\Lambda_\epsilon^{r,L})^M$. Then, under Assumptions 2, 4 and 5 and assuming that $\partial\mathcal{Y} \subseteq f(\partial\mathcal{X})$, with probability at least $1 - \delta_M$,*

$$d_H(\hat{\mathcal{Y}}^M, H(\mathcal{Y})) \leq \epsilon \quad \textit{and} \quad \mathcal{Y} \subseteq \hat{\mathcal{Y}}_\epsilon^M.$$

Proof We refer to Appendix B.3. We first prove that Assumptions 4 and 5 imply that Assumption 3 holds with $\Lambda_\epsilon^L = p_0 \Lambda_\epsilon^{r,L}$. The finite-sample bound then follows by applying Theorem 2. ■

The constant $\Lambda_\epsilon^{r,L}$ corresponds to the p-dimensional Lebesgue volume of two hyperspherical caps and can be computed analytically, see (Li, 2011; Petitjean, 2013) and Appendix C.

5.3. Insights: the difficulty of reachability analysis and algorithmic design

Theorem 2 reveals which characteristics of the problem make reachability analysis challenging:

- **Assuming the smoothness of f is necessary:** given an input set $\mathcal{X}$ and a sampling distribution $\mathbb{P}_{\mathcal{X}}$, one can construct problems for which sampling-based reachability analysis algorithms require arbitrarily many samples to compute an ϵ-accurate approximation of $\mathcal{Y}$, see Section 6.1. To derive finite-sample rates, assuming that the reachability map f is L-Lipschitz (Assumption 2) is necessary if only assumptions on input coverage density (Assumption 3) are available.
- **The smoother the easier**: a smaller Lipschitz constant L and a larger radius parameter r induce tighter bounds in Theorem 2, requiring a smaller number of samples M to obtain a desired accuracy with high probability $1 - \delta_M$. Indeed, such conditions guarantee a lower bound on the probability of sampling outputs $y_i = f(x_i) \in \mathcal{Y}$ that are close to the boundary $\partial\mathcal{Y}$, which is necessary to accurately reconstruct the true convex hull of the reachable set from samples.
- **Scalability**: by Theorem 2, the number of required samples to reach a desired ϵ-accuracy with high probability depends on the covering number. This constant characterizes the size of the parameter space in terms of dimensionality (the number of different parameters) and volume (variations of each parameter). Given any $\mathcal{X} \in \mathcal{K}$ and $d = \sup_{x \in \partial\mathcal{X}} \|x\|$, a simple and general bound for the covering number is $D(\partial\mathcal{X}, \epsilon) \leq (2d\sqrt{n}/\epsilon)^n$ (Shalev-Shwartz and Ben-David, 2009).

6. Results and applications

We perform a sensitivity analysis in Section 6.1 to illustrate the insights from Theorem 2. In Section 6.2, we compute the reachable sets of a dynamical system with a simple neural network policy and compare with prior work. Finally, in Section 6.3, we embed ϵ-RANDUP in a model predictive control (MPC) framework to reliably control a robotic platform. Our code and hardware results are available at https://github.com/StanfordASL/RandUP and https://youtu.be/sDkblTwPuEg. All computation times are measured on a computer with a 3.70GHz Intel Core i7-8700K CPU.

6.1. Sensitivity analysis

We analyze the sensitivity of ϵ-RANDUP to the sampling distribution and the smoothness of the reachability map. We consider a 2-dimensional input ball $\mathcal{X} = B(0, 1)$ and the map $f(x) = (Lx_1, x_2)$ with $L \geq 1$. Clearly, $\mathcal{X}^c$ is 1-convex and f is L-Lipschitz continuous, so Corollary 1 applies for any sampling distribution satisfying Assumption 5. We consider a distribution $\mathbb{P}_{\mathcal{X}}^{\alpha}$ that depends on a parameter $\alpha \geq 1$, such that $\mathbb{P}_{\mathcal{X}}^{\alpha}$ varies from a uniform distribution over $\mathcal{X}$ for $\alpha = 1$ to a uniform distribution over the boundary $\partial\mathcal{X}$ as $\alpha \to \infty$. Given $\delta_M = 10^{-3}$, we determine the minimum padding ϵ guaranteeing $\mathbb{P}(d_H(\hat{\mathcal{Y}}^M, \mathcal{Y}) \leq \epsilon) \geq 1 - \delta_M$ using Corollary 1, see Appendix E.1. We take $M = 1000$ samples and present results in Figure 3. We observe better performance than the predicted finite-sample bounds and that distributions with a higher probability of sampling close to the boundary (i.e., larger values of α) perform better, corresponding to lower Hausdorff distance errors. Also, ϵ-RANDUP performs better on problems with smoother reachability maps, as is visible from our empirical evaluation and theoretical bounds on the Hausdorff distance. This validates the discussion in Section 5.3.

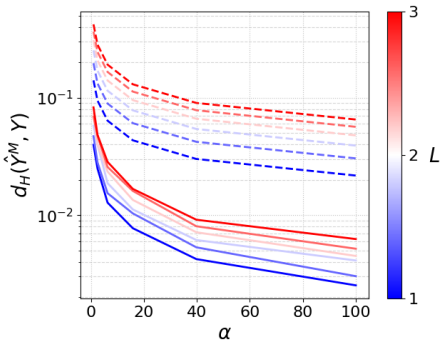

Figure 3: Results for the sensitivity analysis in Section 6.1. Experimental results are shown with continuous lines, theoretical upper bounds with dashed lines.

6.2. Verification of neural network controllers

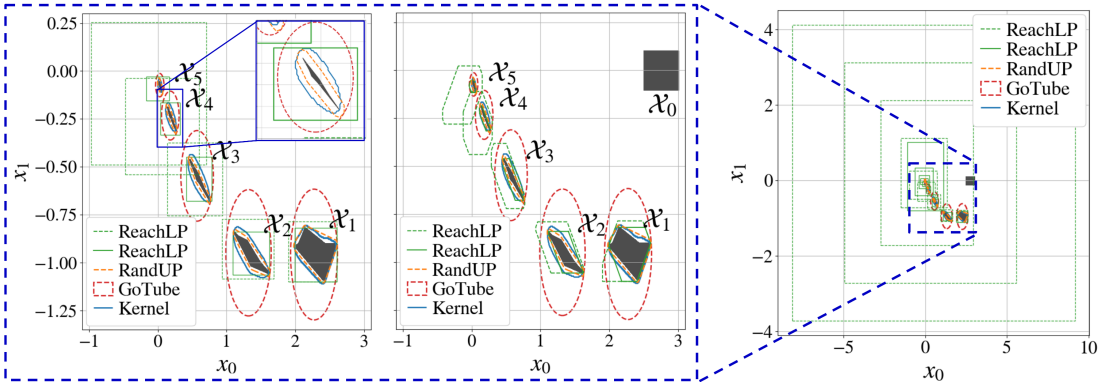

Figure 4: Reachable sets computed in Section 6.2 for a total prediction horizon $N = 9$. Sets from the formal method REACHLP *are shown in green, dashed sets correspond to no input splitting, straight-lines correspond to splitting $\mathcal{X}_0$ into 16 components. We use $M = 10^3$ samples for all sampling-based methods and $\epsilon = 0.02$.*

Next, we consider the verification of a neural network controller $u_t = \pi_{\text{nn}}(x_t)$ for a known linear dynamical system $x_{t+1} = Ax_t + Bu_t$, where $t \in \mathbb{N}$ denotes a time index, and $x_t \in \mathbb{R}^2$ and $u_t \in \mathbb{R}$ denote the state and control input. Given a rectangular set of initial states $\mathcal{X}_0 \subset \mathbb{R}^2$, the problem consists of estimating the reachable set at time $t \in \mathbb{N}$ defined as $\mathcal{X}_t = \{(A(\cdot) + B\pi_{\text{nn}}(\cdot)) \circ \cdots \circ (Ax_0 + B\pi_{\text{nn}}(x_0)) : x_0 \in \mathcal{X}_0\}$. Defining $(\mathcal{X}, \mathcal{Y}) = (\mathcal{X}_0, \mathcal{X}_t)$ and $f(x) = (A(\cdot) + B\pi_{\text{nn}}(\cdot)) \circ \cdots \circ (Ax + B\pi_{\text{nn}}(x))$, we see that this problem fits the mathematical form described in Section 1. We use a ReLU network π_{nn} from (Everett et al., 2021) with two layers of 5 neurons each. We compare ϵ-RANDUP with the formal method REACHLP (Everett et al., 2021)[1] and with two recently-derived sampling-based approaches: the kernel method proposed in (Thorpe et al., 2021)

1. Comparisons with REACHSDP (Hu et al., 2020), which is more conservative than REACHLP, show a similar trend.

and GOTUBE (Gruenbacher et al., 2022). We implement GOTUBE using the ϵ-RANDUP algorithm where we replace the last convex hull bounding step with an outer-bounding ball. As ground-truth, we use the reachable sets from ϵ-RANDUP with $\epsilon = 0$ and $M = 10^6$, which is motivated by the asymptotic results from Theorem 1 and was previously done in (Everett et al., 2021). We refer to Appendix E.2 for details and present results in Figures 4 and 5.

Formal methods that explicitly bound the output of each layer of the neural network can guarantee that their reachable set approximations are always conservative. However, obtaining tight approximations with REACHLP requires splitting the input set: a computationally expensive procedure (Fig. 5, bottom). Figures 4 and 5 show that REACHLP is more conservative than ϵ-RANDUP even when considering polytopic outputs with eight facets. As shown in Figure 4 (right), the conservatism of these methods increases over time. This shows that even when considering small neural networks, verifying safety specifications over long horizons remains an open challenge.

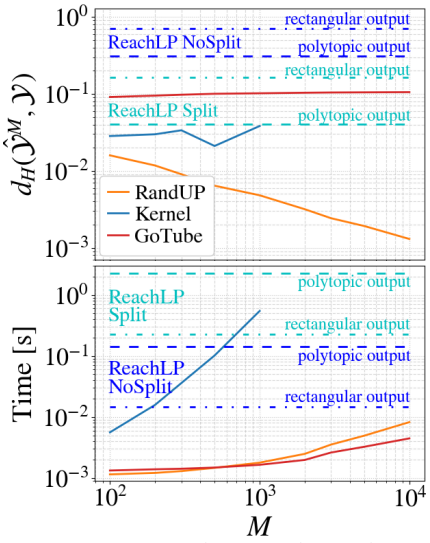

Figure 5: Neural network verification analysis in Section 6.2: we report the computation time of each algorithm and their averaged Hausdorff distance error (with ϵ=0 for ϵ-RANDUP and GOTUBE) over 100 tries when estimating $\mathcal{Y} = \mathcal{X}_4$.

Sampling-based approaches do not suffer from the long-horizon conservatism of formal methods. This comes at the expense of probabilistic guarantees (that rely on knowledge of the Lipschitz constant of the model), as opposed to deterministic conservatism guarantees. ϵ-RANDUP and GOTUBE have comparable computation time[2] and are significantly faster than other approaches. ϵ-RANDUP is significantly more accurate than prior work, especially for larger values of M. Also, the results from Theorem 2 allow for principled hyperparameter selection for ϵ-RANDUP: given $\epsilon = 0.02$, sampling 1400 uniformly-distributed inputs on $\partial\mathcal{X}$ is sufficient for the output sets to be conservative with probability at least $1 - 10^{-4}$ (for $L = 1$, see Section E.2).

These experiments show that for short-horizon problems (5 steps) with relatively simple network architectures, both REACHLP and ϵ-RANDUP return accurate reachable set approximations. For longer-horizon problems (9 steps) with networks of moderate dimensions (which allows using existing methods to pre-compute a Lipschitz constant, see (Fazlyab et al., 2019) and Section D), ϵ-RANDUP is guaranteed to efficiently return non-overly-conservative reachable set approximations with high probability. Finally, though we do not present such results here, the generality of ϵ-RANDUP allows it to tackle complex model architectures (see (Lew et al., 2022) for experiments with longer horizons and more complex networks with uncertain weights) for which no alternative methods exist, albeit without finite-sample accuracy guarantees.

6.3. Application to robust model predictive control

Finally, we show that ϵ-RANDUP can be embedded in a robust MPC formulation to reliably control a planar spacecraft system actuated by cold-gas thrusters. Its state at time $t \geq 0$ is denoted as $x_t \in$

2. Plotting the kernel-based level set estimator in (Thorpe et al., 2021) from M samples requires classifying a dense grid of points. To evaluate the computation time of this method, we only account for the time to classify M new samples.

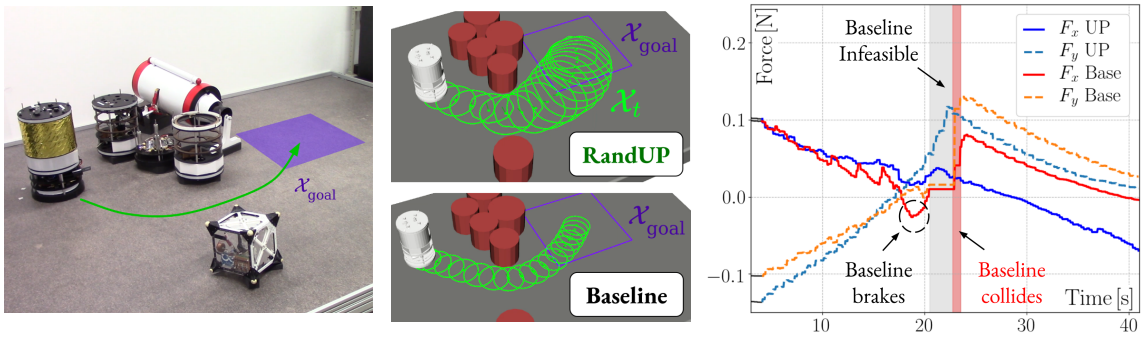

Figure 6: Application of ϵ-RANDUP to safely control a free-flyer robot in a cluttered environment (left). Using a model predictive controller that does not account for the uncertain dynamics (middle) leads to unsafe behavior, colliding with an obstacle and causing the optimization problem to be infeasible at run-time (right).

$\mathbb{R}^6$ and its control inputs are given as $u_t \in \mathbb{R}^3$. We use an auxiliary linear feedback controller (Lew et al., 2022) and an uncertain linear model $x_{t+1} = f(x_t, u_t, m, F)$ that depends on an uncertain mass $m \in [10, 18]$ kg (depending on the payload transported by the robot and the current weight of the gas tanks) and an unknown force $F = (F_x, F_y) \in [-0.015, 0.015]^2$ N that accounts for the tilt of the table. To control the system from an initial state $x_0 \in \mathbb{R}^n$ to a goal region $\mathcal{X}_{\text{goal}} \subset \mathbb{R}^n$ while minimizing fuel consumption and remaining in a feasible set $\mathcal{X}_{\text{free}}$ (i.e., avoiding obstacles and respecting velocity bounds), we consider the following MPC formulation:

$$\min_{(\mu,\nu)} \quad \textstyle\sum_{t=1}^{N} (\mu_t - x_{\text{goal}})^\top Q (\mu_t - x_{\text{goal}}) + \sum_{t=1}^{N} \nu_t^\top R \nu_t, \quad \text{s.t.} \quad \mu_0 = x_0, \tag{4a}$$

$$\mu_{t+1} = f(\mu_t, \nu_t, \bar{m}, \bar{F}), \;\; \nu_t \in \mathcal{U}, \;\; \mathcal{X}_t(\nu) \subset \mathcal{X}_{\text{free}}, \;\; \mathcal{X}_N(\nu) \subset \mathcal{X}_{\text{goal}}, \;\; t = 0, \ldots, N-1. \tag{4b}$$

where $\mu = (\mu_0, \ldots, \mu_N)$ and $\nu = (\nu_0, \ldots, \nu_{N-1})$ are optimization variables representing the nominal state and control trajectories, $(\bar{m}, \bar{F}_x, \bar{F}_y) = (14, 0, 0)$ are nominal parameter values, $x_{\text{goal}} \in \mathcal{X}_{\text{goal}}$ is the center of the goal set, and the reachable sets $\mathcal{X}_t(\nu) \subset \mathbb{R}^n$ are defined as $\mathcal{X}_t(\nu) = \{x_t = f(\cdot, \nu_{t-1}, m, F) \circ \cdots \circ f(x_0, \nu_0, m, F) : (m, F) \in [10, 18] \times [-0.015, 0.015]\}$. The numerical implementation is described in (Lew and Pavone, 2020). With a Python implementation, $\epsilon = 0.03$, and $M = 10^3$, our MPC controller runs at 10Hz which is sufficient for this platform and could be improved, e.g., by parallelizing computations on a GPU. We compare with a MPC baseline that does not consider uncertainty over the parameters (i.e., assumes $(m, F) \in \{14\} \times \{(0, 0)\}$). As shown in Figure 6 and in the attached video, this baseline is unsafe and collides with an obstacle. In contrast, our reachability-aware controller is recursively feasible, satisfies all constraints, and allows safely reaching the goal. These experiments motivate the development of efficient reachability algorithms that can be embedded in generic control frameworks to account for uncertain parameters.

7. Conclusion

We derived new asymptotic and finite-sample statistical guarantees for ϵ-RANDUP, a simple yet efficient algorithm for reachability analysis of general systems. We demonstrated its efficacy for a neural network verification task and its applicability to robust model predictive control. In future work, we will investigate tighter finite-sample bounds by leveraging further information about the smoothness of the input set boundary $\partial\mathcal{X}$. Of practical interest is investigating which sampling distributions enable better sample efficiency, interfacing ϵ-RANDUP with Lipschitz constant computation methods (e.g., (Fazlyab et al., 2019) for neural networks), exploring methods to scale to high-dimensional input spaces, and applying the technique to safety-aware reinforcement learning.

Acknowledgments

The authors thank Robin Brown for her helpful feedback and insightful discussions about neural network verification, Edward Schmerling for his helpful comments and suggestions, and Adam Thorpe for helpful discussions about kernel methods. The NASA University Leadership Initiative (grant #80NSSC20M0163) provided funds to assist the authors with their research, but this article solely reflects the opinions and conclusions of its authors and not any NASA entity. NVIDIA provided funds to assist the authors with their research. L.J. was supported by the National Science Foundation via grant CBET-2112085.

References

E. Aamari. *Rates of Convergence for Geometric Inference*. PhD thesis, Université Paris-Saclay, 2017.

E. Aamari and C. Levrard. Stability and minimax optimality of tangential delaunay complexes for manifold reconstruction. *Discrete & Computational Geometry*, 59(4):923–971, 2018.

M. Althoff, G. Frehse, and A. Girard. Set propagation techniques for reachability analysis. *Annual Review of Control, Robotics, and Autonomous Systems*, 4(1):369–395, 2021.

E. Arias-Castro, B. Pateiro-Lopez, and A. Rodriguez-Casal. Minimax estimation of the volume of a set under the rolling ball condition. *Journal of the American Statistical Association*, 114(527): 1162–1173, 2019.

A. Baillo and A. Cuevas. On the estimation of a star-shaped set. *Advances in Applied Probability*, 33(4):717–726, 2001.

J. D. Boissonnat and A. Ghosh. Manifold reconstruction using tangential delaunay complexes. *Discrete & Computational Geometry*, 51(1):221–267, 2013.

R. T. Q. Chen, Y. Rubanova, J. Bettencourt, and D. Duvenaud. Neural ordinary differential equations. In *Conf. on Neural Information Processing Systems*, 2018.

X. Chen, E. Abraham, and S. Sankaranarayanan. Flow*: An analyzer for non-linear hybrid systems. In *Proc. Int. Conf. Computer Aided Verification*, 2013.

A. Cuevas. Set estimation: Another bridge between statistics and geometry. *Boletin de Estadistica e Investigacion Operativa*, 25(2):71–85, 2009.

E. De Vito, L. Rosasco, and A. Toigo. Learning Sets with Separating Kernels. *Applied and Computational Harmonic Analysis*, 37(2):185–217, 2014.

A. Devonport and M. Arcak. Estimating reachable sets with scenario optimization. In *Proc. of the 2nd Conference on Learning for Dynamics and Control*, 2020.

L. Devroye and G. L. Wise. Detection of abnormal behavior via nonparametric estimation of the support. *SIAM Journal on Applied Mathematics*, 38(3):480–488, 1980.

L. Dumbgen and G. Walther. Rates of convergence for random approximations of convex sets. *Advances in Applied Probability*, 28(2):384–393, 1996.

M. Everett, G. Habibi, S. Chuangchuang, and J. P. How. Reachability analysis of neural feedback loops. *IEEE Access*, 2021. Available at https://arxiv.org/abs/2101.01815.

M. Fazlyab, A. Robey, H. Hassani, M. Morari, and G. J. Pappas. Efficient and accurate estimation of lipschitz constants for deep neural networks. In *Conf. on Neural Information Processing Systems*, 2019.

H. Federer. Curvature measures. *Transactions of the American Mathematical Society*, (93):418–491, 1959.

A. Girard. Reachability of uncertain linear systems using zonotopes. In *Hybrid Systems: Computation and Control*, 2005.

S. Gruenbacher, M. Lechner, R. Hasani, D. Rus, T. A. Henzinger, S. Smolka, and R. Grosu. GoTube: Scalable stochastic verification of continuous-depth models. In *Proc. AAAI Conf. on Artificial Intelligence*, 2022.

B. Hanin and D. Rolnick. Deep relu networks have surprisingly few activation patterns. In *Conf. on Neural Information Processing Systems*, 2019.

R. Harman and V Lacko. On decompositional algorithms for uniform sampling from n-spheres and n-balls. *Journal of Multivariate Analysis*, 101(10):2297–2304, 2010.

H. Hu, M. Fazlyab, M. Morari, and G. J. Pappas. Reach-SDP: Reachability analysis of closed-loop systems with neural network controllers. In *Proc. IEEE Conf. on Decision and Control*, 2020.

R. Ivanov, J. Weimer, R. Alur, G. J. Pappas, and I. Lee. Verisig: verifying safety properties of hybrid systems with neural network controllers. In *Hybrid Systems: Computation and Control*, 2019.

S. Kong, S. Gao, W. Chen, and E. Clarke. dreach: δ-reachability analysis for hybrid systems. In *Int. Conf. on Tools and Algorithms for the Construction and Analysis of Systems* , 2015.

A. B. Kurzhanski and P. Varaiya. Ellipsoidal techniques for reachability analysis. In *Hybrid Systems: Computation and Control*, 2000.

T. Lew and M. Pavone. Sampling-based reachability analysis: A random set theory approach with adversarial sampling. In *Conf. on Robot Learning*, 2020.

T. Lew, A. Sharma, J. Harrison, A. Bylard, and M. Pavone. Safe active dynamics learning and control: A sequential exploration-exploitation framework. *IEEE Transactions on Robotics*, 2022. In Press.

S. Li. Concise formulas for the area and volume of a hyperspherical cap. *Asian Journal of Mathematics and Statistics*, 4(1):66–70, 2011.

C. Liu, T. Arnon, C. Lazarus, C. Strong, C. Barrett, and M. J. Kochenderfer. Algorithms for verifying deep neural networks. *Foundations and Trends in Optimization*, 4(3-4):244–404, 2021.

G. Matheron. *Random sets and integral geometry*. Wiley Series in Probability and Mathematical Statistics, 1975.

M. Matt. How to compute the volume of intersection between two hyperspheres. Mathematics Stack Exchange, available at https://math.stackexchange.com/q/162873, 2013.

I. Molchanov. *Theory of Random Sets*. Springer-Verlag, second edition, 2017.

G. Montufar, R. Pascanu, K. Cho, and Y. Bengio. On the number of linear regions of deep neural networks. In *Conf. on Neural Information Processing Systems*, 2014.

M. Petitjean. Spheres unions and intersections and some of their applications in molecular modeling. In *Distance Geometry: Theory, Methods, and Applications*, pages 61–83. Springer New York, 2013.

B. D. Ripley and J. P. Rasson. Finding the edge of a poisson forest. *Journal of Applied Probability*, 14:483–491, 1977.

A. Rodriguez-Casal and P. Saavedra-Nieves. A fully data-driven method for estimating the shape of a point cloud. *ESAIM: Probability and Statistics*, 20(1):332–348, 2016.

A. Rodriguez-Casal and P. Saavedra-Nieves. Extent of occurrence reconstruction using a new data-driven support estimator. Available at https://arxiv.org/abs/1907.08627, 2019.

A. Rudi, E. De Vito, A. Verri, and F. Odone. Regularized Kernel Algorithms for Support Estimation. *Frontiers in Applied Mathematics and Statistics*, 3:1–15, 2017.

R. Schneider. Random approximation of convex sets. *Journal of Microscopy*, 151(3):211–227, 1988.

R. Schneider. *Convex Bodies: The Brunn-Minkowski Theory*. Cambridge Univ. Press, second edition, 2014.

B. Schürmann, N. Kochdumper, and M. Althoff. Reachset model predictive control for disturbed nonlinear systems. In *Proc. IEEE Conf. on Decision and Control*, 2018.

T. Serra, C. Tjandraatmadja, and S. Ramalingam. Bounding and counting linear regions of deep neural networks. In *Int. Conf. on Machine Learning*, 2018.

S. Shalev-Shwartz and S. Ben-David. *Understanding Machine Learning*. Cambridge University Press, 2009.

Y. S. Shao, C. Chen, S. Kousik, and R. Vasudevan. Reachability-based trajectory safeguard (RTS): A safe and fast reinforcement learning safety layer for continuous control. *IEEE Robotics and Automation Letters*, 6(2):239–261, 2021.

A. Sinha, M. O'Kelly, T. Tedrake, and J. Duchi. Neural bridge sampling for evaluating safety-critical autonomous systems. In *Conf. on Neural Information Processing Systems*, 2020.

A. J. Thorpe, K. R. Ortiz, and Oishi M. M. K. Learning approximate forward reachable sets using separating kernels. In *Proc. of the 3rd Conference on Learning for Dynamics and Control*, 2021.

H.-D. Tran, D. Manzanas Lopez, P. Musau, X. Yang, L. V. Nguyen, W. Xiang, and T. T. Johnson. Star-based reachability analysis of deep neural networks. In *Int. Symp. on Formal Methods*, 2019.

J. A. Vincent and M. Schwager. Reachable polyhedral marching (RPM): A safety verification algorithm for robotic systems with deep neural network components. In *Proc. IEEE Conf. on Robotics and Automation*, 2021.

G. Walther. Granulometric smoothing. *The Annals of Statistics*, 25(6):2273 – 2299, 1997.

S. Webb, T. Rainforth, Y.W. Teh, and M. P. Kumar. A statistical approach to assessing neural network robustness. In *Int. Conf. on Learning Representations*, 2019.

A. Wittig, P. Di Lizia, R. Armellin, K. Makino, Bernelli-Zazzera F., and M. Berz. Propagation of large uncertainty sets in orbital dynamics by automatic domain splitting. *Celestial Mechanics and Dynamical Astronomy*, 122:239–261, 2015.