Neural Closure Certificates

Alireza Nadali, Vishnu Murali, Ashutosh Trivedi, Majid Zamani
University of Colorado Boulder
{a_nadali, vishnu.murali, ashutosh.trivedi, majid.zamani}@colorado.edu

Abstract

Notions of transition invariants and closure certificates have seen recent use in the formal verification of controlled dynamical systems against ω-regular properties. The existing approaches face limitations in two directions. First, they require a closed-form mathematical expression representing the model of the system. Such an expression may be difficult to find, too complex to be of any use, or unavailable due to security or privacy constraints. Second, finding such invariants typically rely on optimization techniques such as sum-of-squares (SOS) or satisfiability modulo theory (SMT) solvers. This restricts the classes of systems that need to be formally verified. To address these drawbacks, we introduce a notion of *neural closure certificates*. We present a data-driven algorithm that trains a neural network to represent a closure certificate. Our approach is formally correct under some mild assumptions, *i.e.*, one is able to formally show that the unknown system satisfies the ω-regular property of interest if a neural closure certificate can be computed. Finally, we demonstrate the efficacy of our approach with relevant case studies.

Introduction

The recent advances in deep learning and neural networks have revolutionized the capability in perception and natural language processing (Voulodimos et al. 2018; Otter, Medina, and Kalita 2020). As a result, control and robotics applications in modern critical infrastructure are progressively incorporating deep neural networks into their operations, leading to the grand challenge of ensuring safety in learning-enabled cyber-physical systems. Examples of safety-critical such systems include implantable medical devices, autonomous vehicles, and surgical robots.

In response to this overarching challenge, substantial efforts have been directed (Haesaert, Van den Hof, and Abate 2017; Ratschan 2017; Nejati et al. 2023) towards adapting successful classical formal and statistical approaches to verify systems with unknown dynamics. Among these approaches, neural network based barrier certificates have shown considerable promise. These certificates capture state invariants providing a certificate for safety ("nothing bad happens"). We extend this line of research by developing neural network based certificates to capture liveness properties ("something good eventually happens") by approximating the transitive closure of state transitions. We dub these certificates *neural closure certificates*.

Neural Barrier Certificates for Safety. A key contribution in the use of neural networks as a verification tool has been that of *neural barrier certificates* (Zhao et al. 2020) to demonstrate the safety of dynamical systems. We say that a dynamical system is *safe* with respect to a set of unsafe states, if it never reaches these states as it evolves. To address the problem of safety verification, Prajna and Jadbabaie (2004) proposed a notion of *barrier certificates*.

A barrier certificate is real-valued function that is nonpositive over the initial states, positive over the unsafe states and non-increasing over the evolution of the system. These conditions together ensure that the system is safe. Observe that a barrier certificate is a functional *inductive state invariant*. Every initial state has a barrier value which is nonpositive. Furthermore, given any state with a nonpositive barrier value, all of its successors must also be nonpositive via an inductive proof. Lastly, as the unsafe states do not have a nonpositive barrier value, they cannot be reached from the initial states. As the search for a barrier certificate involves finding a function that satisfies the above constraints, there has been a widespread adoption of novel techniques and approaches to make use of neural networks to represent such certificates (Abate et al. 2021; Dawson, Gao, and Fan 2023; Nadali, Trivedi, and Zamani 2023).

Beyond Safety. While neural barrier certificates have seen significant success in ensuring the safety of systems, they are inadequate in certifying an important class of specifications known as *liveness*. A system satisfies a liveness property *if something good eventually happens* (Alpern and Schneider 1987). An example of such a property is to ensure that a system visits a set of states infinitely often. While safety can be demonstrated using an inductive argument, one needs to develop a well-foundedness argument (Cook 2009) to establish liveness. To address this question, Podelski and Rybalchenko (2004) proposed a notion of *transition invariants*.

A transition invariant is a set of pairs of states such that it may be possible to reach the second element from the first. This represents a superset of the transitive closure of the transition relation that represents the evolution of the system. Using transition invariants, they developed a well-foundedness argument to refute a liveness property to prove program termination.

Certification Beyond Safety. The ω-regular properties are an expressive and well-behaved class of formal specifications that capture linear-time properties. These are precisely those properties that can be compiled into ω-automata (Vardi 2005). Their compilation into automata, allows for one to make use of algorithms over the automata, which entail automatic proofs. Podelski and Rybalchenko (2004) demonstrate how one may use transition invariants to verify programs against ω-regular properties. Inspired by this approach, Murali, Trivedi, and Zamani (2023) developed the notion of *closure certificates*, functional transition invariants, to verify dynamical systems against ω-regular properties.

Formally, a closure certificate is a real-valued function that is defined over pairs of states of the system. Consider the states x, y and z where the state y is the immediate successor of state x. As a base case, the closure certificate must be nonnegative for every pair (x, y). If the closure certificate is nonnegative for the pair (y, z), then it must also be nonnegative for the pair (x, z). These together ensure that if state z is reachable from state x, then one has the value of the closure certificate to be nonnegative for the pair (x, z). One adds additional constraints to the above to verify safety, refute liveness or prove ω-regular properties of interest.

Though closure certificates hold promise, one must address their shortcomings. Notably, current methods of searching for these certificates require knowing the precise mathematical model of the dynamical system. In many cases finding such closed-form expression is difficult, due to the complexities of the system and requires significant manual effort. Moreover, existing approaches for searching for closure certificates rely on the dynamics of the system to be of a certain form. They either rely on optimization techniques such as Sum-of-Squares programming (SOS) (Parrilo 2003) or Satisfiability Modulo Theory (SMT) (Barrett and Tinelli 2018) solvers. In the former, one requires the dynamics to be specified as a polynomial, while in the latter the dynamics need to be supported by the chosen SMT solver.

Neural Closure Certificates. We propose a novel data-driven algorithm to formally verify unknown dynamical systems against ω-regular properties. To do so, we represent closure certificates as neural networks. We construct a loss function such that it being zero in tandem with a validity condition, is sufficient to ensure that our neural network is a closure certificate, provided our system is Lipschitz-continuous. Unlike earlier SOS and SMT approaches which restrict the class of functions that act as closure certificates, our use of neural networks allows us to consider any Borel-measurable function as a candidate closure certificate.

Organization. The paper is organized as follows. We begin the formal discussion by introducing notations and definitions in the Preliminaries. We then discuss closure certificates and their properties. Next, we introduce a data-driven approach to find *Neural Closure Certificates* assuming the system dynamics are Lipschitz continuous. This ensures that our trained neural network provides a formal guarantee of correctness and proves that the system does in fact satisfy the ω-regular property of interest. Moreover, we illustrate the efficacy of our proposed algorithm with relevant case studies in the corresponding section.

Preliminaries

Notations. We use $\mathbb{N}$ and $\mathbb{R}$ to denote the set of natural numbers and reals, respectively. For any given $a \in \mathbb{R}$, $\mathbb{R}_{\geq a}$ and $\mathbb{R}_{>a}$ refer to intervals $[a, \infty)$ and (a, ∞), respectively. Likewise, for any $b \in \mathbb{N}$, $\mathbb{N}_{\geq b}$ denotes the set of natural numbers greater than or equal to b. We use $\|v\|_\infty$ to represent the infinity norm of vector $v \in \mathbb{R}^n$ for some $n \in \mathbb{N}$.

For a given alphabet A, we write A^* and A^ω for the set of finite and infinite sequences of elements in A, respectively. As usual, $|A|$, $A \setminus B$, and $A \times B$ represent the cardinality, set difference and Cartesian product of sets A and B.

We call a function $f : A \to \mathbb{R}$ *bounded* if there exist $u, l \in \mathbb{R}$ such that $l \leq f(a) \leq u$, for all $a \in A$. We define the $\text{ReLU} : \mathbb{R} \to \mathbb{R}$ activation function as $\text{ReLU}(z) = \max(0, z)$ for all $z \in \mathbb{R}$. A function $f : \mathbb{R}^n \to \mathbb{R}$ is Lipschitz continuous with respect to the infinity norm, and with Lipschitz constant $\mathfrak{L}$, if for all $x, y \in \mathbb{R}^n$, one has $\|f(x) - f(y)\|_\infty \leq \mathfrak{L}\|(x-y)\|_\infty$.

System Definition

Definition 1 (Discrete-time Dynamical System)**.** *A discrete-time dynamical system is a tuple $\mathfrak{S} = (\mathcal{X}, \mathcal{X}_0, f)$, where $\mathcal{X} \subset \mathbb{R}^n$ represents the state set, $\mathcal{X}_0 \subseteq \mathcal{X}$ denotes the initial set, and $f : \mathcal{X} \to \mathcal{X}$ is the transition function (or model) of the system. The evolution of the system can be described as:*

$$\mathfrak{S} : x(t+1) = f(x(t)). \tag{1}$$

A state sequence is an infinite series $x_0 x_1 \ldots \in \mathcal{X}^\omega$ where $x_0 \in \mathcal{X}_0$ and $x_{i+1} = f(x_i)$, for all $i \in \mathbb{N}$. We assume the state set $\mathcal{X}$ to be compact and the function f to be Lipschitz continous. Furthermore, we say that system is *unknown* when we do not have an explicit closed form expression for the function f. To reason about ω-regular properties, we employ a labeling function $\mathcal{L} : \mathcal{X} \to \Sigma$ that maps a state of the system to a letter in a finite alphabet. This labeling function extends naturally to map a state sequence to an infinite trace (or word) $w = \mathcal{L}(x_0)\mathcal{L}(x_1) \ldots \in \Sigma^\omega$. Let $T(\mathfrak{S}, \mathcal{L})$ denote the set of all traces of system $\mathfrak{S}$ under labeling function $\mathcal{L}$.

Specification

Definition 2 (Nondeterministic Büchi Automata (NBA))**.** *A nondeterministic Büchi automata (NBA) $\mathcal{A}$ is a tuple $(Q, \Sigma, q_0, \delta, Q_{acc})$ where Q denotes a finite set of states, Σ denotes a finite alphabet, $\delta : Q \times \Sigma \to 2^Q$ denotes the transition relation, and $q_0 \in Q$ and $Q_{acc} \subseteq Q$ denote the initial state*[1] *and set of accepting states, respectively.*

A *run* of $\mathcal{A}$ on an infinite word $w = \sigma_0 \sigma_1 \ldots \in \Sigma^\omega$ is an infinite sequence $\rho = q_0 q_1 q_2 \ldots \in Q^\omega$ such that $q_{i+1} \in \delta(q_i, \sigma_i)$ for all $i \in \mathbb{N}$. A word $w = w_0 w_1 \ldots$ is accepted by the NBA $\mathcal{A}$ if there exists a corresponding run $\rho = q_0 q_1 \ldots$

[1]Without loss of generality, one can convert an NBA with a set of initial states to an NBA with a single initial state.

of the NBA on w such that for every $j \in \mathbb{N}$, there exists some $i \in \mathbb{N}_{\geq j}$, such that $q_i \in Q_{acc}$.

As the state set of the NBA is finite, one can equivalently think of the set Q as the set $\{0, 1, \ldots, |Q|-1\}$. The language of an NBA $\mathcal{A}$, denoted by $L(\mathcal{A})$, is defined as the set of all words w that are accepted by it, and the language characterizes an ω-regular property. Finally, NBAs are closed under complementation (Safra 1988), *i.e.*, given an NBA $\mathcal{A}'$, one can construct its complement $\mathcal{A}$ where $L(\mathcal{A}) = \Sigma^\omega \setminus L(\mathcal{A}')$.

Verification

We say that a system satisfies an ω-regular property characterized by an NBA $\mathcal{A}'$ under labeling function $\mathcal{L}$, if $T(\mathfrak{S}, \mathcal{L}) \subseteq L(\mathcal{A}')$ and denote it as $\mathfrak{S} \models_{\mathcal{L}} \mathcal{A}'$. In the following, we implicitly infer the labeling function from the context unless it needs to be explicitly specified. Lastly, as NBAs are closed under complement, one can instead verify that $\mathfrak{S} \models_{\mathcal{L}} \mathcal{A}'$ by showing that $T(\mathfrak{S}, \mathcal{L}) \cap L(\mathcal{A}) = \emptyset$ where the NBA $\mathcal{A}$ denotes the complement of $\mathcal{A}'$. A well-known approach (automata-theoretic approach) to determine whether a system satisfies a desired ω-regular property or not is to first construct the synchronous product of the system and the (complement of) the specification.

Definition 3 (Synchronous Product). *We define the synchronous product $\mathfrak{S} \otimes \mathcal{A}$ of a system $\mathfrak{S} = (\mathcal{X}, \mathcal{X}_0, f)$, and an NBA $\mathcal{A} = (Q, \Sigma, q_0, \delta, Q_{acc})$ as the tuple $(\mathcal{X}', \mathcal{X}_0', f')$ where $\mathcal{X}' = \mathcal{X} \times \{0, \ldots, |Q| - 1\}$ and $\mathcal{X}_0' = \mathcal{X}_0 \times q_0$ denote state and initial set, respectively. The transition function $f' : \mathcal{X}' \to 2^{\mathcal{X}'}$ is defined as:*

$$f'((x, i)) = \big\{(f(x), j) \mid q_j \in \delta(q_i, \mathcal{L}(x))\big\}.$$

We should add that the set $f'((x, i))$ is finite for every $x \in \mathcal{X}$, and $q_i \in Q$.

For complex learning-enabled cyber-physical systems, it is impractical and difficult to find a closed form expression of the transition function. To overcome this challenge, our goal is to develop approaches to verify a system $\mathfrak{S}$ against an ω-regular property when the transition function is unknown. To do so, we assume that we have access to the next step evolution of the system based on samples, that is, given a sample state $\hat{x} \in \mathcal{X}$, we can get the value of $f(\hat{x})$. Furthermore, we assume that the function f is Lipschitz continuous.

Now we are in the position to state the central problem studied in this paper.

Problem 1. *Given an unknown system $\mathfrak{S}$, and an ω-regular specification, expressed by an NBA $\mathcal{A}'$, verify whether $\mathfrak{S} \models_{\mathcal{L}} \mathcal{A}'$.*

Closure Certificates

A functional approach to formally verify a dynamical system against a desired ω-regular property, is through the use of transition invariants (Podelski and Rybalchenko 2004) and closure certificates (Murali, Trivedi, and Zamani 2023). As neural closure certificates employ notions of closure certificates to provide guarantees against ω-regular properties, we provide a sketch of the main results as well as the proofs for self-containment. We refer the interested reader to (Murali, Trivedi, and Zamani 2023) for detailed proofs. First, we discuss the use of closure certificates in ensuring safety.

Definition 4 (Closure Certificate for Safety). *A bounded function $\mathcal{T} : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ is a Closure Certificate for $\mathfrak{S} = (\mathcal{X}, \mathcal{X}_0, f)$ with a set of unsafe states $\mathcal{X}_u$ if there exists a value $\delta \in \mathbb{R}_{>0}$ such that for all states $x, y \in \mathcal{X}$, $x_0 \in \mathcal{X}_0$ and $x_u \in \mathcal{X}_u$ we have:*

$$\mathcal{T}(x, f(x)) \geq 0, \tag{2}$$
$$\mathcal{T}(f(x), y) \geq 0 \implies \mathcal{T}(x, y) \geq 0, \qquad \textit{and} \tag{3}$$
$$\mathcal{T}(x_0, x_u) \leq -\delta. \tag{4}$$

Theorem 1 (Closure Certificate imply Safety). *Consider a system $\mathfrak{S} = (\mathcal{X}, \mathcal{X}_0, f)$. The existence of a function $\mathcal{T} : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ satisfying* (2)-(4) *implies that no trace of the system reaches $\mathcal{X}_u$.*

Intuitively, the first two conditions ensure that all the pairs of states that are in the transitive closure of the function f have a nonnegative closure certificate value. The first condition encodes the condition that the successor of a state $x \in \mathcal{X}$ is reachable from x and acts as the base case of induction. The second condition acts as an inductive step, where we require that if some state $y \in \mathcal{X}$ may be reachable from the state $f(x) \in \mathcal{X}$, then it is reachable from the state $x \in \mathcal{X}$. Thus, the value of $\mathcal{T}(x, y)$ is nonnegative for every state $y \in \mathcal{X}$ that is reachable from state $x \in \mathcal{X}$. Lastly, condition (4) ensures that the value of $\mathcal{T}(x_0, x_u)$ is negative for every pair of initial state $x_0 \in \mathcal{X}_0$ and unsafe sate $x_u \in \mathcal{X}_u$. Hence, no unsafe state is reachable from any initial state. Similar to how closure certificates can be used to demonstrate safety, they may also be used to prove a region is visited finitely often. This can be achieved by changing the last condition to a decreasing argument as follows.

Definition 5 (Closure Certificate for Persistence). *Consider a system $\mathfrak{S} = (\mathcal{X}, \mathcal{X}_0, f)$. A bounded function $\mathcal{T} : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ is a Closure Certificate for $\mathfrak{S}$ with a set of states $\mathcal{X}_{VF} \subseteq \mathcal{X}$, which must be visited finitely often, if there exists a value $\delta \in \mathbb{R}_{>0}$ such that for all states $x, z \in \mathcal{X}$, $x_0 \in \mathcal{X}_0$, and $y, y' \in \mathcal{X}_{VF}$ we have:*

$$\mathcal{T}(x, f(x)) \geq 0, \tag{5}$$
$$\mathcal{T}(f(x), z) \geq 0 \implies \mathcal{T}(x, z) \geq 0, \qquad \textit{and} \tag{6}$$
$$\big(\mathcal{T}(x_0, y) \geq 0\big) \wedge \big(\mathcal{T}(y, y') \geq 0\big) \implies \big(\mathcal{T}(x_0, y') \leq \mathcal{T}(x_0, y) - \delta\big). \tag{7}$$

Theorem 2 (Closure Certificate imply Persistence). *Consider a system $\mathfrak{S} = (\mathcal{X}, \mathcal{X}_0, f)$. The existence of a function $\mathcal{T} : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ satisfying conditions* (5) *to* (7) *implies that the traces of the system visit the set $\mathcal{X}_{VF}$ finitely often.*

Observe that similar to the case of safety the first two conditions overapproximate the set of states $z \in \mathcal{X}$ that are reachable from state $x \in \mathcal{X}$. Let us suppose that there exist pairs of states $y, y' \in \mathcal{X}_{VF}$ that may be reachable from an initial state $x_0 \in \mathcal{X}_0$. Then the last condition requires that the value of the closure certificate decreases for the pair (x_0, y') compared to the pair (x_0, y). As the function $\mathcal{T}$ is bounded, there can only be a finite amount of such decreases before the value of $\mathcal{T}(x_0, y')$ becomes negative. This ensures that only a finite number of values $y, y' \in \mathcal{X}_{VF}$ can occur in any given state sequence of the system and, hence, the set $\mathcal{X}_{VF}$ is visited finitely often. We make use of this idea to show how closure certificates can be used to prove that the accepting states of the complement of an NBA is visited finitely often. To do so, let us consider NBA $\mathcal{A}'$ to denote the ω-regular property of interest and the NBA $\mathcal{A}$ its complement.

Definition 6 (Closure Certificate for Büchi Objectives)**.** *Consider a system $\mathfrak{S} = (\mathcal{X}, \mathcal{X}_0, f)$ and a desired ω-regular property specified by an NBA $\mathcal{A}' = (Q', \Sigma, q'_0, \delta', Q'_{acc})$. Let the complement of NBA $\mathcal{A}'$ be an NBA $\mathcal{A} = (Q, \Sigma, q_0, \delta, Q_{acc})$. A bounded function $\mathcal{T} : \mathcal{X} \times \mathbb{N} \times \mathcal{X} \times \mathbb{N} \to \mathbb{R}$ is a closure certificate for $\mathfrak{S}$ and NBA $\mathcal{A}$ if there exists a value $\delta \in \mathbb{R}_{>0}$ such that for all states $x, y \in \mathcal{X}$, and states $q_\zeta, q_\ell \in Q$, and $q_\xi \in \delta(q_\zeta, \mathcal{L}(x))$ we have:*

$$\mathcal{T}(x, \xi, f(x), \zeta) \geq 0, \tag{8}$$

$$\mathcal{T}(f(x), \zeta, y, \ell) \geq 0 \implies \mathcal{T}(x, \xi, y, \ell) \geq 0, \tag{9}$$

and for all states $x_0 \in \mathcal{X}_0$, $y, y' \in \mathcal{X}$, and states $q_j, q_{j'} \in Q_{acc}$, we have:

$$\big(\mathcal{T}(x_0, 0, y, j) \geq 0\big) \wedge \big(\mathcal{T}(y, j, y', j') \geq 0\big) \implies \big((\mathcal{T}(x_0, 0, y', j') \leq \mathcal{T}(x_0, 0, y, j) - \delta\big). \tag{10}$$

Theorem 3 (Closure Certificate imply Büchi Objectives)**.** *Consider a system $\mathfrak{S}$ and an ω-regular property specified by the NBA $\mathcal{A}'$. Let NBA $\mathcal{A}$ denote the complement of $\mathcal{A}'$. The existence of a closure certificate $\mathcal{T}$ satisfying conditions* (8) *to* (10) *implies that $\mathfrak{S} \models_{\mathcal{L}} \mathcal{A}'$.*

The above closure certificate ensures that the accepting state of the NBA $\mathcal{A}$ is visited finitely often by the system $\mathfrak{S}$, thus, the system satisfies the original property of interest.

Neural Closure Certificates

Unfortunately, the search for a suitable closure certificate in many cases is either computationally expensive (Murali, Trivedi, and Zamani 2023) or relies on an expert choice of well-behaved functions (either specified as polynomials or functions that can be reasoned about in SMT solvers). Current methods also require the model of the system to be specified. In many real-world systems the model is unknown either due to security concerns or difficulty in finding a closed-form expression. While such close-form expressions may be unavailable, it is often a reasonable assumption to have access to samples or a black-box model, that is, for a given state $\hat{x} \in \mathcal{X}$, one has access to the value of $f(\hat{x})$. Thus, we develop a novel data-driven algorithm which relies on *sampling* to find a closure certificate for a given system to guarantee the satisfaction of a desired ω-regular property. To do so, we learn the closure certificates as neural networks and we dub our approach *neural closure certificates*.

We provide an outline for our algorithm in Figure 1. First, we partition the state space into finitely many cover elements and pick representative points from these cover elements. We then train the neural network on these sample points until a stopping criterion (the loss function l'_s in equation (18) is zero and condition (20) holds) is met. If we succeed, then the trained neural network is a neural closure certificate. Otherwise, the neural network training continues until we reach a threshold number of maximum iterations. If this fails, we change the hyperparameters of the neural network, such as the architecture of the neural network, or the learning rate, and train again. We should emphasise that the trained neural network provides a formal guarantee for the satisfaction of ω-regular property, under the assumption that the unknown transition function is Lipschitz continuous. To learn a neural network to represent a closure certificate, we first reformulate the implication in conditions (8)-(10). Our reformulation utilizes bilinearity (Gusev and Likhtarnikov 2006; Murali, Trivedi, and Zamani 2023) and is defined as follows:

Definition 7 (Encoding Closure Certificates)**.** *Consider a system $\mathfrak{S} = (\mathcal{X}, \mathcal{X}_0, f)$, and NBA $\mathcal{A}' = (Q', \Sigma, q'_0 \delta', Q'_{acc})$, representing an ω-regular property of interest. Let NBA $\mathcal{A} = (Q, \Sigma, q_0, \delta, Q_{acc})$ represent the complement of $\mathcal{A}'$. A bounded function $\mathcal{T} : \mathcal{X} \times \mathbb{N} \times \mathcal{X} \times \mathbb{N} \to \mathbb{R}$ is a closure certificate for $\mathfrak{S}$ and NBA $\mathcal{A}$ if there exist values $\delta, \lambda_1, \lambda_2 \in R_{>0}$ such that for all states $x, y \in \mathcal{X}$, and states $q_\zeta, q_\ell \in Q$, and $q_\xi \in \delta(q_\zeta, \mathcal{L}(x))$, we have:*

$$g_{1,\mathcal{T}}(x, \zeta, \xi) \geq 0, \tag{11}$$

$$g_{2,\mathcal{T}}(x, \zeta, \xi, y, \ell) \geq 0, \tag{12}$$

and for all $x_0 \in \mathcal{X}_0$, $y, y' \in \mathcal{X}$, and $q_j, q_{j'} \in Q_{acc}$, we have:

$$g_{3,\mathcal{T}}(x, 0, y, j, y', j') \geq 0, \tag{13}$$

where functions $g_{1,\mathcal{T}}$ $g_{2,\mathcal{T}}$, and $g_{3,\mathcal{T}}$ are parameterized over $\mathcal{T}$ and defined as:

$$g_{1,\mathcal{T}}(x, \zeta, \xi) = \mathcal{T}(x, \zeta, f(x), \xi), \tag{14}$$

$$g_{2,\mathcal{T}}(x, \zeta, y, \ell) = \mathcal{T}(x, \zeta, y, \ell) - \mathcal{T}(f(x), \xi, y, \ell), \text{ and} \tag{15}$$

$$g_{3,\mathcal{T}}(x, 0, y, j, y', j') = (1 - \lambda_1)\mathcal{T}(x, 0, y, j) - \mathcal{T}(x, 0, y', j') - \delta - \lambda_2 \mathcal{T}(y, j, y', j'). \tag{16}$$

Observe that a function $\mathcal{T}$ satisfying conditions (11) to (13) satisfies conditions (8) to (10). This follows from the well-known reduction of an implication to a bilinear constraint following the S-procedure (Gusev and Likhtarnikov 2006). However, note that this is only *sufficient*. Namely, one may still be able to find a closure certificate satisfying conditions (8) to (10) even if one is unable to find a closure certificate satisfying conditions (11) to (13). While the above conditions are only sufficient, we found that they ensure the loss is relatively stable while training a candidate neural closure certificate. This is in contrast to encoding the conditions as implications where the loss is much less stable and so finding a neural closure certificate is more difficult.

We now illustrate how one can represent closure certificates as neural networks. To do so, we define a function l_s with respect to a system $\mathfrak{S} = (\mathcal{X}, \mathcal{X}_0, f)$, the NBA $\mathcal{A} = (Q, \Sigma, q_0, \delta, Q_{acc})$ representing the complement of the specification of interest, and a function $\mathcal{T} : \mathcal{X} \times \mathbb{N} \times \mathcal{X} \times \mathbb{N} \to \mathbb{R}$. The function l_s is defined over the states $x, y, y' \in \mathcal{X}$,

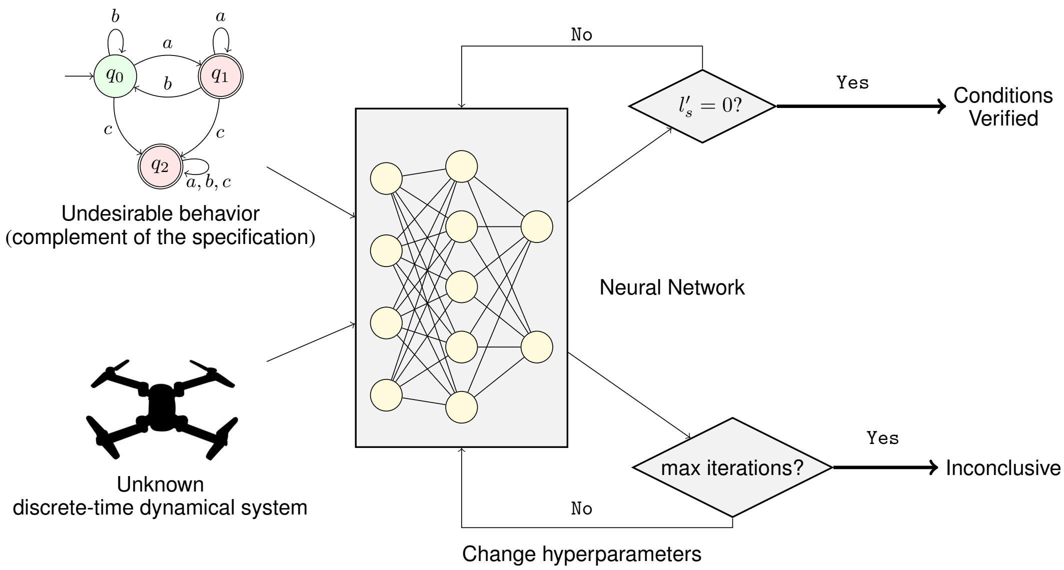


Figure 1: Our proposed algorithm. The property of interest and unknown dynamical system are the inputs. At each iteration, the algorithm checks if function l'_s (18) is zero and if the validity condition holds, if so, training stops. Otherwise, the training of the neural network continues until the maximum number of iterations is reached.

$x_0 \in \mathcal{X}_0$, $q_\ell, q_\zeta \in Q$, $q_\xi \in \delta(q_\zeta, \mathcal{L}(x))$, and $q_j, q_{j'} \in Q_{acc}$ as follows:

$$\begin{aligned} l_s = &\text{ReLU}(-g_{1,\mathcal{T}}(x,\zeta,\xi)) \\ &+ \text{ReLU}(-g_{2,\mathcal{T}}(x,\zeta,\xi,y,\ell)) \\ &+ \text{ReLU}(-g_{3,\mathcal{T}}(x_0,q_0,y,j,y',j')). \end{aligned} \tag{17}$$

Suppose that the above function is zero for all the values in its domain, then we can conclude that the function $\mathcal{T}$ satisfies conditions (11) to (13) and is thus a closure certificate. We make use of this idea to learn closure certificates using neural networks. Let us specify the template for function $\mathcal{T}$ as a neural network $F_\theta : \mathcal{X} \times \mathbb{N} \times \mathcal{X} \times \mathbb{N} \to R$ with parameters θ (the weights and biases). We choose the activation function of F_θ to be ReLU for all its neurons. This ensures that F_θ is Lipschitz-continuous over its domain. We then perturb function l_s by adding a robustness parameter η. Consider function l'_s, defined over the same domain as l_s:

$$\begin{aligned} l'_s = &\text{ReLU}(-g_{1,F_\theta}(x,\zeta,\xi)+\eta) \\ &+ \text{ReLU}(-g_{2,F_\theta}(x,\zeta,\xi,y,\ell)+\eta) \\ &+ \text{ReLU}(-g_{3,F_\theta}(x_0,0,y,j,y',j')+\eta). \end{aligned} \tag{18}$$

Observe that if $\eta \geq 0$ and l'_s is zero for all values in its domain, then function l_s is also zero with $\mathcal{T} = F_\theta$ and, hence, one obtains that the neural network F_θ is a closure certificate satisfying conditions (11) to (13). We now show how one can train the neural network F_θ over only finitely many data points with the loss function specified as l'_s to represent a closure certificate. To do so, let us partition the state set $\mathcal{X}$ into finitely many cells $\mathcal{X}_1, \mathcal{X}_2, \ldots, \mathcal{X}_M$, by picking a discretization parameter $\epsilon > 0$. We then pick sample points $x_i \in \mathcal{X}_i$ from each of these cells such that:

$$\|x - x_i\|_\infty \leq \frac{\epsilon}{2}, \text{ for all } x \in \mathcal{X}_i. \tag{19}$$

Let us denote the set of all those sampled points by $\mathcal{X}_d$. One way of partitioning the state set into such cells, is to partition it into hyperrectangles. We then pick the centers of these hyperrectangles as the representative points. Without loss of generality, we assume that labeling function $\mathcal{L}$ does not change within these hyperrectangles, meaning all points inside a hyperrectangle have the same label. That is, for any hyperrectangle $\mathcal{X}_i$, we have $\mathcal{L}(x_i) = \mathcal{L}(x)$ for all $x \in \mathcal{X}_i$, where x_i is the center of $\mathcal{X}_i$. As the function f is Lipschitz continuous and we make use of the ReLU activation function for the neural network F_θ, we have function g_{1,F_θ} to be Lipschitz continuous with respect to the first argument, function g_{2,F_θ} to be Lipschitz continuous with respect to the first, and fourth argument, and function g_{3,F_θ} to be Lipschitz continuous with respect to the first, third, and fifth arguments. Let their Lipschitz constants be $\mathfrak{L}_1$, $\mathfrak{L}_2$, and $\mathfrak{L}_3$, respectively. Let the maximum of these Lipschitz constants be denoted by $\mathfrak{L}$. We now state the main result of our paper:

Theorem 4 (Correctness of Neural Closure Certificates)**.** *Consider a neural network F_θ where θ represents its parameters, a system $\mathfrak{S} = (\mathcal{X}, \mathcal{X}_0, f)$, and an NBA $\mathcal{A}' = (Q', \Sigma, q'_0, \delta', Q'_{acc})$ representing an ω-regular property of interest. Let NBA $\mathcal{A} = (Q, \Sigma, q_0, \delta, Q_{acc})$ denote the complement of $\mathcal{A}'$. Let $\mathcal{X}_d$ be the set of sample points with a discretization parameter ϵ. Assume that the trained neural network F_θ ensures that the function l'_s is zero for all sample points and:*

$$\frac{\mathfrak{L}\epsilon}{2} - \eta \leq 0. \tag{20}$$

Then we can conclude that $\mathfrak{S} \models_{\mathcal{L}} A'$.

Proof. To prove that the system satisfies the desired property, we show that conditions (11) to (13) hold in their respective domains. First, let us consider condition (11) and states $x \in \mathcal{X}$, $q_\zeta \in Q$, and $q_\xi \in \delta(q_\zeta, \mathcal{L}(x))$. Observe that function l'_s is zero for all sample points and, hence, condition (11) holds for all points in $\mathcal{X}_d$. Now, let x be an out of sample point, *i.e.*, $x \notin \mathcal{X}_d$. By our construction, there exists some sample point $x_d \in \mathcal{X}_d$ such that $\|x - x_d\|_\infty \leq \frac{\epsilon}{2}$. Furthermore, states x and x_d are in the same cell and, hence. $\mathcal{L}(x_d) = \mathcal{L}(x)$, and $\delta(q_\zeta, \mathcal{L}(x_d)) = \delta(q_\zeta, \mathcal{L}(x))$ for all states $q_\zeta \in Q$. We thus obtain the following inequality:

$$g_{1,F_\theta}(x_d, \zeta, \xi) - g_{1,F_\theta}(x, \zeta, \xi) \leq$$
$$\mathfrak{L}\|(x, \zeta, \xi), (x_d, \zeta, \xi)\|_\infty \leq \frac{\mathfrak{L}\epsilon}{2},$$

for all states $q_\zeta \in Q$, and $q_\xi \in \delta(q_\zeta, \mathcal{L}(x_d))$.

Since $\frac{\mathfrak{L}\epsilon}{2} \leq \eta$ due to (20), one has:

$$g_{1,F_\theta}(x_d, \zeta, \xi) - g_{1,F_\theta}(x, \zeta, \xi) \leq \eta, \text{ or,}$$
$$g_{1,F_\theta}(x, \zeta, \xi) \geq g_{1,F_\theta}(x_d, \zeta, \xi) - \eta.$$

Since the function in (18) is zero at the sample point x_d, one get $g_{1,F_\theta}(x_d, \zeta, \xi) \geq \eta$. By combining these two inequalities, one gets: $g_{1,F_\theta}(x, \zeta, \xi) \geq 0$. Thus, condition (11) is satisfied for any out-of-sample point $x \in \mathcal{X} \setminus \mathcal{X}_d$ and sample points in $\mathcal{X}_d$. One can use similar reasoning for conditions (12) and (13). Therefore, the neural network F_θ is a closure certificate for the system $\mathfrak{S}$, and the NBA $\mathcal{A}$. This ensures that the system satisfies the ω-regular property represented by $\mathcal{A}'$. The proof is now complete. □

Observe that one needs to know the Lipschitz constants of functions g_{1,F_θ}, g_{2,F_θ}, and g_{3,F_θ} to formally verify the correctness of the neural closure certificate. While the system designers may not have a closed-form expression of the function f, they may be able to accurately determine the Lipschitz constant of the transition function f. In such cases, one can use existing techniques (Pauli et al. 2021) to determine the Lipschitz constant of the trained neural network F_θ and hence the value of $\mathfrak{L}$. Even if the Lipschitz constant of function f is unknown, one may make use of (Weng et al. 2018, Algorithm I) to estimate an upper bound of it. Finally, one can leverage regularization terms (Goodfellow, Bengio, and Courville 2016) to force the neural network to have a small Lipschitz constant, and hence reduce the value of $\mathfrak{L}$.

Case Studies

We demonstrate the effectiveness of neural closure certificates with relevant case studies, a 1-dimensional Kuramoto oscillator, a two room temperature, and finally an inverted pendulum model where the inverted pendulum is controlled by a neural network controller. In all case studies, we learn a candidate neural closure certificate to verify an ω-regular property whose negation is described by the NBA $\mathcal{A}$ in Figure 2. The property of interest combines both safety and persistence, and requires that the traces of a given system do not visit states with a label c, while also ensuring that they visits states with a label a only finitely often.

To train a candidate neural closure certificate, we first partition the state set $\mathcal{X}$ into finitely many hyperrectangles of diameter ϵ. We then pick the center of each of these hyperrectangles as sample points. We train a candidate neural closure certificate on these points, while minimizing the loss l'_s as specified in equation (18). If (18) is zero for all sampled points, we estimate the Lipschitz constant of the neural network, and hence $\mathfrak{L}$. We then check if condition (20) holds, and if so, we guarantee the neural network is indeed a neural closure certificate and the system satisfies the desired ω-regular property. Moreover, we ensure the neural network has a small Lipschitz constant by adding regularization terms to the loss function. We repeat the above procedure till a maximum number of iterations is reached. If we are unsuccessful, we change the hyperparameters (architecture or the discretization parameter) and train again. If this fails, our approach is inconclusive.

We employ the same neural network architecture of one hidden layer with 80 neurons to train the neural closure certificates for our case studies. The input layer is dependant on the system, and the output layer is always one-dimensional. We used the normalized Adam optimizer (Kingma and Ba 2014; Zhang 2018) to train the neural network in our implementations. All our experiments are conducted on a single Nvidia RTX 4090 graphics card coupled with an Intel 13700k CPU and 32GB of RAM. Finally, we should emphasise that while we state the transition functions in all case studies, they are purely for collecting samples. We do *not* use these functions to encode or verify the conditions, only to get the value of $f(\hat{x})$ for each sample points $\hat{x}$.

Kuramoto Oscillator

The Kuramoto model has been extensively deployed to represent chemical oscillators, with a wide array of applications in neuroscience as well as modern power system analysis (Guo et al. 2021). As the first case study, we consider system $\mathfrak{S} = (\mathcal{X}, \mathcal{X}_0, f)$ as a one dimensional Kuramoto oscillator where $\mathcal{X} = [0, 2\pi]$ and $\mathcal{X}_0 = [\frac{4\pi}{9}, \frac{5\pi}{9}]$. Furthermore, the dynamics governing the system are described as:

$$f(x) = x + \tau\Omega + t_s K \sin(-x) - 0.532x^2 + 1.69,$$

where $x \in \mathcal{X}$ denotes the phase of the oscillator, $t_s = 0.1$ is the sampling time, $\omega = 0.01$ is the natural frequency, and $K = 0.0006$ is the coupling strength.

We verify the above system against an ω-regular specification whose complement is specified by NBA $\mathcal{A}$ in Figure 2. The labeling function $\mathcal{L} : \mathcal{X} \to \Sigma$ is defined as:

$$\mathcal{L}(x) = \begin{cases} b & \text{if } x \in [\frac{11\pi}{18}, \frac{12\pi}{18}] \\ c & \text{if } x \in [\frac{7\pi}{9}, \frac{8\pi}{9}] \\ a & \text{otherwise.} \end{cases}$$

To do so we partition the state set $\mathcal{X}$ into finitely many hyperrectangles of diameter $\epsilon = 0.02$. We then select the centers of these hyperrectangles as representative sample points to train our neural network. We train the candidate neural closure certificate on the points while minimizing the loss l'_s as specified in equation (18). We stop training when equation (18) is zero for all sampled points. When we stop training, we find the values of $\delta = \eta = 0.01$. We then estimate the Lipschitz constant $\mathfrak{L} = 0.2798$, and check if condition (20) holds. For the above values, this is indeed true,

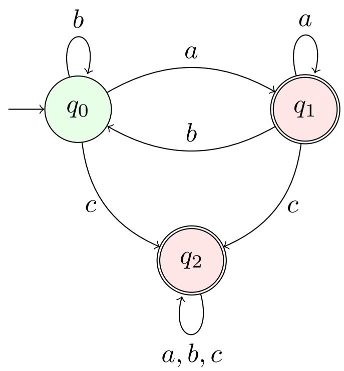
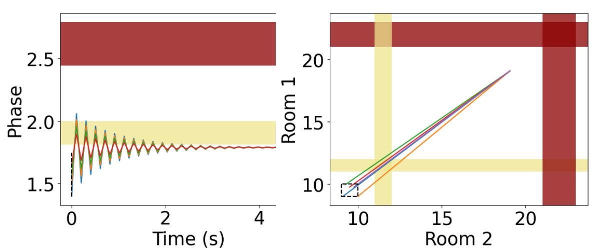


Figure 2: An NBA denoting the complement of the specification for our case studies.

and so our candidate certificate is indeed a neural closure certificate that demonstrates that the system satisfies the desired ω-regular property. The time taken for our training to converge was 5 hours.

We depict some state sequences of the system starting from some initial states in Fig 3a. One observes that these state sequences satisfy the desired ω-regular property.



(a) (b)

Figure 3: Some state sequences showing the evolution of the phase of the oscillator (Figure 3a) and the temperature of two rooms (Figure 3b). The areas marked with red and yellow indicate the unsafe and finite visit sets, respectively. We denote the initial set by the dotted black square and line.

Two Room Temperature Model

In our second case study, we consider a two dimensional room temperature model $\mathfrak{S} = (\mathcal{X}, \mathcal{X}_0, f)$ where $\mathcal{X} = [9, 19] \times [9, 19] \in \mathbb{R}^2$ represents the set of temperatures of both rooms, and $\mathcal{X}_0 = [9, 10] \times [9, 10]$ indicates the initial set. Moreover, transition function is described as:

$$f(T_1, T_2) = A \begin{bmatrix} T_1 \\ T_2 \end{bmatrix} + \mu T_h \begin{bmatrix} u(T_1) \\ u(T_2) \end{bmatrix} + \theta \begin{bmatrix} T_e \\ T_e \end{bmatrix},$$

where T_i denotes the temperature of the room $i \in \{1, 2\}$. The matrix A is:

$$A := \begin{bmatrix} 1 - 2\alpha - \theta - \mu u(T_1) & \alpha \\ \alpha & 1 - 2\alpha - \theta - \mu u(T_2) \end{bmatrix},$$

where $\alpha = 0.01$, $\theta = 0.06$, and $\mu = 0.145$ denote the conduction factors, and $u(T_i) = 0.59 - 0.011 x_i$ represents the controller for each room. Constants $T_h = 40C$ and $T_e = -5C$ indicate the heater and ambient temperatures, respectively. We consider the ω-regular property whose negation is expressed by an NBA in Figure 2 under the labeling function $\mathcal{L} : \mathcal{X} \to \Sigma$ defined as:

$$\mathcal{L}(x) = \begin{cases} b & \text{if } x \in [11, 12] \times [11, 12] \\ c & \text{if } x \in [21, 23] \times [21, 23] \\ a & \text{otherwise.} \end{cases}$$

We repeat the above procedure with a discretization parameter $\epsilon = 0.5$, and find the values of $\delta = \eta = 0.2$. We then estimate the Lipschitz constant of the conditions as $\mathfrak{L} = 0.6814$. This does in fact satisfy condition (20) and so the network is indeed a neural closure certificate. The time taken to learn the neural closure certificate was around than 4 hours. We depict some traces of the system in Figure 3b. One can readily observe that all properties of interest are satisfied.

Inverted Pendulum

As the third case study, we consider the system $\mathfrak{S} = (\mathcal{X}, \mathcal{X}_0, f)$ to be an inverted pendulum where $\mathcal{X} = [\frac{-\pi}{4}, \frac{\pi}{4}] \times [\frac{-\pi}{4}, \frac{\pi}{4}]$ and $\mathcal{X}_0 = [\frac{-\pi}{15}, \frac{\pi}{15}] \times [\frac{-\pi}{15}, \frac{\pi}{15}]$. The transition function is given by:

$$f(x_1, x_2) = \begin{bmatrix} x_1 + \tau x_2 \\ x_2 + \frac{g\tau}{l} \sin\left(x_1 + \frac{1}{ml^2} u(x_1, x_2)\right) \end{bmatrix},$$

where x_1 and x_2 are the angular position and velocity, respectively. Furthermore, $g = 9.8$ is the gravitation acceleration, and constants $l = 1$ and $m = 1$ represent the length and the mass of pendulum, and $\tau = 0.01$ is the sampling time. We represent the control input of the system by $u(x_1, x_2)$, represented by a neural network trained for safety based on (Anand and Zamani 2023, Algorithm I). We consider the ω-regular property whose complement NBA is shown in Figure 2 with the following labelling function $\mathcal{L} : \mathcal{X} \to \Sigma$:

$$\mathcal{L}(x) = \begin{cases} b & \text{if } x \in [0.1, 0.15] \times [0.1, 0.15] \\ c & \text{if } x \in X \setminus \left([-\frac{\pi}{6}, \frac{\pi}{6}] \times [-\frac{\pi}{6}, \frac{\pi}{6}]\right) \\ a & \text{otherwise.} \end{cases}$$

We then partitioning the state set into hyperrectangles of diameter $\epsilon = 0.05$ and train a candidate neural closure certificate. We find the values of $\delta = \eta = 0.2$, and then estimate the value of $\mathfrak{L} = 1.002$. The above values satisfy condition (20) and hence we have a neural closure certificate that certifies the system satisfies the ω-regular property.

Conclusion

This paper introduces a notion of neural closure certificates to verify systems against ω-regular properties. We demonstrated the effectiveness of neural closure certificates using some case studies. A key direction for future research is to leverage compositional reasoning to alleviate the sample complexity inherent to neural closure certificates. Another key challenge is to leverage neural closure certificates to aid in the correct-by-construction design of controllers for systems. A final challenge is to adapt these notions for the verification and synthesis for continuous time systems.

Acknowledgements

This work was supported in part by the NSF through the NSF CAREER awards CCF-2146563, and CNS-2145184, and is supported by grants ECCS-2015403, CNS-2039062, and CNS-2111688.

References

Abate, A.; Ahmed, D.; Edwards, A.; Giacobbe, M.; and Peruffo, A. 2021. FOSSIL: a software tool for the formal synthesis of lyapunov functions and barrier certificates using neural networks. In *Proceedings of the 24th International Conference on Hybrid Systems: Computation and Control*, 1–11.

Alpern, B.; and Schneider, F. B. 1987. Recognizing safety and liveness. *Distributed computing*, 2: 117–126.

Anand, M.; and Zamani, M. 2023. Formally verified neural network control barrier certificates for unknown systems. In *Proceedings of the 22nd World Congress of the International Federation of Automatic Control*, 2742–2747.

Barrett, C.; and Tinelli, C. 2018. *Satisfiability modulo theories*. Springer.

Cook, B. 2009. Principles of program termination. *Engineering Methods and Tools for Software Safety and Security*, 22(161): 125.

Dawson, C.; Gao, S.; and Fan, C. 2023. Safe control with learned certificates: A survey of neural lyapunov, barrier, and contraction methods for robotics and control. *IEEE Transactions on Robotics*.

Goodfellow, I.; Bengio, Y.; and Courville, A. 2016. *Deep Learning*. MIT Press. http://www.deeplearningbook.org.

Guo, Y.; Zhang, D.; Li, Z.; Wang, Q.; and Yu, D. 2021. Overviews on the applications of the Kuramoto model in modern power system analysis. *International Journal of Electrical Power & Energy Systems*, 129.

Gusev, S. V.; and Likhtarnikov, A. L. 2006. Kalman-Popov-Yakubovich lemma and the S-procedure: A historical essay. *Automation and Remote Control*, 67: 1768–1810.

Haesaert, S.; Van den Hof, P. M.; and Abate, A. 2017. Data-driven and model-based verification via Bayesian identification and reachability analysis. *Automatica*, 79: 115–126.

Kingma, D. P.; and Ba, J. 2014. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*.

Murali, V.; Trivedi, A.; and Zamani, M. 2023. Closure Certificates. *arXiv preprint arXiv:2305.17519*.

Nadali, A.; Trivedi, A.; and Zamani, M. 2023. Transfer Learning for Barrier Certificates. In *62nd IEEE Conference on Decision and Control (CDC)*, 7994–7999.

Nejati, A.; Lavaei, A.; Jagtap, P.; Soudjani, S.; and Zamani, M. 2023. Formal Verification of Unknown Discrete-and Continuous-Time Systems: A Data-Driven Approach. *IEEE Transactions on Automatic Control*.

Otter, D. W.; Medina, J. R.; and Kalita, J. K. 2020. A survey of the usages of deep learning for natural language processing. *IEEE transactions on neural networks and learning systems*, 32(2): 604–624.

Parrilo, P. A. 2003. Semidefinite programming relaxations for semialgebraic problems. *Mathematical programming*, 96: 293–320.

Pauli, P.; Koch, A.; Berberich, J.; Kohler, P.; and Allgöwer, F. 2021. Training robust neural networks using Lipschitz bounds. *IEEE Control Systems Letters*, 6: 121–126.

Podelski, A.; and Rybalchenko, A. 2004. Transition invariants. In *Proceedings of the 19th Annual IEEE Symposium on Logic in Computer Science, 2004.*, 32–41. IEEE.

Prajna, S.; and Jadbabaie, A. 2004. Safety verification of hybrid systems using barrier certificates. In *International Workshop on Hybrid Systems: Computation and Control*, 477–492.

Ratschan, S. 2017. Simulation based computation of certificates for safety of dynamical systems. In *Formal Modeling and Analysis of Timed Systems: 15th International Conference, FORMATS 2017, Berlin, Germany, September 5–7, 2017, Proceedings 15*, 303–317. Springer.

Safra, S. 1988. On the complexity of ω-automata. In *Proc. 29th IEEE Symp. Found. of Comp. Sci*, 319–327.

Vardi, M. Y. 2005. An automata-theoretic approach to linear temporal logic. *Logics for concurrency: structure versus automata*, 238–266.

Voulodimos, A.; Doulamis, N.; Doulamis, A.; Protopapadakis, E.; et al. 2018. Deep learning for computer vision: A brief review. *Computational intelligence and neuroscience*, 2018.

Weng, T.-W.; Zhang, H.; Chen, P.-Y.; Yi, J.; Su, D.; Gao, Y.; Hsieh, C.-J.; and Daniel, L. 2018. Evaluating the Robustness of Neural Networks: An Extreme Value Theory Approach. In *International Conference on Learning Representations (ICLR)*.

Zhang, Z. 2018. Improved adam optimizer for deep neural networks. In *2018 IEEE/ACM 26th international symposium on quality of service (IWQoS)*, 1–2. Ieee.

Zhao, H.; Zeng, X.; Chen, T.; and Liu, Z. 2020. Synthesizing barrier certificates using neural networks. In *Proceedings of the 23rd international conference on hybrid systems: computation and control*, 1–11.