

A Survey of Algorithms for Black-Box Safety Validation of Cyber-Physical Systems

Anthony Corso

*Aeronautics and Astronautics, Stanford University,
Stanford, CA 94305, USA*

ACORSO@STANFORD.EDU

Robert J. Moss

*Computer Science, Stanford University,
Stanford, CA 94305, USA*

MOSSR@CS.STANFORD.EDU

Mark Koren

*Aeronautics and Astronautics, Stanford University,
Stanford, CA 94305, USA*

MKOREN@STANFORD.EDU

Ritchie Lee

*NASA Ames Research Center,
Moffett Field, CA 94035, USA*

RITCHIE.LEE@NASA.GOV

Mykel J. Kochenderfer

*Aeronautics and Astronautics, Stanford University,
Stanford, CA 94305, USA*

MYKEL@STANFORD.EDU

Abstract

Autonomous cyber-physical systems (CPS) can improve safety and efficiency for safety-critical applications, but require rigorous testing before deployment. The complexity of these systems often precludes the use of formal verification and real-world testing can be too dangerous during development. Therefore, simulation-based techniques have been developed that treat the system under test as a black box operating in a simulated environment. Safety validation tasks include finding disturbances in the environment that cause the system to fail (falsification), finding the most-likely failure, and estimating the probability that the system fails. Motivated by the prevalence of safety-critical artificial intelligence, this work provides a survey of state-of-the-art safety validation techniques for CPS with a focus on applied algorithms and their modifications for the safety validation problem. We present and discuss algorithms in the domains of optimization, path planning, reinforcement learning, and importance sampling. Problem decomposition techniques are presented to help scale algorithms to large state spaces, which are common for CPS. A brief overview of safety-critical applications is given, including autonomous vehicles and aircraft collision avoidance systems. Finally, we present a survey of existing academic and commercially available safety validation tools.

1. Introduction

Increasing levels of autonomy in cyber-physical systems (CPS) promise to revolutionize industries such as automotive transportation (U.S. Department of Transportation, 2018) and aviation (Federal Aviation Administration, 2019; Kochenderfer et al., 2012) by improving convenience and efficiency while lowering cost. Innovations have been driven by recent progress in artificial intelligence, particularly in machine learning (LeCun et al., 2015; Rus-

sell & Norvig, 2020) and planning (Kochenderfer, 2015; Sutton & Barto, 2018). Machine learning has recently achieved human-competitive performance in board games (Silver et al., 2016; Silver et al., 2017), video games (Mnih et al., 2015; Vinyals et al., 2019), and visual perception (He et al., 2017; Pillai et al., 2019). However, applying machine learning technologies to safety-critical domains has been challenging. Safety-critical CPS differ from conventional autonomous systems in that their failures can have serious consequences, such as loss of life and property. As a result, these systems must undergo extensive validation and testing prior to certification and deployment.

Safety validation is the process of ensuring the correct and safe operation of a system operating in an environment. Desired safety properties are stipulated in a specification language and a failure is any violation of that specification. Typically, simulation is used to find failures of a system caused by disturbances in the environment, and a model of the disturbances can then be used to determine the probability of failure. A system is deemed safe if no failure has been found after adequate exploration of the space of possible disturbances, or if the probability of failure is found to be below an acceptable threshold. The procedure of proving that a system is safe to all disturbances is known as formal verification (Clarke et al., 2018; Fitting, 2012; Katoen, 2016; Platzer & Quesel, 2008; Schumann, 2001) and is outside the scope of this survey.

In this paper, we focus on CPS, which involve software and physical systems interacting over time. This broad definition includes systems such as robots, cars, aircraft, and planetary rovers. There are several reasons why validating cyber-physical systems is challenging. First, many of these systems contain complex components, including those produced by machine learning. The safety properties of these systems may not be well-understood and subtle and emergent failures can go undetected (Yeh, 2018). Second, many systems, such as autonomous cars and aircraft, interact with complex and stochastic environments that are difficult to model. Third, safety properties are generally defined over both the system under test and its environment. For example, the requirement that “the test vehicle shall not collide with pedestrians” involves both the system under test (test vehicle) and actors in its environment (pedestrians). As a result, safety validation must be performed over the combined system. Another challenge is that sequential interactions between the system and the environment means that failure scenarios are trajectories over time, and therefore the search space is combinatorially large. Finally, safety validation is often applied to mature safety-critical systems later in development, where failures can be extremely rare.

Traditional methods for ensuring safety—through safety processes, engineering analysis, and conventional testing—though necessary, do not scale to the complexity of next-generation systems. Advanced validation techniques are needed to build confidence in these systems. Many validation approaches have been proposed in the literature. They can be broadly categorized by the information they use for analysis. White-box methods use knowledge of the internals of the system. For example, formal verification in the form of model checking (Clarke et al., 2018; Katoen, 2016) and automated theorem proving (Fitting, 2012; Platzer & Quesel, 2008; Schumann, 2001), represents the system using mathematical models. Because the model is known, formal verification methods can find failure examples when they exist or prove the absence of failures when they do not. However, because formal verification considers all execution possibilities, it often has difficulty scaling to large problems (see Alur, 2015 for a discussion of verification applied to CPS).

In contrast to white-box methods, black-box techniques do not assume that the internals of the system are known. They consider a general mapping from input to output that can be sampled. Black-box methods can be applied to a much broader class of systems because they do not require a system specification. Although model-checking techniques can be applied to some black-box systems (Peled et al., 1999), a prohibitively large (or possibly infinite) number of samples may be required to provide complete coverage and prove the absence of failures. Instead, black-box methods often aim to quickly and efficiently find failure examples. If no failures are found, confidence in the safety of the system will increase with additional sampling. Due to its flexibility and scalability, black-box validation is often the only feasible option for large complex systems, and is the focus of this survey.

We consider three safety validation tasks for a system with safety properties. First, falsification aims to find an example disturbance in the environment that causes the system to violate the property. This formulation is useful for discovering previously unknown failure modes and finding regions where the system can operate safely. The second safety validation task is to find the most-likely failure according to a probabilistic model of the disturbances. The model can be created through expert knowledge or data to reflect the probabilities in the real environment. The third safety validation task is to estimate the probability that a failure will occur. Failure probability estimation is important for acceptance and certification.

There are many algorithms that have been used for these safety validation tasks. This survey categorizes and presents many of them. Falsification and most-likely failure analysis are related tasks in that they involve finding failures of an autonomous system. Categories of algorithms that are suited for these tasks include optimization, path planning, and reinforcement learning. Optimization approaches seek to find a trajectory of disturbances that cause the system to fail. Path planning approaches use the environment’s state to aid in the exploration of possible failure modes. Reinforcement learning frames the problem as a Markov decision process and searches for a policy that maps environment states to disturbances that cause the system to fail. When the goal is to estimate the probability of failure and failures are rare, importance sampling techniques generate scenarios and translate their results into a probability estimate. For all of the safety validation tasks, a major challenge is scalability, so problem-decomposition techniques are presented that can allow for better scalability of the presented algorithms.

This paper is organized as follows. Section 2 introduces notation, formally defines common terms such as safety validation or black box, and formulates three safety validation tasks. Section 3 gives an overview of the safety validation process, which involves defining safety properties, choosing a cost function and algorithm, and determining when sufficient testing has been performed. Section 4 summarizes the optimization-based algorithms for safety validation. Section 5 describes how to use the environment state to find failures of the system through several path-planning algorithms. Section 6 shows how to apply reinforcement learning to safety validation. Section 7 introduces importance sampling algorithms for estimating the probability of failure. Section 8 presents several ways to address problem scalability through decomposition. Section 9 surveys the various applications and discusses common strategies and adaptations of approaches for each domain. Finally, section 10 surveys existing tools in the literature and compares their basic features.

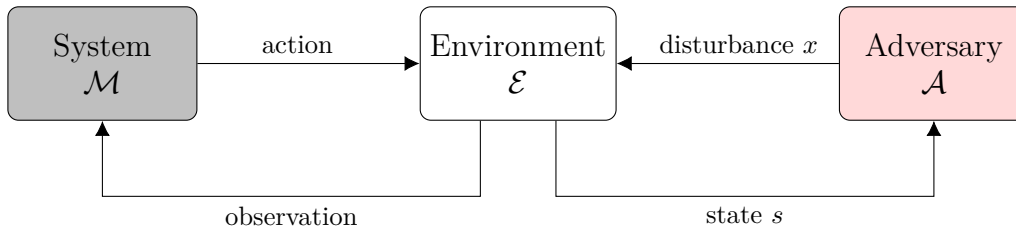


Figure 1: Model of the safety validation problem.

2. Preliminaries

This section first describes the notation used for the description of safety validation algorithms. It then defines *safety validation* and *black box*, and then summarizes several safety validation tasks.

2.1 Notation

A safety validation problem (fig. 1) consists of a system \mathcal{M} , an environment \mathcal{E} , and a safety property ψ that the system should have. The safety property is defined over state trajectories $\mathbf{s} = [s_1, \dots, s_t]$, where $s_t \in S$ is the state of the environment at time $t \in \{1, \dots, t_{\max}\}$. If the state trajectory \mathbf{s} satisfies property ψ we write $\mathbf{s} \in \psi$, and write $\mathbf{s} \notin \psi$, otherwise.

The environment is perturbed by an adversary \mathcal{A} through disturbances $x \in X$, where disturbances are chosen in order to induce behavior in the system which violates the safety property. Disturbance trajectories \mathbf{x} can be chosen freely by the adversary, but they have an associated probability density $p(\mathbf{x})$, $p(x)$, or $p(x | s)$, which models their likelihood in the environment. The disturbance likelihood can be constructed through expert knowledge or learned from data.

The environment transitions between states according to a dynamics function f that depends on the system, environment, and disturbances. In this work, we assume that the environment and the system are fixed, making disturbances the only way to affect the system. Therefore, f maps disturbance trajectories into state trajectories

$$\mathbf{s} = f(\mathbf{x}). \tag{1}$$

Some algorithms require the ability to simulate a disturbance x_t for a single timestep from a state s_t , denoted

$$s_{t+1} = f(s_t, x_t). \tag{2}$$

2.2 Definitions

Safety Validation. A *safety* property specifies that a certain “bad event” will not occur. In contrast, a *liveness* property specifies that a certain “good event” will eventually occur. The safety-liveness distinction is important because a safety property can be shown to be violated with a concrete counterexample (the primary goal of the surveyed algorithms), while

the violation of a liveness property requires formal argumentation (Alpern & Schneider, 1987). The definition of a “bad event” is domain specific.

Verification is the process of proving that the system meets its requirements while *validation* is the process of ensuring that a system fulfills its intended purpose in its operational environment (Hirshorn et al., 2017). Although many of the algorithms presented in this survey can be applied to both verification and validation, we choose the term *validation* to emphasize the focus on testing full-scale system prototypes in simulated operational environments. *Safety validation* is therefore the processes of investigating the adherence of a system to a safety property in its operational domain.

Black-Box Assumption. A system is said to be a *black box* if the system model \mathcal{M} is not known or is too complex to explicitly reason about. In contrast, a *white-box* system can be described analytically or specified in a formal modeling language, and a *gray-box* system lies in between. Some white-box systems may be treated as a black box if knowledge of their design does not help the validation process. For example, while small neural networks can have properties formally verified by analyzing the network weights (Katz et al., 2017), large neural networks with millions or billions of parameters are generally too large for such techniques, and they would need to undergo black-box validation. In some cases, both the system and the environment are treated as a black-box, which precludes the use of validation algorithms that require the environment state (see section 3.3 for more details).

2.3 Safety Validation Goals

Three safety validation tasks are considered in this work and are defined below.

Falsification. Falsification is the process of finding a disturbance trajectory that causes the outputs to violate a specification ψ . Such a trajectory is known as a *counterexample*, *failure trajectory*, or *falsifying trajectory*. Falsification finds

$$\mathbf{x} \text{ s.t. } f(\mathbf{x}) \notin \psi. \quad (3)$$

Most-Likely Failure Analysis. Most-likely failure analysis tries to find the failure trajectory with maximum likelihood

$$\arg \max_{\mathbf{x}} p(\mathbf{x}) \text{ s.t. } f(\mathbf{x}) \notin \psi. \quad (4)$$

Failure Probability Estimation. Failure probability estimation tries to compute the probability that a specification will be violated. Failure probability is given by the expectation of observing a failure under the disturbance model

$$P_{\text{fail}} = \mathbb{E}[\mathbb{1}\{f(\mathbf{x}) \notin \psi\}]. \quad (5)$$

3. Overview of Solution Techniques

Solving a safety validation problem for a given CPS requires the following steps: 1) define a safety property to validate, 2) define an appropriate cost function to guide the search, 3) choose a safety validation algorithm, which depends on the system, environment and safety validation task and 4) run it until a counterexample is discovered (for falsification and most

likely failure analysis) or the space of possible scenarios has been sufficiently covered. This section provides an overview of each of these steps.

3.1 Safety Specification with Temporal logic

In safety validation, the first step is to define a safety property to validate. Although safety properties could be defined using natural language or human preferences (both of which are highly expressive), formal specification languages are often preferred because they reduce ambiguity and permit efficient numerical evaluation of state trajectories. The most common formal specification languages are based on temporal logic.

Temporal logic is a logical framework for describing properties of signals over time. It enables reasoning about time and temporal information. Properties are stated as formulas that evaluate to a Boolean value and the syntax of these formulas is governed by a grammar. Various temporal logics have been proposed for different domains, including linear temporal logic (Pnueli, 1977), metric temporal logic (Koymans, 1990), and computation tree logic (Clarke & Emerson, 1981). Signal temporal logic (STL) is a temporal logic for real-valued signals that is widely used as a specification language for the safety validation of cyber-physical systems (Donzé & Maler, 2010; Kapinski et al., 2016). The basic unit of an STL formula is an atomic formula of the form $\mu(x_i) \geq 0$, where x_i is a real-valued signal and μ is an arbitrary function from \mathbb{R}^n to \mathbb{R} . A combination of Boolean and temporal operators can be applied to one or more atomic formulas to form more complex formulas. Boolean operators can include unary or binary Boolean operators, such as *negation* \neg , *conjunction* \wedge , and *disjunction* \vee . Temporal operators reason over the temporal aspect of signals. Examples of temporal operators include *always* $\Box\psi$ (ψ holds on the entire subsequent path), *eventually* $\Diamond\psi$ (ψ holds somewhere on the subsequent path), and *until* $\psi_1\mathcal{U}\psi_2$ (ψ_1 holds at least until ψ_2 becomes true on the subsequent path). Temporal operators may be indexed by a time interval I over which the property is considered.

The following are some examples of STL formulas representing safety properties:

$$\psi_1 : \quad \Diamond_{[0,100]}(d_{\text{goal}} < 10) \tag{6}$$

$$\psi_2 : \quad \Box_{[0,\infty]}(\neg(d_h < 500 \wedge d_v < 100)) \tag{7}$$

The specification ψ_1 (eq. (6)) states that at some point in the next 100 seconds, the distance to the goal d_{goal} becomes less than 10 feet. Such a specification could be used, for example, to require that an autonomous vehicle reach a goal location within a certain time limit. The specification ψ_2 (eq. (7)) states that for all time, the horizontal separation d_h and the vertical separation d_v between two aircraft shall not be simultaneously less than 500 feet and 100 feet, respectively. This specification describes the absence of a near mid-air collision, which is an important safety event for aircraft collision avoidance systems (Kochenderfer et al., 2012).

3.2 Cost Functions

A naive approach to safety validation is to search randomly over disturbance trajectories until a counterexample is discovered. If counterexamples are rare, this process can be inefficient or even intractable. A cost function $c_{\text{state}}(\mathbf{s})$ that measures the level of safety of

the system over the state trajectory \mathbf{s} can be used to guide the search. A well-designed cost function will help bias the search over disturbance trajectories toward those that are less safe. Additionally, if the goal is to find the most-likely failure, then the cost function can incorporate the likelihood of the disturbance trajectory.

Once a cost function is defined, safety validation becomes an optimization problem over disturbance trajectories

$$\mathbf{x}^* = \arg \min_{\mathbf{x}} c(\mathbf{x}), \quad (8)$$

where $c(\mathbf{x}) = c_{\text{state}}(f(\mathbf{x}))$. The cost function is designed such that $c(\mathbf{x}) \geq \epsilon \iff f(\mathbf{x}) \in \psi$, where ϵ is a safety threshold. Therefore, if a disturbance \mathbf{x} causes $c(\mathbf{x}) < \epsilon$, then \mathbf{x} is a counterexample.

The design of the cost function c is specific to the application, but can be done ad-hoc or using a formal measure of the satisfaction of a safety property. For simple safety properties such as collision avoidance, the miss distance (point of closest approach between agents) is a common choice (Koren et al., 2018; Lee et al., 2020). Balkan et al. (2017) propose several possible functions for finding non-convergence behaviors in control systems, including Lyapunov-like functions, M -step Lyapunov-like functions, neural networks, and support vector machines.

When the specification is represented by a temporal logic expression, a natural choice for c is the temporal logic robustness $\rho(\mathbf{s}, \psi)$. The robustness is a measure of the degree to which the trajectory \mathbf{s} satisfies the property ψ . Large values of robustness mean that at no point does the trajectory come close to violating the specification, while low but positive values of robustness mean that the trajectory is close to violating the specification. A robustness value less than zero means that the specification has been violated and gives an indication of by how much. The robustness for space-time signals can be computed from a recursive definition (Donzé & Maler, 2010; G. E. Fainekos & Pappas, 2009; Leung et al., 2020; H. Yang, 2013). The derivative of the robustness with respect to the state trajectory can be approximated, which may help derive gradient-driven coverage metrics (Leung et al., 2020; Pant et al., 2017) or with doing gradient-based optimization if the black-box assumption is relaxed. Causal information in the form of a Bayesian network can be incorporated into the robustness for improved falsification (Akazaki et al., 2017). Connections have also been made between robustness and delta-reachability to relate falsification and exhaustive search through approximations (Abbas et al., 2017).

Robustness can be a challenging objective to optimize because it can be non-smooth (Leung et al., 2020) and be dominated by state variables with larger magnitude values. Leung et al. (2020) propose smooth approximations to the robustness and Akazaki et al. (2018) optimize a convex function of the robustness $-\exp(-\rho(\mathbf{s}))$, which bounds the maximum cost. State variables may be normalized to alleviate differences in scale between the variables, but normalization requires the range of the state variables which may not be known a priori. Zhang et al. (2019) propose measuring the robustness of each state variable independently and using a multi-armed bandit algorithm to decide which robustness value to optimize on each iteration.

In addition to measuring safety, cost functions may include other search heuristics. For most likely failure analysis, the cost function includes the likelihood of the disturbance

trajectory $p(\mathbf{x})$ for counterexamples

$$c'(\mathbf{x}) = \begin{cases} c(\mathbf{x}) & \text{if } c(\mathbf{x}) \geq \epsilon \\ -p(\mathbf{x}) & \text{if } c(\mathbf{x}) < \epsilon. \end{cases} \quad (9)$$

An objective that achieves a similar goal but may be easier to optimize includes a penalty term for low-likelihood disturbances

$$c'(\mathbf{x}) = c(\mathbf{x}) - \lambda p(\mathbf{x}), \quad (10)$$

where the penalty $\lambda > 0$ is user-defined. Additional penalty terms related to coverage (see section 3.4 for an overview) can be included to encourage exploration. Domain-specific heuristics are also possible. For example, Qin et al. (2019) penalize disturbances that do not follow a domain-specific set of constraints, which is useful for getting adversarial agents in a driving scenario to follow traffic laws. Other examples of domain-specific heuristics are described in section 9.

Solution techniques that build trajectories sequentially (such as those based on planning and reinforcement learning) can require the evaluation of trajectory segments or states. Upper and lower bounds on robustness can be computed for incomplete trajectory segments (Dreossi et al., 2015). An approach for most-likely failure analysis called adaptive stress testing (AST) (Koren et al., 2019; Lee et al., 2020) defines a cost for individual state-disturbance pairs

$$c(s_t, x_t) = \begin{cases} 0 & \text{if } s_t \in S_{\text{fail}} \\ \lambda & \text{if } s_t \notin S_{\text{fail}}, t \geq t_{\text{max}} \\ \log p(x_t | s_t) & \text{if } s_t \notin S_{\text{fail}}, t < t_{\text{max}}, \end{cases} \quad (11)$$

where the specification ψ is for the system to avoid reaching a set of failure states S_{fail} , and λ penalizes disturbances that do not end in a failure state. The log probability of the disturbance is awarded at each time step to encourage the discovery of likely failures.

3.3 Overview of Algorithms

As demonstrated in the previous section, safety validation algorithms solve an optimization problem to discover counterexamples. This section outlines four categories of algorithms distinguished by the information they use for optimization, the requirements on the simulated environment, and the desired safety validation task. For each category, we summarize the approach and describe its strengths, weaknesses, and requirements. Specific algorithms are discussed in more detail in sections 4 to 8.

3.3.1 BLACK-BOX OPTIMIZATION.

Many safety validation algorithms attempt to solve eq. (8) directly over the space of disturbance trajectories. Although gradient-based approaches cannot be used due to the black-box assumption on the system, there are many existing algorithms that can be applied with no modification (Kochenderfer & Wheeler, 2019). Due to the complexity of many autonomous systems and environments, however, the optimization problem is generally non-convex and

can have many local minima. Therefore, algorithms that can escape local minima and have adequate exploration over the space of disturbance trajectories are preferred.

Two algorithms that have been used out of the box in falsification software (Y. Annapureddy et al., 2011; Donzé, 2010) (see section 10 for more details) are covariance matrix adaptation evolution strategy (CMA-ES) (Hansen & Ostermeier, 1996) and globalized Nelder-Mead (GNM) (Luersen & Le Riche, 2004), both of which are effective for global optimization. Another way to escape local minima is to combine global and local search (Adimoolam et al., 2017; Deshmukh et al., 2015; Mathesen et al., 2019; Yaghoubi & Fainekos, 2019), where one optimization routine is used to explore the space of disturbances trajectories to find regions of interest, and another algorithm does local optimization to find the best disturbance trajectory in a region.

The primary benefit of optimization-based safety validation is the minimal set of restrictions placed on the simulator of the system and the environment. Black-box optimizers do not need access to the state of the environment and only needs to return the value of a safety metric for a given disturbance trajectory. The simulation state may be unavailable for implementation or privacy reasons so optimization-based approaches would be a good choice in those cases. If, however, the state is available and would be useful for finding failures, then optimization approaches may not function as well as path planning or reinforcement learning approaches. If an environment has stochasticity in state transitions (in addition to the applied disturbances), then optimization techniques such as Bayesian optimization can be used to account for it. The biggest drawback to optimization strategies is the need to optimize over the entire space of disturbance trajectories, which scales exponentially with the time horizon.

3.3.2 PATH PLANNING.

Safety validation may be framed as a planning problem through the state space of the environment using the disturbances as control inputs. Planning algorithms construct a disturbance trajectory \mathbf{x} from an initial state s_0 to a set of failure states S_{fail} , with a corresponding cost for each transition. Classical planning algorithms (Ghallab et al., 2004) typically make the assumption that there exists a model in a formal language such as PDDL (McDermott et al., 1998) or STRIPS (Fikes & Nilsson, 1971), and that the size of the state space is discrete and not too large. Both assumptions are violated when dealing with black-box autonomous systems operating in continuous state spaces so traditional planning algorithms such as forward and backward chaining cannot be directly applied. Other planning algorithms such as value iteration (Sutton & Barto, 2018), Dijkstra’s algorithm (Dijkstra, 1959), and their variants can work with a black-box transition function but do not scale well to large or continuous state spaces. Heuristics such as state novelty (Lipovetzky & Geffner, 2012) have been successfully applied to black-box planning settings with large state spaces (Francès et al., 2017; Lipovetzky et al., 2015), and could therefore be applicable to safety validation of CPS.

Planning problems with continuous and high-dimensional state spaces commonly occur in the field of robotics and path planning (see the overview by LaValle (2006)). Optimal control algorithms (Lewis et al., 2012) are common in this domain but typically violate the black-box assumption by requiring gradients of the transition function. Sampling based

approaches such as the rapidly exploring random tree (RRT) algorithm and the probabilistic roadmap planner (Kavraki et al., 1996) can work with a black-box simulator and have been shown to be effective in high-dimensional continuous planning problems. The probabilistic roadmap planner, however, requires the ability to connect independent state trajectory segments making it difficult to use in a completely black-box setting. RRT on the other hand, grows a search tree forward from the set of reachable states and can therefore be used with a black-box simulator. RRT has been used extensively for safety validation of CPS and variations on the algorithm will be discussed in greater detail in section 5.

Path planning algorithms rely heavily on the environment state to discover failures, which can be a strength and weakness. Planning algorithms can efficiently search high-dimensional state spaces by reusing trajectory segments, and often provide a natural way to compute state space coverage which can be used to determine when sufficient testing has been done. If the reachable set of states is small compared to the state space, however, planning algorithms may perform poorly (J. Kim et al., 2005). A drawback to path planning algorithms is their inability to naturally handle stochasticity. Most path planning algorithms rely on the ability to deterministically replay trajectory segments or initialize a simulator into a predefined state, which may be challenging for some simulators. Additionally, if the problem has a long time horizon, then prohibitively large trees may be required to find failure trajectories.

3.3.3 REINFORCEMENT LEARNING.

In reinforcement learning (RL), the safety validation is modeled as a Markov decision process (MDP). An MDP is defined by a transition model $P(s' | s, x)$ that gives the probability of arriving in state s' given the current state s , a disturbance (or action) x , a reward $R(s, x)$, and a discount factor $\gamma \in [0, 1]$ that decreases the value of future rewards. RL algorithms learn a policy π (a function that maps states to disturbances $x = \pi(s)$) that maximizes expected future rewards. Typically the reward function is chosen to be $R(s, x) = -c(s, x)$ so that maximizing reward minimizes cost. For an overview of MDPs and their solvers see the texts by Kochenderfer (2015) or Sutton and Barto (2018).

Reinforcement learning algorithms are similar to path-planning approaches because they also rely on the environment state (unless specifically formulated otherwise as in Koren et al. (2019)). While planning algorithms search for full trajectories, RL algorithms learn a policy that generates disturbances from the current state. Since policies do not explicitly represent the entire disturbance trajectory, they may be easier to optimize and can be applied to long time horizon problems. Uncontrolled stochasticity in the environment is naturally handled by RL algorithms, which are designed to function in stochastic environments. Additionally, many RL algorithms can be used with episodic simulators that only require the ability to reset and step forward in time, rather than the ability to initialize to a particular state. The downside to RL algorithms is that they may be sample inefficient and require complex (and sometimes brittle) training procedures compared to optimization and path-planning approaches.

3.3.4 IMPORTANCE SAMPLING.

For many cyber-physical applications, it is impossible to design an autonomous system that never violates a safety property. In that case, failure probability is a useful metric of safety. If failure events are rare, then Monte Carlo approaches will require a large number of samples before converging to the true probability of failure (Hahn, 1972). To address this problem, importance sampling approaches artificially increase the likelihood of failure with a proposal distribution $q(\mathbf{x})$, and then weight observed failures to get an unbiased estimate of the probability of failure with fewer samples.

In addition to causing more failures, the proposal distribution has the property that $q(\mathbf{x}) > 0$ everywhere that $p(\mathbf{x})\mathbb{1}\{c(\mathbf{x}) < \epsilon\} > 0$ (so all disturbances that lead to failure can be sampled from q). The proposal distribution is also referred to as the biased, sampled, or importance distribution. The importance sampling estimate of the probability of failure is done by taking N samples drawn from q and computing the weighted average

$$\hat{P}_{\text{fail}} = \frac{1}{N} \sum_{i=1}^N \frac{p(\mathbf{x}_i)}{q(\mathbf{x}_i)} \mathbb{1}\{c(\mathbf{x}_i) < \epsilon\}. \quad (12)$$

The variance of the importance sampling estimate is given by

$$\text{Var}(\hat{P}_{\text{fail}}) = \frac{1}{N} \mathbb{E}_q \left[\frac{(p(\mathbf{x})\mathbb{1}\{c(\mathbf{x}) < \epsilon\} - q(\mathbf{x})P_{\text{fail}})^2}{q(\mathbf{x})} \right]. \quad (13)$$

The goal of a good importance sampling distribution is to minimize the variance of the estimator \hat{P}_{fail} so that fewer samples are needed for a good estimate. It is clear from eq. (13) that a zero variance estimate can be obtained with the optimal importance sampling distribution

$$q^*(\mathbf{x}) = \frac{p(\mathbf{x})\mathbb{1}\{c(\mathbf{x}) < \epsilon\}}{P_{\text{fail}}}. \quad (14)$$

Generating this distribution is not possible in practice because $c(\mathbf{x})$ is a black box and the normalization constant P_{fail} is the very quantity being estimated. Importance sampling algorithms seek to estimate the optimal importance sampling distribution $q^*(\mathbf{x})$.

Importance sampling approaches require finding many failure examples to learn a distribution over failures. Therefore, failure examples can, in principle, be found using any of the three previous approaches. The most common importance sampling approaches such as multilevel splitting and the cross-entropy method function most similarly to optimization-based techniques because they search directly over the space of disturbance trajectories and do not require state information. These techniques therefore carry the same strengths and weakness as optimization techniques. Likewise, importance sampling techniques that involve the use of environment state have the same benefits and limitation as the corresponding path planning or RL algorithm.

3.4 Coverage Metrics

A natural question that arises in safety validation is “when is testing complete?” There are an infinite number of possible tests, so measures of *coverage* are used as a principled way of determining when sufficient testing has been performed to declare a system safe. We

overview three types of testing coverage: the probability of failure, coverage of the space of disturbance trajectories, and coverage of the reachable set of states.

3.4.1 PROBABILISTIC COVERAGE.

Testing of a safety critical system may be complete when the estimated probability of failure is below a specified threshold with high confidence. The outcome of each simulation of a disturbance trajectory is a Bernoulli random variable,

$$\mathbb{1} \{f(\mathbf{x}) \notin \psi\} \tag{15}$$

where a positive outcome (failure) occurs with probability P_{fail} . These samples can be used with the statistical tools of hypothesis testing and estimation to determine when sufficient testing has been performed.

In hypothesis testing, two complimentary hypotheses are compared based on the evidence of sample trajectories. For example, the null hypothesis H_0 may be that the probability of failure is less than an acceptable threshold p_0 , $H_0 := P_{\text{fail}} < p_0$, and the alternative hypothesis is the complement $H_1 := P_{\text{fail}} \geq p_0$. The level of confidence in a hypothesis test is specified by the probability of type I and type II errors (with lower probability of error requiring more data). For an overview of hypothesis testing techniques for statistical model checking see the work of Agha and Palmkog (2018).

Estimation techniques compute a confidence interval over the probability of failure. The frequentist approach involves taking the mean of N independent samples to produce an estimate of the probability of failure

$$\hat{P}_{\text{fail}} = \frac{1}{N} \sum_{i=1}^N \mathbb{1} \{f(\mathbf{x}_i) \notin \psi\}, \tag{16}$$

and uses a concentration inequality to estimate the bounds. Suppose after N samples no failures have been observed, then Hoeffding’s inequality states that

$$\Pr [P_{\text{fail}} \geq p_0] \leq 2e^{-2Np_0^2}. \tag{17}$$

The number of samples N can be chosen so the right-hand side of the inequality is sufficiently small. In contrast, the Bayesian approach involves maintaining a belief over the probability of failure that gets updated with each new sample. For Bernoulli random variables, the Beta distribution can be used to represent the belief. Benefits of the Bayesian approach include the ability to incorporate a prior estimate of the probability of failure and the ability to explicitly compute confidence bounds.

When validating safety-critical systems, failures are often rare and safety thresholds are strict, so many samples are required to make probabilistic arguments for safety. To alleviate this burden, importance sampling approaches are used (see section 7 for details), where samples are drawn from a proposal distribution $q(\mathbf{x})$, under which failures are more likely. The samples are then appropriately weighted when performing hypothesis tests (Harrison, 2012) or estimation.

A poorly chosen, or *inefficient*, proposal distribution can increase the variance of an estimator, which in turn reduces the confidence in the estimate. An inefficient proposal

distribution can be identified when the importance weights are large, or equivalently, when the effective sample size is small compared to the actual number of samples. When samples with large weights are being drawn, Y. Kim and Kochenderfer (2016) suggest limiting the maximum weight by clipping the proposal distribution in regions with large weights. Uesato et al. (2019) suggest combining the importance sampling estimator with a basic Monte Carlo estimator to minimize the downside of a bad proposal distribution and Neufeld et al. (2014) provide a principled way of choosing the best estimator from several possibilities.

3.4.2 DISTURBANCE-SPACE COVERAGE.

If failure events have extremely low probability, or if no probability model is available, coverage can be measured by how well the sampled disturbance trajectories fill the space of possible trajectories. Let $X \times T$ be the space of disturbance trajectories, and let V be a set of sampled trajectories. Informally, the coverage $C(V) \in [0, 1]$ is the degree to which V represent the space $X \times T$. Testing can be concluded when the coverage reaches close to 1. Although a variety of coverage metrics can be employed, *dispersion* (Esposito et al., 2004) and *star discrepancy* (Dang et al., 2008; Dreossi et al., 2015; Nahhal & Dang, 2007) have previously been used for safety validation.

For a distance metric between states d , the dispersion is the radius of the largest ball in $X \times T$ that contains no points in V , i.e. $\max_{\mathbf{x} \in X \times T} \min_{\mathbf{x}_i \in V} d(\mathbf{x}, \mathbf{x}_i)$. Dispersion is challenging to compute in many dimensions and can be overly conservative (Esposito et al., 2004). Instead, a coverage metric based on average dispersion can be computed on a grid with n points and a spacing of δ as

$$C_{\text{disp}}(V) = 1 - \frac{1}{\delta} \sum_{j=1}^n \frac{\min(d_j(V), \delta)}{n}, \tag{18}$$

where $d_j(V)$ is the shortest distance from grid point j to any node in V according to the metric d (Esposito et al., 2004). The value $\min(d_j(V), \delta)$ is therefore the radius of the largest ball that can sit at grid point j and not touch any nodes in the tree or another grid point. Note that $C_{\text{disp}} \rightarrow 1$ when V contains points at each grid point. The coverage is therefore related to the fidelity of the grid, with a finer grid giving a more comprehensive value at greater computational expense.

The star discrepancy coverage metric measures how evenly a set of points is distributed. The discrepancy D of a set over a region $B \subseteq X \times T$ is

$$D(V, B) = \frac{|V \cap B|}{|V|} - \frac{\text{vol}(B)}{\text{vol}(X \times T)}, \tag{19}$$

where vol is the volume of a region and $|V|$ is the number of sample points. The first term is the fraction of the points that are in B and the second term is the ratio of the volume of B to the volume of space of disturbance trajectories. The star discrepancy D^* is the largest discrepancy over all possible subregions of S

$$D^*(V) = \max_B D(V, B). \tag{20}$$

Note that $D^* \rightarrow 0$ when all subregions of S have their “fair share” of the sample points based on volume, and $D^* \rightarrow 1$ when all the sample points overlap. To turn discrepancy

into a coverage metric, define

$$C_{\text{disc}}(V, B) = 1 - D(V, B) \tag{21}$$

and

$$C_{\text{star}}(V) = 1 - D^*(V). \tag{22}$$

It is possible to approximate $D^*(V)$ using a finite partitioning of the state space into regions $\{B_1, \dots, B_n\}$ such that $\cup_{i=1}^n B_i = X \times T$. The discrepancy coverage can be computed for each box $C_{\text{disc}}(V \cap B_i, B_i)$ (Dreossi et al., 2015) or a lower and upper bound of the star discrepancy coverage can be computed based on B_i (Nahhal & Dang, 2007; Thiémond, 2001).

3.4.3 STATE-SPACE COVERAGE.

When the space of disturbance trajectories is prohibitively large (e.g. for long time horizon problems), it can be more efficient to define coverage metrics for the state space. In fact, both dispersion and star-discrepancy were first used as state-space coverage metrics for the rapidly exploring random tree algorithm (Dang et al., 2008; Dreossi et al., 2015; Esposito et al., 2004; Nahhal & Dang, 2007).

The challenge to using state-space coverage metrics is that not every state in the state space may be reachable. A state s is *reachable* if there is a disturbance trajectory that can be applied that will cause the environment to reach s . In many safety validation problems, the reachable states are a subset of the full state space due to the limited control via disturbances (Esposito et al., 2004). If the reachable set is small compared to S , then the maximum possible coverage value will also be small, and testing will never terminate. To mitigate this problem, Esposito et al. (2004) proposed a growth metric on the set of samples defined as

$$g(V) = \Delta C(V) / \Delta |V|. \tag{23}$$

The growth g measures how much the coverage metric increases with the increase in sample points in V . In addition to stopping based on adequate coverage, testing can be terminated if the growth is below a specified threshold, suggesting that adding more sample points to V will not improve the coverage.

4. Black-Box Optimization

This section surveys algorithms that use black-box optimization techniques to solve safety validation problems, as well as the modifications that the techniques require to be effective for safety validation. Safety validation has been performed with simulated annealing, evolutionary algorithms, Bayesian optimization, and extended ant colony optimization.

4.1 Simulated Annealing

An approach to stochastic global optimization known as simulated annealing uses a random walk around the disturbance space to minimize a cost function c . A temperature parameter β is used to control the amount of stochasticity in the method over time and a transition function $P(\mathbf{x}' | \mathbf{x})$ describes the probability distribution over the next disturbance trajectory

\mathbf{x}' . If the new trajectory is a counterexample, then it is returned, otherwise it is adopted as \mathbf{x} with probability $\exp(-\beta(c(\mathbf{x}') - c(\mathbf{x})))$. Due to the stochastic nature of simulated annealing, it is often a suitable algorithm for solving a global optimization problem with many local minima and can therefore be effective for safety validation (Abbas et al., 2013; Aerts et al., 2018).

A common choice for transition function is to use a Gaussian distribution around the current point \mathbf{x} with a standard deviation that is adjusted based on the ratio of accepted points (Kochenderfer & Wheeler, 2019). This approach may not work well when the disturbance space has constraints that must be satisfied, such as lower and upper bounds on the possible disturbances (Abbas et al., 2013). Abbas et al. (2013) proposes the use of a *hit and run* approach to transitioning that respects constraints. It follows three steps:

1. Sample a random direction \mathbf{d} in the disturbance trajectory space.
2. Perform a line search in the direction of \mathbf{d} to determine the range of α such that $\mathbf{x} + \alpha\mathbf{d}$ does not violate any constraints.
3. Sample α from this range according to a chosen distribution. The standard deviation of this distribution can be adjusted using the acceptance ratio to improve convergence.

Aerts et al. (2018) improved the hit and run scheme by suggesting that α be chosen for each disturbance dimension separately so that highly constrained dimensions do not restrict the step size of less constrained dimensions (Aerts et al., 2018).

Typically, the size of \mathbf{x} remains fixed throughout the optimization, meaning the temporal discretization of the disturbance trajectory is never adjusted. Aerts et al. (2018) note that the frequency content of the disturbance trajectory is salient for some falsification problems, and therefore the temporal discretization of the disturbance trajectory should itself be optimized. An approach called input-signal-space optimization uses a two-layered approach, where an outer loop uses SA to select the length of the disturbance trajectory, and an inner loop finds the lowest cost trajectory for that length. This approach is able to adapt the fidelity of the time domain, getting improved results for some falsification problems (Aerts et al., 2018).

4.2 Evolutionary Algorithms

Evolutionary algorithms approach global optimization by mimicking the biological process of evolution. A population of individuals is sampled and then evolved using the processes of *crossover*, where low cost individuals are combined to form new individuals, and *mutation*, where individuals are randomly modified to incorporate new traits into the population.

Q. Zhao et al. (2003) applied a genetic algorithm to generate test cases for embedded control systems. Each individual is represented by a real-valued vector for all continuous variables and a binary encoding for all discrete variables. Selection occurs with a probability inversely related to the cost function. The crossover between two individuals is done using either a randomized concatenation of subsequences or an arithmetic combination of the individuals. Although evolutionary algorithms lack convergence guarantees and are largely based on heuristics, they can be effective for problems with very large input spaces, as demonstrated by their successful use in the safety validation of multi-agent sense and avoid algorithms (Zou et al., 2014).

Evolutionary algorithms can optimize more complex inputs such as temporal logic expressions represented as trees. Corso and Kochenderfer (2020) hypothesize that counterexamples can often be described by a simple temporal logic property of the disturbance trajectory, called a *failure description*. Failure descriptions lend interpretability to automated testing and have been shown to be effective at finding counterexamples as well as giving engineering insight into the failure modes of the system. Genetic programming is used to optimize failure descriptions φ that satisfy

$$\mathbf{x} \in \varphi \implies c(\mathbf{x}) < \epsilon. \tag{24}$$

The cost function for a given failure description is the average cost of trajectories that satisfy it. Sampling satisfying trajectories is, in general, a hard problem, but Corso and Kochenderfer (2020) provide an algorithm to do so under a set of assumptions.

4.3 Bayesian Optimization

In Bayesian optimization (Mockus, 2012), a surrogate model (such as a Gaussian process) is used to represent the cost function over the space of disturbance trajectories. The model is used to select disturbance trajectories that are likely to lower the cost function. Maintaining a surrogate model can be beneficial when evaluations of $c(\mathbf{x})$ are costly or when the cost is stochastic. For these reasons, Bayesian optimization is a natural choice for safety validation (Abeyirigoonawardena et al., 2019; Akazaki et al., 2017; Deshmukh et al., 2017; Mullins et al., 2018; Silveti et al., 2017; X. Yang et al., 2020).

Bayesian optimization iterates over two steps: 1) updating the surrogate model with new evaluations, and 2) choosing the next disturbance trajectory to evaluate. The update step computes a posterior distribution given the new evaluations (the details depend upon the employed surrogate model). Then the next disturbance trajectory is chosen using an inner optimization loop that optimizes a metric such as prediction-based exploration, error-based exploration, lower confidence bound exploration, probability of improvement, or expected improvement (see Kochenderfer and Wheeler (2019) for details). The inner optimization can be performed using simulated annealing (Akazaki et al., 2017) or a sampling plan with good coverage of the space of disturbance trajectories (Silveti et al., 2017).

One drawback to using Gaussian processes is their inability to scale to large dimensions and many sample points. To improve scalability for safety validation, Mathesen et al. (2019) introduced an algorithm called stochastic optimization with adaptive restarts (SOAR) that uses a two-layered Bayesian optimization approach to trade off between global and local search. A Gaussian process model is maintained over the global search space, and a stochastic method is used to select regions for further exploration. Local Gaussian process models are used for refined searching of local regions. When a local region is no longer seeing improvement, a new region is selected. Deshmukh et al. (2017) addressed scaling to large dimensions using a dimensionality reduction technique based on random embeddings (Wang et al., 2016).

4.4 Extended Ant-Colony Optimization

Ant colony optimization is a probabilistic technique initially used for finding optimal paths through graphs (Dorigo et al., 1996). Ant colony optimization was extended to continuous

spaces by G. E. Fainekos and Giannakoglou (2003) and later applied to falsification (Y. S. R. Annapureddy & Fainekos, 2010). Extended ant-colony optimization works by treating each disturbance trajectory \mathbf{x} as a path through a graph with edges between adjacent temporal points (x_t, x_{t+1}) . At each time t , the space of disturbances is discretized into N equally spaced cells. Ants traverse the graph by selecting their next cell i at time t based on the amount of pheromone present, where high pheromone implies low cost. After an ant selects their next cell, it then selects a point uniformly at random inside that cell and moves to it. At the end of each trajectory, pheromone is deposited at each cell for low-cost trajectories.

5. Path Planning

This section surveys safety validation algorithms based on planning algorithms (and their required modifications). The presented algorithms are variants of the rapidly exploring random tree algorithm, multiple shooting methods, and Las Vegas tree search.

5.1 Rapidly Exploring Random Tree

Rapidly-exploring random tree (RRT) is a path planning technique for efficiently finding failure trajectories (Lavalle, 1998). A tree is iteratively constructed by sampling the state space and growing in the direction of unexplored regions. RRT has been applied to the safety validation of black-box systems (Branicky et al., 2006; Dang et al., 2008; Dreossi et al., 2015; Esposito et al., 2004; J. Kim et al., 2005; Koschi et al., 2019; Nahhal & Dang, 2007; Plaku et al., 2009; Tuncali & Fainekos, 2019).

Algorithm 1 Rapidly-exploring random tree.

```

1: function RRT( $s_0, S_{\text{fail}}$ )
2:    $T \leftarrow \text{INITIALIZETREE}(s_0)$ 
3:   loop
4:      $s_{\text{goal}} \leftarrow \text{SAMPLESTATE}()$ 
5:      $s_{\text{near}} \leftarrow \text{NEARESTNEIGHBOR}(T, s_{\text{goal}})$ 
6:      $x_{\text{new}} \leftarrow \text{GETDISTURBANCE}(s_{\text{near}}, s_{\text{goal}})$ 
7:      $s_{\text{new}} \leftarrow f(s_{\text{near}}, x_{\text{new}})$ 
8:      $\text{ADDDNODE}(T, s_{\text{near}}, s_{\text{new}})$ 
9:   return  $\text{COUNTEREXAMPLES}(T, S_{\text{fail}})$ 

```

The basic approach is presented in algorithm 1. On each iteration, a random point s_{goal} in the state space is generated, which acts as the goal state for the next node to be added (line 4). The tree is searched for the node s_{near} that is closest to the goal state (line 5) based on some distance metric d . This node will act as the starting point when attempting to reach the goal. A disturbance x_{new} is generated that drives s_{near} toward s_{goal} (line 6). Since the system is a black-box, an x_{new} that causes $s_{\text{near}} = s_{\text{goal}}$ can only be approximated through random sampling or a more advanced optimization procedure. Lastly, the disturbance x_{new} is simulated, starting from s_{near} , resulting in a new state s_{new} (line 7) which is then added to the tree as a child of s_{near} (line 8). Note that if the simulator cannot be initialized to any state, then the trace can be simulated by starting at the root

and simulating the disturbances through the branch containing s_{near} . The algorithm stops when the maximum number of iterations is reached, a suitable falsifying trajectory is found, or tree coverage reaches a specified threshold. Variants of RRT (Branicky et al., 2006; Dang et al., 2008; Dreossi et al., 2015; Esposito et al., 2004; J. Kim et al., 2005; Koschi et al., 2019; Nahhal & Dang, 2007; Tuncali & Fainekos, 2019) typically differ in their approach to state space sampling, choice of distance metric for nearest neighbor selection, or by adding additional steps that reconfigure the tree for improved performance.

5.1.1 ADAPTIVE SAMPLING.

When the reachable set is only a subset of the state space, a uniform sampling of goal states may be inefficient, leading to slow tree growth and low coverage values. There are several techniques for biasing goal samples toward the reachable set.

An approach known as guided-RRT (g-RRT) (Nahhal & Dang, 2007) biases the selection of s_{goal} to regions with low coverage. In g-RRT, the state space is segmented into n regions $\{B_1, \dots, B_n\}$ and each region is assigned a weight $w(B_i)$. Regions are sampled according to

$$P(B_{\text{goal}} = B_i) = \frac{w(B_i)}{\sum_j w(B_j)}, \quad (25)$$

and then s_{goal} is sampled uniformly from B_{goal} . In the original version of g-RRT (Dang et al., 2008; Nahhal & Dang, 2007), the weight $w(B_i)$ is related to the increase in the lower and upper bounds of C_{star} when a new point is added to B_i . In later work (Dreossi et al., 2015), the weight is $w(B_i) = \sigma(D(T, B_i))$, where σ is the sigmoid function. The goal in both cases is to sample points that increase the coverage of the tree.

J. Kim et al. (2005) maintain a distribution over s_{goal} that has a mean biased toward low values of the cost function and a standard deviation that adaptively changes to maximize sampling in the reachable set. As the algorithm progresses, they keep track of the ratio β of successful expansions of the tree. A successful expansion is one where

$$d(s_{\text{new}}, s_{\text{goal}}) > d(s_{\text{near}}, s_{\text{goal}}), \quad (26)$$

meaning the tree was able to grow in the direction of s_{goal} . The ratio β is updated after a user-specified number of iterations and is used to update the standard deviation between bounds $[\sigma_{\text{min}}, \sigma_{\text{max}}]$ as

$$\sigma = (1 - \beta)(\sigma_{\text{max}} - \sigma_{\text{min}}) + \sigma_{\text{min}}. \quad (27)$$

Thus, when the number of successful expansions is large, the standard deviation is reduced to continue sampling in that region, but when the tree frequently cannot grow toward s_{goal} the range of values is increased to better search for the reachable set.

5.1.2 NEIGHBOR SELECTION.

The choice of distance metric for finding the nearest neighbor s_{near} is critical for the performance of RRT. In some applications, it is acceptable to minimize the Euclidean distance between s_{near} and s_{goal} (Koschi et al., 2019), but this approach may be insufficient for some problems. First, Euclidean distance does not take into account whether s_{goal} can be reached from s_{new} based on the dynamics of the system. Secondly, if the reachable set is only a small

subset of the state space, then points on the boundary of the reachable set will frequently be chosen as s_{near} , limiting the coverage of the tree. Lastly, if there is a cost associated with each trajectory, it should be considered when picking which node to expand.

To encode reachability in the selection of node neighbors, J. Kim et al. (2005) use an estimate of the time to go from s_{near} to s_{goal} defined by

$$t(s_{\text{near}}, s_{\text{goal}}) = \begin{cases} d(s_{\text{near}}, s_{\text{goal}})/v & \text{if } v > 0 \\ \infty & \text{otherwise} \end{cases}, \quad (28)$$

where the velocity is

$$v = \max_{x \in X} \left(- \frac{\partial d(s, s_{\text{goal}})}{\partial s} f(s_{\text{near}}, x) \Big|_{s=s_{\text{near}}} \right). \quad (29)$$

To compute the derivative $\partial d(s, s_{\text{goal}})/\partial s$ exactly requires white-box knowledge of the system, but it could be estimated with domain knowledge or from simulations (J. Kim et al., 2005).

Another modification to neighbor selection is known as history-based selection, which penalizes the selection of nodes that fail to expand (J. Kim et al., 2005). A node fails to expand if it is selected as s_{near} and creates an output s_{new} that is already in the tree. Failure to expand typically occurs when s_{near} is on the boundary of the reachable set. The number of times a node has failed to expand is stored as n_f and the distance between nodes is modified as

$$d_h(s_{\text{near}}, s_j) = d(s_{\text{near}}, s_j) + \lambda n_f \quad (30)$$

for any choice of distance metric d and a user specified constant $\lambda > 0$. J. Kim et al. (2005) chose λ to balance the contribution of d and n_f .

To incorporate a state-dependent cost function $c(s)$ into the RRT algorithm, Karaman and Frazzoli (2011) developed the approach known as RRT*. In RRT*, the neighbor selection is done by finding a set S_{near} of nodes in an ϵ -radius of s_{goal} and then selecting the node in S_{near} that has the lowest cost, i.e.

$$s_{\text{near}} = \arg \min_{s_i \in S_{\text{near}}} c(s_i) \quad (31)$$

where

$$S_{\text{near}} = \{s \mid d(s, s_{\text{goal}}) < \epsilon\}. \quad (32)$$

For the cost function, Dreossi et al. (2015) use the partial robustness of the system specification while Tuncali and Fainekos (2019) use a heuristic cost for autonomous driving.

5.1.3 OTHER RRT VARIANTS.

In the basic version of RRT, a single tree is maintained and grown until termination, but under some circumstances multiple trees may be used. When searching over a space of static parameters θ , a different tree can be grown for each choice of parameter in an approach called rapidly exploring forest of trees (RRFT) (Esposito et al., 2004). In RRFT, any tree that has reached a threshold coverage value or has stopped growing will be terminated, and

a new tree (with a new value of θ) is added to the forest. This process continues until a counterexample is found or until the parameter space has been adequately covered with fully grown trees. Another example of maintaining multiple trees is the approach of Koschi et al. (2019) where a fixed number of nodes K are added at each iteration, including the first (so K trees are maintained). Each new node is added to the tree that has the closest node. Trees that are not being grown can be removed for memory efficiency (Koschi et al., 2019).

Tuncali and Fainekos (2019) combined stochastic global optimization techniques with RRT by including a transition test (similar to simulated annealing) and a similarity test for the addition of new nodes. New nodes are only added if they pass the transition test (based on their cost function) and if they are sufficiently different from all of the other nodes in the tree. These two techniques enhance exploration and coverage of the state space.

Koschi et al. (2019) introduced the backwards algorithm for RRT which connects nodes in the tree backward in time without breaking the black-box assumption. The algorithm starts by sampling a state s_0 in the failure set. At each iteration, a state is randomly sampled as s_{goal} , and the nearest neighbor s_{near} in the tree is determined. A disturbance is selected that drives s_{goal} toward s_{near} (notice this is opposite from the basic algorithm), and a new state s_{new} is generated by simulating $f(s_{\text{goal}}, x_{\text{new}})$. To maintain continuity in the tree, s_{new} replaces s_{near} and the entire branch connecting s_{near} to s_0 is re-simulated with the same disturbances. If the resulting state (the new s_0) is no longer in the failure set, then the branch is terminated and the algorithm repeats. This approach, although more computationally expensive, showed improved performance for finding rare failure events of an adaptive cruise control system (Koschi et al., 2019). Note, however, that unlike the basic RRT algorithm, this approach requires a simulator that can be initialized based only on the observed state of the environment.

5.2 Multiple Shooting Methods

Multiple shooting methods (Bock & Plitt, 1984) solve non-linear initial value problems and have been used for robotic motion planning (Diehl et al., 2006) and falsification (Zutshi et al., 2014; Zutshi et al., 2013). The idea is to sample trajectory segments using random initial states and random disturbances. A shortest path problem is solved to find a *candidate trajectory* through these segments that connects initial states and failure states, where two segments are connected if the terminal state of one segment is sufficiently close to the starting state of another. The discovered path may not be feasible due to the gaps between segments, so an optimization routine is used to find disturbances that minimize those gaps and find an actual failure trajectory. In most multiple-shooting algorithms, the procedure to connect trajectory segments involves using white-box information such as dynamical equations or gradients (Zutshi et al., 2013), and therefore multiple-shooting algorithms are not in general applicable to the black-box setting.

Zutshi et al. (2014) introduces two ideas that make multiple shooting methods applicable for black-box falsification in large state spaces (Zutshi et al., 2014):

1. The state space should only be explored around the reachable set.
2. State space discretization and refinement can be used to connect segments.

To achieve this, the state space is implicitly discretized into disjoint cells C , each of size δ . The cells are not explicitly stored, but they are constructed so it is efficient to find which cell contains a given state. One efficient representation is a Cartesian grid with a fixed cell size. The reachable set is estimated by running simulations forward in time with random disturbances and recording which cells are connected together. The cell connections form a graph that can be searched for candidate falsifying trajectories. Cells in these trajectories are refined until an actual counterexample is found. Since the algorithm relies on a graph search over a discretized grid, it may be useful to provide a set of initial states S_0 to have more than one starting cell.

Algorithm 2 Black-box multiple shooting method.

```

1: function MULTIPLESHOOTING( $S_0, S_{\text{fail}}, \delta, \gamma, N_{\text{max}}, K$ )
2:    $T \leftarrow \emptyset$ 
3:   loop
4:      $G \leftarrow \text{SAMPLESEGMENTS}(S_0, T, \delta, N_{\text{max}}, K)$ 
5:      $T \leftarrow \text{FINDCANDIDATETRAJECTORIES}(G, S_0, S_{\text{fail}})$ 
6:     if  $T = \emptyset$ 
7:       return  $\emptyset$ 
8:     if  $\text{ACTUALTRAJECTORIES}(T) \neq \emptyset$ 
9:       return  $\text{ACTUALTRAJECTORIES}(T)$ 
10:     $\delta \leftarrow \gamma\delta$ 
11: function SAMPLESEGMENTS( $S_0, T, \delta, N_{\text{max}}, K$ )
12:    $G \leftarrow \emptyset$ 
13:    $Q \leftarrow \text{SAMPLECELLS}(S_0, \delta)$ 
14:   while  $Q \neq \emptyset$  and  $|G| < N_{\text{max}}$ 
15:      $C \leftarrow \text{POP}(Q)$ 
16:     Sample  $\{s_1, \dots, s_K\}$  uniformly from  $C$ 
17:     Sample  $\{x_1, \dots, x_K\}$  from  $p(x)$ 
18:     for  $i \in 1 : K$ 
19:        $s_{\text{new}} = f(s_i, x_i)$ 
20:        $C_{\text{new}} \leftarrow \text{FINDCELL}(s_{\text{new}}, \delta)$ 
21:       if  $C_{\text{new}} \in T$  or  $T = \emptyset$ 
22:          $G \leftarrow G \cup (C, C_{\text{new}}, x_i)$ 
   return  $G$ 

```

The approach is presented in algorithm 2. It takes as input a set of initial states S_0 and a set of failure states S_{fail} , a discrete cell size δ , a refinement factor γ , a maximum number of segments per iteration N_{max} , and the number of samples per cell K . A set of candidate trajectories T is initialized to the empty set (line 2). On each iteration, a graph G is constructed with edges (C_i, C_j, x_{ij}) , where x_{ij} is the disturbance that transformed the system from a state $s_i \in C_i$ to a state $s_j \in C_j$ (line 4). An initially empty graph is constructed as follows:

- Some starting cells are sampled by sampling states in the initial set S_0 , finding the corresponding cell, and adding the cell to a queue Q (line 13).

- On each iteration, a cell C is popped off the queue and K states are sampled uniformly at random from within C (line 16). Then, a disturbance is randomly sampled for each state (line 17).
- Each sampled state s_i is simulated forward one timestep based on the corresponding random disturbance x_i (line 19), resulting in a new state s_{new} in cell C_{new} (line 20).
- To ensure the search focuses on promising regions of the state space, the edge (C, C_{new}, x_i) is only added to the graph if C_{new} is in the set of candidate trajectories T (line 21). On the first iteration, when the set of candidate trajectories is empty, all edges are added.

Once the graph is constructed, it is searched for candidate trajectories that connect cells in the initial set to cells in the failure set (line 5). If no candidate trajectories are found, then the algorithm failed. Each candidate trajectory in T is simulated from its starting state using the disturbance applied at each segment to get an actual trajectory (line 8). The actual trajectory will not match that candidate trajectory, since the candidate trajectory is made up of disjoint segments. If the actual trajectory is a counterexample, then return it, otherwise, reduce the grid size by a factor of γ (line 10) and repeat the procedure.

5.3 Las Vegas Tree Search

Las Vegas tree search (LVTS) (Ernst, Sedwards, et al., 2019) is a tree-based falsification algorithm based on two ideas:

1. The disturbance trajectory \mathbf{x} should only be discretized in time as finely as it needs to be.
2. The system is more likely to be falsified by extreme values of the disturbance space than less extreme values.

To achieve the first goal, the disturbance trajectory is constructed piecewise from disturbances of different durations. Longer duration disturbances can be implemented as a repetition of a disturbance. Let the notation $x^{[\ell]}$ indicate that x is applied ℓ times. In LVTS, the set of possible disturbances at each step is discretized and represented by the set Y and the probability of selecting disturbances $x^{[\ell]}$ is $P(x^{[\ell]})$. To achieve the second goal, P is defined to favor disturbances with more extreme. LVTS (algorithm 3) takes as input Y , P , cost function c , and safety threshold ϵ .

The algorithm proceeds by growing a tree of disturbance trajectories \mathbf{x} . Each unfinished disturbance trajectory \mathbf{x} has associated with it a set of explored disturbances (initialized to the empty set in line 2) and a set of unexplored disturbances (initialized to the set of all possible segments Y in line 3). For each iteration of the algorithm, a trajectory is stochastically expanded by sampling a new disturbance from P (line 7) and then concatenating it to the current trajectory (line 8). If the new disturbance has not been explored before, it is removed from the unexplored set (line 10) and a lower and upper bound on the cost is computed from the function `COSTBOUNDS`. If the cost function is the robustness, then the bounds can be computed by the method of Dreossi et al. (2015). If the upper bound is lower than the safety threshold, then a counterexample is found and returned (line 13). If

Algorithm 3 Las Vegas tree search.

```

1: function LVTS( $Y, P, c, \epsilon$ )
2:   explored( $\mathbf{x}$ )  $\leftarrow \emptyset$  for all  $\mathbf{x}$ 
3:   unexplored( $\mathbf{x}$ )  $\leftarrow Y$  for all  $\mathbf{x}$ 
4:   loop
5:      $\mathbf{x} \leftarrow \emptyset$ 
6:     while unexplored( $\mathbf{x}$ )  $\neq \emptyset$  or explored( $\mathbf{x}$ )  $\neq \emptyset$ 
7:       Sample  $x$  from  $P(x)$ 
8:        $\mathbf{x}_{\text{new}} \leftarrow [\mathbf{x} \ x]$ 
9:       if  $x \in \text{unexplored}(\mathbf{x})$ 
10:        unexplored( $\mathbf{x}$ )  $\leftarrow \text{unexplored}(\mathbf{x}) \setminus \{x\}$ 
11:         $\underline{c}, \bar{c} \leftarrow \text{COSTBOUNDS}(f(\mathbf{x}_{\text{new}}))$ 
12:        if  $\bar{c} < \epsilon$ 
13:          return  $\mathbf{x}_{\text{new}}$ 
14:        if  $\underline{c} > \epsilon$ 
15:          break 4
16:        explored( $\mathbf{x}$ )  $\leftarrow \text{explored}(\mathbf{x}) \cup \{x\}$ 
17:       $\mathbf{x} \leftarrow \mathbf{x}_{\text{new}}$ 
    
```

the lower bound is greater than the safety threshold, then a counterexample is not possible for this disturbance trajectory and the algorithm restarts (line 15).

Disturbance segments are grouped by an integer $\ell \in \{1, \dots, \ell_{\max}\}$ that controls the duration of the segment. For duration ℓ , the set Y_ℓ is given by

$$Y_\ell = \{x^{[\ell]} \mid x_i = x_{i,\min} + \alpha_{i,j}(x_{i,\max} - x_{i,\min})\}, \quad (33)$$

where x_i is the i th component of the disturbance space and $x_{i,\min}$ and $x_{i,\max}$ are the minimum and maximum values of that component. The factor

$$\alpha_{i,j} = (2j + 1)/2^{b_i} \quad (34)$$

for any integer $j < (2^{\ell_{\max} - \ell} - 1)/2$ and $b_1 + \dots + b_n = 1$. The construction of these sets ensures that segments with larger values of ℓ have longer duration and more extreme disturbance values. The full set of segments Y is the union of all the Y_ℓ up to a maximum level of refinement ℓ_{\max} :

$$Y = \bigcup_{\ell=1}^{\ell_{\max}} Y_\ell \quad (35)$$

The segment sampling distribution P ensures that disturbances with larger values of ℓ and lower costs are chosen more often. The distribution is constructed implicitly by assigning a weight to each level of refinement

$$w_\ell = \frac{|\text{unexplored}(\mathbf{x}) \cap Y_\ell| + |\text{explored}(\mathbf{x}) \cap Y_\ell|}{2^{\ell_{\max} - \ell} |Y_\ell|} \quad (36)$$

and choosing ℓ with probability $w_\ell / \sum_{k=1}^{\ell_{\max}} w_k$. The weight is constructed to favor lower values of ℓ (due to the exponential factor in the denominator) until the number of unexplored

edges goes to zero without finding many plausible trajectories (note that the explored set only increases if the trajectory remains plausible for falsification). Once the refinement level is selected, one of the following four options is selected uniformly at random:

1. Sample x from $\text{unexplored}(\mathbf{x}) \cup Y_l$.
2. Sample x from $\text{explored}(\mathbf{x}) \cup Y_l$.
3. Choose the x from $\text{explored}(\mathbf{x}) \cup Y_l$ that minimizes the cost of $[\mathbf{x} x]$.
4. Choose the x from $\text{explored}(\mathbf{x}) \cup Y_l$ that minimizes the cost of $[\mathbf{x} x x']$ for all $x' \in \text{explored}([\mathbf{x} x])$.

Option 1 ensures exploration over the unexplored set, while options 2–4 increasingly exploit knowledge of trajectories with low costs.

6. Reinforcement Learning

The algorithms presented in this section are based on techniques from reinforcement learning. We present variants of Monte Carlo tree search and deep reinforcement learning.

6.1 Monte Carlo Tree Search

Monte Carlo tree search (MCTS) is an online planning algorithm for sequential decision making that has seen success for long time-horizon problems (Silver et al., 2016), making it useful for safety validation (Delmas et al., 2019; Julian et al., 2020; Lee et al., 2020; Moss et al., 2020; Wicker et al., 2018; Zhang et al., 2018). Monte Carlo tree search uses online planning to determine the best actions to take from a starting state to maximize reward. A search tree is maintained over all of the paths tried as well as an estimate of the value function at each step. Each iteration in MCTS consists of four steps:

1. *Selection*: Starting from the root of the search tree, disturbances are selected sequentially until reaching a leaf node, with the goal of choosing disturbances that are more likely to lead to high reward.
2. *Expansion*: Once the algorithm arrives at a leaf node, a new disturbance is chosen, and a node is added to the tree with zero visits.
3. *Rollout*: The value of the new node is estimated by running a simulation from the new node while choosing disturbances according to a rollout policy until the episode terminates or the finite planning horizon is reached.
4. *Backpropagation*: The value of the new node is used to update the value of all of its ancestors in the tree.

These four steps are repeated until a stopping criterion is met such as computational budget, wall-clock time limit, or a threshold for the reward of the best solution found so far. At that point, the best trace can be returned, or, for longer time-horizon problems, the best candidate disturbance is selected and the process restarts, possibly retaining the subtree associated with the selected disturbance.

In problems where there is a discrete state or observation space, it makes sense to maintain separate state and disturbance nodes so if a state is repeated, the corresponding node in the tree can be reused, increasing the ability of the algorithm to reuse prior information. In most safety validation problems, however, the state space of the simulator will be continuous (or unavailable to the algorithm entirely), in which case the same state will rarely be sampled twice. To save memory, it is common to only include disturbance nodes in the search tree, which is equivalent to setting the state equal to the concatenation of all disturbances that led to it ($s_t = \mathbf{x}_{1:t}$).

The discrete nature of nodes in the search tree are incompatible with a continuous disturbance space X . Delmas et al. (2019) discretize the disturbance space into a small number of discrete disturbances that are representative of the continuous space. As an alternative to discretization, when adding a new node to the tree, Lee et al. (2020) sample a new disturbance x_t from a known distribution over disturbances by selecting a random seed uniformly at random and using it to produce disturbances from $p(x)$.

Zhang et al. (2018) combined MCTS with global optimization, using MCTS for exploration of the disturbance trajectory space and global optimization for refinement of the disturbance trajectories. The disturbance space is first discretized into $L_1 \times \dots \times L_n$ equal-sized regions. Each node in the tree represents one of the regions in the disturbance space denoted B_t . When a new node is added to the tree at depth d , its value is estimated by solving the following constrained optimization problem:

$$\begin{aligned} \max_{\mathbf{x}} \quad & \mathbb{E} \left[\sum_t \gamma^t R(s_t, x_t) \right] \\ \text{s.t.} \quad & x_1 \in B_1, \quad \dots, \quad x_d \in B_d. \end{aligned} \tag{37}$$

This approach can be combined with progressive widening (Chaslot et al., 2008; Coulom, 2007) when the disturbance space is large. Instead of randomly sampling a new region, optimization can be used to find the optimal region to expand. When adding a new disturbance at depth $d + 1$, solve eq. (37) with the added constraint that \mathbf{x}_{d+1} exists in the set of regions that have yet to be expanded. For ease of optimization, that subset may be further restricted to its convex subset (Zhang et al., 2018). Once the MCTS budget is spent, traces with high reward can be found by solving the constrained optimization problem of the highest performing branch of the tree.

The use of optimization for each new node of the search tree increases the computational cost of the algorithm. Zhang et al. (2018) note that balancing the computational budget between optimization iterations and tree expansions is critical to the success of this approach. They compare simulated annealing (SA) (Abbas et al., 2013), Globalized Nelder-Mead (GNM) (Luersen & Le Riche, 2004), and covariance matrix adaptation evolution strategy (CMA-ES) (Hansen & Ostermeier, 1996) for the global optimization algorithms. It is noted that since CMA-ES has a built-in exploration strategy, it sees less improvement when combined for MCTS than SA and GNM. GNM, on the other hand, does not have an exploration strategy (beyond probabilistic restarts) and therefore sees a significant improvement with MCTS.

6.2 Deep Reinforcement Learning

Deep reinforcement learning (DRL) is a category of reinforcement learning that uses deep neural networks to represent the value function $V(s)$, the state-action value function $Q(s, x)$, or the policy $\pi(s)$. DRL has shown state-of-the-art results in playing Atari games (Mnih et al., 2015), playing chess (Silver et al., 2017), and robot manipulation from camera input (Gu et al., 2017). In recent years, different DRL techniques have been applied to falsification and most-likely failure analysis (Akazaki et al., 2018; Behzadan & Munir, 2019; Corso et al., 2019; Koren et al., 2018; Koren & Kochenderfer, 2019, 2020; Kuutti et al., 2020; Qin et al., 2019).

DRL algorithms are broadly split between value-function approaches, where the neural network is used to represent the value function, and policy search approaches, where the neural network represents the policy. There are advantages and disadvantages to both, and several algorithms are discussed below. The reader is referred to the original papers for implementation details.

If the disturbance space X is discrete, then the deep Q-network (DQN) (Mnih et al., 2015) algorithm can be used for falsification (Akazaki et al., 2018; Qin et al., 2019). In DQN, the optimal Q -function is estimated by a neural network that takes as input the state s and outputs a value for each discrete disturbance. Disturbances are selected greedily as

$$x = \arg \max_x Q(s, x) \tag{38}$$

with probability $1 - \epsilon$ and chosen at random with probability ϵ to encourage exploration. The parameters of the Q -network are updated to minimize the mean squared error between the current value and a target

$$Q_{\text{target}}(s, x) = r(s, x) + \gamma \max_{x'} Q(s', x'). \tag{39}$$

Since the target value greedily selects the highest value disturbance, the Q -network can be trained on any sample of a state, disturbance, and reward (a feature known as off-policy learning). Many implementations of DQN therefore maintain a replay buffer that stores previous (s, x, r, s') tuples which are repeatedly used to update the network parameters. This reuse of data makes DQN a relatively sample-efficient algorithm. The main drawback of DQN is that it is incompatible with large or continuous disturbance spaces.

For large or continuous disturbance spaces, the policy itself is represented by a neural network (with parameters θ) that takes the state as input and either outputs a disturbance directly (e.g. $x = \pi_\theta(s)$) or outputs parameters of a distribution from which a disturbance can be sampled (e.g. for a normal distribution $[\mu, \sigma^2] = \pi_\theta(s)$ and $x \sim \mathcal{N}(\mu, \sigma^2)$). The policy is optimized to produce higher rewards using the policy gradient method (Sutton et al., 2000). Policy gradient methods can suffer from high variance and can be unstable during optimization. To improve optimization stability, an approach known as trust region policy optimization (TRPO) (Schulman et al., 2015) restricts the amount a policy can change at each step. TRPO has previously been used for falsification of autonomous vehicles (Corso et al., 2019; Koren et al., 2018; Koren & Kochenderfer, 2019).

Another drawback of policy gradient methods is their inability to learn off-policy. Without data reuse, these methods can require a large number of simulations to converge. Newer

approaches combine policy gradient methods with value function methods to create the actor-critic paradigm, which can perform well on problems with continuous disturbance spaces while also using previous simulation data to improve sample efficiency. Actor-critic methods use two neural networks, one for the policy (the actor network) and one for the value function (the critic network) and come in several varieties. Advantage actor critic (A2C) was used for falsification by Kuutti et al. (2020). Its more scalable counterpart, asynchronous advantage actor critic (A3C), was used by Akazaki et al. (2018). Behzadan and Munir (2019) use another actor-critic method known as deep deterministic policy gradient (DDPG) combined with Ornstein-Uhlenbeck exploration (Lillicrap et al., 2016).

One potential drawback of using DRL for black-box falsification is the requirement for a simulator that can be observed after each disturbance is applied. In practice, high-fidelity simulators may be large, complex software projects, so it may be difficult to access the true simulator state at each timestep, whereas getting the results or only some partial data on the final state may be easier. Koren and Kochenderfer (2019) developed a DRL technique (also used by Corso et al. (2019)) that does not require access to the simulator state. The technique uses a recurrent neural network (RNN) with long-short term memory (LSTM) layers as the policy (Hochreiter & Schmidhuber, 1997). The RNN maintains a hidden state akin to the state of the simulator that is used to make future decisions. The input to each layer is the previous disturbance and the initial state of the simulation, allowing the approach to generalize across initial conditions.

The advantages of DRL and tree search methods can be combined in an approach called go-explore (GE) (Ecoffet et al., 2019), which has been effective for hard-exploration problems with long time horizons and no reward shaping. GE has two phases, a tree search exploration phase, and a DRL robustification phase. While the original version of GE uses the state of the simulator when building the tree and training the robust policy, Koren and Kochenderfer (2020) modified the algorithm to use the history of disturbances instead, reducing the access requirements of the simulator.

7. Importance Sampling Algorithms

This section summarizes algorithms for estimating the probability of failure based on importance sampling. Algorithms include the cross-entropy method, multilevel splitting, classification-based importance sampling, and state-dependent importance sampling.

7.1 Cross-Entropy Method

The cross-entropy method iteratively learns the optimal importance sampling distribution from a family of distributions $q(\mathbf{x}; \theta)$ parameterized by θ (see De Boer et al. (2005) for an overview and Rubinstein and Kroese (2013) for a deeper examination). The optimal distribution parameters θ^* are found by minimizing the KL-divergence between a proposal distribution $q(\mathbf{x}; \theta)$ and the optimal distribution $q^*(\mathbf{x})$, i.e.

$$\theta^* = \arg \min_{\theta} D_{\text{KL}}(q^*(\mathbf{x}) \parallel q(\mathbf{x}; \theta)), \quad (40)$$

where D_{KL} calculates the KL-divergence. Equation (40) can be cast as a stochastic optimization problem as

$$\theta^* \approx \arg \max_{\theta} \frac{1}{N} \sum_{i=1}^N \left[\mathbf{1}\{c(\mathbf{x}_i) < \epsilon\} \frac{p(\mathbf{x}_i)}{q(\mathbf{x}_i; \theta)} \log q(\mathbf{x}_i; \theta) \right], \quad (41)$$

where φ is any set of parameters and \mathbf{x}_i are sampled from $q(\mathbf{x}; \varphi)$. Equation (41) can be solved analytically when the family of algorithms is in the natural exponential family (i.e. normal, exponential, Poisson, gamma, binomial, and others), and the solution corresponds to the maximum likelihood estimate of the parameters (De Boer et al., 2005).

For an iterative solution to finding θ^* , the initial parameters θ_0 are chosen so that $q(\mathbf{x}; \theta_0)$ is close to $p(\mathbf{x})$. Then, for $k = 0, 1, \dots$

1. Set $\varphi = \theta_k$.
2. Draw samples $\{\mathbf{x}_1, \dots, \mathbf{x}_N\}$ from $q(\mathbf{x}; \varphi)$.
3. Solve eq. (41) for θ_{k+1} .

One major challenge to this approach is the rarity of failure events. If all samples have $c(\mathbf{x}) > \epsilon$, then $\hat{P}_{\text{fail}} = 0$ and the algorithm may not converge to the optimal proposal distribution. One solution is to adaptively update the safety threshold ϵ at each iteration. At iteration k , a safety threshold ϵ_k and a rarity parameter ρ is chosen so that the fraction of samples that have $c(\mathbf{x}) < \epsilon_k$ is ρ . The parameter ρ is also known as the quantile level and is often set in the range $\rho = [0.01, 0.2]$ (Y. Kim & Kochenderfer, 2016; O’Kelly et al., 2018).

The cross-entropy method has been used to estimate the probability of failure for aircraft collision avoidance systems (Y. Kim & Kochenderfer, 2016) and autonomous vehicles (Z. Huang et al., 2017; O’Kelly et al., 2018; D. Zhao et al., 2016). Typically, probability distributions in the natural exponential family are used (Y. Kim & Kochenderfer, 2016; O’Kelly et al., 2018; D. Zhao et al., 2016) so that cross-entropy updates can be performed analytically. Z. Huang et al. (2017) propose a method for using piecewise exponential distributions for more flexibility while retaining the ability to compute updates analytically. Sankaranarayanan and Fainekos (2012) discuss piecewise uniform distributions over the disturbance space, and techniques for factoring the space to reduce the number of parameters needed.

7.2 Multilevel Splitting

Multilevel splitting (Kahn & Harris, 1951) is a non-parametric approach to estimating the optimal importance sampling distribution. Rather than relying on an explicit probability distribution (as in the cross-entropy method), multilevel splitting relies on Markov chain Monte Carlo (MCMC) estimation and scales better to larger dimensions (Botev & Kroese, 2008). It has been applied to the estimation of probability of failure of autonomous driving policies with a large number of parameters (Norden et al., 2019).

The idea of multilevel splitting is to define a set of threshold levels $\infty = \epsilon_0 > \epsilon_1 > \dots > \epsilon_K = \epsilon$ and assume that the probability of failure can be computed as a Markov chain of

the form

$$P(c(\mathbf{x}) < \epsilon) = \prod_{k=1}^K P(c(\mathbf{x}) < \epsilon_k \mid c(\mathbf{x}) < \epsilon_{k-1}). \quad (42)$$

Given enough levels (i.e. a large enough K), each conditional probability

$$P_k = P(c(\mathbf{x}) < \epsilon_k \mid c(\mathbf{x}) < \epsilon_{k-1}) \quad (43)$$

is much larger than $P(c(\mathbf{x}) < \epsilon)$ and can therefore be computed efficiently with basic Monte Carlo methods. Algorithm 4 outlines the algorithm. At iteration k , N samples of \mathbf{x} are sorted by their cost function (line 8) and used to estimate the conditional probability P_k (line 10). The total probability is updated based on the Markov assumption (line 11). All samples with $c(\mathbf{x}) > \epsilon_k$ are discarded, and the remaining samples are resampled to get back up to N samples (line 12). Those samples perform a random walk of M steps using a kernel $T(\mathbf{x}' \mid \mathbf{x})$ (line 13). The process repeats until the level reaches the true safety threshold. The choice of levels ϵ_k can be done adaptively (C erou & Guyader, 2007) by selecting a rarity parameter ρ such that ρN samples are kept at each iteration (line 9).

Algorithm 4 Adaptive multilevel splitting.

```

1: function MULTILEVELSPLITTING( $p, N, M, T, c, \epsilon$ )
2:    $P_{\text{fail}} \leftarrow 1$ 
3:    $k \leftarrow 1$ 
4:    $\epsilon_0 = \infty$ 
5:   Sample  $\{\mathbf{x}_1, \dots, \mathbf{x}_N\}$  from  $p(\mathbf{x})$ 
6:   while  $\epsilon_k > \epsilon$ 
7:      $k \leftarrow k + 1$ 
8:     Sort  $\{\mathbf{x}_1, \dots, \mathbf{x}_N\}$  by  $c(\mathbf{x}_i)$ 
9:      $\epsilon_k \leftarrow \max(c(\mathbf{x}_{\rho N}), \epsilon)$ 
10:     $P_k \leftarrow \frac{1}{N} \sum_{i=1}^N \mathbb{1}\{c(\mathbf{x}_i) < \epsilon_k\}$ 
11:     $P_{\text{fail}} \leftarrow P_{\text{fail}} P_k$ 
12:    Resample  $\{\mathbf{x}_1, \dots, \mathbf{x}_N\}$  from  $\{\mathbf{x}_1, \dots, \mathbf{x}_{\rho N}\}$ 
13:    Random walk  $M$  steps using  $T(\mathbf{x}' \mid \mathbf{x})$  for each  $\{\mathbf{x}_1, \dots, \mathbf{x}_N\}$ 
14:  return  $P_{\text{fail}}$ 

```

7.3 Classification Methods

Supervised learning can be used to classify disturbances as safe or unsafe. That classification can then be combined with importance sampling to estimate the probability of failure of the system. Supervised learning often requires observing more than one example of a failure, so these approaches are most applicable when failures can be found relatively easily. One approach is to use a space-filling sampling plan for the disturbance (Z. Huang et al., 2018) to ensure good coverage of the disturbance space, but this fails to scale to high dimensions. Another approach is to use previous versions of the system that are far less safe (Uesato et al., 2019). Known as a *continuation approach*, this technique is well suited to black-boxes that have learned behavior. During the learning process, the system will fail more

easily (but in related ways) to the version that is ultimately being tested. Therefore, earlier versions of the system can be used for classification of disturbances.

One way to combine supervised learning with importance sampling was explored by Z. Huang et al. (2018). They built a proposal distribution centered on the boundary between safe and unsafe disturbances. If $q(\mathbf{x}; \theta)$ is a proposal distribution with parameters θ , then they search for the point \mathbf{x}^* in the set of failures that maximizes the probability of the proposal distribution:

$$\mathbf{x}^* = \arg \max_{\mathbf{x}} \mathbb{1}\{c(\mathbf{x}) < \epsilon\}q(\mathbf{x}; \theta) \tag{44}$$

Then, they adjust the parameters of the proposal distribution such that the mean is located at \mathbf{x}^* . This approach has shown to construct an efficient proposal distribution when the unsafe set is convex (Sadowsky & Bucklew, 1990). In practice, it is unlikely that the set is convex, but a linear boundary between safe and unsafe examples can be constructed using the support vector machine (SVM) algorithm (Suykens & Vandewalle, 1999) if the disturbance space is lifted to a higher dimension through a mapping $\phi(\mathbf{x})$.

Once a boundary is found, the choice of distribution q determines the feasibility of computing \mathbf{x}^* . One approach is to represent q by a mixture of K Gaussians with weights α :

$$q_{\text{GMM}}(\mathbf{x}; \mu_i, \sigma_i) = \sum_{i=1}^K \alpha_i \mathcal{N}(\mathbf{x}; \mu_i, \sigma_i) \tag{45}$$

The optimal mean for each Gaussian model \mathbf{x}_i^* is determined separately and the proposal distribution is reconstructed by

$$q_{\text{GMM}}^*(\mathbf{x}; \mathbf{x}_i^*, \sigma_i) = \sum_{i=1}^K \alpha_i \mathcal{N}(\mathbf{x}; \mathbf{x}_i^*, \sigma_i). \tag{46}$$

Note that a Gaussian mixture model (GMM) can be obtained in the lifted space $\tilde{q}_{\text{GMM}}(\phi(\mathbf{x}))$ and then reduced to the true disturbance space by integrating over the extra dimensions. The approach is outlined in algorithm 5.

Algorithm 5 Classification-based importance sampling.

- 1: **function** CLASSIFICATIONIMPORTANCESAMPLING(p, M, ϕ, c, ϵ)
 - 2: Sample $\{\mathbf{x}_1, \dots, \mathbf{x}_M\}$ from $p(\mathbf{x})$ and fit a GMM $\tilde{q}_{\text{GMM}}(\phi(\mathbf{x}); \mu_i, \sigma_i)$
 - 3: Sample $\{\mathbf{x}_1, \dots, \mathbf{x}_N\}$ and compute $\{\mathbb{1}\{c(\mathbf{x}_1) < \epsilon\}, \dots, \mathbb{1}\{c(\mathbf{x}_N) < \epsilon\}\}$
 - 4: Lift the disturbances by computing $\{\phi(\mathbf{x}_1), \dots, \phi(\mathbf{x}_N)\}$
 - 5: Apply SVM on pairs $\{(\phi(\mathbf{x}_1), \mathbb{1}\{c(\mathbf{x}_1) < \epsilon\}), \dots, (\phi(\mathbf{x}_N), \mathbb{1}\{c(\mathbf{x}_N) < \epsilon\})\}$
 - 6: Determine the optimal points ϕ_i^*
 - 7: Construct $\tilde{q}_{\text{GMM}}^*(\phi(\mathbf{x}); \phi_i^*, \sigma_i)$
 - 8: Marginalize to $q^*(\mathbf{x})$
 - 9: **return** ESTIMATEPROBABILITY(p, q^*)
-

An alternative approach by Uesato et al. (2019) uses supervised learning to estimate the probability of failure $\hat{P}_{\text{fail}}(\mathbf{x})$ for a disturbance \mathbf{x} , and therefore does not make the assumption that failures are deterministic given \mathbf{x} . The function $\hat{P}_{\text{fail}}(\mathbf{x})$ can be represented using a neural network or some other model and is trained on failure examples which may

come from weaker versions of the system. The optimal importance sampling distribution was proven to be

$$q^*(\mathbf{x}) = \frac{\sqrt{\hat{P}_{\text{fail}}(\mathbf{x})p(\mathbf{x})}}{\mathbb{E}_p \left[\sqrt{\hat{P}_{\text{fail}}(\mathbf{x})} \right]}, \quad (47)$$

which can be sampled from using rejection sampling.

7.4 State-Dependent Importance Sampling

While most importance sampling approaches focus on the entire space of disturbance trajectories, a sequential decision making framework can be used to find the optimal importance sampling policy $q(x | s)$ for a simulation state s (Chryssanthacopoulos et al., 2010; Corso et al., 2020). For a Markovian system and simulator, the optimal importance sampling policy is given by

$$q(x | s) = \frac{p(x | s)P_{\text{fail}}(s')}{P_{\text{fail}}(s)}, \quad (48)$$

where s' is deterministically reached after disturbance x is applied in state s . The probability of failure $P_{\text{fail}}(s)$ can be estimated through the approximate solution of the Bellman equation

$$P_{\text{fail}}(s) = \begin{cases} 1 & \text{if } s \in S_{\text{fail}} \\ 0 & \text{if } s \notin S_{\text{fail}}, s \in S_{\text{term}} \\ \sum_a p(x | s)P_{\text{fail}}(s') & \text{otherwise,} \end{cases} \quad (49)$$

where S_{term} is the set of terminal states that are not failure states. Local approximation dynamic programming and Monte Carlo policy evaluations are two successful approaches for solving eq. (49) (Corso et al., 2020).

8. Problem Decomposition Techniques

One of the most significant challenges to overcome for the safety validation of autonomous systems is that algorithms often scale poorly to large disturbance spaces and state spaces. A common approach to dealing with scalability is the use of decomposition approaches to simplify a larger problem into more tractable subproblems. For a survey of decomposition approaches in MDPs see the survey of Daoui et al. (2010). Decomposition approaches that rely on the factorization of the transition function (Guestrin et al., 2003) are not applicable due to the black-box assumption, but techniques that generalize over the state and action spaces could be applied to safety validation. This section will present two approaches that have been used for decomposing the falsification problem into more manageable subproblems to accelerate finding counterexamples.

8.1 State Space Decomposition

The approach presented by Corso et al. (2020) involves decomposing the simulator state and disturbance spaces into independent components, and finding failures for each component. Originally done in the context of multiple distinct actors in a simulated driving

environment, this approach can be used with any simulation that has components that can be simulated individually. To separate the state space into M different components, define a decomposition operator D such that

$$\{s^{(1)}, \dots, s^{(M)}\} = D(s) \tag{50}$$

and the disturbance space $X^{(i)}$ associated with the state $s^{(i)}$ is smaller than the full disturbance space (i.e. $|X^{(i)}| < |X|$). For each subproblem, solve for a policy that finds counterexamples

$$x^{(i)} = \pi^{(i)}(s^{(i)}) \tag{51}$$

and then combine the policies with a fusion function F such that

$$x = F(\pi^{(1)}, \dots, \pi^{(M)})(s). \tag{52}$$

An approach for solving the subproblems that is amenable to policy fusion is approximate dynamic programming (see section 7.4). In that case, the subproblem policy is defined by computing the probability of failure for each state component $P_{\text{fail}}^{(i)}$. The fusion function can then apply simple arithmetic operations (like mean, max, or min) to arrive at a joint policy

$$\tilde{P}_{\text{fail}}(s) = F\left(P_{\text{fail}}^{(1)}(s^{(1)}), \dots, P_{\text{fail}}^{(M)}(s^{(M)})\right) \tag{53}$$

$$x = \frac{p(\mathbf{x} | s)P_{\text{fail}}(s')}{P_{\text{fail}}(s)}, \tag{54}$$

where s' is the deterministic state reached after applying disturbance x at state s .

If the fusion function is simple, then it may not capture the joint interactions between multiple disturbance components (Corso et al., 2020). To address this problem, a global correction factor can be learned from the full simulation. The estimate probability of failure is given by

$$P_{\text{fail}}(s) \approx \tilde{P}_{\text{fail}}(s) + \delta P_{\theta}(s). \tag{55}$$

The correction factor δP_{θ} is trained using rollouts $\{s_1, \dots, s_N\}$ from the full simulation and minimizing the difference between the estimate $\tilde{P}_{\theta}(s_i)$ and the actual discounted return $G(s_i)$ where

$$G(s_i) = \mathbb{1}\{s_N \in S_{\text{fail}}\} \frac{\prod_{t=i}^N p(x_t | s_{t-1})}{\pi(x_t | s_{t-1})}. \tag{56}$$

This approach has been shown to increase the number of failures found in a complex driving environment with a large disturbance space (Corso et al., 2020).

8.2 Compositional Approach with Machine Learning Components

The approach of Dreossi et al. (2019) is applicable when the system contains a machine-learned component that performs classification (such as a neural network based perception system). Knowing something about the structure of the system makes this approach gray-box, but due to the prevalence of ML components, it is still widely applicable.

The algorithm is performed with the following steps:

- The ML component of interest $f(\mathbf{x})$ is partitioned from the rest of the system and environment. The ML component, which has a large disturbance space, is replaced with an idealized abstraction with a much smaller input space.
- The simulator with the idealized ML component is separated into two versions: 1) $f^+(\mathbf{x})$ where the ML component behaves as well as possible (i.e. classifying all inputs correctly), and 2) $f^-(\mathbf{x})$ where the ML component behaves poorly.
- For each version of the simulator, a traditional falsification algorithm is used to partition the disturbance space into the regions that satisfy the specification $\text{safe}(\mathbf{x})$ and regions of failures $\text{fail}(\mathbf{x})$. The notation $\text{safe}(\mathbf{x})^\pm$ indicates the safe disturbance set for $f^\pm(\mathbf{x})$. Similar notation is used for the failure set.
- To find counterexamples that were caused by the ML component, find falsifying examples in the set $\text{safe}(\mathbf{x})^+ \setminus \text{safe}(\mathbf{x})^-$. The set $\text{safe}(\mathbf{x})^-$ is removed because no failure is possible in this region even with the ML component functioning as poorly as possible. This part of the disturbance space is referred to as the region of interest.
- Finally, use a separate analyzer (possibly white-box) to identify the high-dimensional inputs that lead to failures (i.e. misclassifications) in the region of interest.

This approach has been able to find counterexamples in a neural network perception system used by an autonomous vehicle (Dreossi et al., 2019). A similar approach by Julian et al. (2020) uses a white-box neural network analyzer combined with Monte Carlo tree search to find failure trajectories of an image-based control system.

9. Applications

Autonomous cars and aircraft are two major application domains of black-box safety validation methods. While there are a variety of scenarios within these application domains, a common underlying principle is that of *miss distance* as a reward heuristic. For most scenarios, it is possible to use some sort of distance metric to measure how close the system came to a failure during the scenario. This distance is then used to allow optimization to find counterexamples. The autonomous vehicles and aircraft application domains are covered in more detail below.

9.1 Autonomous Driving

Within the field of autonomous driving, there are multiple scenarios that are commonly used for testing:

Lane-Following. Lane-following is one of the most common examples in the autonomous vehicle field, and has been used to test systems ranging from full-stack systems (O’Kelly et al., 2018) to systems with ideal perception (Behzadan & Munir, 2019) to systems that are not fully autonomous like advanced driver-assistance systems (ADAS) (Koschi et al., 2019). The system under test tries to maintain a desired speed in the current lane and may (Tuncali & Fainekos, 2019) or may not (Kuutti et al., 2020) have the ability to change lanes. The goal is often to find collisions, most commonly rear-end collisions. Scenario variations include both highway driving (Norden et al., 2019) and local road driving (Tuncali et al., 2019).

Intersection Scenarios. In an intersection scenario, the system under test approaches a stoplight (Tuncali et al., 2019), stop sign (Abeyirigoonawardena et al., 2019), crosswalk (Koren et al., 2018), or other form of intersection and must proceed through without a failure. Failures can include collisions with pedestrians (Koren & Kochenderfer, 2020) or other vehicles (Tuncali et al., 2019). Failures may also include violations of traffic laws (Kress-Gazit & Pappas, 2008) or other rule-sets, such as those designed to prevent at-fault collisions (Hekmatnejad et al., 2020).

Lane Change Scenarios. Lane change scenarios can involve the test vehicle initiating a lane change or reacting to one (Qin et al., 2019; D. Zhao et al., 2016). Existing work has considered ADAS and other driver aid systems (Z. Huang et al., 2018). Failure modes include rear-end collisions as well as side collisions caused by turning into an occupied lane.

Platooning Vehicles. Hu et al. (2000) present a platooning scenario, where a number of cars are following a lead car while keeping a specific trailing distance. The vehicles are subjected to various stochastic disturbances arising from road conditions, wind conditions, or the presence of human operators. The goal of the system is to minimize the time that platoon vehicles spend in “chasing” mode, where they attempt to catch up to the vehicle ahead. The system fails when any of the vehicles gets too far from or too close to the leading vehicle.

Within these applications, researchers have developed domain specific techniques to improve performance. The most common approach is to use miss distance, the physical distance between the test vehicle and other vehicles in the simulation, as a reward heuristic to allow easier optimization (Koren & Kochenderfer, 2020). Miss distance is also the underlying principle for robust semantics of temporal formulas when applied to autonomous vehicles (Zhang et al., 2018). Some work has been done on identifying situations where a collision is imminent instead of the collision itself, sometimes called “unsafe states” (Koschi et al., 2019) or “boundary cases” (Tuncali & Fainekos, 2019) since these regions may be easier to find.

9.2 Autonomous Flying and Aircraft Collision Avoidance

Black-box validation techniques have also been applied to aircraft in multiple ways:

Flight Control Software. Delmas et al. (2019) present an application where the controller must keep an aircraft in steady flight in response to disturbances such as wind or pilot inputs. Failures include reduced flight quality, autopilot disengagements, and overshoots of expert-defined thresholds. A second application is presented by Ernst, Arcaini, et al. (2019), where an F-16 controller performs automatic maneuvers to avoid ground collisions. Validation algorithms search for violations of a minimum altitude from various initial conditions. Julian et al. (2020) perform safety validation on a neural network controller with camera inputs for aircraft taxiing.

Collision Avoidance. There are several examples of applications to collision avoidance problems for aircraft. Lee et al. (2020) validate the next-generation aircraft collision avoidance system (ACAS X), which makes recommendations to pilots to avoid mid-air collisions.

The system may or may not coordinate its recommendations with the other aircraft. A failure occurs when the aircraft pass too close to one another, known as a near mid-air collision (NMAC). Esposito et al. (2004) validate a control system that guides a group of planes from an initial location to a final destination in the presence of stochastic wind disturbances. A failure in this case is a collision between any two aircraft.

Flight Path-Planning. Systems are given a mission, which is a series of waypoints that the system must navigate (Moss et al., 2020; Tuncali et al., 2018), or a target location and keep-out zones along the path (Lee et al., 2019). Falsifying trajectories may be different mission parameters or disturbances during mission execution, such as wind and sensor noise. Failures occur when the system enters areas defined as off-limits, collides with an obstacle, or produces a software error. X. Yang et al. (2020) validate scheduling and trajectory planning for urban air mobility and package delivery systems.

9.3 General Systems

Black-box safety validation has also been applied to systems that are not tied specific applications such as hybrid systems, neural network controllers and planning algorithms.

Hybrid Systems. Algorithms for black-box safety validation have also been applied to a number of hybrid systems outside of autonomous driving and autonomous aircraft. Hoxha et al. (2015) present an automatic transmission system model, which is widely used as a benchmark problem in the literature. The system under test selects a gear based on throttle and brake inputs as well as state information such as current engine load and car speed. Failure events include violations of speed thresholds and changing gears too often. Jin et al. (2014) present another widely used benchmark problem from the automotive field. The system is an abstract fuel controller for an automotive powertrain. The controller receives inputs such as fuel-flow measurements and throttle and must output a fuel command. Failures may be either steady-state, such as the violation of an air-to-fuel ratio, or transient, such as pulses that violate a settling-time constraint.

Non-automotive systems have also been considered as well. J. Kim et al. (2005) analyze a thermostat model under various environmental conditions and test whether the heater will be active for longer than a certain proportion of time. Schuler et al. (2017) present a simplified model of a wind turbine, which attempts to generate as much power as possible based on the current wind speed. Failures are violations of safety criteria, which include violations of thresholds on tower base moment and rotor speed.

Controllers. Systems with neural network controllers have also become common case studies for validation techniques. Yaghoubi and Fainekos (2019) test a steam condenser controlled by a recurrent neural network (RNN). The system modulates steam flow rate based on energy balance and cooling water mass balance. A failure occurs when the pressure falls outside the acceptable range. Yaghoubi and Fainekos (2019) also present a generalized non-linear system controlled by a feed-forward neural network. A failure is defined as a constraint on the reference signal value. Ernst, Arcaini, et al. (2019) study a neural network controlled magnet levitation system. The controller takes a reference position as input and attempts to move the magnet to track the given reference position. A failure occurs if the

magnet does not stabilize to a position close enough to the reference position within some time limit.

Planning Modules. Falsification of planning modules is common among the autonomous vehicle and aircraft applications, but there are examples in other domains as well. J. Kim et al. (2005) present a hovercraft navigation application. The validation task tests whether the hovercraft can successfully navigate to some goal region while subjected to stochastic wind disturbances. Similarly, Zhang et al. (2018) validate a free-floating robot system that must navigate to a desired target location. Fehnker and Ivani (2004) present a generalization of this task, where an agent is moving in a discrete 2D environment, sometimes called a *gridworld* task. In their formulation, there are states that must be reached and states that must be avoided. A violation of either constraint is considered a failure.

10. Existing Tools

Implementations of various safety validation algorithms have been made available as tools for others to use. The existing tools range from open-source academic-based toolboxes to closed-source commercial software. A detailed survey that includes falsification tools is provided in (Kapinski et al., 2016). Each of the tools described in this section create falsifying disturbance trajectories to a system given a set of system requirements that should be satisfied. Certain tools also perform most-likely failure analysis and failure probability estimation. Many of the existing tools interface with the MATLAB programming language/environment to stress industry standard Simulink models. The tools surveyed in this section focus on black or gray-box testing of cyber-physical systems and although some of these tools include additional functionality, this section focuses on features of the safety validation components. A brief overview of the existing safety validation tools will be discussed and their benefits and restrictions will be highlighted.

10.1 Academic Tools

Many of the existing safety validation tools are products of academic research and released as experimental prototypes. Although these tools are prototypes, several have matured enough to gain wide-spread usage and acceptance (Diwakaran et al., 2017; Dreossi et al., 2015; Tuncali et al., 2018; Zhang et al., 2019; Zutshi et al., 2014). Two particular falsification tools have become benchmark standards in the field: S-TALIRO (G. Fainekos et al., 2019) and Breach (Donzé, 2010). Both are open-sourced MATLAB toolboxes for optimization-based falsification. While most of the tools are open-sourced, two other tools referenced in the falsification literature are not publicly available but still discussed. Although their respective papers indicate how to recreate their work, none of those tools have code or executables available online. The following collection is organized into optimization-based and reinforcement learning-based tools.

10.2 Optimization-based Tools

The tools in this section employ a standard optimization-based technique to search for counterexamples. Section 4 outline the approaches implemented by the following tools.

S-TaLiRo. S-TALiRO (Systems Temporal Logic Robustness) (Y. Annapureddy et al., 2011; G. Fainekos et al., 2019) is a simulation-based MATLAB toolbox for temporal logic falsification of non-linear hybrid systems. S-TALiRO parameterizes the disturbance space to reformulate the falsification problem as an optimization problem. S-TALiRO instructs the user to specify system requirements in temporal logic formulas and then constructs the optimization cost function to minimize a global robustness metric. Various optimization techniques are included in the S-TALiRO toolbox, such as simulated annealing (described in section 4.1), genetic algorithms (described in section 4.2), stochastic optimization with adaptive restarts (described in section 4.3), the cross-entropy method (described in section 7.1), and uniform random sampling. S-TALiRO is designed to analyze arbitrary MATLAB functions or Simulink/Stateflow models. S-TALiRO is open-source and available under the GNU General Public License (GPL).¹

Specific add-ons to S-TALiRO have been implemented that extend the core falsification functionalities. These add-ons generally provide other solution methods or provide additional simulation environments that interface with S-TALiRO.

Breach. Breach (Donzé, 2010; Dreossi et al., 2019) is a simulation-based MATLAB toolbox for falsification of temporal logic specifications for hybrid dynamical systems, similar to S-TALiRO. Breach uses optimization-based techniques including simulated annealing (described in section 4.1), genetic algorithms (described in section 4.2), globalized Nelder-Mead (Luersen & Le Riche, 2004), and CMA-ES (Hansen & Ostermeier, 1996). The user-defined system requirements are input using temporal logic formulas. These requirements, i.e. specifications, are used to construct a cost function to be minimized based on a robustness metric. Breach is designed to test arbitrary MATLAB functions and Simulink models and includes a MATLAB graphical user interface (GUI) that gives the user access to the input parameter sets, temporal logic formulas, and trajectory visualizations. Breach is open-source and available under the BSD license.²

RRT-Rex. RRT-REX (Rapidly-exploring Random Tree Robustness-guided Explorer) (Dreossi et al., 2015) is a MATLAB falsification tool that focuses on coverage given a computational budget. A Simulink model and user-defined requirements written in temporal logic are taken as input. RRT path planning algorithms are used to search the disturbance space for falsifying cases, guided by a combined state space coverage metric and a robustness satisfaction metric. Section 5 discusses the RRT approach in detail. RRT-REX is not currently publicly available.

10.3 Reinforcement Learning-Based Tools

As a direct replacement to optimization algorithms, reinforcement learning can be used as the central idea behind searching for falsifying trajectories. The following tools implement reinforcement learning algorithms as solvers for the falsification problem. Sections 6.1 and 6.2 describe the reinforcement learning algorithms implemented in the tools.

FalStar. FALSTAR (Zhang et al., 2018) is a prototype Scala tool for falsification of cyber-physical systems that interfaces with MATLAB through a Java API. FALSTAR uses re-

1. <https://sites.google.com/a/asu.edu/s-taliro>

2. <https://github.com/decyphir/breach>

inforcement learning combined with optimization techniques to generate counterexamples. Techniques include Monte Carlo tree search with stochastic optimization as described in section 6.1 and adaptive Las Vegas tree search (aLVTS) as described in section 5.3. FALSTAR requires a Simulink model as input and uses the above techniques to generate counterexamples to the temporal logic specifications. FALSTAR can also interface directly with the Breach toolbox to use the available solvers implemented in Breach. FALSTAR is open-source and available under the BSD license.³

falsify. falsify (Akazaki et al., 2018) is a prototype simulation-based falsification tool that uses deep reinforcement learning. Common among the academic tools, falsify can interface directly with MATLAB functions and Simulink models. Implementing a robustness-guided approach, falsify defines the reward as a convex function of the robustness, as described in section 6. This robustness-guided reward function is used by two deep reinforcement learning algorithms implemented in falsify: asynchronous advantage actor critic (A3C) and double deep-Q network (double DQN), both described in section 6.2. The falsify tool is not currently publicly available.

AST Toolbox. The AST Toolbox (Adaptive Stress Testing Toolbox) (Koren et al., 2018) is a Python toolbox for safety validation, which includes falsification and most-likely failure analysis. The AST Toolbox uses reinforcement learning to find the most likely failures of black-box systems. Two reinforcement learning techniques used as solvers are included: Monte Carlo tree search (described in section 6.1) and deep reinforcement learning (described in section 6.2). The AST Toolbox is built on top of two popular reinforcement learning packages, namely, OpenAI Gym (Brockman et al., 2016) and Garage.⁴ Building off of these packages gives the user access to widely used reinforcement learning benchmarking problems.

To test a system, users must provide the definitions for three basic interfacing functions. These interfacing functions allow the tool to interact with the black-box system or simulator. The interface defines how to *initialize* the system, how to *step* the system forward (returning indications of found failures or a real-valued measure of distance to a failure), and finally a means to determine if the system is in a *terminal* state. While the implementation of the interface is restricted to Python, the user can call out to existing executables or other languages from within Python. As an example, Python can interface with MATLAB through the MATLAB Engine API for Python.⁵ The AST Toolbox is open-source and available under the MIT license.⁶

Two related toolboxes, called AdaptiveStressTesting.jl (Lee et al., 2020) and POMDP-StressTesting.jl (Moss et al., 2020), follow a similar paradigm as the AST Toolbox but are implemented in the Julia programming language (Bezanson et al., 2017). Julia can interface directly with many other programming languages,⁷ and notably, Julia can interface with MATLAB through the MATLAB.jl package.⁸ AdaptiveStressTesting.jl⁹ and POMDP-

3. <https://github.com/ERATOMMSD/falstar>

4. <https://github.com/rlworkgroup/garage>

5. <https://www.mathworks.com/help/matlab/matlab-engine-for-python.html>

6. <https://github.com/sisl/AdaptiveStressTestingToolbox>

7. <https://github.com/JuliaInterop>

8. <https://github.com/JuliaInterop/MATLAB.jl>

9. <https://github.com/sisl/AdaptiveStressTesting.jl>

StressTesting.jl¹⁰ are open-source and available under the Apache License Version 2.0 and the MIT license, respectively.

10.4 Commercial Tools

Certain techniques for safety validation of cyber-physical systems have become available as commercial toolboxes. This section briefly describes these commercially available tools relating to black-box safety validation.¹¹

Reactis. Reactis (Reactive Systems Inc., 2013) is a simulation-based MATLAB tool from Reactive Systems for falsification of Simulink models. Reactis has three components: Reactis Tester, Reactis Simulator, and Reactis Validator. The Reactis Tester controls the generation of falsifying trajectories. It uses a patented technique called *guided simulation* which uses proprietary algorithms and heuristics to generate trajectories to maximize coverage for falsification. Then, the Reactis Simulator runs the system under test given the purposed falsifying trajectory. The last component, the Reactis Validator, uses proprietary techniques to search for violations of user-defined model specifications. Reactive Systems also has a version of Reactis for C (Reactive Systems Inc., 2011) that has analogous Reactis components for systems developed in the C programming language. Reactis is commercially available through Reactive Systems, Inc.¹²

TestWeaver. TestWeaver (Junghanns et al., 2008) is a simulation-based falsification tool from Synopsys. TestWeaver uses proprietary search algorithms to generate falsifying trajectories while maximizing coverage. User-defined system requirements and worst-case quality indicators are used to guide the search. Extensive knowledge of the underlying system may be required to provide useful worst-case quality indicators, which may limit the application. TestWeaver can interface with other simulation frameworks to control the disturbance trajectories via libraries in Simulink, Modelica, Python, and C. TestWeaver is commercially available through Synopsys, Inc.¹³

TrustworthySearch API. TrustworthySearch API (Norden et al., 2019) is a risk-based falsification and probability estimation framework from Trustworthy AI. TrustworthySearch API is general for black-box validation of systems and has applications in autonomous vehicle safety validation. The tool uses a proprietary sequential importance sampling and sequential Monte Carlo approach. A version of the probabilistic-based approach was shown using adaptive importance sampling techniques described in section 7. Importance sampling is a stand-in for traditional optimization algorithms used to search the disturbance space for falsification. The use of a proposal distribution biased towards rare failure events ensures that these low probability events are sampled more frequently. Adaptive multilevel splitting (AMS), described in section 7.2, is implemented to estimate this biased distribution from data. Along with falsification, the tool can also perform failure event probability estimation, unlike other commercially available products and most academic tools. To estimate the failure probability, the probability from the biased distribution is reweighted

10. <https://github.com/sisl/POMDPStressTesting.jl>

11. The authors are not affiliated with any of the companies.

12. <https://www.reactive-systems.com/>

13. <https://www.synopsys.com/verification/virtual-prototyping/virtual-ecu/testweaver.html>

according to the likelihood from the original unbiased distribution. Although specific to autonomous vehicle safety validation, we include TrustworthySearch API to highlight recent advancements of black-box safety validation tools. TrustworthySearch API is commercially available through Trustworthy AI, Inc.¹⁴

10.5 Toolbox Competition

An academically-driven friendly competition for systems verification, called ARCH-COMP, has been held annually since 2017 (Ernst, Arcaini, et al., 2019). One category within the competition is the falsification of temporal logic specifications for cyber-physical systems. Falsification researchers compare their tools against common benchmark problems and use the competition to track state-of-the-art falsification tools.

As detailed in their 2019 report (Ernst, Arcaini, et al., 2019), S-TALIRO, Breach, FALSTAR, and falsify participated in the most recent competition. Six benchmark problems from the literature were used to evaluate each tool. The tools were evaluated based on their falsification rate and statistics on number of simulations required to find a falsifying trajectory. The outcome of the 2019 competition showed that falsify had the most success, only requiring a single simulation (after training) in certain benchmark problems. A notable emphasis on repeatability of results was made during this recent competition. For future competitions, we suggest a comparison metric for tool runtime to help assess relative computational timing complexities. Continuing to mature the competition as more falsification methods arise will help drive discussion around falsification tool design decisions. The competition’s benchmarks and results are available online.¹⁵

10.6 Tools Discussion

The available tools can be categorized based on which aspects of the safety validation process they implement, as described in section 2.3: falsification, most-likely failure analysis, and failure probability estimation. Note that all of the available tools perform falsification as their core task. In the context of the safety validation problem, the academic tools S-TALIRO, Breach, FALSTAR, RRT-REX, and falsify are falsification-only tools. As for commercial tools, Reactis and TestWeaver also fall into the falsification-only category. TrustworthySearch API is the only commercial tool that also performs failure probability estimation and the AST Toolbox is the only academic tool that performs most-likely failure analysis. A full comparison is outlined in table 1.

As for solution techniques, the tools S-TALIRO and Breach take a standard optimization-based approach, while RRT-REX reformulates the falsification problem to solve it using path planning algorithms. The AST Toolbox and falsify are based on reinforcement learning and FALSTAR combines reinforcement learning and global optimization techniques for further refinement of the disturbance trajectory search. Common among all of the tools is an interface to MATLAB, with a particular emphasis on testing Simulink models. This is evident in the survey and can be attributed to an industrial emphasis on the use of MATLAB/Simulink models for prototyping. A common component specific to academic falsification-only tools is the use of temporal logic to specify system requirements—encoded

14. <http://trustworthy.ai/>

15. <https://gitlab.com/goranf/ARCH-COMP>

Table 1: Black-box Safety Validation Tools

Tool	Safety Validation Problem			Technique	Source
	Falsification	Most-Likely Failure	Probability Estimation		
S-TALIRO (Y. Annapureddy et al., 2011) ^{† *}	✓	·	·	Optimization	Open
Breach (Donzé, 2010) [*]	✓	·	·	Optimization	Open
RRT-REX (Dreossi et al., 2015) [†]	✓	·	·	Path Planning	Closed
FALSTAR (Zhang et al., 2018) ^{† *}	✓	·	·	Optimization, RL	Open
falsify (Akazaki et al., 2018) ^{† *}	✓	·	·	Reinforcement Learning	Closed
AST Toolbox (Koren et al., 2018)	✓	✓	·	Reinforcement Learning	Open
Reactis (Reactive Systems Inc., 2013)	✓	·	·	Proprietary	Commercial
TestWeaver (Junghanns et al., 2008)	✓	·	·	Proprietary	Commercial
TrustworthySearch API (Norden et al., 2019)	✓	·	✓	Importance Sampling	Commercial

^{*} Competed in ARCH-COMP 2019 (Ernst, Arcaini, et al., 2019).

[†] Accepts system specification in temporal logic (STL or MTL).

in signal temporal logic (STL) or metric temporal logic (MTL). Although expressive, requiring the strict usage of a temporal logic to encode system requirements could also limit the applicability of these tools. With a common goal of ensuring that safety-critical systems are in fact safe, the availability of these tools allows users to provide feedback given their specific use-cases and experience. Continued tool and technique development is further encouraged by academically-driven competitions such as ARCH-COMP (Ernst, Arcaini, et al., 2019).

11. Conclusion

With the rapid increase of safety-critical autonomous systems operating with humans, it is important that we develop robust testing procedures that can ensure the safety of these systems. Due to the high level of system complexity, we generally need black-box validation strategies to find failures of the autonomous system. This work described the problems of falsification (where a single failure example is searched for), most-likely failure analysis (where the most likely failure is searched for), and failure probability estimation (where we seek a good estimate of the likelihood of failure).

With these goals defined, we outlined a wide array of algorithms that have been used to accomplish these tasks. Global optimization, path planning, and reinforcement learning algorithms have been used to find falsifying examples, while importance sampling methods have been used to estimate the probability of failure even when it is close to zero. To address the problem of scalability, we described approaches for decomposing the safety validation problem into more manageable components. We gave a brief overview of the main applications for black-box safety validation including autonomous driving and flight. Finally, we provided an overview of the existing tools that can be used to tackle these validation tasks.

References

- Abbas, H., Fainekos, G., Sankaranarayanan, S., Ivani, F., & Gupta, A. (2013). Probabilistic temporal logic falsification of cyber-physical systems. *ACM Transactions on Embedded Computing Systems (TECS)*, 12(2s), 1–30.
- Abbas, H., O’Kelly, M., & Mangharam, R. (2017). Relaxed decidability and the robust semantics of metric temporal logic. *ACM International Conference on Hybrid Systems: Computation and Control (HSCC)*, 217–225.
- Abeyesirigoonawardena, Y., Shkurti, F., & Dudek, G. (2019). Generating adversarial driving scenarios in high-fidelity simulators. *IEEE International Conference on Robotics and Automation (ICRA)*, 8271–8277.
- Adimoolam, A., Dang, T., Donzé, A., Kapinski, J., & Jin, X. (2017). Classification and coverage-based falsification for embedded control systems. *International Conference on Computer Aided Verification (CAV)*, 483–503.
- Aerts, A., Minh, B. T., Mousavi, M. R., & Reniers, M. A. (2018). Temporal logic falsification of cyber-physical systems: An input-signal-space optimization approach. *IEEE International Conference on Software Testing, Verification and Validation Workshops (ICSTW)*, 214–223.
- Agha, G., & Palmkog, K. (2018). A survey of statistical model checking. *ACM Transactions on Modeling and Computer Simulation (TOMACS)*, 28(1), 1–39.
- Akazaki, T., Kumazawa, Y., & Hasuo, I. (2017). Causality-aided falsification. *Electronic Proceedings in Theoretical Computer Science*, 257.
- Akazaki, T., Liu, S., Yamagata, Y., Duan, Y., & Hao, J. (2018). Falsification of cyber-physical systems using deep reinforcement learning. *International Symposium on Formal Methods (FM)*, 456–465.
- Alpern, B., & Schneider, F. B. (1987). Recognizing safety and liveness. *Distributed Computing*, 2(3), 117–126.
- Alur, R. (2015). *Principles of cyber-physical systems*. MIT Press.
- Annapureddy, Y., Liu, C., Fainekos, G., & Sankaranarayanan, S. (2011). S-TaLiRo: A Tool for Temporal Logic Falsification for Hybrid Systems. *International Conference on Tools and Algorithms for the Construction and Analysis of Systems (TACAS)*, 254–257.
- Annapureddy, Y. S. R., & Fainekos, G. E. (2010). Ant colonies for temporal logic falsification of hybrid systems. *Annual Conference on IEEE Industrial Electronics Society (IECON)*, 91–96.
- Balkan, A., Tabuada, P., Deshmukh, J. V., Jin, X., & Kapinski, J. (2017). Underminer: A framework for automatically identifying nonconverging behaviors in black-box system models. *ACM Transactions on Embedded Computing Systems (TECS)*, 17(1), 1–28.
- Behzadan, V., & Munir, A. (2019). Adversarial reinforcement learning framework for benchmarking collision avoidance mechanisms in autonomous vehicles. *IEEE Intelligent Transportation Systems Magazine*.

- Bezanson, J., Edelman, A., Karpinski, S., & Shah, V. B. (2017). Julia: A fresh approach to numerical computing. *SIAM Review*, 59(1), 65–98.
- Bock, H. G., & Plitt, K.-J. (1984). A multiple shooting algorithm for direct solution of optimal control problems. *IFAC Proceedings Volumes*, 17(2), 1603–1608.
- Botev, Z. I., & Kroese, D. P. (2008). An efficient algorithm for rare-event probability estimation, combinatorial optimization, and counting. *Methodology and Computing in Applied Probability*, 10(4), 471–505.
- Branicky, M. S., Curtiss, M. M., Levine, J., & Morgan, S. (2006). Sampling-based planning, control and verification of hybrid systems. *IEE Proceedings D–Control Theory and Applications*, 153(5), 575–590.
- Brockman, G., Cheung, V., Pettersson, L., Schneider, J., Schulman, J., Tang, J., & Zaremba, W. (2016). OpenAI Gym. *arXiv e-prints*, Article arXiv:1606.01540, arXiv:1606.01540.
- C erou, F., & Guyader, A. (2007). Adaptive multilevel splitting for rare event analysis. *Stochastic Analysis and Applications*, 25(2), 417–443.
- Chaslot, G. M. J.-B., Winands, M. H. M., Herik, H. J. V. D., Uiterwijk, J. W. H. M., & Bouzy, B. (2008). Progressive strategies for Monte Carlo tree search. *New Mathematics and Natural Computation*, 4(03), 343–357.
- Chryssanthacopoulos, J. P., Kochenderfer, M. J., & Williams, R. E. (2010). Improved Monte Carlo sampling for conflict probability estimation. *AIAA Non-Deterministic Approaches Conference*.
- Clarke, E. M., & Emerson, E. A. (1981). Design and synthesis of synchronization skeletons using branching time temporal logic. *Workshop on Logic of Programs*, 52–71.
- Clarke, E. M., Henzinger, T. A., Veith, H., & Bloem, R. (2018). *Handbook of model checking*. Springer.
- Corso, A., Du, P., Driggs-Campbell, K., & Kochenderfer, M. J. (2019). Adaptive stress testing with reward augmentation for autonomous vehicle validation. *IEEE International Conference on Intelligent Transportation Systems (ITSC)*, 163–168.
- Corso, A., & Kochenderfer, M. J. (2020). Interpretable safety validation for autonomous vehicles. *IEEE International Conference on Intelligent Transportation Systems (ITSC)*.
- Corso, A., Lee, R., & Kochenderfer, M. J. (2020). Scalable autonomous vehicle safety validation through dynamic programming and scene decomposition. *IEEE International Conference on Intelligent Transportation Systems (ITSC)*.
- Coulom, R. (2007). Computing ‘ELO ratings’ of move patterns in the game of go. *ICGA Journal*, 30(4), 198–208.
- Dang, T., Donz e, A., Maler, O., & Shalev, N. (2008). Sensitive state-space exploration. *IEEE Conference on Decision and Control (CDC)*, 4049–4054.
- Daoui, C., Abbad, M., & Tkiouat, M. (2010). Exact decomposition approaches for Markov decision processes: A survey. *Advances in Operations Research*, 2010, 1–19.

- De Boer, P.-T., Kroese, D. P., Mannor, S., & Rubinstein, R. Y. (2005). A tutorial on the cross-entropy method. *Annals of Operations Research*, 134(1), 19–67.
- Delmas, R., Loquen, T., Boada-Bauxell, J., & Carton, M. (2019). An evaluation of Monte Carlo tree search for property falsification on hybrid flight control laws. *International Workshop on Numerical Software Verification*, 45–59.
- Deshmukh, J., Horvat, M., Jin, X., Majumdar, R., & Prabhu, V. S. (2017). Testing cyber-physical systems through bayesian optimization. *ACM Trans. Embed. Comput. Syst.*, 16(5s).
- Deshmukh, J., Jin, X., Kapinski, J., & Maler, O. (2015). Stochastic local search for falsification of hybrid systems. *International Symposium on Automated Technology for Verification and Analysis (ATVA)*, 500–517.
- Diehl, M., Bock, H. G., Diedam, H., & Wieber, P.-B. (2006). Fast direct multiple shooting algorithms for optimal robot control. *Fast motions in biomechanics and robotics* (pp. 65–93). Springer.
- Dijkstra, E. W. (1959). A note on two problems in connexion with graphs. *Numerische mathematik*, 1(1), 269–271.
- Diwakaran, R. D., Sankaranarayanan, S., & Trivedi, A. (2017). Analyzing neighborhoods of falsifying traces in cyber-physical systems. *International Conference on Cyber-Physical Systems (ICCP)*, 109–119.
- Donzé, A. (2010). Breach, a toolbox for verification and parameter synthesis of hybrid systems. *International Conference on Computer Aided Verification (CAV)*, 167–170.
- Donzé, A., & Maler, O. (2010). Robust satisfaction of temporal logic over real-valued signals. *International Conference on Formal Modeling and Analysis of Timed Systems (FORMATS)*, 92–106.
- Dorigo, M., Maniezzo, V., & Colorni, A. (1996). Ant system: Optimization by a colony of cooperating agents. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, 26(1), 29–41.
- Dreossi, T., Dang, T., Donzé, A., Kapinski, J., Jin, X., & Deshmukh, J. V. (2015). Efficient guiding strategies for testing of temporal properties of hybrid systems. *NASA Formal Methods Symposium (NFM)*, 127–142.
- Dreossi, T., Donzé, A., & Seshia, S. A. (2019). Compositional falsification of cyber-physical systems with machine learning components. *Journal of Automated Reasoning*, 63(4), 1031–1053.
- Ecoffet, A., Huizinga, J., Lehman, J., Stanley, K. O., & Clune, J. (2019). Go-Explore: A new approach for hard-exploration problems. *arXiv e-prints*, Article arXiv:1901.10995, arXiv:1901.10995.
- Ernst, G., Arcaini, P., Donze, A., Fainekos, G., Mathesen, L., Pedrielli, G., Yaghoubi, S., Yamagata, Y., & Zhang, Z. (2019). Arch-comp 2019 category report: Falsification. *EPiC Series in Computing*, 61, 129–140.

- Ernst, G., Sedwards, S., Zhang, Z., & Hasuo, I. (2019). Fast falsification of hybrid systems using probabilistically adaptive input. *International Conference on Quantitative Evaluation of Systems (QEST)*, 165–181.
- Esposito, J. M., Kim, J., & Kumar, V. (2004). Adaptive RRTs for validating hybrid robotic control systems. *Algorithmic foundations of robotics vi* (pp. 107–121). Springer.
- Fainekos, G., Hoxha, B., & Sankaranarayanan, S. (2019). Robustness of specifications and its applications to falsification, parameter mining, and runtime monitoring with S-TaLiRo. *International Conference on Runtime Verification (RV)*, 27–47.
- Fainekos, G. E., & Giannakoglou, K. C. (2003). Inverse design of airfoils based on a novel formulation of the ant colony optimization method. *Inverse Problems in Engineering*, 11(1), 21–38.
- Fainekos, G. E., & Pappas, G. J. (2009). Robustness of temporal logic specifications for continuous-time signals. *Theoretical Computer Science*, 410(42), 4262–4291.
- Federal Aviation Administration. (2019). *FAA aerospace forecast fiscal years 2020–2040* (tech. rep.). Federal Aviation Administration.
- Fehnker, A., & Ivani, F. (2004). Benchmarks for hybrid systems verification. *ACM International Conference on Hybrid Systems: Computation and Control (HSCC)*, 326–341.
- Fikes, R. E., & Nilsson, N. J. (1971). Strips: A new approach to the application of theorem proving to problem solving. *Artificial intelligence*, 2(3-4), 189–208.
- Fitting, M. (2012). *First-order logic and automated theorem proving*. Springer Science & Business Media.
- Francès, G., Ramirez, M., Lipovetzky, N., & Geffner, H. (2017). Purely declarative action representations are overrated: Classical planning with simulators. *International Joint Conference on Artificial Intelligence (IJCAI)*, 4294–4301.
- Ghallab, M., Nau, D., & Traverso, P. (2004). *Automated planning: Theory and practice*. Elsevier.
- Gu, S., Holly, E., Lillicrap, T., & Levine, S. (2017). Deep reinforcement learning for robotic manipulation with asynchronous off-policy updates. *IEEE International Conference on Robotics and Automation (ICRA)*, 3389–3396.
- Guestrin, C., Koller, D., Parr, R., & Venkataraman, S. (2003). Efficient solution algorithms for factored mdps. *Journal of Artificial Intelligence Research*, 19, 399–468.
- Hahn, G. (1972). Sample sizes for Monte Carlo simulation. *IEEE Transactions on Systems, Man, and Cybernetics*.
- Hansen, N., & Ostermeier, A. (1996). Adapting arbitrary normal mutation distributions in evolution strategies: The covariance matrix adaptation. *IEEE International Conference on Evolutionary Computation*, 312–317.
- Harrison, M. T. (2012). Conservative hypothesis tests and confidence intervals using importance sampling. *Biometrika*, 99(1), 57–69.

- He, K., Gkioxari, G., Dollár, P., & Girshick, R. (2017). Mask R-CNN. *International Conference on Computer Vision (ICCV)*, 2961–2969.
- Hekmatnejad, M., Hoxha, B., & Fainekos, G. (2020). Search-based test-case generation by monitoring responsibility safety rules. *IEEE International Conference on Intelligent Transportation Systems (ITSC)*, 1–8.
- Hirshorn, S. R., Voss, L. D., & Bromley, L. K. (2017). *Nasa systems engineering handbook*. NASA Special Publication.
- Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural Computation*, 9(8), 1735–1780.
- Hoxha, B., Abbas, H., & Fainekos, G. (2015). Benchmarks for temporal logic requirements for automotive systems. *International Workshop on Applied Verification for Continuous and Hybrid Systems (ARCH)*, 25–30.
- Hu, J., Lygeros, J., & Sastry, S. (2000). Towards a theory of stochastic hybrid systems. *ACM International Conference on Hybrid Systems: Computation and Control (HSCC)*, 160–173.
- Huang, Z., Guo, Y., Arief, M., Lam, H., & Zhao, D. (2018). A versatile approach to evaluating and testing automated vehicles based on kernel methods. *American Control Conference (ACC)*, 4796–4802.
- Huang, Z., Lam, H., LeBlanc, D. J., & Zhao, D. (2017). Accelerated evaluation of automated vehicles using piecewise mixture models. *IEEE Transactions on Intelligent Transportation Systems*, 19(9), 2845–2855.
- Jin, X., Deshmukh, J. V., Kapinski, J., Ueda, K., & Butts, K. (2014). Powertrain control verification benchmark. *ACM International Conference on Hybrid Systems: Computation and Control (HSCC)*, 253–262.
- Julian, K. D., Lee, R., & Kochenderfer, M. J. (2020). Validation of image-based neural network controllers through adaptive stress testing. *IEEE International Conference on Intelligent Transportation Systems (ITSC)*, 1–7.
- Junghanns, A., Mauss, J., & Tatar, M. (2008). TestWeaver: A tool for simulation-based test of mechatronic designs. *International Modelica Conference*.
- Kahn, H., & Harris, T. E. (1951). Estimation of particle transmission by random sampling. *National Bureau of Standards Applied Mathematics Series*, 12, 27–30.
- Kapinski, J., Deshmukh, J. V., Jin, X., Ito, H., & Butts, K. (2016). Simulation-based approaches for verification of embedded control systems: An overview of traditional and advanced modeling, testing, and verification techniques. *IEEE Control Systems Magazine*, 36(6), 45–64.
- Karaman, S., & Frazzoli, E. (2011). Sampling-based algorithms for optimal motion planning. *The International Journal of Robotics Research*, 30(7), 846–894.
- Katoen, J.-P. (2016). The probabilistic model checking landscape. *ACM/IEEE Symposium on Logic in Computer Science (LICS)*, 31–45.

- Katz, G., Barrett, C., Dill, D. L., Julian, K., & Kochenderfer, M. J. (2017). Reluplex: An efficient SMT solver for verifying deep neural networks. *International Conference on Computer Aided Verification (CAV)*, 97–117.
- Kavraki, L. E., Svestka, P., Latombe, J.-C., & Overmars, M. H. (1996). Probabilistic roadmaps for path planning in high-dimensional configuration spaces. *IEEE Transactions on Robotics and Automation*, 12(4), 566–580.
- Kim, J., Esposito, J. M., & Kumar, V. (2005). *An RRT-based algorithm for testing and validating multi-robot controllers* (tech. rep.). Moore School of Electrical Engineering GRASP Lab.
- Kim, Y., & Kochenderfer, M. J. (2016). Improving aircraft collision risk estimation using the cross-entropy method. *Journal of Air Transportation*, 24(2), 55–62.
- Kochenderfer, M. J. (2015). *Decision making under uncertainty: Theory and application*. MIT Press.
- Kochenderfer, M. J., Holland, J. E., & Chryssanthacopoulos, J. P. (2012). Next-generation airborne collision avoidance system. *Lincoln Laboratory Journal*, 19(1), 17–33.
- Kochenderfer, M. J., & Wheeler, T. A. (2019). *Algorithms for optimization*. MIT Press.
- Koren, M., Alsaif, S., Lee, R., & Kochenderfer, M. J. (2018). Adaptive stress testing for autonomous vehicles. *IEEE Intelligent Vehicles Symposium (IV)*, 1–7.
- Koren, M., Corso, A., & Kochenderfer, M. J. (2019). The adaptive stress testing formulation. *Workshop on Safe Autonomy, Robotics: Science and Systems*.
- Koren, M., & Kochenderfer, M. J. (2019). Efficient autonomy validation in simulation with adaptive stress testing. *IEEE International Conference on Intelligent Transportation Systems (ITSC)*, 4178–4183.
- Koren, M., & Kochenderfer, M. J. (2020). Adaptive stress testing without domain heuristics using Go-Explore. *IEEE International Conference on Intelligent Transportation Systems (ITSC)*.
- Koschi, M., Pek, C., Maierhofer, S., & Althoff, M. (2019). Computationally efficient safety falsification of adaptive cruise control systems. *IEEE International Conference on Intelligent Transportation Systems (ITSC)*, 2879–2886.
- Koymans, R. (1990). Specifying real-time properties with metric temporal logic. *Real-Time Systems*, 2, 255–299.
- Kress-Gazit, H., & Pappas, G. J. (2008). Automatically synthesizing a planning and control subsystem for the DARPA urban challenge. *International Conference on Automation Science and Engineering (CASE)*, 766–771.
- Kuutti, S., Fallah, S., & Bowden, R. (2020). Training adversarial agents to exploit weaknesses in deep control policies. *IEEE International Conference on Robotics and Automation (ICRA)*, 108–114.
- LaValle, S. M. (2006). *Planning algorithms*. Cambridge University Press.

- Lavalle, S. M. (1998). *Rapidly-exploring random trees: A new tool for path planning* (tech. rep.). Iowa State University.
- LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, *521*(7553), 436–444.
- Lee, R., Mengshoel, O. J., Agogino, A. K., Giannakopoulou, D., & Kochenderfer, M. J. (2019). Adaptive stress testing of trajectory planning systems. *AIAA Scitech Intelligent Systems Conference (IS)*.
- Lee, R., Mengshoel, O. J., Saksena, A., Gardner, R., Genin, D., Silbermann, J., Owen, M., & Kochenderfer, M. J. (2020). Adaptive stress testing: Finding likely failure events with reinforcement learning. *Journal of Artificial Intelligence Research*, *69*, 1165–1201.
- Leung, K., Aréchiga, N., & Pavone, M. (2020). Back-propagation through signal temporal logic specifications: Infusing logical structure into gradient-based methods. *Workshop on Algorithmic Foundations of Robotics*.
- Lewis, F. L., Vrabie, D., & Syrmos, V. L. (2012). *Optimal control*. John Wiley & Sons.
- Lillicrap, T. P., Hunt, J. J., Pritzel, A., Heess, N., Erez, T., Tassa, Y., Silver, D., & Wierstra, D. (2016). Continuous control with deep reinforcement learning. *International Conference on Learning Representations*.
- Lipovetzky, N., & Geffner, H. (2012). Width and serialization of classical planning problems. *European Conference on Artificial Intelligence (ECAI)*, 540–545.
- Lipovetzky, N., Ramirez, M., & Geffner, H. (2015). Classical planning with simulators: Results on the Atari video games. *International Joint Conference on Artificial Intelligence (IJCAI)*, 1610–1616.
- Luersen, M. A., & Le Riche, R. (2004). Globalized Nelder–Mead method for engineering optimization. *Computers & Structures*, *82*(23-26), 2251–2260.
- Mathesen, L., Yaghoubi, S., Pedrielli, G., & Fainekos, G. (2019). Falsification of cyber-physical systems with robustness uncertainty quantification through stochastic optimization with adaptive restart. *International Conference on Automation Science and Engineering (CASE)*, 991–997.
- McDermott, D., Ghallab, M., Howe, A., Knoblock, C., Ram, A., Veloso, M., Weld, D., & Wilkins, D. (1998). *PDDL—the planning domain definition language* (tech. rep.). Yale Center for Computational Vision and Control.
- Mnih, V., Kavukcuoglu, K., Silver, D., Rusu, A. A., Veness, J., Bellemare, M. G., Graves, A., Riedmiller, M., Fidjeland, A. K., Ostrovski, G., et al. (2015). Human-level control through deep reinforcement learning. *Nature*, *518*(7540), 529–533.
- Mockus, J. (2012). *Bayesian approach to global optimization: Theory and applications* (Vol. 37). Springer Science & Business Media.
- Moss, R. J., Lee, R., Visser, N., Hochwarth, J., Lopez, J. G., & Kochenderfer, M. J. (2020). Adaptive stress testing of trajectory predictions in flight management systems. *Digital Avionics Systems Conference (DASC)*, 1–10.

- Mullins, G. E., Stankiewicz, P. G., Hawthorne, R. C., & Gupta, S. K. (2018). Adaptive generation of challenging scenarios for testing and evaluation of autonomous vehicles. *Journal of Systems and Software*, 137, 197–215.
- Nahhal, T., & Dang, T. (2007). Test coverage for continuous and hybrid systems. *International Conference on Computer Aided Verification (CAV)*, 449–462.
- Neufeld, J., Gyorgy, A., Szepesvari, C., & Schuurmans, D. (2014). Adaptive Monte Carlo via bandit allocation. *International Conference on Machine Learning (ICML)*, 1944–1952.
- Norden, J., O’Kelly, M., & Sinha, A. (2019). Efficient black-box assessment of autonomous vehicle safety. *arXiv e-prints*, Article arXiv:1912.03618, arXiv:1912.03618.
- O’Kelly, M., Sinha, A., Namkoong, H., Tedrake, R., & Duchi, J. C. (2018). Scalable end-to-end autonomous vehicle testing via rare-event simulation. *Advances in Neural Information Processing Systems (NIPS)*, 9827–9838.
- Pant, Y. V., Abbas, H., & Mangharam, R. (2017). Smooth operator: Control using the smooth robustness of temporal logic. *IEEE Conference on Control Technology and Applications (CCTA)*, 1235–1240.
- Peled, D., Vardi, M. Y., & Yannakakis, M. (1999). Black box checking. *Formal methods for protocol engineering and distributed systems* (pp. 225–240). Springer.
- Pillai, S., Ambru, R., & Gaidon, A. (2019). Superdepth: Self-supervised, super-resolved monocular depth estimation. *IEEE International Conference on Robotics and Automation (ICRA)*, 9250–9256.
- Plaku, E., Kavragi, L. E., & Vardi, M. Y. (2009). Hybrid systems: From verification to falsification by combining motion planning and discrete search. *Formal Methods in System Design*, 34(2), 157–182.
- Platzer, A., & Quesel, J.-D. (2008). KeYmaera: A hybrid theorem prover for hybrid systems (system description). *International Joint Conference on Automated Reasoning (IJ-CAR)*, 171–178.
- Pnueli, A. (1977). The temporal logic of programs. *Foundations of Computer Science, 1977*, 46–57.
- Qin, X., Aréchiga, N., Best, A., & Deshmukh, J. (2019). Automatic testing and falsification with dynamically constrained reinforcement learning. *arXiv e-prints*, Article arXiv:1910.13645, arXiv:1910.13645.
- Reactive Systems Inc. (2011). *Finding bugs in C programs with Reactis for C* (tech. rep. RSITR 2.5). Reactive Systems Inc.
- Reactive Systems Inc. (2013). *Testing and validation of Simulink models with Reactis* (tech. rep. RSITR 1.11). Reactive Systems Inc.
- Rubinstein, R. Y., & Kroese, D. P. (2013). *The cross-entropy method: A unified approach to combinatorial optimization, Monte Carlo simulation and machine learning*. Springer Science & Business Media.
- Russell, S., & Norvig, P. (2020). *Artificial intelligence: A modern approach* (4th ed.). Prentice Hall.

- Sadowsky, J. S., & Bucklew, J. A. (1990). On large deviations theory and asymptotically efficient Monte Carlo estimation. *IEEE Transactions on Information Theory*, 36(3), 579–588.
- Sankaranarayanan, S., & Fainekos, G. (2012). Falsification of temporal properties of hybrid systems using the cross-entropy method. *ACM International Conference on Hybrid Systems: Computation and Control (HSCC)*, 125–134.
- Schuler, S., Adegas, F. D., & Anta, A. (2017). Hybrid modelling of a wind turbine (benchmark proposal). In G. Frehse & M. Althoff (Eds.), *International workshop on applied verification for continuous and hybrid systems (arch)* (pp. 18–26).
- Schulman, J., Levine, S., Abbeel, P., Jordan, M., & Moritz, P. (2015). Trust region policy optimization. *International Conference on Machine Learning (ICML)*, 1889–1897.
- Schumann, J. M. (2001). *Automated theorem proving in software engineering*. Springer Science & Business Media.
- Silver, D., Huang, A., Maddison, C. J., Guez, A., Sifre, L., Van Den Driessche, G., Schrittwieser, J., Antonoglou, I., Panneershelvam, V., Lanctot, M., et al. (2016). Mastering the game of Go with deep neural networks and tree search. *Nature*, 529(7587), 484–489.
- Silver, D., Schrittwieser, J., Simonyan, K., Antonoglou, I., Huang, A., Guez, A., Hubert, T., Baker, L., Lai, M., Bolton, A., et al. (2017). Mastering the game of go without human knowledge. *Nature*, 550(7676), 354–359.
- Silvetti, S., Policriti, A., & Bortolussi, L. (2017). An active learning approach to the falsification of black box cyber-physical systems. *International Conference on Integrated Formal Methods (iFM)*, 3–17.
- Sutton, R. S., & Barto, A. G. (2018). *Reinforcement learning: An introduction*. MIT Press.
- Sutton, R. S., McAllester, D. A., Singh, S. P., & Mansour, Y. (2000). Policy gradient methods for reinforcement learning with function approximation. *Advances in Neural Information Processing Systems (NIPS)*, 1057–1063.
- Suykens, J. A. K., & Vandewalle, J. (1999). Least squares support vector machine classifiers. *Neural Processing Letters*, 9(3), 293–300.
- Thiérmard, E. (2001). An algorithm to compute bounds for the star discrepancy. *Journal of Complexity*, 17(4), 850–880.
- Tuncali, C. E., & Fainekos, G. (2019). Rapidly-exploring random trees for testing automated vehicles. *IEEE Intelligent Transportation Systems Conference (ITSC)*, 661–666.
- Tuncali, C. E., Fainekos, G., Prokhorov, D., Ito, H., & Kapinski, J. (2019). Requirements-driven test generation for autonomous vehicles with machine learning components. *IEEE Transactions on Intelligent Vehicles*.
- Tuncali, C. E., Hoxha, B., Ding, G., Fainekos, G., & Sankaranarayanan, S. (2018). Experience report: Application of falsification methods on the uxas system. *NASA Formal Methods Symposium (NFM)*, 452–459.

- Uesato, J., Kumar, A., Szepesvári, C., Erez, T., Ruderman, A., Anderson, K., Dvijotham, K., Heess, N., & Kohli, P. (2019). Rigorous agent evaluation: An adversarial approach to uncover catastrophic failures. *International Conference on Learning Representations*.
- U.S. Department of Transportation. (2018). *Automated vehicles 3.0: Preparing for the future of transportation* (tech. rep.). U.S. Department of Transportation.
- Vinyals, O., Babuschkin, I., Czarnecki, W. M., Mathieu, M., Dudzik, A., Chung, J., Choi, D. H., Powell, R., Ewalds, T., Georgiev, P., et al. (2019). Grandmaster level in starcraft II using multi-agent reinforcement learning. *Nature*, 575(7782), 350–354.
- Wang, Z., Hutter, F., Zoghi, M., Matheson, D., & de Freitas, N. (2016). Bayesian optimization in a billion dimensions via random embeddings. *Journal of Artificial Intelligence Research*, 55, 361–387.
- Wicker, M., Huang, X., & Kwiatkowska, M. (2018). Feature-guided black-box safety testing of deep neural networks. *International Conference on Tools and Algorithms for the Construction and Analysis of Systems (TACAS)*, 408–426.
- Yaghoubi, S., & Fainekos, G. (2019). Gray-box adversarial testing for control systems with machine learning components. *ACM International Conference on Hybrid Systems: Computation and Control (HSCC)*, 179–184.
- Yang, H. (2013). *Dynamic programming algorithm for computing temporal logic robustness* (Master’s thesis). Arizona State University.
- Yang, X., Egorov, M., Evans, A., Munn, S., & Wei, P. (2020). Stress testing of UAS traffic management decision making systems. *AIAA AVIATION Forum*, 2868.
- Yeh, D. (2018). Autonomous systems and the challenges in verification, validation, and test. *IEEE Design & Test*, 35(3), 89–97.
- Zhang, Z., Ernst, G., Sedwards, S., Arcaini, P., & Hasuo, I. (2018). Two-layered falsification of hybrid systems guided by Monte Carlo tree search. *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, 37(11), 2894–2905.
- Zhang, Z., Hasuo, I., & Arcaini, P. (2019). Multi-armed bandits for Boolean connectives in hybrid system falsification. In I. Dillig & S. Tasiran (Eds.), *Computer aided verification (cav)* (pp. 401–420). Springer International Publishing.
- Zhao, D., Lam, H., Peng, H., Bao, S., LeBlanc, D. J., Nobukawa, K., & Pan, C. S. (2016). Accelerated evaluation of automated vehicles safety in lane-change scenarios based on importance sampling techniques. *IEEE Transactions on Intelligent Transportation Systems*, 18(3), 595–607.
- Zhao, Q., Krogh, B. H., & Hubbard, P. (2003). Generating test inputs for embedded control systems. *IEEE Control Systems Magazine*, 23(4), 49–57.
- Zou, X., Alexander, R., & McDermid, J. (2014). Safety validation of sense and avoid algorithms using simulation and evolutionary search. *International Conference on Computer Safety, Reliability, and Security (SafeComp)*, 33–48.

- Zutshi, A., Deshmukh, J. V., Sankaranarayanan, S., & Kapinski, J. (2014). Multiple shooting, CEGAR-based falsification for hybrid systems. *International Conference on Embedded Software (ICESS)*, 1–10.
- Zutshi, A., Sankaranarayanan, S., Deshmukh, J. V., & Kapinski, J. (2013). A trajectory splicing approach to concretizing counterexamples for hybrid systems. *IEEE Conference on Decision and Control (CDC)*, 3918–3925.