VERIFYING OMEGA-REGULAR PROPERTIES OF NEURAL NETWORK-CONTROLLED SYSTEMS VIA PROOF CERTIFICATES*

Anonymous authors
Paper under double-blind review

ABSTRACT

With the recent advances of deep reinforcement learning techniques, nowadays in control systems, neural networks are widely adopted as control policies. However, new concerns about whether the temporal behaviors of neural network-controlled systems (NNCSs) are consistent with user-defined specifications are caused when NNCSs are deployed in the real world. In this work, we consider formal verification of temporal properties including linear temporal logic (LTL) and more generally, ω-regular properties in NNCSs, which leverages proof certificates whose existence can certify that NNCSs satisfy the properties. Concretely, given an NNCS and an ω-regular property, (i) we formally verify the satisfaction of the property; (ii) when the NNCS operates in a noisy environment and becomes stochastic, we verify whether the NNCS satisfies the property almost surely (i.e., with probability 1), both of which are achieved via closure certificates that are real-valued functions over pairs of states of NNCSs. We build our approaches into a prototype and showcase its efficacy on several popular benchmarks. Further results about quantitative verification in stochastic NNCSs are also outlined in this work.

1 INTRODUCTION

The recent advances in deep reinforcement learning (DRL) and neural networks (NNs) have revolutionized the capability in handling complex tasks where human-level intelligence is required. As a consequence, NNs are widely adopted as control policies in control systems, including autonomous vehicles (Lin et al., 2018), robotics (Xu et al., 2018), racing drones (Kaufmann et al., 2023), etc. While this facilitates the application of control systems, new concerns are caused when such neural network-controlled systems (NNCSs) are deployed in the real world, especially in safety-critical domains, e.g., aircraft collision avoidance systems (Julian et al., 2016). The reasons are that: (i) DRL techniques like reward-objective optimization lack formal guarantees for infinite-horizon behaviors of systems (König et al., 2024); (ii) NNs can be perturbed by environmental noises, adversarial attacks, etc, and thus make incorrect decisions (Wu et al., 2024), both of which can result in damages and loss. Therefore, it is non-trivial to formally ensure that the temporal behaviors of NNCSs are (highly) consistent with the user-defined specifications.

Previous work considers formal verification of temporal properties like reachability, safety, and reach-avoidance in dynamical systems (Xue et al., 2021; Zikelic et al., 2023; Ansaripour et al., 2023; Zikelic et al., 2024; Neustroev et al., 2024). However, specifications of correctness for complex systems contain complex temporal behaviors, which can be described using *linear temporal logic* (LTL) (Pnueli, 1977) and more generally, *ω-regular languages* (Thomas, 1990). As shown below, formal verification of LTL and ω-regular properties for NNCSs is highly challenging. First, these properties specify temporal behaviors over infinite system trajectories (Baier & Katoen, 2008), and verifying such infinite-horizon properties becomes especially difficult once the NNCS is subject to environmental noise, which introduces stochastic behaviors. Second, control policies in NNCSs are typically implemented by neural networks, which are complex and opaque decision-making models (Samek et al., 2021). Finally, the problem is further compounded by the continuous, potentially infinite state space and the non-linear dynamics characteristic of NNCSs, which together create significant barriers to conventional verification methods.

*An extended version of this work has been submitted to CAV 2025.

In this work, we consider formal verification of LTL and ω-regular properties in NNCSs via proof certificates which, as abstraction-free approaches, have been popular in verification and static analysis (Chakarov & Sankaranarayanan, 2013; Chatterjee et al., 2016; 2024b;a). (Murali et al., 2024) present a kind of proof certificates named *closure certificates* (CCs) for persistence, which are real-valued functions over pairs of states of dynamical systems and can be used to prove a system against ω-regualr properties. Inspired by it, we propose CCs for recurrence, which can directly prove a system satisfying (part of) ω-regular properties and thus complements CCs for persistence. Combined with the two CCs, we further propose CCs for persistence and recurrence, which can prove general ω-regular properties represented by deterministic Rabin automata (DRAs). Based on our new CCs, we formally verify the satisfaction of an ω-regular property in both deterministic and stochastic NNCSs. Note that at least two (deterministic) ω-automata are needed to generate before using CCs for persistence (see Section 4.1), hence our CCs for recurrence relieve this by requiring only one deterministic ω-automaton. Moreover, CCs for persistence are restricted to the Büchi acceptance condition which is not suitable to represent ω-regular properties in stochastic processes due to the non-determinism (Vardi, 1991; Cook et al., 2007), so our CCs for persistence and recurrence fill this gap when considering stochastic NNCSs. We use NNs to represent our proof certificates due to the powerful expressivity of NNs, and synthesize them by a counter-example guided inductive synthesis (CEGIS) approach whose correctness is also formally proved. We implement our approaches into a prototype named Veri-ω and showcase its efficacy on several popular benchmarks. Further results about quantitative verification in stochastic NNCSs (i.e., computing lower and upper bounds on the satisfaction probabilities of ω-regular properties) are outlined due to the page limit.

2 Related Work

Temporal Logic Verification via Proof Certificates. To prove a discrete-time system against ω-regular properties, (Wongpiromsarn et al., 2015) propose a conservative state-triplet approach to find barrier certificates (Prajna et al., 2007) between edges of the automaton to disallow the system from visiting an accepting state, while (Murali et al., 2024) present CCs for persistence that establish disjunctively well-founded transition invariants of systems (Podelski & Rybalchenko, 2004). (Abate et al., 2024) propose Streett supermartingales to qualitatively verify ω-regular properties of discrete-time dynamical systems. There is also a variety of work w.r.t. verification of specific temporal properties like reachability, safety and avoidance (Xue et al., 2021; Zikelic et al., 2023; Ansaripour et al., 2023; Zikelic et al., 2024; Neustroev et al., 2024).

Formal Verification of Neural Network-Controlled Systems. (Schilling et al., 2022) study the verification problem for closed-loop NNCSs and propose a reachability algorithm based on set representations. (Mandal et al., 2024) present methods for verifying safety and liveness properties for DRL systems using k-induction, and neural Lyapunov barrier certificates. (Zhi et al., 2024b) give a unified framework for both qualitative and quantitative safety verification of DNN-controlled systems via barrier certificates. (Gracia et al., 2024) consider LTL_f verification in stochastic systems, which reduces the problem into learning an uncertain Markov decision process (UMDP).

Safe Deep Reinforcement Learning. DRL techniques combined with LTL goals are proposed to shape reward functions to synthesize safe policies satisfying the LTL specifications with maximal probabilities (Yuan et al., 2019; Hasanbeig et al., 2020; 2023). There is recent work (Zhu et al., 2019; Wang & Zhu, 2023) about the combination of verification-based RL methods and programming reasoning techniques, which produces formally verified controllers. Shield learning in RL (Alshiekh et al., 2018; Carr et al., 2023) and LTL modulo theories (Rodriguez et al., 2024) allows generating shields conforming to complex safety specifications in expressive logic.

3 Preliminaries

We denote by $\mathbb{N}$, $\mathbb{Z}$ and $\mathbb{R}$ the sets of all natural numbers, integers, and real numbers, respectively.

3.1 Neural Network-controlled Systems

We consider *neural network-controlled systems* (NNCSs) (Saerens & Soquet, 1991) where the control policies are implemented by neural networks that are trained for specific tasks. Formally, an NNCS can be modeled as a tuple $M = (S, S_0, A, \pi, f, R)$, where $S \subseteq \mathbb{R}^m$ is the set of (possibly

continuous and infinite) system states, $S_0 \subseteq S$ is the set of initial states, A is the set of actions, $\pi : S \to A$ is the trained policy implemented by a neural network, $f : S \times A \to S$ is the system dynamics, and $R : S \times A \times S \to \mathbb{R}$ is the reward function. A trained NNCS $M = (S, S_0, A, \pi, f, R)$ is a decision-making system that continuously interacts with the environment. At each time step $t \in \mathbb{N}_0$, it observes a state s_t and feeds s_t into the associated neural network to compute the optimal action $a_t = \pi(s_t)$ that shall be taken. Action a_t is then performed, transitioning s_t to the successor state $s_{t+1} = f(s_t, a_t)$ via the system dynamics f and earning a reward $r_{t+1} = R(s_t, a_t, s_{t+1})$.

Stochastic NNCSs & Trajectories. Since NNCSs typically operate in an open and noisy environment (Cheng et al., 2019; Zhang et al., 2020), there is some randomness in their system dynamics. The uncertainty is captured by a global update function $g : S \times A \times W \to S$ such that the successor state is $s_{t+1} = g(s_t, a_t, w_t)$ where $w_t \in W$ is a stochastic disturbance that follows a predefined probability distribution, i.e., $w_t \sim \mu$ and the support of μ is $W = \mathsf{supp}(\mu)$. We denote such a stochastic NNCS by $M_\mu = (S, S_0, A, \pi, f, R, W, g)$.[1] A sequence $\zeta = \{s_t\}_{t \in \mathbb{N}_0}$ of states is called a *trajectory* of an NNCS, if for every $t \in \mathbb{N}_0$ we have $a_t = \pi(s_t)$, $w_t \in W$, and $s_{t+1} = g(s_t, a_t, w_t)$.

Probability Space. Given an initial state $s_0 \in S_0$, the stochastic NNCS M_μ induces a Markov process which gives rise to the probability space $(\Omega_{s_0}, \mathcal{F}_{s_0}, \mathbb{P}_{s_0})$ over the set of all trajectories that start from s_0. That is, Ω_{s_0} is the set of all trajectories starting from s_0 by the environmental interaction, $\mathcal{F}_{s_0}$ is a σ-algebra over Ω_{s_0} and $\mathbb{P}_{s_0} : \mathcal{F}_{s_0} \to [0, 1]$ is a probability measure on $\mathcal{F}_{s_0}$.

Assumptions. We assume that the state space S is compact in the Euclidean topology of $\mathbb{R}^m$, its system dynamics f (and thus global update function g) and trained policy π are Lipschitz continuous. We further assume that the system has forward invariance (Xue et al., 2021), i.e., all the states fall into the state space. These assumptions are common in control theory (Ames et al., 2019; Zikelic et al., 2023). Moreover, we require that the noise distribution μ either has bounded support or be a product of independent univariate distributions (Ansaripour et al., 2023; Zhi et al., 2024a).

3.2 Temporal Properties

ω-regular Properties. *ω-regular properties* (Thomas, 1990) provide a mathematically rich framework that encompasses LTL (see details of LTL in Appendix B) and can specify more complex behaviors that LTL cannot express. Typically, ω-regular properties are represented using ω-automata which are defined as sets of infinite words (languages) satisfying particular acceptance conditions. Below we introduce several ω-automata that can recognize ω-regular languages.

NBAs. A *nondeterministic Büchi automaton* (NBA) is a tuple $\mathcal{A} = (Q, q_0, \Sigma, \delta, Acc)$, consisting of a finite set Q of states, the initial state $q_0 \in Q$,[2] a finite alphabet Σ, the transition relation $\delta : Q \times \Sigma \to 2^Q$, and the finite set $Acc \subseteq Q$ of accepting states.

Given an NBA $\mathcal{A} = (Q, q_0, \Sigma, \delta, Acc)$, a *word* of $\mathcal{A}$, $\sigma = (\sigma_0, \sigma_1, \dots) \in \Sigma^\omega$, is an infinite sequence of letters. A *run* on a word σ, $\rho = (q_0, q_1, \dots) \in Q^\omega$, is an infinite sequence of states such that $q_{i+1} \in \delta(q_i, \sigma_i)$ for all $i \in \mathbb{N}_0$. Let $\mathrm{Inf}(\rho)$ be the set of states in ρ that are visited infinitely often. A run ρ is *accepting* if $\mathrm{Inf}(\rho) \cap Acc \neq \emptyset$. A word σ is *accepted* by $\mathcal{A}$ if there exists an accepting run ρ on the word σ. The *language* of $\mathcal{A}$, denoted by $\mathcal{L}(\mathcal{A})$, is the set of all words accepted by $\mathcal{A}$.

DRAs. A *deterministic Rabin Automaton* (DRA) is a tuple $\mathcal{A} = (Q, q_0, \Sigma, \delta, Acc)$ where Q, q_0, Σ are the same as those in NBAs, $\delta : Q \times \Sigma \to Q$ is the transition relation, and $Acc = \{(E_i, F_i) \mid E_i, F_i \subseteq Q, E_i \cap F_i = \emptyset, 1 \leq i \leq n\}$ is the acceptance condition such that a run $\rho = (q_0, q_1, \dots)$ on a word $\sigma = (\sigma_1, \sigma_1, \dots)$ is accepting iff for some $i \in [1, n]$, $\mathrm{Inf}(\rho) \cap E_i = \emptyset$ and $\mathrm{Inf}(\rho) \cap F_i \neq \emptyset$.

Note that sets E_i's shall be finitely often visited (FOV), which refers to persistence, while sets F_i's shall be infinitely often visited (IOV), which is related to recurrence. We assume that $E_i \cap F_i = \emptyset$, as it is straightforward to see (from first principles) that given a Rabin automaton with the acceptance condition $\{(E_i, F_i) \mid i \in [1, n]\}$, there is an equivalent Rabin automaton with the acceptance condition $\{(E_i^-, F_i^-) \mid i \in [1, n]\}$ where E_i^- and F_i^- are disjoint for each $i \in [1, n]$ (Casares et al., 2022).

[1] When there is no noise, i.e., $W = \emptyset$, M_μ is reduced to M and functions $g = f$.

[2] One can always convert an NBA with a set of initial states to an NBA with a single initial state.

3.3 Problem Statement

Given an NNCS $M = (S, S_0, A, \pi, f, R)$ and a finite set $\Pi = \{p_0, p_1, \ldots, p_N\}$ of atomic propositions where for each $p_i \in \Pi$, $[\![p_i]\!] \subseteq S$ denotes a subset of states of M that satisfy p_i. Define a labeling function $L : S \to 2^\Pi$ such that for any trajectory $\zeta = \{s_t\}_{t\in\mathbb{N}_0}$ of M, there is a corresponding word $\sigma = (L(s_0), L(s_1), \ldots)$ where each $L(s_t) \in 2^\Pi$. Let $Words(M)$ be the set of all words induced by M. Any LTL or ω-regular property φ over Π can be translated into an ω-automaton $\mathcal{A}_\varphi = (Q, q_0, 2^\Pi, \delta, Acc)$ that accepts the same language of φ, i.e., $\mathcal{L}(\varphi) = \mathcal{L}(\mathcal{A}_\varphi)$ (Duret-Lutz et al., 2022). Note that $M \models \varphi$ iff $Words(M) \subseteq \mathcal{L}(\varphi)$. Our problem has two parts as follows.

Problem Statement. Part I: Does M satisfy φ, i.e., $M \models \varphi$? **Part II**: In case the environment is noisy, does the stochastic NNCS $M_\mu = (S, S_0, A, \pi, f, R, W, g)$ satisfy φ almost surely?

4 Proof Certificates & Automata-based Methods

In this section, we present proof certificates and automata-based methods for the verification of LTL and ω-regular properties in NNCSs. We first introduce two types of subsets that are used throughout the work, which match sets E_i, F_i's in the Rabin acceptance condition (Section 3.2). Below we fix an NNCS $M = (S, S_0, A, \pi, f, R)$.

Definition 1 (FOV and IOV Sets)**.** A subset $S_{\text{FOV}} \subseteq S$ is an *FOV-set* (or simply FOV) just if elements of S_{FOV} must occur only finitely often in any M-trajectory, i.e., $\forall\zeta\ .\ \text{Inf}(\zeta) \cap S_{\text{FOV}} = \emptyset$. A subset $S_{\text{IOV}} \subseteq S$ is an *IOV-set* (or simply IOV) just if for every M-trajectory ζ, some element of S_{IOV} must occur infinitely often in ζ, i.e., $\forall\zeta\ .\ \text{Inf}(\zeta) \cap S_{\text{IOV}} \neq \emptyset$.

4.1 Theoretical Results of Closure Certificates

Definition 2 (Closure Certificates for Persistence)**.** A bounded function $T : S \times S \to \mathbb{R}$ is a *persistence closure certificate for a set* $S_{\text{FOV}} \subseteq S$ if there exists a constant $\epsilon > 0$ such that for all states $x, y \in S$ with $x' = f(x, \pi(x))$, and all states $x_0 \in S_0, y', y'' \in S_{\text{FOV}}$, the following conditions hold:

$$T(x, x') \geq 0 \tag{1}$$

$$T(x', y) \geq 0 \Rightarrow T(x, y) \geq 0 \tag{2}$$

$$(T(x_0, y') \geq 0 \wedge T(y', y'') \geq 0) \Rightarrow T(x_0, y'') \leq T(x_0, y') - \epsilon \tag{3}$$

Theorem 1 ((Murali et al., 2024))**.** *The existence of a persistence closure certificate in Definition 2 implies that the system M visits the set* S_{FOV} *finitely often.*

Intuition. Condition (1) and Condition (2) tell us that the non-negative-valued part of T restricts to a transitive closure over the reachable states. Condition (3) requires that for every initial state $x_0 \in S_0$ and states y', y'' in S_{FOV}, if the system can reach from y' to y'', then $T(x_0, y'')$ is less than $T(x_0, y')$ by a certain amount. Therefore, if a trajectory of M visits S_{FOV} infinitely often, the boundedness of T will be broken, which contradicts the fact.

Note that Theorem 1 only verifies the persistence of a system, i.e., a system visits a given set of states finitely often. To use it to verify that a system M satisfies an ω-regular property φ represented by an NBA $\mathcal{A}_\varphi$, one needs to (i) first construct an intermediate DRA D_φ via Safra's construction (Kozen, 2006) from $\mathcal{A}_\varphi$, flip D_φ to get another DRA $D_{\neg\varphi}$ and translate $D_{\neg\varphi}$ to $A_{\neg\varphi}$, and (ii) then prove $M \not\models \neg\varphi$ by showing that no trajectory is accepted by $\mathcal{A}_{\neg\varphi}$ via a persistence CC (see (Murali et al., 2024)). As the above automata construction is costly (i.e., the worst-case complexity is $2^{O(a \log a)}$ where a is the number of states of $\mathcal{A}_\varphi$) and unavoidable (i.e., Büchi automata are not determinizable (Thomas, 1990), so in general, given an NBA, there is not an equivalent deterministic Büchi automaton (DBA)[3]), a natural question arises: *can we directly verify the acceptance of* $\mathcal{D}_\varphi$ *holds in a system M?* Recall the Rabin acceptance condition (Section 3.2), one of the pairs conditions refers to recurrence, which cannot be proved by persistence CCs. To complement this, we propose the notion of closure certificates for recurrence, which can be leveraged to verify the recurrence of a system.

[3]One can construct the complement NBA directly by the negation LTL $\neg\varphi$. However, without the loss of generality, this does not hold for complex properties that can only be represented by ω-automata.

Table 1: The recurrence closure certificate for Cartpole (Definition 3).

Formula	Input	Output
Condition (4) - $T(x, x')$	(3.75, -0.65, -0.42, -1.73, 3.74, -0.84, -0.45, -1.59)	0.2398
Condition (4) - $T(x, x')$	(0.96, -1.55, 0.32, 0.70, 0.93, -1.36, 0.33, 0.52)	0.0648
Condition (5) - $T(x, y)$	(0.76, -0.63, 0.01, 1.39, 0.69, 0.16, 0.24, 1.69)	0.0433
Condition (5) - $T(x', y)$	(0.78, -0.83, -0.02, 1.69, 0.69, 0.16, 0.24, 1.69)	0.058
Condition (6) - $T(x_0, z)$	(-0.05, 0.04, 0.01, 0.04, 2.94, 0.43, 0.33, 0.70)	-0.0094
Condition (6) - $T(x_0, z')$	(-0.05, -0.05, -0.01, 0.04, -2.73, 1.78, 0.25, -0.38)	-0.0041

Definition 3 (Closure Certificates for Recurrence)**.** A bounded function $T : S \times S \to \mathbb{R}$ is a *recurrence closure certificate for a set* $S_{\text{IOV}} \subseteq S$ if there is a constant $\epsilon > 0$ such that for all states $x, y \in S$ with $x' = f(x, \pi(x))$, $x_0 \in S_0$, $z \in S \setminus S_{\text{IOV}}$ with $z' = f(z, \pi(z))$, the following conditions hold:

$$T(x, x') \geq 0 \tag{4}$$

$$T(x', y) \geq 0 \Rightarrow T(x, y) \geq 0 \tag{5}$$

$$T(x_0, z) \geq 0 \Rightarrow T(x_0, z') \geq T(x_0, z) + \epsilon \tag{6}$$

Theorem 2. *The existence of a recurrence closure certificate in Definition 3 implies that the system M visits the set S_{IOV} infinitely often.*

Intuition. The first two conditions in Definition 3 establish a transitive closure over the reachable states. Condition (6) requires that for every initial state $x_0 \in S_0$ and state z outside the set S_{IOV}, if z is reachable from x_0 and z' is the successor state from z, then $T(x_0, z')$ is greater than $T(x_0, z)$ by a certain amount. We can prove Theorem 2 by contradiction: if a trajectory of M stays outside S_{IOV} forever, then the boundedness of T will be violated.

Example 1. *Consider a classic RL task named Cartpole (Brockman et al., 2016). A pole is attached by an un-actuated joint to a cart. The goal of training is to learn a controller that prevents the pole from falling over by applying a force of +1 or -1 to the cart. The state space $S = (S_1, S_2, S_3, S_4)$ consists of four continuous dimensions where $S_1 = [-4.8, 4.8]$ represents the cart position, $S_2 = [-3, 3]$ means the cart velocity, $S_3 = [-0.42, 0.42]$ denotes the pole angle, and $S_4 = [-3, 3]$ means the pole velocity at tip. The initial set $S_0 = (S_0^1, S_0^2, S_0^3, S_0^4)$ where all $S_0^i = [-0.05, 0.05]$. To keep the pole upright, S_{IOV} is restricted to $S_{\text{IOV}}^1 = [-2.4, 2.4], S_{\text{IOV}}^3 = [-0.21, 0.21]$. We find a neural recurrence CC (see Section 5) and part of it is shown in Table 1. As we can see, each row in Table 1 corresponds to the CC condition in Definition 3, e.g., $T(x, x')$ is non-negative, and $T(x_0, z) < 0$ so that the pole does not fall over.*

Next, we propose a new CC to prove the existence of both persistence and recurrence, which will show its non-triviality in proving ω-regular properties represented by DRAs.

Definition 4 (Closure Certificates for Persistence and Recurrence)**.** A bounded function $T : S \times S \to \mathbb{R}$ is a closure certificate for a set $S_{\text{FOV}} \subseteq S$ and a set $S_{\text{IOV}} \subseteq S$ if there exist two constants $\epsilon, \epsilon' > 0$ such that for all states $x, y \in S$ with $x' = f(x, \pi(x))$, and all states $x_0 \in S_0$, $y', y'' \in S_{\text{FOV}}$, $z \in S \setminus (S_{\text{FOV}} \cup S_{\text{IOV}})$ with $z' = f(z, \pi(z))$, the following conditions hold:

$$T(x, x') \geq 0 \tag{7}$$

$$T(x', y) \geq 0 \Rightarrow T(x, y) \geq 0 \tag{8}$$

$$(T(x_0, y') \geq 0 \wedge T(y', y'') \geq 0) \Rightarrow T(x_0, y'') \leq T(x_0, y') - \epsilon \tag{9}$$

$$T(x_0, z) \geq 0 \Rightarrow T(x_0, z') \geq T(x_0, z) + \epsilon' \tag{10}$$

Theorem 3. *The existence of a closure certificate in Definition 4 implies that the system M visits the set S_{FOV} finitely often and the set S_{IOV} infinitely often.*

Intuition. It is straightforward to see that the first two conditions make the transitive closure, and the last two conditions refer to persistence and recurrence. Theorem 3 holds via Theorems 1 and 2.

4.2 Automata-based Verification

Given an NNCS $M = (S, S_0, A, \pi, f, R)$, a stochastic NNCS $M_\mu = (S, S_0, A, \pi, f, R, W, g)$ and an DRA $\mathcal{A}_\varphi = (Q, q_0, \Sigma, \delta, Acc)$ whose acceptance condition $Acc = \{(E_i, F_i) \mid E_i, F_i \subseteq Q, E_i \cap F_i = \emptyset, 1 \leq i \leq n\}$. To address our problem (see Section 3.3), we build two product systems as follows.

Product Systems. A product system $M \otimes \mathcal{A}_\varphi = \{X_t\}_{t\in\mathbb{N}_0}$ satisfies that: (1) $X_0 = (s_0, q_0)$ and $s_0 \in S_0$; and (2) $X_t = (s_t, q_t)$ for all $t \geq 0$, where $X_{t+1} = (f(s_t, \pi(s_t)), \delta(q_t, L(s_t)))$. A new product system $M_\mu \otimes \mathcal{A}_\varphi$ is constructed by replacing $f(s_t, \pi(s_t))$ with $g(s_t, \pi(s_t), w_t)$ in $M \otimes \mathcal{A}_\varphi$.

Recall that a Rabin acceptance condition consists of a finite set $\{(E_i, F_i) \mid i \in [i, n]\}$ of Rabin pairs, and a trajectory satisfies the acceptance condition if it satisfies one of the pairs. To establish that an NNCS satisfies a Rabin acceptance condition, it suffices to find a CC for one of the Rabin pairs.

Theorem 4. *If there is a bounded closure certificate $T : S \times Q \times S \times Q \to \mathbb{R}$ over $M \otimes \mathcal{A}_\varphi$ satisfying that for some Rabin pair (E_i, F_i), there exist two constants $\epsilon, \epsilon' > 0$ such that for all states $x, y, y', y'', z \in S$ with $x' = f(x, \pi(x))$ and $z' = f(z, \pi(z))$, $x_0 \in S_0$, $q_i, q_j \in Q$ with $q_i' = \delta(x, L(x))$, $q_k, q_l \in E_i$, and $q_m \in Q \setminus (E_i \cup F_i)$ with $q_m' = \delta(q_m, L(z))$, the following conditions hold:*

$$T((x, q_i), (x', q_i')) \geq 0 \tag{11}$$

$$T((x', q_i'), (y, q_j)) \geq 0 \Rightarrow T((x, q_i), (y, q_j)) \geq 0 \tag{12}$$

$$\begin{aligned}&(T((x_0, q_0), (y', q_k)) \geq 0 \wedge T((y', q_k), (y'', q_l)) \geq 0)\\ &\quad \Rightarrow T((x_0, q_0), (y'', q_l)) \leq T((x_0, q_0), (y', q_k)) - \epsilon\end{aligned} \tag{13}$$

$$T((x_0, q_0), (z, q_m)) \geq 0 \Rightarrow T((x_0, q_0), (z', q_m')) \geq T((x_0, q_0), (z, q_m)) + \epsilon', \tag{14}$$

then we can conclude that M satisfies the ω-regular property φ.

For a stochastic NNCS M_μ, we extend Theorem 4 by considering the effect of stochastic disturbance.

Theorem 5. *If there is a bounded closure certificate $T : S \times Q \times S \times Q \to \mathbb{R}$ over $M_\mu \otimes \mathcal{A}_\varphi$ satisfying that for some Rabin pair (E_i, F_i), there exist two constants $\epsilon, \epsilon' > 0$ such that for all states $x, y, y', y'', z \in S$ and noises $w \in W$ with $x' = g(x, \pi(x), w)$ and $z' \in g(z, \pi(z), w)$, states $x_0 \in S_0$, $q_i, q_j \in Q$ with $q_i' = \delta(x, L(x))$, $q_k, q_l \in E_i$, and $q_m \in Q \setminus (E_i \cup F_i)$ with $q_m' = \delta(q_m, L(z))$, the following conditions hold:*

$$T((x, q_i), (x', q_i')) \geq 0 \tag{15}$$

$$T((x', q_i'), (y, q_j)) \geq 0 \Rightarrow T((x, q_i), (y, q_j)) \geq 0 \tag{16}$$

$$\begin{aligned}&(T((x_0, q_0), (y', q_k)) \geq 0 \wedge T((y', q_k), (y'', q_l)) \geq 0)\\ &\Rightarrow T((x_0, q_0), (y'', q_l)) \leq T((x_0, q_0), (y', q_k)) - \epsilon\end{aligned} \tag{17}$$

$$T((x_0, q_0), (z, q_m)) \geq 0 \Rightarrow T((x_0, q_0), (z', q_m')) \geq T((x_0, q_0), (z, q_m)) + \epsilon', \tag{18}$$

then we can conclude that the system M_μ satisfies the ω-regular property φ almost surely.

Omitted proofs and other automata-based verification via NBAs or DBAs are put in Appendix C.

5 Synthesis of Neural Proof Certificates

In this section, we use NNs to represent our CCs and synthesize neural closure certificates (NCCs) by the CEGIS approach (Abate et al., 2018). We encode the validation conditions to the training loss functions. If the loss value is zero, then the training stops and it produces a valid certificate. Otherwise, it collects counter-examples and refines the partition of state space to construct larger sample sets for re-training, until a valid certificate is found or timeout. For brevity, here we show the synthesis of NCCs in Theorem 4. Other certificates can be handled in the same manner.

Training Data Construction. Since the state space S of an NNCS M can be continuous and infinite, we partition S into finitely many cells $S_1, \ldots, S_K$ via a granularity $\tau > 0$, and for each cell S_i pick sample points $s_i \in S_i$ such that $\|s_i - s\|_1 \leq \tau$ for any $s \in S_i$. Denote the finite set of all these sample points by $\tilde{S}$. As S is compact, this can be achieved by partitioning S into hyperrectangles, and collecting points in these hyperrectangles. Without the loss of generality, we assume that each point in the same cell has the same label, i.e., for any S_i, $L(s_i) = L(s)$ for all $s \in S_i$.

Loss Function Generation. A candidate NCC is initialized as a neural network T_θ w.r.t. the network parameters θ. Then T_θ is trained over the finite sample set $\tilde{S}$ by minimizing the loss functions:

$$\mathcal{L}_{CC}(\theta) := \sum_{i=1}^{4} k_i \cdot ReLU(-g_i + \beta_i) \tag{19}$$

where $k_i > 0$ are the algorithmic parameters and the constants $\beta_i \geq 0$ are used to ensure the correctness of the training (see Theorem 6). Recall Theorem 4, there are implications in the CC conditions. To ease the synthesis of NCCs, we encode these implications to their sufficient conditions via S-procedure (Gusev & Likhtarnikov, 2006) (see details in Appendix D). Here we abuse the subscripts to represent automata states, e.g., use i to stand for q_i. The functions g_i's in Eq. (19) are as follows:

$$\begin{aligned}
g_1(x, i, i') =& T((x, i), (f(x, \pi(x)), i')) \geq 0 \\
g_2(x, i, i', y, j) =& T((x, i), (y, j)) - \lambda_1 \cdot T((f(x, \pi(x)), i'), (y, j)) \geq 0 \\
g_3(x_0, 0, y', k, y'', l) =& (1 - \lambda_2) \cdot T((x_0, 0), (y', k)) - \epsilon - T((x_0, 0), (y'', l)) \\
& - \lambda_3 \cdot T((y', k), (y'', l)) \geq 0 \\
g_4(x_0, 0, z, m, m') =& T((x_0, 0), (f(z, \pi(z)), m')) - (1 + \lambda_4) \cdot T((x_0, 0), (z, m)) - \epsilon' \geq 0
\end{aligned}$$

where $\lambda_i > 0$ for any $i \in [1, 4]$. Intuitively, a loss will incur if any $g_i - \beta_i \geq 0$ is violated.

As NCCs are trained over finite sample points, even though the loss function is zero, we cannot formally guarantee that the certificate is valid over the whole state space. To address this issue, we propose the correctness theorem by leveraging the Lipschitz continuities of system dynamics, policy networks and neural certificates to provide formal guarantees (see proofs in Appendix D).

Theorem 6 (Correctness of Neural Closure Certificates)**.** *If the loss function in Eq.* (19) *is zero, and for any $i \in [1, 4]$, $\mathfrak{L}_i \tau - \beta_i \leq 0$ where $\mathfrak{L}_i$ is the Lipschitz constant of the function g_i, then the synthesized NCC is valid, i.e., all the CC conditions hold over the whole state space.*

6 Evaluation

We implement our approaches into a prototype named Veri-ω, and our experimental goals include evaluating the effectiveness of our closure certificates in both deterministic and stochastic NNCSs.

Benchmarks and Experimental Setup. We evaluate the effectiveness of our approaches on five classic NNCSs. Concretely, CartPole and MountainCar are drawn from OpenAI's Gym (Brockman et al., 2016), while B1 and B2 are widely used in state-of-the-art verification tools (Ivanov et al., 2021). Additionally, the Mars Rover task originates from (Yuan et al., 2019). Among these, CartPole is used to verify the safety property (SF), whereas the remaining tasks focus on reach-avoid verification (RV). All experiments are executed on a workstation running Ubuntu 22.04, with a 32-core AMD Ryzen Threadripper CPU, 128GB RAM, and a single 24564MiB GPU.

For stochastic NNCSs, we consider two types of state perturbations: (i) Gaussian noises with zero means and different deviations $\upsilon > 0$, and (ii) Uniform noises with zero means and different radii $r > 0$. Specifically, for each state $s = (\mathfrak{s}_1, \ldots, \mathfrak{s}_m)$, we add noises $w = (\mathfrak{w}_1, \ldots, \mathfrak{w}_m)$ to each dimension of s and obtain the perturbed state $\hat{s} = (\mathfrak{s}_1 + \mathfrak{w}_1, \ldots, \mathfrak{s}_m + \mathfrak{w}_m)$, where $\mathfrak{w}_i \sim \mathbf{G}(0, \upsilon)$ $(1 \leq i \leq m)$ is some Gaussian distributed noise or $\mathfrak{w}_i \sim \mathbf{U}(-r, r)$ $(1 \leq i \leq m)$ is some uniformly distributed noise. Due to the data sparsity of only one initial state, we randomly choose multiple initial states (instead of a single one) from the initial set S_0. We simulate 10,000 episodes starting from each of these initial states under different perturbations and use the statistical results as the baseline.

Table 2: Verification results in deterministic NNCSs.

Task	T.P.	Automata	NCC V.R.		# Vio.
			V.R.	S.T. (s)	
CP	SF	NBA	✓	18974	0
		DBA	✓	14587	0
		DRA	✓	19026	0
MC	RV	NBA	✓	19156	0
		DBA	✓	13255	0
		DRA	✓	19238	0
B1	RV	NBA	✓	18337	0
		DBA	✓	11763	0
		DRA	✓	18452	0
B2	RV	NBA	✓	19351	0
		DBA	✓	18100	0
		DRA	✓	19399	0
MR	RV	NBA	✓	22918	0
		DBA	✓	23201	0
		DRA	✓	22508	0

Remarks. T.P.: Temporal Property; **S.T.**: Synthesis Time (in seconds); **V.R.**: Verification Result; **# Vio.**: the number of property violations in simulation.

6.1 Effectivenss of ω-regular Verification in Deterministic NNCSs

We first evaluate the effectiveness of closure certificates in deterministic NNCSs, which corresponds to **Part I** of our problem (Section 3.3). For each benchmark in Table 2, we represent the same ω-regular property by different ω-automata (see the column **Automata**). As we can see, even if we use three ω-automata to represent the same

property, all three corresponding closure certificates are found (i.e., ✓ in the column **V.R.**), which shows the validity of our theoretical results. Moreover, as there is no violation in simulation, i.e., **#Vio**=0, our verification results are consistent with the simulation results. In all, results in Table 2 show the effectiveness of our method for ω-regular verification in deterministic NNCSs.

6.2 Effectivenss of ω-regular Verification in Stochastic NNCSs

Table 3 shows the qualitative verification results in stochastic NNCSs, which corresponds to **Part II** of our problem (Section 3.3). The layout is similar to that of Table 2 except that the third column specifies the magnitude of the noises. For each benchmark, when there is no noise applied to the system (e.g., **CP** with $r = 0$), we observe that a valid NCC is found, and there are no violations in the simulation results, indicating consistency between the verification and simulation outcomes. This holds true even when a small noise is introduced to the system (e.g., **CP** with $r = 0.01$), where the approach still successfully identifies a valid NCC with no violations in the simulation. However, as the perturbation magnitude increases (e.g., **CP** with $r = 0.05$), our method fails to find a valid NCC within the specified time threshold, i.e., 25,000 seconds. In this case, some violations are detected in the simulation results (e.g., **#Vio=50**), which implies that a valid NCC cannot exist. This consistency is evident across all tasks and levels of noise, demonstrating the effectiveness of our qualitative verification in stochastic NNCSs.

Table 3: Verification results in stochastic NNCSs.

Task	T.P.	Pert.	NCC V.R.		# Vio.
			V.R.	S.T. (s)	
CP	SF	$r = 0$	✓	13385	0
		$r = 0.01$	✓	13900	0
		$r = 0.05$	*N/A*	*T.O.*	**50**
MC	RV	$r = 0$	✓	18987	0
		$r = 0.01$	✓	19482	0
		$r = 0.05$	*N/A*	*T.O.*	**70**
B1	RV	$\upsilon = 0$	✓	19262	0
		$\upsilon = 0.01$	✓	19631	0
		$\upsilon = 0.1$	*N/A*	*T.O.*	**469**
B2	RV	$\upsilon = 0$	✓	18570	0
		$\upsilon = 0.1$	✓	19123	0
		$\upsilon = 0.3$	*N/A*	*T.O.*	**432**
MR	RV	$\upsilon = 0$	✓	23897	0
		$\upsilon = 0.01$	✓	21130	0
		$\upsilon = 0.15$	*N/A*	*T.O.*	**885**

Remarks. T.P.: Temporal Property; **Pert.**: perturbations; **V.R.**: Verification Result; **S.T.**: Synthesis Time (in seconds); **# Vio.**: the number of violations in simulation; *N/A*: Unknown; *T.O.*: Timeout.

7 Conclusion and Further Results

In this work, we consider temporal logic verification including LTL and ω-regular properties of NNCSs. We propose variants of closure certificates to qualitatively verify whether an NNCS formally satisfies its specification represented by some LTL or ω-regular property. When the NNCS operates in a noisy environment and the temporal behaviors become stochastic, we also verify whether the property holds almost surely. Future work could be considering other temporal properties like branch-time properties.

Further Results about Quantitative Verification. Theorem 5 is too strict to stochastic NNCSs as it requires a system M_μ satisfies an ω-regular property φ almost surely (i.e., with probability 1). When it fails to find a CC in Theorem 5, we turn to consider quantitative verification of ω-regular properties, i.e., computing the lower and upper bounds $l, u \in [0, 1]$ on the satisfaction probability of the property such that $\mathbb{P}[M_\mu \models \varphi] \in [l, u]$. Due to the page limit, here we outline the further results about quantitative verification of ω-regular properties in stochastic NNCSs, which to our knowledge, is the first time presented via proof certificates.

Given an DRA $\mathcal{A}_\varphi = (Q, q_0, \Sigma, \delta, Acc)$ whose acceptance condition $Acc = \{(E_i, F_i) \mid E_i, F_i \subseteq Q, E_i \cap F_i = \emptyset, 1 \leq i \leq n\}$. Each set E_i is an FOV-set which refers to persistence, and each set F_i is an IOV-set w.r.t. recurrence (see Definition 1). As E_i and F_i are disjoint, we handle them separately. To realize quantitative verification, we use two counters to record the number of visiting times to E_i and consecutive visiting times to $Q \setminus F_i$, respectively. By devising *stochastic persistence barrier certificates*, we have that $\mathbb{P}[E_i \text{ is FOV}] \in [l_i^{\text{fin}}, u_i^{\text{fin}}]$, while by developing *stochastic recurrence barrier certificates*, we have that $\mathbb{P}[F_i \text{ is IOV}] \in [l_i^{\text{inf}}, u_i^{\text{inf}}]$. The final probability bounds are obtained according to the Rabin acceptance condition in use.

References

Alessandro Abate, Cristina David, Pascal Kesseli, Daniel Kroening, and Elizabeth Polgreen. Counterexample guided inductive synthesis modulo theories. In *CAV'18*, pp. 270–288. Springer, 2018.

Alessandro Abate, Mirco Giacobbe, and Diptarko Roy. Stochastic omega-regular verification and control with supermartingales. In *CAV'24*, pp. 395–419. Springer, 2024.

Mohammed Alshiekh, Roderick Bloem, Rüdiger Ehlers, Bettina Könighofer, Scott Niekum, and Ufuk Topcu. Safe reinforcement learning via shielding. In *Proceedings of the AAAI conference on artificial intelligence*, volume 32, 2018.

Aaron D. Ames, Samuel Coogan, Magnus Egerstedt, Gennaro Notomista, Koushil Sreenath, and Paulo Tabuada. Control barrier functions: Theory and applications. In *ECC*, pp. 3420–3431, 2019.

Matin Ansaripour, Krishnendu Chatterjee, Thomas A Henzinger, Mathias Lechner, and Djordje Zikelic. Learning provably stabilizing neural controllers for discrete-time stochastic systems. In *ATVA'23*, pp. 357–379. Springer, 2023.

Christel Baier and Joost-Pieter Katoen. *Principles of model checking*. MIT press, 2008.

Greg Brockman, Vicki Cheung, Ludwig Pettersson, Jonas Schneider, John Schulman, Jie Tang, and Wojciech Zaremba. OpenAI Gym, 2016. arXiv:1606.01540.

Steven Carr, Nils Jansen, Sebastian Junges, and Ufuk Topcu. Safe reinforcement learning via shielding under partial observability. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 37, pp. 14748–14756, 2023.

Antonio Casares, Thomas Colcombet, and Karoliina Lehtinen. On the size of good-for-games rabin automata and its link with the memory in muller games. In *ICALP*, 2022.

Aleksandar Chakarov and Sriram Sankaranarayanan. Probabilistic program analysis with martingales. In *CAV'13*, pp. 511–526. Springer, 2013.

Krishnendu Chatterjee, Hongfei Fu, Petr Novotnỳ, and Rouzbeh Hasheminezhad. Algorithmic analysis of qualitative and quantitative termination problems for affine probabilistic programs. In *POPL'16*, pp. 327–342, 2016.

Krishnendu Chatterjee, Amir Goharshady, Ehsan Goharshady, Mehrdad Karrabi, and Djordje Zikelic. Sound and complete witnesses for template-based verification of ltl properties on polynomial programs. In *FM'24*, pp. 600–619. Springer, 2024a.

Krishnendu Chatterjee, Ehsan Kafshdar Goharshady, Petr Novotnỳ, and Djordje Zikelic. Equivalence and similarity refutation for probabilistic programs. In *PLDI'24*, volume 8, pp. 2098–2122. ACM New York, NY, USA, 2024b.

Richard Cheng, Gábor Orosz, Richard M Murray, and Joel W Burdick. End-to-end safe reinforcement learning through barrier functions for safety-critical continuous control tasks. In *AAAI'19*, volume 33, pp. 3387–3395, 2019.

Byron Cook, Alexey Gotsman, Andreas Podelski, Andrey Rybalchenko, and Moshe Y Vardi. Proving that programs eventually do something good. *ACM SIGPLAN Notices*, 42(1):265–276, 2007.

Alexandre Duret-Lutz, Etienne Renault, Maximilien Colange, Florian Renkin, Alexandre Gbaguidi Aisse, Philipp Schlehuber-Caissier, Thomas Medioni, Antoine Martin, Jérôme Dubois, Clément Gillard, et al. From spot 2.0 to spot 2.10: what's new? In *International Conference on Computer Aided Verification*, pp. 174–187. Springer, 2022.

Ibon Gracia, Luca Laurenti, Manuel Mazo Jr, Alessandro Abate, and Morteza Lahijanian. Temporal logic control for nonlinear stochastic systems under unknown disturbances. *arXiv preprint arXiv:2412.11343*, 2024.

Sergei V Gusev and Andrey Leonidovich Likhtarnikov. Kalman-popov-yakubovich lemma and the s-procedure: A historical essay. *Automation and Remote Control*, 67:1768–1810, 2006.

Hosein Hasanbeig, Daniel Kroening, and Alessandro Abate. Certified reinforcement learning with logic guidance. *Artificial Intelligence*, 322:103949, 2023.

Mohammadhosein Hasanbeig, Daniel Kroening, and Alessandro Abate. Deep reinforcement learning with temporal logics. In *FORMATS'20*, pp. 1–22. Springer, 2020.

Radoslav Ivanov, Taylor Carpenter, James Weimer, Rajeev Alur, George Pappas, and Insup Lee. Verisig 2.0: Verification of neural network controllers using taylor model preconditioning. In *CAV*, pp. 249–262, 2021.

Kyle D Julian, Jessica Lopez, Jeffrey S Brush, Michael P Owen, and Mykel J Kochenderfer. Policy compression for aircraft collision avoidance systems. In *DASC'16*, pp. 1–10. IEEE, 2016.

Elia Kaufmann, Leonard Bauersfeld, Antonio Loquercio, Matthias Müller, Vladlen Koltun, and Davide Scaramuzza. Champion-level drone racing using deep reinforcement learning. *Nature*, 620(7976):982–987, 2023.

Lukas König, Christian Heinzemann, Alberto Griggio, Michaela Klauck, Alessandro Cimatti, Franziska Henze, Stefano Tonetta, Stefan Küperkoch, Dennis Fassbender, and Michael Hanselmann. Towards safe autonomous driving: Model checking a behavior planner during development. In *International Conference on Tools and Algorithms for the Construction and Analysis of Systems*, pp. 44–65. Springer, 2024.

Dexter C Kozen. *Theory of computation*, volume 170. Springer, 2006.

Lei Lin, Siyuan Gong, Tao Li, and Srinivas Peeta. Deep learning-based human-driven vehicle trajectory prediction and its application for platoon control of connected and autonomous vehicles. In *The Autonomous Vehicles Symposium*, volume 2018, 2018.

Udayan Mandal, Guy Amir, Haoze Wu, Ieva Daukantas, Fletcher Lee Newell, Umberto Ravaioli, Baoluo Meng, Michael Durling, Kerianne Hobbs, Milan Ganai, et al. Safe and reliable training of learning-based aerospace controllers. In *2024 AIAA DATC/IEEE 43rd Digital Avionics Systems Conference (DASC)*, pp. 1–10. IEEE, 2024.

Vishnu Murali, Ashutosh Trivedi, and Majid Zamani. Closure certificates. In *Proceedings of the 27th ACM International Conference on Hybrid Systems: Computation and Control*, pp. 1–11, 2024.

Grigory Neustroev, Mirco Giacobbe, and Anna Lukina. Neural continuous-time supermartingale certificates. *arXiv preprint arXiv:2412.17432*, 2024.

Amir Pnueli. The temporal logic of programs. In *18th annual symposium on foundations of computer science (sfcs 1977)*, pp. 46–57. ieee, 1977.

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

Stephen Prajna, Ali Jadbabaie, and George J Pappas. A framework for worst-case and stochastic safety verification using barrier certificates. *IEEE Trans. Automat. Contr.*, 52(8):1415–1428, 2007.

Andoni Rodriguez, Guy Amir, Davide Corsi, Cesar Sanchez, and Guy Katz. Shield synthesis for ltl modulo theories. *arXiv preprint arXiv:2406.04184*, 2024.

Marco Saerens and Alain Soquet. Neural controller based on back-propagation algorithm. In *IEE Proceedings F (Radar and Signal Processing)*, volume 138, pp. 55–62. IET, 1991.

Wojciech Samek, Grégoire Montavon, Sebastian Lapuschkin, Christopher J Anders, and Klaus-Robert Müller. Explaining deep neural networks and beyond: A review of methods and applications. *Proceedings of the IEEE*, 109(3):247–278, 2021.

Christian Schilling, Marcelo Forets, and Sebastián Guadalupe. Verification of neural-network control systems by integrating taylor models and zonotopes. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 36, pp. 8169–8177, 2022.

Wolfgang Thomas. Automata on infinite objects. In *Formal Models and Semantics*, pp. 133–191. Elsevier, 1990.

Moshe Y Vardi. Verification of concurrent programs: The automata-theoretic framework. *Annals of Pure and Applied Logic*, 51(1-2):79–98, 1991.

Yuning Wang and He Zhu. Verification-guided programmatic controller synthesis. In *International Conference on Tools and Algorithms for the Construction and Analysis of Systems*, pp. 229–250. Springer, 2023.

Tichakorn Wongpiromsarn, Ufuk Topcu, and Andrew Lamperski. Automata theory meets barrier certificates: Temporal logic verification of nonlinear systems. *IEEE Transactions on Automatic Control*, 61(11):3344–3355, 2015.

Haoze Wu, Omri Isac, Aleksandar Zeljić, Teruhiro Tagomori, Matthew Daggitt, Wen Kokke, Idan Refaeli, Guy Amir, Kyle Julian, Shahaf Bassan, et al. Marabou 2.0: a versatile formal analyzer of neural networks. In *International Conference on Computer Aided Verification*, pp. 249–264. Springer, 2024.

Jing Xu, Zhimin Hou, Wei Wang, Bohao Xu, Kuangen Zhang, and Ken Chen. Feedback deep deterministic policy gradient with fuzzy reward for robotic multiple peg-in-hole assembly tasks. *IEEE Transactions on Industrial Informatics*, 15(3):1658–1667, 2018.

Bai Xue, Renjue Li, Naijun Zhan, and Martin Fränzle. Reach-avoid analysis for stochastic discrete-time systems. In *ACC*, pp. 4879–4885, 2021.

Lim Zun Yuan, Mohammadhosein Hasanbeig, Alessandro Abate, and Daniel Kroening. Modular deep reinforcement learning with temporal logic specifications. *arXiv preprint arXiv:1909.11591*, 2019.

Huan Zhang, Hongge Chen, Chaowei Xiao, Bo Li, Mingyan Liu, Duane Boning, and Cho-Jui Hsieh. Robust deep reinforcement learning against adversarial perturbations on state observations. *Advances in Neural Information Processing Systems*, 33:21024–21037, 2020.

Dapeng Zhi, Peixin Wang, Cheng Chen, and Min Zhang. Robustness verification of deep reinforcement learning based control systems using reward martingales. In *AAAI'24*, volume 38, pp. 19992–20000, 2024a.

Dapeng Zhi, Peixin Wang, Si Liu, C-H Luke Ong, and Min Zhang. Unifying qualitative and quantitative safety verification of dnn-controlled systems. In *CAV'24*, pp. 401–426. Springer, 2024b.

He Zhu, Zikang Xiong, Stephen Magill, and Suresh Jagannathan. An inductive synthesis framework for verifiable reinforcement learning. In *Proceedings of the 40th ACM SIGPLAN conference on programming language design and implementation*, pp. 686–701, 2019.

Dorde Zikelic, Mathias Lechner, Thomas A. Henzinger, and Krishnendu Chatterjee. Learning control policies for stochastic systems with reach-avoid guarantees. In *AAAI*, pp. 11926–11935, 2023.

Dorde Zikelic, Mathias Lechner, Abhinav Verma, Krishnendu Chatterjee, and Thomas Henzinger. Compositional policy learning in stochastic control systems with formal guarantees. *NeurIPS'24*, 36, 2024.

Appendix

A Probability Theory

We start by reviewing some notions from probability theory.

Random Variables and Stochastic Processes. A probability space is a triple $(\Omega, \mathcal{F}, \mathbb{P})$, where Ω is a non-empty sample space, $\mathcal{F}$ is a σ-algebra over Ω, and $\mathbb{P}(\cdot)$ is a probability measure over $\mathcal{F}$, i.e. a function $\mathbb{P}\colon \mathcal{F} \to [0, 1]$ that satisfies the following properties: (1) $\mathbb{P}(\emptyset) = 0$, (2)$\mathbb{P}(\Omega - A) = 1 - \mathbb{P}[A]$, and (3) $\mathbb{P}(\cup_{i=0}^{\infty} A_i) = \sum_{i=0}^{\infty} \mathbb{P}(A_i)$ for any sequence $\{A_i\}_{i=0}^{\infty}$ of pairwise disjoint sets in $\mathcal{F}$.

Given a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, a random variable is a function $X : \Omega \to \mathbb{R} \cup \{\infty\}$ that is $\mathcal{F}$-measurable, i.e., for each $a \in \mathbb{R}$ we have that $\{\omega \in \Omega | X(\omega) \leq a\} \in \mathcal{F}$. Moreover, a discrete-time stochastic process is a sequence $\{X_n\}_{n=0}^{\infty}$ of random variables in $(\Omega, \mathcal{F}, \mathbb{P})$.

Conditional Expectation. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and X be a random variable in $(\Omega, \mathcal{F}, \mathbb{P})$. The expected value of the random variable X, denoted by $\mathbb{E}[X]$, is the Lebesgue integral of X wrt $\mathbb{P}$. If the range of X is a countable set A, then $\mathbb{E}[X] = \sum_{\omega \in A} \omega \cdot \mathbb{P}(X = \omega)$. Given a sub-sigma-algebra $\mathcal{F}' \subseteq \mathcal{F}$, a conditional expectation of X for the given $\mathcal{F}'$ is a $\mathcal{F}'$-measurable random variable Y such that, for any $A \in \mathcal{F}'$, we have:

$$\mathbb{E}[X \cdot \mathbb{I}_A] = \mathbb{E}[Y \cdot \mathbb{I}_A] \tag{20}$$

Here, $\mathbb{I}_A : \Omega \to \{0, 1\}$ is an indicator function of A, defined as $\mathbb{I}_A(\omega) = 1$ if $\omega \in A$ and $\mathbb{I}_A(\omega) = 0$ if $\omega \notin A$. Moreover, whenever the conditional expectation exists, it is also almost-surely unique, i.e., for any two $\mathcal{F}'$-measurable random variables Y and Y' which are conditional expectations of X for given $\mathcal{F}'$, we have that $\mathbb{P}(Y = Y') = 1$.

Filtrations and Stopping Times. A filtration of the probability space $(\Omega, \mathcal{F}, \mathbb{P})$ is an infinite sequence $\{\mathcal{F}_n\}_{n=0}^{\infty}$ such that for every n, the triple $(\Omega, \mathcal{F}_n, \mathbb{P})$ is a probability space and $\mathcal{F}_n \subseteq \mathcal{F}_{n+1} \subseteq \mathcal{F}$. A stopping time with respect to a filtration $\{\mathcal{F}_n\}_{n=0}^{\infty}$ is a random variable $T : \Omega \to \mathbb{N}_0 \cup \{\infty\}$ such that, for every $i \in \mathbb{N}_0$, it holds that $\{\omega \in \Omega | T(\omega) \leq i\} \in \mathcal{F}_i$. Intuitively, T returns the time step at which some stochastic process shows a desired behavior and should be "stopped".

A discrete-time stochastic process $\{X_n\}_{n=0}^{\infty}$ in $(\Omega, \mathcal{F}, \mathbb{P})$ is adapted to a filtration $\{\mathcal{F}_n\}_{n=0}^{\infty}$, if for all $n \geq 0$, X_n is a random variable in $(\Omega, \mathcal{F}_n, \mathbb{P})$.

B Supplementary Materials for Section 3

Linear Temporal Logic. *Linear temporal logic* (LTL) is a specification language that allows users to define linear-time properties over infinite system traces Pnueli (1977) including safety (i.e., bad things never happen) and liveness (i.e., something good eventually happens). The syntax of LTL can be given via the following grammar:

$$\varphi := \text{true} \mid p \mid \varphi \wedge \varphi \mid \neg\varphi \mid \bigcirc\varphi \mid \varphi \,\mathcal{U}\, \varphi,$$

where $p \in \Pi$ denotes an atomic proposition in a finite set Π of atomic propositions, and symbols $\wedge, \neg$ represent the logical AND and NOT operators, respectively. The temporal operators next and until are denoted by $\bigcirc$ and $\mathcal{U}$, respectively. Note that the above operators are sufficient to derive the logical OR ($\vee$) and implication ($\Rightarrow$), and the temporal operators eventually ($\Diamond$) and always ($\Box$), respectively. For example, we have that $\Diamond\varphi = \text{true}\mathcal{U}\varphi$ and $\Box\varphi = \neg\Diamond\neg\varphi$.

Let $\sigma = (P_0, P_1, \dots) \in (2^{\Pi})^{\omega}$ be an infinite sequence over the propositions Π, and $\sigma[i+] = (P_i, P_{i+1}, \dots)$ be the suffix of σ from index i on. Then given an LTL formula φ, $\sigma \models \varphi$ means σ satisfies φ, and the satisfaction relation $\models$ is given by:

- $\sigma \models \text{true}$;
- $\sigma \models p$ iff $p \in P_0$;
- $\sigma \models \varphi_1 \wedge \varphi_2$ iff $\sigma \models \varphi_1$ and $\sigma \models \varphi_2$;
- $\sigma \models \neg\varphi$ iff $\sigma \not\models \varphi$;
- $\sigma \models \bigcirc\varphi$ iff $\sigma[1+] \models \varphi$;
- $\sigma \models \varphi_1 \mathcal{U} \varphi_2$ iff $\exists j.\ \sigma[j+] \models \varphi_2$ and $\sigma[i+] \models \varphi_1$ for all $0 \leq i < j$.

Furthermore, we express the semantics of several complicated LTL formulas that are widely used in the literature Baier & Katoen (2008).

- (Eventually) $\sigma \models \Diamond\varphi$ iff $\exists i \geq 0.\ \sigma[i+] \models \varphi$;
- (Always) $\sigma \models \Box\varphi$ iff $\forall i \geq 0.\ \sigma[i+] \models \varphi$;
- (Persistence) $\sigma \models \Diamond\Box\varphi$ iff $\exists i \geq 0\ \forall j \geq i.\ \sigma[j+] \models \varphi$;
- (Recurrence) $\sigma \models \Box\Diamond\varphi$ iff $\forall i \geq 0\ \exists j \geq i.\ \sigma[j+] \models \varphi$.

C Supplementary Materials for Section 4

C.1 Proofs of Closure Certificates

Theorem 2. Consider an NNCS $M = (S, S_0, A, \pi, f, R)$. The existence of a recurrence closure certificate in Definition 3 implies that the system M visits the set S_{IOV} infinitely often (i.e. every M-trajectory contains at least one infinitely occurring element from S_{IOV}).

Proof. We prove Theorem 2 by contradiction. Assume we find a recurrence closure certificate T as in Definition 3 but there is an M-trajectory $\zeta = (x_0, x_1, \dots)$ that starts from $x_0 \in S_0$ and visits the set S_{IOV} only finitely often, i.e., for some $j > 0$, for all $k > j$, we have $x_k \notin S_{\text{IOV}}$.

According to condition (4) and condition (5), we have $T(x_0, x_i) \geq 0$ for any index $i > 0$. As the function T is bounded, there exists a constant T^* such that $T(x, y) < T^*$ for any pairs of states $x, y \in S$. By condition (6) and induction, we have that for all index $k > j$,

$$\begin{aligned} T(x_0, x_k) &\geq T(x_0, x_{k-1}) + \epsilon \\ &\geq T(x_0, x_j) + (k - j) \cdot \epsilon \\ &\geq (k - j) \cdot \epsilon \end{aligned}$$

Therefore, there must exist an index $k' > j > 0$ such that $T(x_0, x_{k'}) > T^*$, which contradicts the assumption. □

Definition 4 (Closure Certificates for Persistence and Recurrence). Consider an NNCS $M = (S, S_0, A, \pi, f, R)$ with two subsets $S_{\text{FOV}}, S_{\text{IOV}} \subseteq S$ such that $S_{\text{FOV}} \neq S_{\text{IOV}}$. A bounded function $T : S \times S \to \mathbb{R}$ is a closure certificate for the set S_{FOV} that must be visited finitely often and the set S_{IOV} that must be visited infinitely often if there exist two constants $\epsilon, \epsilon' > 0$ such that for all states $x, y \in S$ with $x' = f(x, \pi(x))$, and all states $x_0 \in S_0$, $y', y'' \in S_{\text{FOV}}$, $z \in S \setminus (S_{\text{FOV}} \cup S_{\text{IOV}})$ with $z' = f(z, \pi(z))$, the following conditions hold:

$$T(x, x') \geq 0 \tag{7}$$

$$T(x', y) \geq 0 \Rightarrow T(x, y) \geq 0 \tag{8}$$

$$(T(x_0, y') \geq 0 \wedge T(y', y'') \geq 0) \Rightarrow T(x_0, y'') \leq T(x_0, y') - \epsilon \tag{9}$$

$$T(x_0, z) \geq 0 \Rightarrow T(x_0, z') \geq T(x_0, z) + \epsilon' \tag{10}$$

Possible Issue from Definition 4. Recall Definition 4, one may argue that the last two conditions seem to have a restriction to the allowed system trajectories, i.e., for any $y_i, y_j \in S_{\text{FOV}}$ where y_j is visited later than y_i, it implicitly requires that the length of consecutive visited x_k's in $S \setminus (S_{\text{FOV}} \cup S_{\text{IOV}})$ before y_j is smaller than that before y_i. Actually, by setting T^*, ϵ, ϵ' properly, this restriction does not hold. An alternative way to avoid the two issues is to compute T_1 for persistence and T_2 for recurrence separately, but it can cost more time. Anyway, Theorem 3 is sound.

Theorem 3 Consider an NNCS $M = (S, S_0, A, \pi, f, R)$ with two subsets $S_{\text{FOV}}, S_{\text{IOV}} \subseteq S$ such that $S_{\text{FOV}} \neq S_{\text{IOV}}$. The existence of a closure certificate in Definition 4 implies that the system M visits the set S_{FOV} finitely often and the set S_{IOV} infinitely often.

Proof. Assume we can find a closure certificate T as in Definition 4. Then suppose there exists a trajectory $\zeta = (x_0, x_1, \dots)$ of M that starts from $x_0 \in S_0$, and the trajectory ζ either (i) visits the set S_{FOV} infinitely often or (ii) visits the set S_{IOV} finitely often; or both.
For Case (i), let the infinite sequence $(y_0, y_1, \dots)$ be the states in S_{FOV} that are visited in the order, so $\zeta = (x_0, \dots, x_i, y_0, \dots, y_1, \dots)$. According to condition (7) and condition (8), we have $T(y_i, y_{i+1}) \geq 0$ and $T(x_0, y_i) \geq 0$ for all indices $i \geq 0$. By condition (9) and induction,

$$\begin{aligned} T(x_0, y_j) &\leq T(x_0, y_{j-1}) - \epsilon \\ &\leq T(x_0, y_0) - j \cdot \epsilon \\ &\leq T^* - j \cdot \epsilon \end{aligned}$$

Since the function T is bounded, there is a constant $T^* > 0$ such that $T(x, y) < T^*$ for any pairs of states $x, y \in S$. As $\epsilon > 0$, there must exist an index $j > 0$ such that $T(x_0, y_j) < 0$, which contradicts

the fact that $T(x_0, y_i) \geq 0$ for all indices $i \geq 0$. Therefore, M must visit S_{FOV} finitely often.
For Case (ii), let the finite sequence $(z_0, \dots, z_n)$ be the states in S_{IOV} that are visited in the order. Since we prove that M must visit S_{FOV} finitely often, denote by $(y_0, \dots, y_m)$ the finite sequence of states in S_{FOV} that are visited in the order. Then the whole trajectory is $\zeta = (x_0, \dots, x_i, y_0, \dots, y_m, z_0, \dots, z_n, x_j, \dots, x_k, \dots)$ where all x_i's belong to $S \setminus (S_{\text{FOV}} \cup S_{\text{IOV}})$. By condition (10) and induction, we have

$$\begin{aligned} T(x_0, x_k) &\geq T(x_0, x_{k-1}) + \epsilon' \\ &\geq T(x_0, x_j) + (k-j) \cdot \epsilon' \\ &\geq (k-j) \cdot \epsilon' \end{aligned}$$

As $\epsilon' > 0$, there must exist an index $k > j > 0$ such that $T(x_0, x_k) > T^*$, which contradicts the definition of the closure certificate. □

C.2 Other Automata-based Methods

When using NBAs to represent ω-regualr properties, one can use closure certificates for persistence to verify NNCSs against ω-regular properties.

Definition 5 (Closure Certificates for FOVs in Deterministic Systems)**.** *Consider an NNCS $M = (S, S_0, A, \pi, f, R)$ and an NBA $\mathcal{A}_{\neg\varphi} = (Q, q_0, \Sigma, \delta, Acc)$ that represents the negation of a property φ. A bounded function $T : S \times Q \times S \times Q \to \mathbb{R}$ is a closure certificate for FOV in $M \otimes \mathcal{A}_{\neg\varphi}$ if there exists a constant $\epsilon > 0$ such that for all states $x, y, y', y'' \in S$ with $x' = f(x, \pi(x))$, $x_0 \in S_0$, $q_i, q_j \in Q$ with $q_i' \in \delta(q_i, L(x))$, and $q_k, q_l \in Acc$, the following conditions hold:*

$$T((x, q_i), (x', q_i')) \geq 0 \tag{21}$$

$$T((x', q_i'), (y, q_j)) \geq 0 \Rightarrow T((x, q_i), (y, q_j)) \geq 0 \tag{22}$$

$$\begin{aligned} &(T((x_0, q_0), (y', q_k)) \geq 0 \wedge T((y', q_k), (y'', q_l)) \geq 0) \\ &\Rightarrow T((x_0, q_0), (y'', q_l)) \leq T((x_0, q_0), (y', q_k)) - \epsilon \end{aligned} \tag{23}$$

Theorem 7. *Consider an NNCS $M = (S, S_0, A, \pi, f, R)$ and a(n) (LTL) property φ. Let the NBA $\mathcal{A}_{\neg\varphi}$ represent the negation of φ. Then the existence of a closure certificate in Definition 5 implies that M satisfies the property φ.*

Proof. The proof is straightforward. By Theorem 1, a closure certificate in Definition 5 implies that the product system $M \otimes \mathcal{A}_{\neg\varphi}$ visits the accepting states finitely often and thus the system does not satisfy the property $\neg\varphi$, which implies that it satisfies φ. The concrete proof is as follows.
Suppose there exists a trace $\{(s_t, q_t)\}_{t \in \mathbb{N}_0}$ induced by $M_\mu \otimes \mathcal{A}$, where its NNCS trajectory is $\zeta = (s_0, s_1, \dots)$ and its automaton path $\rho = (q_0, q_1, \dots)$ visits the set E_i infinitely often. Let the infinite sequence $(q_0', q_1', \dots)$ be the states in E_i that are visited infinitely often in the order so that the whole path $\rho = (q_0, q_1, \dots, q_0', \dots, q_1', \dots)$. According to condition (21) and condition (22), we have that $T((s_0, q_0), (s_i', q_i')) \geq 0$ and $T((s_i', q_i'), (s_j', q_j')) \geq 0$ for any $j > i$. By condition (23) and induction, we can derive that

$$\begin{aligned} T((x_0, q_0), (s_i', q_i')) &\leq T((s_0, q_0), (s_{i-1}', q_{i-1}') - \epsilon \\ &\leq T((s_0, q_0), (s_0', q_0')) - i \cdot \epsilon \\ &\leq T^* - i \cdot \epsilon \end{aligned}$$

where the last inequality is obtained due to the fact that T is bounded from above by a constant $T^* \in \mathbb{R}$. Since $\epsilon > 0$, there should exist an index $j \in \mathbb{N}$ such that $T((x_0, q_0), (s_j', q_j')) < 0$, which contradicts the assumption. □

By closure certificates for recurrence, we can verify temporal properties represented by DBAs (though they cannot express all the ω-regualr languages).

Definition 6 (Closure Certificates for IOVs in Deterministic Systems)**.** *Consider an NNCS $M = (S, S_0, A, \pi, f, R)$ and an DBA $\mathcal{A}_\varphi = (Q, q_0, \Sigma, \delta, Acc)$ that represents the temporal property φ. A bounded function $T : S \times Q \times S \times Q \to \mathbb{R}$ is a closure certificate for IOV in $M \otimes \mathcal{A}_\varphi$ if there exists a constant $\epsilon' > 0$ such that for all states $x, y, z \in S$ with $x' = f(x, \pi(x))$ and $z' = f(z, \pi(z))$, states*

$x_0 \in S_0$, $q_i, q_j \in Q$ *with* $q_i' = \delta(q_i, L(x))$, *and* $q_m \in Q \setminus Acc$ *with* $q_m' = \delta(q_m, L(z))$, *the following conditions hold:*

$$T((x, q_i), (x', q_i')) \geq 0 \tag{24}$$

$$T((x', q_i'), (y, q_j)) \geq 0 \Rightarrow T((x, q_i), (y, q_j)) \geq 0 \tag{25}$$

$$T((x_0, q_0), (z, q_m)) \geq 0 \Rightarrow T((x_0, q_0), (z', q_m')) \geq T((x_0, q_0), (z, q_m)) + \epsilon' \tag{26}$$

Theorem 8. *Consider an NNCS* $M = (S, S_0, A, \pi, f, R)$ *and a temporal property* φ. *Let the DBA* $\mathcal{A}_\varphi$ *represent the property* φ. *Then the existence of a closure certificate in Definition 6 implies that* M *satisfies the property* φ.

Proof. The proof is straightforward. □

D Supplementary Materials for Section 5

Implication Encoding. Note that there are some implications in the definitions of our closure certificates. To ease the synthesis of these certificates, we first encode these implications to their sufficient conditions via a classical technique named S-procedure Gusev & Likhtarnikov (2006). That is, given an implication in the following form

$$f_1(x) \geq 0 \wedge f_2(x) \geq 0 \cdots \wedge f_m(x) \geq 0 \Rightarrow g(x) \geq 0, \tag{27}$$

where f_i, g are predicates over variables x, we can construct its bilinear form,

$$g'(x) = g(x) - \lambda_1 \cdot f_1(x) - \lambda_2 \cdot f_2(x) - \cdots - \lambda_m \cdot f_m(x), \tag{28}$$

where $\lambda_i > 0$ for all $i \in [1, m]$. The satisfaction of Eq. (28) implies the satisfaction of Eq. (27).

Theorem 6 (Correctness of Neural Closure Certificates). Consider an NNCS M. If the loss term $g_i - \beta_i \geq 0$ holds over the sample set $\tilde{S}$ and

$$\mathfrak{L}_i \tau - \beta_i \leq 0, \tag{29}$$

where $\mathfrak{L}_i$ is the Lipschitz constant of the function g_i, then $g_i \geq 0$ holds over the whole state space S. Moreover, if the loss function in Eq. (19) is zero over the sample set and Eq. (29) holds for all loss terms, then the neural closure certificate is valid, i.e., all the CC conditions hold over the whole state space.

Proof. Let $\mathfrak{L}_T, \mathfrak{L}_f, \mathfrak{L}_\pi$ be the Lipschitz constants of the NCC T_θ, the system dynamics f, and the system policy π, respectively. Given a state $\tilde{x} \in \tilde{S}$, and let $x \in S \setminus \tilde{S}$ be a state not in the finite sample set $\tilde{S}$ but satisfying $\|x - \tilde{x}\| \leq \tau$, so that $L(x) = L(\tilde{x})$ and $\delta(q, L(x)) = \delta(q, L(\tilde{x}))$ for all sates $q \in Q$.

- By the Lipschitz continuity of g_1 and Eq. (29), we have that for any $x \in S$, $\tilde{x} \in \tilde{S}$, $i, i' \in Q$ with $i' \in \delta(i, L(x))$,

$$\begin{aligned}
g_1(\tilde{x}, i, i') - g_1(x, i, i') &= T((\tilde{x}, i), (f(\tilde{x}, \pi(\tilde{x})), i')) - T((x, i), (f(x, \pi(x)), i')) \\
&\leq \mathfrak{L}_T \cdot \|((\tilde{x}, i), (f(\tilde{x}, \pi(\tilde{x})), i')) - ((x, i), (f(x, \pi(x)), i'))\|_1 \\
&= \mathfrak{L}_T \cdot (\|\tilde{x} - x\|_1 + \|f(\tilde{x}, \pi(\tilde{x})) - f(x, \pi(x))\|_1 \\
&\leq \mathfrak{L}_T \cdot (\|\tilde{x} - x\|_1 + \mathfrak{L}_f \cdot \|(\tilde{x}, \pi(\tilde{x})) - (x, \pi(x))\|_1) \\
&\leq \mathfrak{L}_T \cdot (\|\tilde{x} - x\|_1 + \mathfrak{L}_f \cdot (1 + \mathfrak{L}_\pi) \cdot \|\tilde{x} - x\|_1) \\
&= \mathfrak{L}_T \cdot (1 + \mathfrak{L}_f \cdot (1 + \mathfrak{L}_\pi)) \cdot \|\tilde{x} - x\|_1 \\
&\leq \mathfrak{L}_T \cdot (1 + \mathfrak{L}_f \cdot (1 + \mathfrak{L}_\pi)) \cdot \tau \\
&= \mathfrak{L}_1 \cdot \tau \leq \beta_1
\end{aligned}$$

  Therefore, we can derive that

$$g_1(x, i, i') \geq g_1(\tilde{x}, i, i') - \beta_1$$

  Because $g_1(\tilde{x}, i, i') - \beta_1 \geq 0$ for any sample points $\tilde{x} \in \tilde{S}$, we have $g_1(x, i, i') \geq 0$ for all states $x \in S \setminus \tilde{S}$, which implies $g_1(x, i, i') \geq 0$ holds over the whole state space.

- By the Lipschitz continuity of g_2 and Eq. (29), we have that for any $x, y \in S$, $\tilde{x}, \tilde{y} \in \tilde{S}$, $i, j \in Q$ with $i' \in \delta(i, L(x))$,

$$\begin{aligned}
& g_2(\tilde{x}, i, i', \tilde{y}, j) - g_2(x, i, i'y, j) \\
= \; & T((\tilde{x}, i), (\tilde{y}, j)) - T((x, i), (y, j)) \\
& +\lambda_1 \cdot [T((f(x, \pi(x)), i'), (y, j)) - T((f(\tilde{x}, \pi(\tilde{x})), i'), (\tilde{y}, j))] \\
\leq \; & \mathfrak{L}_T \cdot (\|\tilde{x} - x\|_1 + \|\tilde{y} - y\|_1) + \lambda_1 \cdot \mathfrak{L}_T \cdot (1 + \mathfrak{L}_f \cdot (1 + \mathfrak{L}_\pi)) \cdot \tau \\
\leq \; & 2\mathfrak{L}_T \cdot \tau + \lambda_1 \cdot \mathfrak{L}_T \cdot (1 + \mathfrak{L}_f \cdot (1 + \mathfrak{L}_\pi)) \cdot \tau \\
= \; & \mathfrak{L}_T \cdot (2 + \lambda_1 \cdot (1 + \mathfrak{L}_f \cdot (1 + \mathfrak{L}_\pi))) \cdot \tau \\
= \; & \mathfrak{L}_2 \cdot \tau \leq \beta_2
\end{aligned}$$

  Therefore, we can derive that

$$g_2(x, i, i'y, j) \geq g_2(\tilde{x}, i, i', \tilde{y}, j) - \beta_2$$

  Because $g_2(\tilde{x}, i, i', \tilde{y}, j) - \beta_2 \geq 0$ for any sample points $\tilde{x}, \tilde{y} \in \tilde{S}$, we have $g_2(x, i, i'y, j) \geq 0$ for all states $x, y \in S \setminus \tilde{S}$, which implies $g_2(x, i, i'y, j) \geq 0$ holds over the whole state space.

- By the Lipschitz continuity of g_3 and Eq. (29), we have that for all states $x_0 \in S_0, y', y'' \in S$, $\tilde{x}_0 \in \tilde{S}_0, \tilde{y}', \tilde{y}'' \in \tilde{S}$, $k, l \in Acc$,

$$\begin{aligned}
& g_3(\tilde{x}_0, 0, \tilde{y}', k, \tilde{y}'', l) - g_3(x_0, 0, y', k, y'', l) \\
= \; & (1 - \lambda_2) \cdot [T((\tilde{x}, 0), (\tilde{y}', k)) - T((x, 0), (y', k))] + [T((x_0, 0), (y'', l)) - T((\tilde{x}_0, 0), (\tilde{y}'', l))] \\
& +\lambda_3 \cdot [T((y', k), (y'', l)) - T((\tilde{y}', k), (\tilde{y}'', l))] \\
\leq \; & (1 - \lambda_2) \cdot \mathfrak{L}_T \cdot \|(\tilde{x} - x, 0), (\tilde{y}' - y', 0)\|_1 + \mathfrak{L}_T \cdot \|(x_0 - \tilde{x}_0, 0), (y'' - \tilde{y}'', 0)\|_1 \\
& +\lambda_3 \cdot \mathfrak{L}_T \cdot \|(y' - \tilde{y}', 0), (y'' - \tilde{y}'', 0)\|_1 \\
\leq \; & (1 - \lambda_2) \cdot \mathfrak{L}_T \cdot 2\tau + \mathfrak{L}_T \cdot 2\tau + +\lambda_3 \cdot \mathfrak{L}_T \cdot 2\tau \\
= \; & (2(1 - \lambda_2) + 2 + 2\lambda_3)\mathfrak{L}_T \cdot \tau \\
= \; & \mathfrak{L}_3 \cdot \tau \leq \beta_3
\end{aligned}$$

  Therefore, we can derive that

$$g_3(x_0, 0, y', k, y'', l) \geq g_3(\tilde{x}_0, 0, \tilde{y}', k, \tilde{y}'', l) - \beta_3$$

  Because $g_3(\tilde{x}_0, 0, \tilde{y}', k, \tilde{y}'', l) - \beta_3 \geq 0$ for any sample points, we have $g_3(x_0, 0, y', k, y'', l) \geq 0$ for all states outside the sample set, which implies $g_3(x_0, 0, y', k, y'', l) \geq 0$ holds over the whole state space.

- By the Lipschitz continuity of g_4 and Eq. (29), we have that for all states $x_0 \in S_0, z \in S$, $\tilde{x}_0 \in \tilde{S}_0, \tilde{z} \in \tilde{S}$, $m \in Q \setminus Acc$ with $m' \in \delta(m, L(z))$,

$$\begin{aligned}
& g_4(\tilde{x}_0, 0, \tilde{z}, m, m') - g_4(x_0, 0, z, m, m') \\
= \; & [T((\tilde{x}_0, 0), (f(\tilde{z}, \pi(\tilde{z})), m')) - T((x_0, 0), (f(z, \pi(z)), m'))] \\
& +(1 + \lambda_4) \cdot [T((x_0, 0), (z, m)) - T((\tilde{x}_0, 0), (\tilde{z}, m))] \\
\leq \; & \mathfrak{L}_T \cdot \|(\tilde{x}_0 - x_0, 0), (f(\tilde{z}, \pi(\tilde{z})) - f(z, \pi(z)), 0)\|_1 \\
& +(1 + \lambda_4) \cdot \mathfrak{L}_T \cdot \|(x_0 - \tilde{x}_0, 0), (z - \tilde{z}, 0)\|_1 \\
\leq \; & \mathfrak{L}_T \cdot (\|(\tilde{x}_0 - x_0\|_1 + \mathfrak{L}_f \cdot (1 + \mathfrak{L}_\pi) \cdot \|\tilde{z} - z\|_1) \\
& +(1 + \lambda_4) \cdot \mathfrak{L}_T \cdot 2\tau \\
\leq \; & \mathfrak{L}_T \cdot (1 + \mathfrak{L}_f \cdot (1 + \mathfrak{L}_\pi)) \cdot \tau + (1 + \lambda_4) \cdot \mathfrak{L}_T \cdot 2\tau \\
= \; & (\mathfrak{L}_T \cdot (1 + \mathfrak{L}_f \cdot (1 + \mathfrak{L}_\pi)) + 2(1 + \lambda_4) \cdot \mathfrak{L}_T) \cdot \tau \\
= \; & \mathfrak{L}_4 \cdot \tau \leq \beta_4
\end{aligned}$$

  Therefore, we can derive that

$$g_4(x_0, 0, z, m, m') \geq g_4(\tilde{x}_0, 0, \tilde{z}, m, m') - \beta_4$$

  Because $g_4(\tilde{x}_0, 0, \tilde{z}, m, m') - \beta_4 \geq 0$ for any sample points, we have $g_4(x_0, 0, z, m, m') \geq 0$ for all states outside the sample set, which implies $g_4(x_0, 0, z, m, m') \geq 0$ holds over the whole state space.

□