Electronic Research Archive

ERA, 34(5): 3193–3222.
DOI: 10.3934/era.2026144
Received: 07 December 2025
Revised: 16 March 2026
Accepted: 30 March 2026
Published: 13 April 2026

https://www.aimspress.com/journal/ERA

Research article

Safe reinforcement learning for fixed-time stability of a class of nonlinear systems

Musayyab Ali[1], Fahad Mumtaz Malik[1,*], Naveed Mazhar[2] and Nadia Sultan[3]

[1] Department of Electrical Engineering, College of Electrical & Mechanical Engineering, National University of Science and Technology, Islamabad, Pakistan
[2] Interdisciplinary Research Center for Sustainable Energy Systems, KFUPM, Dhahran, Saudi Arabia
[3] Center of Excellence in AI (CoE-AI), Department of Electrical Engineering, Bahria University, Islamabad, Pakistan

* **Correspondence:** Email: malikfahadmumtaz@ceme.nust.edu.pk; Tel: +925154444200.

Abstract: This paper presents novel safe reinforcement-learning-based, fixed-time control framework for a class of discrete time uncertain nonlinear systems which guarantees fixed-time stability (FxTS). This framework is developed to ensure fixed-time convergence for nonlinear systems in the absence of exact model. The main contribution is to present a unified framework by integration of fixed-time Lyapunov-based constraints and Gaussian process-based uncertainty modeling into the safe reinforcement learning loop in a manner that each candidate policy must ensure the fixed-time Lyapunov decrement condition within a uniform bounded convergence time in the entire learning process. The Lyapunov function for the nominal fixed-time stable closed loop system is used for enlarging the safe set for FxTS with accumulation of system data through refining the traditional Gaussian process model for uncertain system dynamics Gaussian process model is used to learn the unknown system dynamics online and to provide probabilistic confidence bounds. These bounds are directly incorporated into the fixed-time Lyapunov condition, leading to safety guarantees and monotonic expansion of the Lyapunov certified safe set. The control policy is then optimized for the enlarged safe set through exploration while ensuring FxTS. The proposed approach is validated through simulated results of inverted pendulum and reduced-order vehicle dynamic model for lateral vehicle dynamics. The validation for both nonlinear systems confirms fixed-time convergence, certified expansion of safe operating region, and improved control performance under safety and Lyapunov constraints.

Keywords: fixed-time stability; nonlinear systems; reinforcement learning; safe learning

1. Introduction

Fixed-time stability (FxTS) enforces a stronger convergence behavior than classical finite-time stability. In case of FxTS, the state of a dynamic system converges to equilibrium within a fixed settling time regardless of the initial conditions. FxTS provides uniform settling time, making it very significant for applications requiring precise and guaranteed performance. The property is beneficial in robotics, autonomous vehicles, and power systems applications, where even small delays or uncertain initial conditions could compromise safety and efficiency. By guaranteeing convergence within a fixed-time, FxTS significantly improves robustness, enables time-critical control applications, and ensures performance in real-world environments.

FxTS particularly finds significant utility in the field of autonomous driving, as it provides a framework which ensures safety, robustness against uncertainties, and a guaranteed response time. FxTS provides powerful assurance that the vehicle dynamics reach safe operating conditions within a predetermined time horizon completely independent of initial conditions or uncertainties. This unique property addresses one of the foremost challenges in autonomous driving, guaranteeing robustness and certainty under critical situations such as sudden collision avoidance or emergency lane changes in dense traffic.

FxTS has certain definitive advantages over comparative control techniques. In finite-time stability, convergence to the equilibrium point is ensured by dynamics of the form $\dot{V} \leq -cV^{\alpha},\ 0 < \alpha < 1,$ which guarantees $V(t) = 0$ for all $t \geq T_s(x_0)$ within finite time. The settling time function, on the other hand, is explicitly dependent on initial conditions and is unbounded over the state space, which makes it less suitable for safety critical systems where uniform safety certification is required. The foundational results of [1] provide finite-time stability properties and establish the mathematical framework. The work presented in [2] establishes a mechanism of fixed-time stabilization for strict feedback nonlinear systems. It demonstrates how dynamic gain feedback can ensure robustness even under nonlinearities and model uncertainties. Fixed-time stability strengthens this notion by employing composite decay rates, for example, $\dot{V} \leq -a_1V^{\alpha} - a_2V^{\beta}, 0 < \alpha < 1 < \beta$, which guarantees convergence within a uniform upper bound T_s that is independent of initial conditions, making it more appropriate for control applications requiring precise and guaranteed convergence.

Predefined time stability further restricts the controller parameters such that the upper bound on the settling time is explicitly specified by the user. Although this ensures convergence within a desired horizon, it restricts the admissible gain combinations, thereby reducing flexibility in shaping transient and near-equilibrium behavior. In a predefined control scheme, guaranteeing convergence within a prescribed settling time demands gain scaling that is dependent on a desired convergence horizon, which may result in saturation, aggressive control effort, and reduced robustness. Prescribed time performance approaches either ensure convergence at user specified terminal time or constraint-based designs using barrier functions to confine the system evolution within predefined bounds. Exact-time (or prescribed time) stability typically rely on time transformations that become singular at the terminal instant, whereas constraint-based approaches often guarantee convergence only to an ε-neighborhood of the equilibrium. Consequently, prescribed performance methods emphasize timing or constraint satisfaction, often at the expense of global convergence flexibility. From the above discussion, it is evident that the fixed-time control provides a framework with initial-condition-independent

convergence guarantees, continuous solutions beyond convergence, and greater freedom in transient shaping. These properties make fixed-time control particularly suitable for safe reinforcement learning.

Fixed-time controller based on nominal system models is typically designed by defining a Lyapunov function tailored to the system dynamics, such that it decreases in a manner that ensures convergence within a fixed upper bound on time, independent of initial conditions. Recent studies have extended this framework by introducing stronger Lyapunov-based conditions that provide sufficient criteria for FxTS, resulting in less conservative and more precise settling time estimates [3]. For robust implementations, authors have proposed the integration of a fixed-time Lyapunov function with control barrier functions to ensure both safety and fast convergence, even in the presence of modeling uncertainties [4]. These advancements, together with foundational work in Lyapunov-based fixed-time stabilization of nonlinear systems [5], collectively prove the effectiveness of fixed-time Lyapunov techniques in real-world time-critical applications. The contributions in [6,7] present comprehensive deterministic and stochastic FxTS results for a class of discrete time autonomous systems guaranteeing convergence with in a bounded, state-independent time horizon.

In practical applications, the exact system dynamics of nonlinear systems are not fully known because of the presence of disturbances, modeling errors, and unmodeled nonlinearities. The nominal fixed-time results no longer remain valid under larger uncertainties, because the Lyapunov decrease condition may be violated for the unmodeled dynamics. Due to the complexity and uncertain dynamics of nonlinear systems, learning-based control schemes such as data-driven control, neural-network-based approaches, and reinforcement learning have emerged as a promising solutions for handling unknown dynamics. These schemes compensate for modeling errors using observed data. The study in [8] presents a constrained reinforcement learning approach for cooperative control of multi-unmanned aerial vehicles (UAV) operating in dense obstacle environments. A noise tolerant fuzzy type zeroing neural network is proposed in [9] to achieve robust synchronization of chaotic systems in the presence of external disturbances. A fixed-time equivalent input disturbance is introduced in [10] to achieve disturbance rejection under system uncertainties. The work presented in [11] is an adaptive dynamic programming-based finite-time sliding mode control for robust stabilization of fractional order chaotic systems.

Several recent contributions such as the study presented in [12] formulates robust predictive control architecture for wind turbines operating under full load conditions. An adaptive nonlinear integral backstepping control framework is implemented in [13] with double deep Q learning for frequency stabilization in microgrid systems. A reliable shared steering control strategy based on deep reinforcement learning is introduced in [14] for emergency obstacle avoidance conditions. The survey developed in [15] provides the integration of transformer models within learning for better decision-making performance.

Neural-network-based system identification methods such as those in [16–18] have been widely used to approximate unknown system dynamics from the observed data. These learned models are often combined with classical controllers to ensure stability properties and performance guarantees. Thereafter, the learning-based controllers are integrated with model-based designs to deal with unknown system dynamics and improve control performance. However, these methods generally do not provide probabilistic uncertainty bounds, which limits their direct integration into robust Lyapunov-based fixed-time certification.

The study presented in [19] develops a GAN based approach highlighting the capability of models to extract structured visual information. A deep convolution neural network (CNN) is

introduced in [20] for accurate surface detection reinforcing the effectiveness of learning-based feature extraction. A light weight CNN architecture is established in [21] to support efficient implementation of deep learning in real world applications. An attention-enhanced lightweight network is developed for object detection, illustrating the importance of advanced perception models in complex autonomous environments [22].

Recent research has focused on advanced control and estimation methodologies under uncertainties and external disturbances for complex dynamical systems. The work presented in [23] introduced a robust antidisturbance interval type 2 fuzzy control framework for interconnected nonlinear partial differential equation (PDE) systems by incorporating a conjunct observer to enhance disturbance compensation and system robustness. The study in [24] presented a nonlifted, norm-optimal, iterative learning control framework for networked dynamical systems, achieving better learning performance than previous methods while effectively reducing computational complexities. In [25], a finite region asynchronous H_∞ filtering scheme is proposed for two dimensional Markov jump systems modeled in the Roesser framework to guarantee robust state estimation in the presence of stochastic switching conditions. These contributions showed the growing interest in robust control, filtering, and learning-based schemes for handling uncertainties in complex dynamical systems. However, many approaches in the literature have focused on robust control design or state estimation and do not explicitly address stability guarantees in learning-based control methods.

As learning-base methods are progressively integrated into control and decision-making frameworks, predictable transient response and specific convergence guarantees have become an essential requirement. This motivates the integration of FxTS and reinforcement learning to achieve guaranteed fixed-time performance in nonlinear systems with uncertain dynamics. A reinforcement learning based secure control scheme is proposed in [26] for nonlinear interconnected systems. The study in [27] introduces an observer-based fault-tolerant control method with a novel event triggered mechanism to deal with input saturation and full state constraints in nonlinear systems. A barrier critic-based robust control framework is presented in [28] for solving nonzero sum differential games in uncertain nonlinear systems with state constraints. An adaptive safety critical framework is formulated in [29] for nonlinear systems with asymmetric input constraints. However, these frameworks guarantee only asymptotic or performance-driven stability without providing fixed-time convergence certification under the probabilistic bounds.

Gaussian process (GP) modeling offers a structured probabilistic mechanism for approximating unknown system dynamics while explicitly quantifying model uncertainty. By using posterior mean and variance estimates, GP regression provides high-probability confidence bounds based on learned residual dynamics. These bounds can be integrated into Lyapunov-based conditions, ensuring robust stability certification in the presence of uncertain system dynamics. On this basis, the proposed framework incorporates a GP-based uncertainty bound directly into a fixed-time Lyapunov-based safety condition, thereby facilitating safe policy learning with guaranteed fixed-time convergence.

Recent developments in reinforcement learning (RL) have introduced learning-based control schemes which effectively deal with complex nonlinear systems without requiring the exact system knowledge. These characteristics makes RL well-suited, especially in autonomous vehicles, robotics, and other safety-critical systems in which system dynamics are not fully known or challenging to model accurately. However, applying RL directly to a control system introduces significant challenges related to stability and safety guarantees, especially during the exploration phase. To address these challenges, safe reinforcement learning frameworks have been introduced that integrate Lyapunov -

based constraints to guarantee stability while ensuring policy improvement within the certified safe set. However, many existing methods are based on exponential or asymptotic stability, which ensure convergence when $t \to \infty$ and do not provide explicit bounds on the system convergence time. Therefore, more rigorous stability guarantees are required to ensure safe and stable system response during the learning process.

Reinforcement learning provides a systematic framework that supports direct policy optimization in the presence of nonlinearities allowing the controller to refine its control policy online during its interaction with the system. RL is particularly effective in partially known environments, as it is capable of online policy improvement. Standard RL methods have the limitation of enforcing stability and safety constraints which restrict their applicability in safety critical applications. This motivates the integration of reinforcement learning with a fixed-time Lyapunov certified structure to ensure guaranteed convergence and safety throughout the learning process.

Modern advancements in safe reinforcement learning have emphasized ensuring constraint satisfaction throughout the process of learning and execution. The study [30] provided a comprehensive survey of reinforcement learning frameworks evolving from single agent to multiagent systems along with their theoretical foundations and industrial applications. The study highlights the importance of cooperative learning in complex environments.

In [31], a mechanism presented for safe exploration that corrects potentially unsafe actions selection in continuous action spaces, guaranteeing real-time safety by projecting every action back into the verified safe set region. In parallel with these stability results, the comprehensive survey in [32] reviewed emerging safe RL techniques, theoretical foundations, and practical applications, focusing on the importance of integration of stability guarantees, uncertainty modeling, and constraint enforcement in safety critical domains. The work in [33] focused on the learning of discrete-time, uncertain, nonlinear systems under probabilistic safety and stability constraints, ensuring that policy updates remain safe even in the presence of model uncertainty. However, the above-mentioned works have several limitations. In particular, most existing approaches rely on approximate safety models and probabilistic constraints formulation or lack unified theoretical stability guarantees for learning -based control methods. Therefore, guaranteeing safety and stability during the learning process remains a challenging issue.

Foundational works in safe learning and reinforcement learning establish the theoretical basis of safety control of uncertain systems. In the context of learning safe control regions, a GP-based framework for safely estimating and expanding safe region of attraction is formulated in [34]. A safe exploration strategy is presented in [35] ensuring safe and efficient data acquisition under safety constraints. The study in [36] presented a comprehensive survey of safe reinforcement learning, highlighting emerging directions and challenges in guaranteeing safety under uncertainty. A policy gradient estimation method for infinite horizon problem is established in [37], providing a theoretical key foundation for optimization in continuous time tasks. Recent research efforts have strengthened the groundwork for safe reinforcement learning, focusing on safe exploration and stability-aware policy learning. The studies in [38,39] collectively introduced the framework that avoids unsafe actions and safe exploration, and they integrate Lyapunov barrier-based safety guarantees for nonlinear control.

Existing safe reinforcement learning methods have incorporated safety and stability directly into the learning process through Lyapunov-based certification such as [41,42]. More recently, learning-based barrier functions have been used as a safety filter mechanism [43], providing stability in nonlinear systems with uncertain dynamics. These methods successfully guarantee constraint enforcement while optimizing performance. However, these methods only ensure asymptotic

convergence guarantees, with settling time totally dependent on initial conditions. Compared to these methods, our framework incorporates fixed-time Lyapunov-based stability condition under GP-modeled uncertainty into a safe reinforcement learning loop for nonlinear systems with uncertain dynamics, guaranteeing fixed-time convergence in the Lyapunov certified region totally independent of initial conditions while ensuring Lyapunov certified safety and monotonic safe set expansion. The design of fixed-time Lyapunov-based controllers becomes more challenging when the system dynamics are unknown, because the Lyapunov decrease condition must be satisfied without having access to the true model of the system. Recent work by [41] proposed a safe reinforcement learning approach that uses a GP to model uncertainty, where Lyapunov conditions were enforced during the policy update process. However, this framework ensures only asymptotic convergence, and settling time remains dependent on the initial conditions. This gap highlights the need for developing a fixed - time safe learning framework that offers safety and convergence guarantees within a prespecified time, independent of initial conditions. To address this gap, we present a fixed-time safe reinforcement learning framework for nonlinear systems with unknown dynamics.

The proposed study presents a novel safe reinforcement learning framework by integrating discrete-time fixed-time stability guarantees explicitly into the learning and control mechanism. The proposed formulation differs from conventional control methods that primarily focus on asymptotic or exponential stability notions. Within this framework, a Lyapunov-based fixed-time decrement condition is incorporated, thereby guaranteeing fixed-time convergence in a uniform bounded time that is independent of initial conditions and plays a vital role in safety critical systems.GP-based uncertainty modeling is introduced to approximate the unknown system dynamics and to capture the parametric uncertainty by probabilistic confidence bounds. Using these confidence bounds, the certified safe set is expanded by including those states which satisfy the fixed-time stability condition. The safe set expansion is monotonic and remains conservative. This framework unifies the fixed-time stability, GP-based uncertainty modeling, and neural-network-based policy optimization which develops a new pathway toward learning controllers that achieve high performance within fixed-time bounded convergence guarantees.

The proposed framework is of significant importance because it addresses several key limitations of the learning-based approaches in the literature within a unified framework. Many prior studies focused on treating the safety considerations, stability guarantees, and uncertainity handling, independently which limits their applicability in safety critical environments. The proposed framework overcomes these limitations by the integration of fixed-time Lyapunov-based safety certification, GP-based uncertainity modeling, and safe policy optimization within a safe reinforcement learning architecture. This integration allows safety-constrained exploration under uncertainity while ensuring the Lyapunov-based stability guarantees during the learning process. Specifically, the incorporation of fixed-time stability ensures convergence guarantees within a uniform bound that is independent of initial conditions, making it well-suited for safety-critical control systems. Therefore, the proposed framework provides a theoretically rigorous and practically applicable approach for guaranteeing safe, reliable, and performance-driven control of nonlinear systems in the presence of uncertain dynamics. This framework integrates fixed-time Lyapunov constraints into the reinforcement learning process, ensuring that each policy update enforces both performance improvement and guaranteed convergence within a predefined time horizon independent of initial conditions. Reinforcement learning is more suitable for this framework, as it provides direct policy optimization without the requirement of exact system models, making it more effective for the nonlinear and uncertain dynamics considered in this

work. Reinforcement learning aims to optimize control policies based on long-term performance objectives through interaction with the true system dynamics. Within the proposed framework, RL is embedded within the Lyapunov-based certified safe set under fixed-time stability constraints, ensuring that each policy update preserves safety and fixed-time convergence guarantees. As a result, reinforcement learning provides an effective mechanism for enhancing transient response, reducing control effort, and providing monotonic expansion of the admissible operating region without violating fixed-time stability conditions.

The main contributions of this paper are outlined below:

- A discrete-time Lyapunov-based stability framework is developed for nonlinear control affine systems, ensuring convergence within a fixed time horizon independent of initial conditions in learning-based control.
- Incorporation of Gaussian process modeling of unknown dynamics into Lyapunov-based stability framework to ensure that only safe policies for FxTS are selected during the exploration process.
- Monotonic safe set expansion because all the admissible states satisfy the FxTS condition under GP uncertainty, guaranteeing the expansion of the safe region.
- The fusion of Lyapunov-based FxTS with learning-based control ensures safety, a guaranteed fixed-time convergence, and robust performance, even under model uncertainty.

The subsequent sections are organized as follows: The problem formulation is presented in Section 2, Section 3 presents the safe set expansion, the safe policy learning and optimization is presented in Section 4, Section 5 includes the simulation results, and the paper is concluded in Section 6.

2. Problem formulation

We present a mathematical description of the problem in this section. To this effect, the actual and nominal system, the fixed-time control law, and the accompanying Lyapunov function and Gaussian process model representation of system dynamics are first presented, followed by the problem statement.

2.1. System model

We consider the following class of discrete-time, control affine nonlinear dynamic systems:

$$x_{k+1} = f(x_k) + g(x_k)u_k \triangleq h(x_k, u_k), \tag{1}$$

where $x_k \in R^n$ and $u_k \in R^m$ represent the system state and input, respectively. The model (1) represents the nonlinear system at discrete points in time and is considered to have originated from the integration of the continuous time model $h_c(x(t), u(t))$ such that

$$h(x_k, u_k) = x_k + \int_{kT}^{(k+1)T} h_c(x, u)dt. \tag{2}$$

Assumption 1: The nonlinear system Eq (1) is locally Lipschitz in its arguments over the domain of interest.

It is pertinent to mention that a locally Lipschitz function over a domain, satisfies the Lipschitz condition with the *same* Lipschitz constant over any compact subset of the domain [39]. Mathematically,

$$\left||h(x_1, u_1) - h(x_2, u_2)|\right| \leq L_1\|x_1 - x_2\| + L_2\|u_1 - u_2\| \ , \ \forall x_1, x_2 \in \boldsymbol{S}, \quad u_1, u_2 \ \boldsymbol{\epsilon} \ \boldsymbol{U}, \tag{3}$$

where S and $\boldsymbol{U}$ are compact subsets of the domain of interest for state and input with same values of Lipschitz constant L_1 and L_2. Lipschitz properties of (1) follow naturally from that of (2). The a priori knowledge of the system dynamics is represented by the following nominal model:

$$\hat{h}(x_k, u_k) \triangleq \hat{f}(x_k) + \hat{g}(x_k)u_k, \tag{4}$$

where the nominal nonlinear functions $\hat{f}(.)$ and $\hat{g}(.)$ are also Lipschitz in their argument over the domain of interest. The modeling error is consequently defined by the following for compactness of notation:

$$\delta(x_k, u_k) \triangleq h(x_k, u_k) - \hat{h}(x_k, u_k). \tag{5}$$

2.2. Fixed-time stability

Consider the following state feedback control law:

$$u_k \triangleq \pi(x_k), \tag{6}$$

where the function $\pi(.)$ is Lipschitz over the domain of interest. The nominal closed loop system dynamics are then represented as

$$\hat{h}(x_k, \pi(x_k)) = \hat{f}(x_k) + \hat{g}(x_k)\pi(x_k). \tag{7}$$

Assumption 2: The nominal closed-loop system admits a continuously differentiable Lyapunov function $V(.)$ such that,

$$\alpha_1(||x_k||) \leq V(x_k) \leq \alpha_2(||x_k||), \tag{8}$$

and

$$\Delta V_{nom} \triangleq V(\hat{h}(x_k, \pi(x_k)).) - V(x_k) \leq -min\{a_1\ V(x_k)^{\alpha} + a_2\ V(x_k)^{\beta}\ , V(x_k)\}, \tag{9}$$

where $a_1, a_2 \in (0,1) > 0$, and $\alpha \in (0\ ,1)$, $\beta > 1$, for all $x \in V(c_0) = \{x \in \boldsymbol{S} | V(x) \leq c_0\}$. Under this condition, Eq (9), the origin is fixed-time stable and settling time is uniformly bounded by an explicit integer K_{max}. By Theorem (3.1) [7], $K(x_0)$ is uniformly bounded,

$$K(x_0) \leq K_{max}. \tag{10}$$

Notably, K_{max} is independent of initial condition x_0, which is actually the fixed-time property, ensuring that regardless of the initial state, convergence is guaranteed in at most the same number of finite steps. The settling time function $K(x_0)$ satisfies the uniform upper bound [7],

$$K(x_0) \leq \left[\frac{lna_1}{(1-\alpha)\, ln(1-a_1)}\right] + \left[\frac{lna_2}{(\beta-1)\, ln(1-a_2)}\right] + 1. \tag{11}$$

Qualitative description of the condition Eq (11) provides an explicit uniform upper bound on the settling time $K(x_0)$ of the system. The inequality shows that under the Lyapunov decrement condition Eq (9), the maximum number of steps required for convergence is uniformly bounded by a finite integer that is the function of constants a_1, a_2, α, β only. The derived bound includes logarithmic and ceiling functions to quantify the steps in both phases that is large V and small V, thus ensuring that every trajectory converges in a fixed uniform time establishing the FxTS property.

2.3. Gaussian process modeling of unknown system dynamics

To address the issue of partially known dynamics, Gaussian process regression is integrated into the learning framework as presented in [34,43]. The mechanism reduces the uncertainty of system dynamics by providing probabilistic bounds of safe sets of increasing size as more data is obtained. If the size of the safe set increases, a broader region for the synthesis of new candidate policies becomes available, consequent to which optimization can be carried out under the fixed-time Lyapunov framework such that each policy update achieves better performance and maintains safety throughout the learning process.

In the proposed framework, safety is ensured by the construction of a certified safe set in the state space based on the fixed-time Lyapunov stability condition. Because the system dynamics are partially unknown, a Gaussian process model is used to learn the unknown dynamics and to provide high-confidence bounds on the modeling error. These bounds enable the Lyapunov decrease condition to be satisfied in a worst-case sense, ensuring that the admissible states remain confined within the certified safe set region. As the learning process progresses, more data is obtained through the safe rollouts, which in turn makes the GP model progressively more accurate, allowing the Lyapunov condition to be verified for previously unexplored states. As a result, the certified safe set expands monotonically while ensuring that all system trajectories remain within the certified safe region where fixed-time stability is guaranteed.

Assumption 3: (GP-based uncertainty modeling with confidence bounds). The GP model for residual dynamics $\delta(x,u)$ satisfies:

$$\|\delta(x_k,u_k)\| \leq \mu_k(x_k,u_k) \pm \beta_k\sigma_k(x_k,u_k), \tag{12}$$

with probabilistic bounds such that $\mu_k(x_k,u_k)$ is the mean prediction of the GP for model error, and $\sigma_k(x_k,u_k)$ is the standard deviation.

The Gaussian process modeling of the unknown system dynamics formulates a stochastic environment in the jargon of reinforcement learning paradigm as discussed in [45]. The state transition (1) upon rearranging (5) is described by

$$x_{k+1} = \hat{h}(x_k,u_k) + \delta(x_k,u_k), \tag{13}$$

where $\delta(x_k,u_k)$ is a random variable as characterized in (12). The expression (13) can be represented as follows, using abridged notation

$$x_{k+1} = \mathbf{h}(x_k,u_k,\delta_k), \tag{14}$$

where $\delta_k = \delta(x_k,u_k)$. The expression (14) is equivalent to the classical stochastic state transition expression presented in reinforcement learning literature consequent to the following definition, attributable to the randomness of δ_k;

$$p(x_{k+1}|x_k,u_k) \triangleq P\{\Theta(x_k,u_k,\delta_k)|x_k,u_k\}, \tag{15}$$

where the set $\Theta(x_k,u_k,\delta_k)$is defined as

$$\Theta(x_k,u_k,\delta_k) = \{\delta_k|x_{k+1} = \mathbf{h}(x_k,u_k,\delta_k)\}. \tag{16}$$

In the context of the discussion above, we state the problem of safe reinforcement learning with FxTS, as one of establishing a safe set with FxTS guarantee for the actual nonlinear system model in the

probabilistic sense with GP bounds using a Lyapunov function for the nominal fixed-time stable system.

In the nutshell, we address two aspects: (i) safety, as the state trajectories remain confined to the Lyapunov certified invariant set, and (ii) performance, by achieving convergence within a fixed-time rather than asymptotically. The safe set for the actual nonlinear system is then enlarged as more system data is obtained. The control policy guaranteeing fixed-time stability is subsequently optimized for larger safe sets through reinforcement learning.

The detailed block diagram of the proposed controller is given in Figure 1.

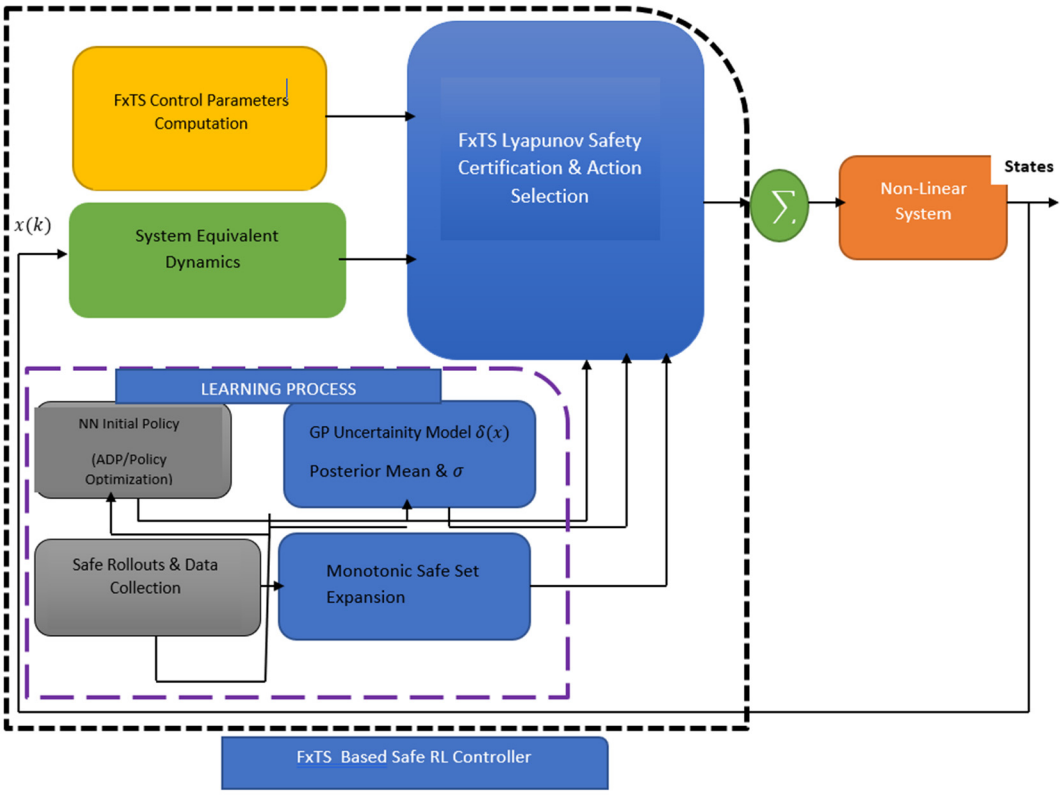

Figure 1. Block diagram of the FxTS-based safe RL controller.

3. Safe set expansion

In this section we discuss the expansion of the safe set for the FxTS of nonlinear system (1). Prior to safe set expansion, we require determination of sufficient conditions for the fixed-time stability of Model (1) under the control law (6) designed for the nominal closed-loop system, which in turn provides an estimate of the region of state space which is safe for exploration in terms of fixed-time stability.

The benchmark of stability analysis in nonlinear control theory is the establishment of bounds on the decrement rate of the positive definite Lyapunov function for a level set. The guaranteed decrement ensures that the compact set is positively invariant for system trajectories. The rate of convergence is determined by the analytical characteristics of the bounds on the decrement rate. The positively invariant Lyapunov level set can therefore be classified as a safe set from the perspective of exploration in reinforcement learning because every state in this set is safe to visit from the perspective of stability.

The preceding paragraph is the guiding principle for analytical contribution in this paper

presented in Theorem 1, which follows below. The Lyapunov function for the nominal closed-loop system under Assumption 2 is used to analyze the dynamics of the uncertain actual nonlinear system (1). The decrement rate bound is required to guarantee fixed-time stability.

The challenge in establishing safety is that data of Model (1) will be available at discrete points of the state space, whereas the exploration guarantees are required for the continuum of points defining the invariant set. Therefore, inclusion of discretization error bounds for nonlinear system is required for the analysis. To this end, we consider the nearby state x_k', which lies within a ball of radius r around x_k to represent the continuum of points, that is,

$$\|x_k' - x_k\| \leq r.$$

The transition of the nonlinear system from x_k' is represented using (2) as follows:

$$x_{k+1}' = x_k' + \int_{kT}^{(k+1)T} h_c(x,u)dt. \tag{17}$$

Based on the above discussion, we present the following theorem which provides sufficient conditions for determination of invariant set for fixed-time stability.

Theorem 1: The nonlinear system Eq (1) is fixed-time-stable with state feedback control policy Eq (6) if there exists $c_k \leq c$ such that,

$$\Delta V(x_k) \triangleq V(f(x_k) + g(x_k)u_k) - V(x_k), \tag{18}$$

satisfies

$$\left[\Delta V(x_k) - L_v\left(\mu_n(x_k,u_k) + \beta_k\sigma_k(x_k,u_k) + r + \varepsilon(T)\right)\right] \leq -\min\left\{a_1 V(x_k)^{\alpha} + a_2 V(x_k)^{\beta}, V(x_k)\right\}, \tag{19}$$

$\forall x \in V(c) \triangleq \{x \in \mathrm{R}^n \mid V(x) \leq c\}$ with positive constant r and $\varepsilon(T)$ related to discretization of state space.

Proof: The difference of the Lyapunov function with control policy (6) under Assumption 2, for system dynamics (1) starting from point x_k in the proximity of x_k', is given by:

$$\Delta V(x) = V(x_{k+1}') - V(x_k').$$

Adding and subtracting $V(x_{k+1})$ and $V(x_k)$,

$$\Delta V(x) = V(x_{k+1}) - V(x_k) + V(x_{k+1}') - V(x_{k+1}) - V(x_k') + V(x_k).$$

Using (4) and (6) leads to

$$\Delta V(x) = V\left(\hat{f}(x_k) + \hat{g}(x_k) * \pi(x_k) + \delta(x_k,u_k)\right) - V(x_k) + V(x_{k+1}') - V(x_{k+1}) - V(x_k') + V(x_k). \tag{20}$$

Equation (20) can be compactly written as

$$\Delta V(x) = V\left(\hat{h}(x_k,u_k) + \delta(x_k,u_k)\right) - V(x_k) + V(x_{k+1}') - V(x_{k+1}) + V(x_k') + V(x_k). \tag{21}$$

The continuous differentiability of $V(.)$ over the domain of interest guarantees the existence of Lipschitz constant L_v over the compact subset $\boldsymbol{S}$.

$$\Delta V(x) \leq V\left(\hat{h}(x_k,u_k)\right) - V(x_k) + L_v(\| \delta(x_k,u_k) \|) + \|x_{k+1}' - x_{k+1}\|) + \|x_k' - x_k\|). \tag{22}$$

As the initial conditions for x'_{k+1} and x_{k+1} are sufficiently close, and the nonlinear system (1) is Lipschitz, the difference between x'_{k+1} and x_{k+1} is a function of r, as described in Theorem 3.4 of [40]. The bound on the difference $x'_{k+1} - x_{k+1}$ is given by

$$\|x'_{k+1} - x_{k+1}\| \leq r.exp\left(\frac{L_x T}{1+L_x}\right) = \varepsilon(T), \tag{23}$$

where L_x is a Lipschitz constant of a continuous time function of $h(x,u)$.

Using Eqs (9) and (23), Eq (22) can be written as:

$$\Delta V(x) \leq \ \Delta V_{nom} + \ L_v\big(\| \delta(x_k,u_k) \| + r + \varepsilon(T)\big) \tag{24}$$

$$\leq -\min\{a_1 V(x_k)^{\alpha} + a_2 V(x_k)^{\beta}, V(x_k)\} + \ L_v\big(\| \delta(x_k,u_k) \| + r + \varepsilon(T)\big), \tag{25}$$

Under Assumption 3, Eq (25) becomes:

$$\Delta V(x) \leq \ -\min\{a_1 V(x_k)^{\alpha} + a_2 V(x_k)^{\beta}, V(x_k)\} + \ L_v\big(\ \ \mu_n(x_k,u_k) + \beta_k \sigma_k(x_k,u_k) + r + \varepsilon(T)\big). \tag{26}$$

Rearranging Eq (26) completes the proof.

The sufficient condition of Theorem 1 provides an estimate of the safe region which is available for exploration and determination of optimal policy. The GP posterior mean $\mu_k(.,.)$ and the standard deviation $\sigma_k(.,.)$ are computed from the data set $D_k = \{[x_i, u_i, x_{i+1}]\}_{i=1}^{n}$, which consists of state-input successor triplets obtained from interaction with the system. The confidence parameter β_k relates to the probability, for example, 95%, so that the actual value of residual lies within the interval $[\mu_k(x,u) \pm \beta_k \sigma_k(x,u)]$. The high probabilistic bounds obtained from the Gaussian process are directly integrated into the fixed-time Lyapunov condition, resulting in a robust Lyapunov inequality that holds valid for all the admissible dynamics within GP confidence bounds. In this approach, the GP uncertainty term is incorporated as the worst-case margin in the Lyapunov condition, thereby ensuring that the fixed-time decrement condition is not valid only for the nominal learned dynamics but for all admissible uncertain dynamics. At each learning iteration, the safe set is selected as the largest Lyapunov sublevel set that satisfies the robust fixed-time Lyapunov stability condition. As more data is collected through exploration, the GP posterior variance decreases, thereby reducing the conservatism of the Lyapunov bound while fully ensuring the fixed-time stability. As a result, larger Lyapunov sublevel sets are provably certified, thus guaranteeing monotonic safe set expansion. As the dataset expands, the GP variance decreases, which in turns tightens the supremum, thereby reducing safety margins, and the certified Lyapunov level set increases. As a result, expansion of the certified safe set is enabled.

$$= \{\, x \in \mathrm{R}^n \mid V(x) \leq c_k \,\}. \tag{27}$$

States that were previously considered unsafe due to larger model uncertainty can then be certified safe as the GP confidence interval shrinks with additional data. The safe set thus expands monotonically with each iteration:

$$S_0 \subseteq S_1 \subseteq S_2 \subseteq S_3 \subseteq \cdots \subseteq S_k. \tag{28}$$

The structured methodology discussed above guarantees that exploration is both safe and guided, ensuring that the agent remains within the certified safe set during the training process.

4. Safe policy learning and optimization under the FxTS constraint

To guarantee improved performance while maintaining safety, policy learning is carried out under FxTS conditions. At each iteration, a neural-network-based policy $\pi_\theta: \mathrm{SK} \to \mathcal{U}$ is optimized to maximize task performance while ensuring FxTS within the certified safe set.

The policy optimization is carried out per the following optimization objective:

$$sup_{||\Delta x(x,\pi_\theta(x))|| \le \beta_k \sigma_k(x,\pi_\theta(x))} \left[\Delta V(x_k) - L_v\big(\quad \mu_n(x_k,u_k) + \beta_k \sigma_k(x_k,u_k) + r + \varepsilon(T)\big)\right]$$

$$\le -\min\{a_1 V(x)^\alpha + a_2 V(x)^\beta \ , V(x_k)\}, \tag{29}$$

where $a_1, a_2 > 0, 0 < \alpha < 1 < \beta$ are user defined constants which guarantee the nonlinear decay condition required for fixed-time convergence.

This condition ensures that for all admissible realization of the model error are within the confidence region, the Lyapunov function decays with a fixed-time rate which ensures forward invariance and fixed-time convergence with high probability.

The policy optimization problem is then formulated as maximizing the expected cumulative reward, ensuring that the FxTS condition is satisfied under the probabilistic uncertainty bounds provided by the GP model:

$$\theta = \quad \arg\ \max_{\theta} \ \mathrm{E}_{x_0 \sim D} \ \left[\sum_{k=0}^{T} l\big(x_k, \pi_\theta(x_k)\big)\right], \tag{30}$$

where $l(x,u)$ represents the instantaneous performance function, $u_k = \pi_\theta(x_k)$ subject to the safety constraints, and $\forall x \in S_k$

$$V\left(\hat{h}\big(x,\pi_\theta(x)\big)\right) - V(x)) - L_v\big(\mu_n(x_k,u_k) + \beta_k \sigma_k(x_k,u_k) + r + \varepsilon(T)\big) \le -\min\{a_1 V(x_k)^\alpha +$$

$$a_2 V(x_k)^\beta, V(x_k)\}. \tag{31}$$

It is necessary to enlarge the admissible region for improving control performance and reducing model uncertainty: however, unconstrained exploration can easily violate the stability of nonlinear system with uncertain dynamics. In the proposed framework, as exploration is directly incorporated within a fixed-time Lyapunov certified condition under the worst-case margin, thereby ensuring that expansion is subject to satisfy the Lyapunov certified fixed-time stability condition. As the learning progresses, and more data is collected, model uncertainty reduces, which results in larger Lyapunov sublevel sets in the admissible region. Consequently, the safe set grows in a monotonic manner, ensuring that every admissible new state satisfies fixed-time convergence guarantees. This mechanism is very useful, as it avoids both overconservatism and unsafe exploration, guaranteeing robust learning that improves control performance while satisfying the fixed-time convergence and safety guarantees.

To ensure FxTS throughout the optimization process, each candidate policy is evaluated using the Gaussian process posterior to check whether the proposed policy satisfies the FxTS requirement Eq (31). Policies that maintain the guaranteed decrease of the Lyapunov function under the GP's worst-case uncertainty bounds are retained [44]. This results in a safe policy improvement mechanism, where exploration remains confined to fixed-time stable actions.

Thus, the proposed policy learning framework integrates reinforcement learning with fixed-time Lyapunov-based safety conditions. The Gaussian process provides probabilistic safety guarantees,

even under model uncertainty, whereas the FxTS condition guarantees uniform convergence to the equilibrium within a predefined bound K independent of the initial conditions. This learning framework admits only those policies that improve task performance and satisfy global FxTS under model uncertainty in nonlinear systems.

The three-line diagram of the proposed algorithm is presented below.

Algorithm 1: Fixed-time Safe Reinforcement Learning

Input: $S_0, \mathrm{D}_0, \mathrm{V(x)}, \mathrm{GP}_0, \beta_\mathrm{k}, \mathrm{a}_1, \mathrm{a}_2, \alpha, \beta$

Output: π_θ, S_k

$i \leftarrow 0$

While not converged do

// Data Collection

Initialize safe set S_0 and initial policy π_{θ_0}

//Apply π_{θ_n} within the certified safe set S_k

Collect transitions data $\mathrm{D_n}$

Update Data set $\mathrm{D_{n+1}}$

// GP Model Update

Train the GP model using D_0

Compute $\mu_\mathrm{k}(\mathrm{x}), \sigma_\mathrm{k}(\mathrm{x}), \mathrm{Residual\ bound}\ \delta(\mathrm{x})$

//Safe Set Expansion

Determine largest c_k such that

$S_{k+1} = \{x: V(x) \leq c_k\}$ satisfies the robust condition

//Policy updation

Update policy $\pi_{\theta_n} \rightarrow \pi_{\theta_{n+1}}$ by maximizing the performance objective subject to the safety constraint condition on safe set

Policy converges or $S_{k+1} = S_k$ (Terminating Condition)

$\mathrm{i} \leftarrow \mathrm{i} + 1$

end while

Return optimized policy π_θ and safe set S_k

5. Simulation results

We evaluate the effectiveness of proposed FxTS-based safe reinforcement learning framework using the classical control problem of an inverted pendulum mounted on a cart as well as vehicle dynamic model.

5.1. Inverted pendulum mounted on a cart

The goal is to stabilize the pendulum in the upright position from a range of initial positions and

velocities, with safety guarantees enforced during the entire learning process. The pendulum dynamics are modeled as a discrete-time nonlinear system of the form:

$$\theta_{k+1} = \theta_k + \Delta t \omega_k \tag{32}$$

$$\omega_{k+1} = \omega_k + \Delta t * \frac{1}{I} * (\tau_k - b\omega_k - mgl_c \sin\theta_k), \tag{33}$$

where $x_k = [\theta_k, \omega_k]^T \in \mathcal{R}^2$ is the state vector with pendulum angular displacement and angular velocity, and $u_k \in \mathcal{R}$ is the control input (force applied to the cart). The unknown dynamics are modeled using GP regression, trained on observed state transitions.

We employ neural network policy to determine actions, which is trained using approximate dynamic programming (ADP). ADP is implemented to improve the policy parameters by using the Bellman updates, and the fixed-time Lyapunov-based safety condition ensures the safety and stability at each policy update. The learning algorithm enforces that every selected action satisfies the FxTS Lyapunov condition under the model uncertainty described by the GP framework. The initial safe set is constructed using a baseline controller with known stability properties, such as linear-quadratic regulator (LQR). The role of the LQR is restricted to offer an initial stabilizing baseline for safe exploration. The set is iteratively expanded by verifying the Lyapunov decrease condition for new states under the uncertainty bounds provided by the GP. This results in a monotonic and data-efficient safe set expansion.

The controller parameters are systematically designed to ensure the fixed-time lyapunov decrement condition and to guarantee admissibility of the safety constraints during the entire learning process. Specifically, the fixed-time controller parameters are initially selected offline to satisfy the convergence behavior and theoretical fixed-time guarantees. These parameters selection is totally based on stability requirement. The policy parameters are then refined online through reinforcement learning within the certified set, with each update satisfying the fixed-time Lyapunov decrease condition and GP based admissible safe set. Thus, all the policy parameter updates enhance the performance without compromising fixed-time convergence guarantees.

In the proposed framework, safety constraints are considered in the context of the certified operating region over which Gaussian process-based uncertainty and fixed-time Lyapunov stability are guaranteed. These safety constraints restrict the system and learning process to operate in the unsafe and unstable conditions. In the inverted pendulum model simulations, state constraints are implemented on the angular displacement and angular velocity, confining the system to a bounded region around the equilibrium point to guarantee the fixed-time stability certification.

The parameters in the Table 1 are selected to simultaneously guarantee safety and improve control performance. The control parameters are selected to satisfy the admissible conditions, that is, $a_1 > 0,\ a_2 > 0,\ 0 < \alpha < 1 < \beta$, which are sufficient conditions for fixed-time stability and to ensure that the discrete-time Lyapunov inequality satisfied. Within the admissible safe set, the values of a_1 and a_2 accelerate the Lyapunov decay rate and reduce the converge time at the cost of high control effort. The parameter α impacts the global decay behavior for larger Lyapunov values,β regulates the local convergence, ensuring the fast convergence. The RL module parameters, which include neural network weights and all the hyper parameters (i.e., learning rates, exploration coefficients, and discount factor) are automatically adjusted through approximate dynamic programming using Bellman consistent policy evaluation and improvement.

To evaluate the control performance, the quadratic cost function defined in the safe policy

optimization section is used. As instantaneous reward is defined as negative of this cost, which penalizes the deviation of the pendulum states from the equilibrium point as well as large control actions. Specifically, the cost in this system depends on angular displacement, angular velocity, and applied torque. Therefore, the performance indicators presented in the results such as states convergence, Lyapunov function evolution, and cumulative reward are consistent with this cost structure and represent the optimization objective while satisfying the discrete-time fixed-time Lyapunov safety condition.

The detailed parameters used in the inverted pendulum simulations are listed in Table 1.

Table 1. Parameter values used in inverted pendulum simulations.

Parameter	Description	Values	Units
Δt	Sampling time interval	0.05	Seconds
M	Mass of pendulum	0.15	Kg (Kilograms)
L	Length of pendulum pole	0.5	m(meters)
B	Coefficient of rotational friction	0.1	N.m.s/rad
G	Acceleration due to gravity	9.8	m/sec^2
a_1, a_2	FxTS Lyapunov function constants	$a_1 = 1.5, a_2 = 2$	–
α, β	FxTS power exponents	$\alpha = 0.5, \beta = 1.5$	–
$V(x)$	Fixed-time Lyapunov function	Quadratic + cross terms	–
GP kernel	Gaussian process kernel type	Squared exponential (ARD)	–
Activation function	NN layer activation	ReLU	–

Based on the control parameters specified in Table 1, the stability analysis and effectiveness of the proposed framework are illustrated in the figures below.

The experimental results show the effectiveness of the proposed FxTS-based, safe RL framework in stabilizing the pendulum while ensuring safety throughout the learning process. The proposed algorithm initializes a certified safe set and expands it monotonically, thereby allowing the states which satisfy the Lyapunov condition in the presence of model uncertainty.

This characterization of the initial and expanded safe set obtained by the FxTS criteria is shown in Figure 2. It is evident that all samples are confined strictly within the certified safe sets throughout the training, verifying the effectiveness of the proposed safely optimized policy learning mechanism and ensuring that data is acquired from regions verified as safe under the stability framework. This behavior guarantees safe exploration and learning without violating safety constraints.

The angular displacement trajectory evolution demonstrates the improvement introduced by the FxTS-based, safe RL policy optimization. The initial policy results in slower convergence as the angular displacement deviates from the steady-state value before reaching equilibrium. The FxTS-optimized policy drives the pendulum to the equilibrium and smooth convergence significantly faster as it approaches the equilibrium point (upright position). The refined policy keeps the angular displacement state with in the verified safe region. This improved convergence behavior is clearly visible in Figure 3. Under the FxTS framework, the learned policy attains higher cumulative rewards as compared to the initial policy, showcasing the enhanced performance as illustrated in Figure 4. The reward accumulation improves reasonably as the learning progresses.

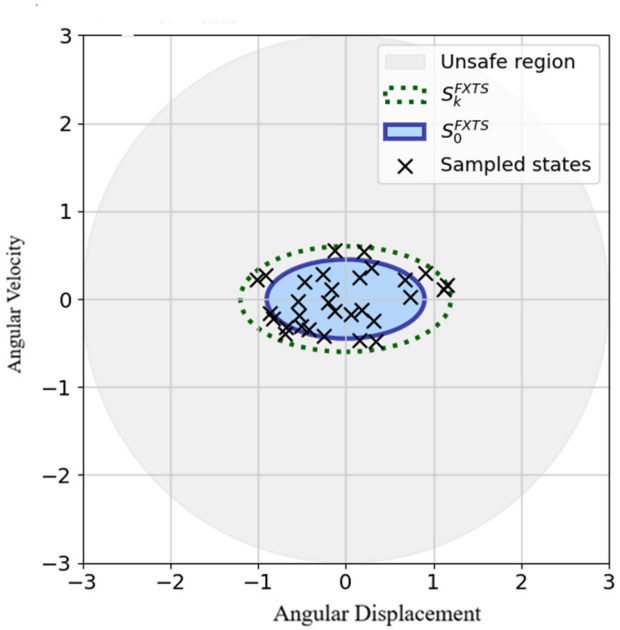

Figure 2. Safe set visualization under the FxTS framework.

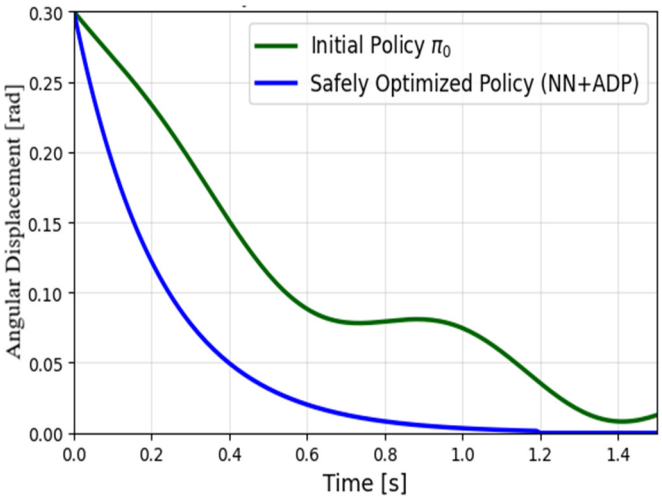

Figure 3. State trajectory under the FxTS framework.

The Lyapunov function converges to zero with in a predetermined time bound, illustrating that the controller ensures FxTS condition and safe stabilization, as illustrated in Figure 5. This complements the state convergence shown in Figure 3 and reinforces the theoretical stability guarantees of the proposed learning framework. The Gaussian process model updates its mean predictions and uncertainty bounds during policy rollouts. As learning proceeds, confidence bounds of the GP model tighten, reflecting reduced uncertainty in the system dynamics and ensuring safer learning, as illustrated in Figure 6. The $\pm 2\sigma$ uncertainty band reflects model confidence, which progressively tightens as more data is collected. The true system dynamics remain enclosed within these bounds, validating the safety margins used during learning.

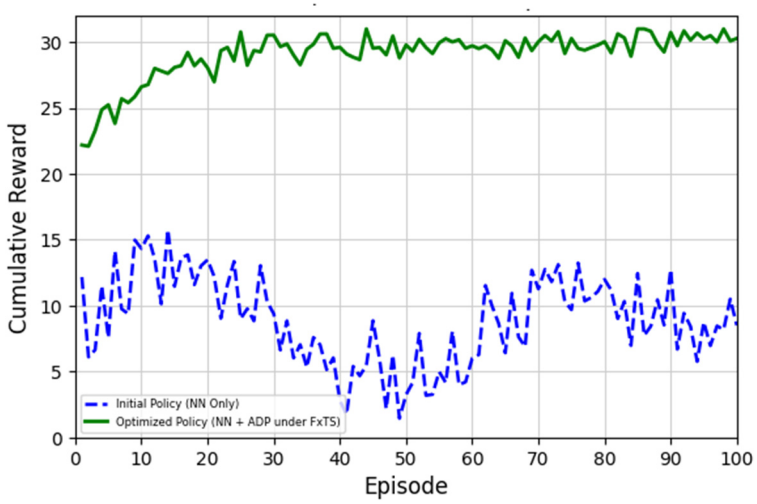

Figure 4. Cumulative reward per episode.

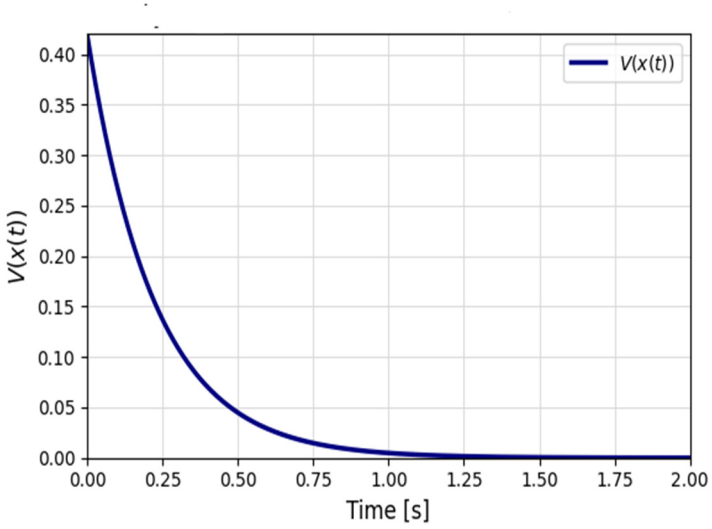

Figure 5. Lyapunov function decay.

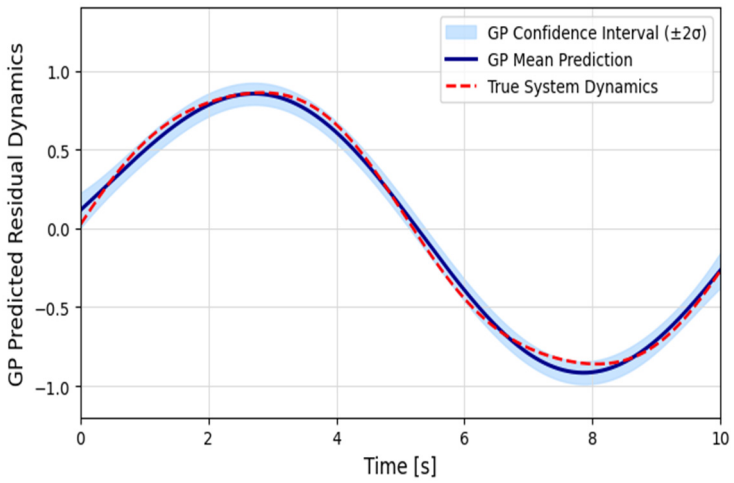

Figure 6. Gaussian process model confidence bounds.

The safe RL algorithm ensures that starting from the initial set, each successive set is enlarged only when the fixed-time Lyapunov decrease condition is satisfied under worst case GP-modeled dynamics. At each iteration, sampled states are confined within the certified boundary, ensuring provably safe exploration. The safe and monotonic learning behavior embedded in our policy optimization strategy as additional candidate states are being verified safe is shown in Figure 7.

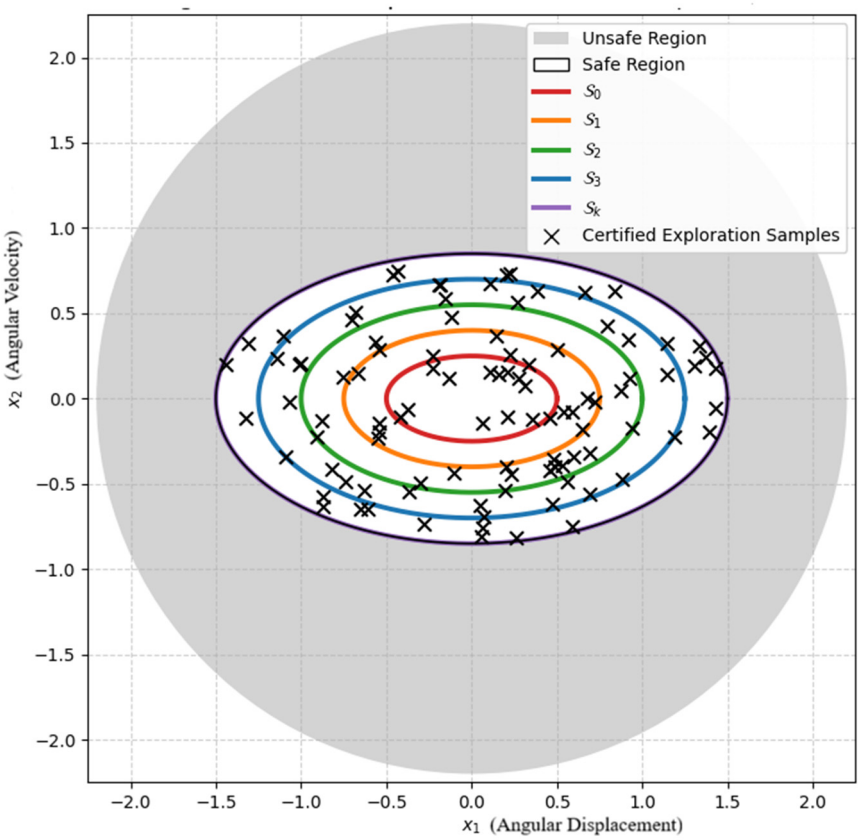

Figure 7. Safe set expansion via certified exploration.

5.2. Reduced-order vehicle dynamic model

To evaluate the applicability of the proposed FxTS-based safe reinforcement learning framework, we implement it on a discrete-time, reduced-order vehicle dynamic model tailored for lane-keeping control. The model states are $x = [y_{pos,k}\theta_k \ \ \dot{y}_{pos,k}\dot{\theta}_k]^T$, capturing the lateral deviation (y_{pos}), heading angle θ_k, and $\dot{y}_{pos}$, $\dot{\theta}_k$ as their respective rates of change. Control inputs are acceleration a and steering angle φ. The safety mechanism is enforced by defining a nonlinear quadratic Lyapunov function $V(z) = \ z^T P z, \ z = [y, \theta]^T$.

The matrix P is chosen in a way to capture the coupling between lateral deviation and heading error. The certified safe region is the sublevel set $S_k = \{z: V(z) \leq c_k\}$. The residual uncertainty is learning online using a GP, and high probability bounds are integrated into the Lyapunov decrease condition. This guarantees that policy updates incorporate the worst-case model error, ensuring the fixed-time convergence property under the uncertain dynamics.

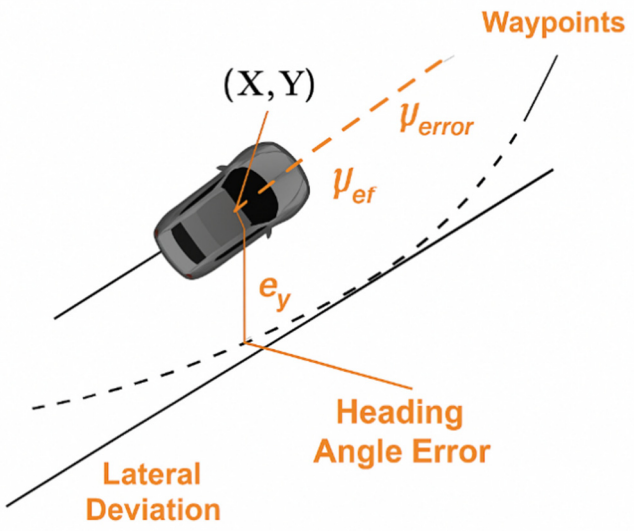

Figure 8. Vehicle model with lateral deviation and heading error.

Through iterative updates, the safe set expands monotonically, and the policy continually operates within the certified safe region while improving closed loop performance over episodes. The detailed parameters used in the vehicle dynamic model simulations are listed in Table 2.

Table 2. Parameters values used in vehicle dynamic model simulations.

Parameter	Symbol & Equation	Value	Units
Wheelbase	L	0.5	Meters
Time step [Episodes]	Δt	0.05	Seconds
Acceleration	a	$a \in [-2\ ,2]$	m/s^2
Steering	φ	$\delta \in [-0.4, 0.4]$	rad
Speed setpoint	v_{ref}	2.0	m/sec
Exponents	(α, β)	(0.5,1.5)	
	$0 < \alpha < 1 < \beta$		
Gains	(a_1, a_2)	(0.35,0.25)	
Certified Bound	T_f	4.0	Seconds

For a vehicle dynamic model platform, safety is dominated by keeping lateral deviation small (staying in the lane) and aligning the heading angle. The initial certified safe set S_0^{FxTS} and the expanded safe set in (y_{pos}, θ) computed using the proposed FxTS-based Lyapunov framework. All the sampled states (black crosses) remain within the certified boundaries, showing that the exploration under the learned policy remains strictly within the verified region, as shown in Figure 9.

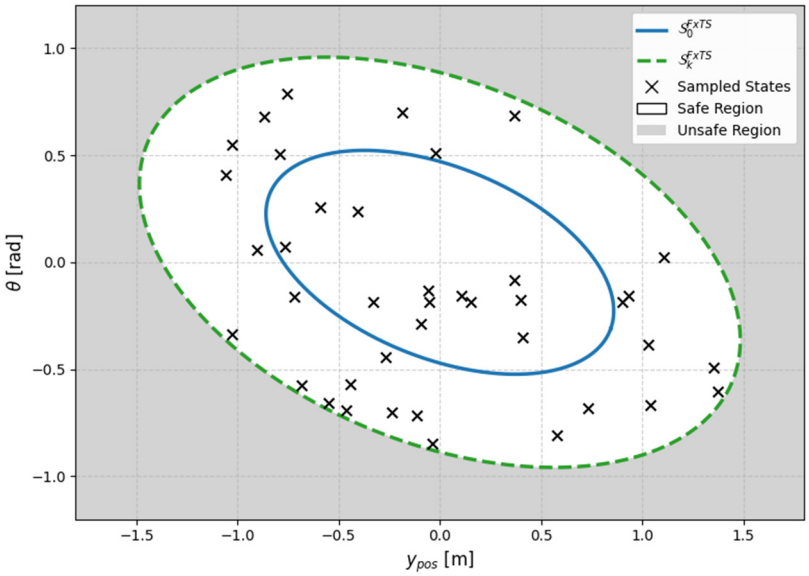

Figure 9. Safe set visualization under the FxTS framework.

The lateral deviation under the safe optimized policy converges to zero more rapidly and smoothly, whereas initial policy shows slower decay and residual offset. The heading angle $\theta(t)$, with the safe optimized policy exhibits improved stability and reduced oscillations. In both cases, the trajectories remain entirely within the certified set, verifying that the learned policy enforces safety and fixed-time convergence guarantees as reflected in Figure 10. The safe optimized policy achieves higher cumulative rewards than the initial policy, showing improved tracking of lateral deviation and heading angle while minimizing the control effort. The clear performance gap reflects the effect of policy optimization under FxTS constraints, which guarantees convergence to the desired state without violating safety constraints, as shown in the Figure 11.

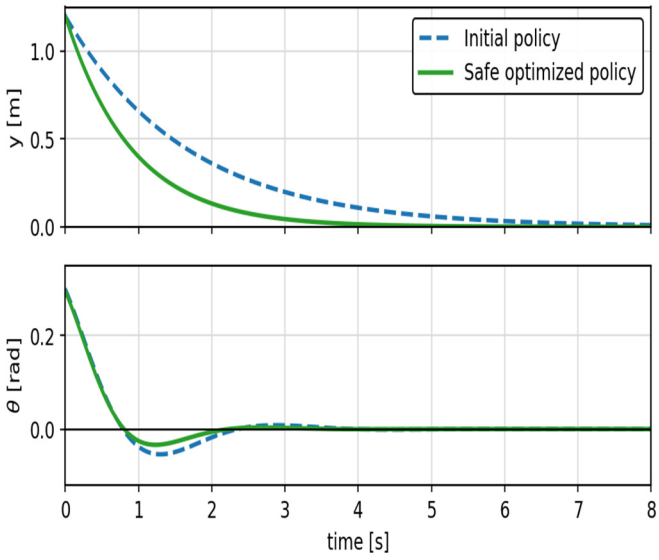

Figure 10. State trajectories under the FxTS framework.

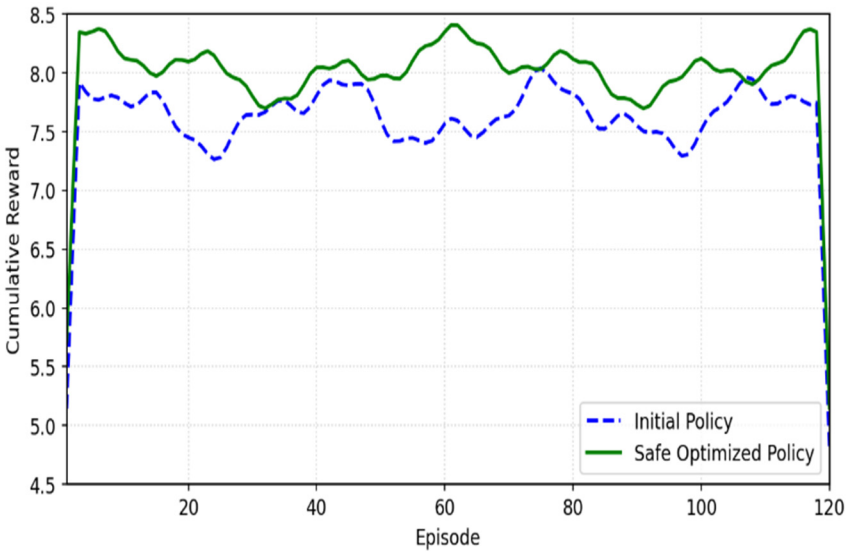

Figure 11. Cumulative reward per episode.

The Lyapunov function exhibits a steady decline from its initial condition to zero, verifying the global FxTS for the defined safe set. The fast convergence shows that the controller effectively regulates lateral deviation and heading angle while ensuring safety constraints, as shown in Figure 12. The GP-based residual dynamics prediction for the vehicle dynamic model states under the proposed fixed-time stability framework reflects that the confidence bounds tighten as more data are collected. This indicates improved model uncertainty, whereas the true residual remains bounded, validating the safety margin in the learning process, as shown in Figure 13.

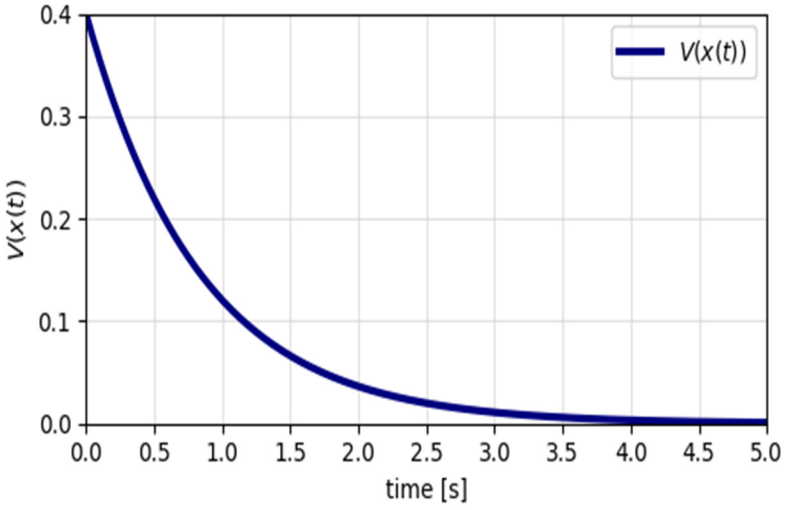

Figure 12. Lyapunov function decay (vehicle model).

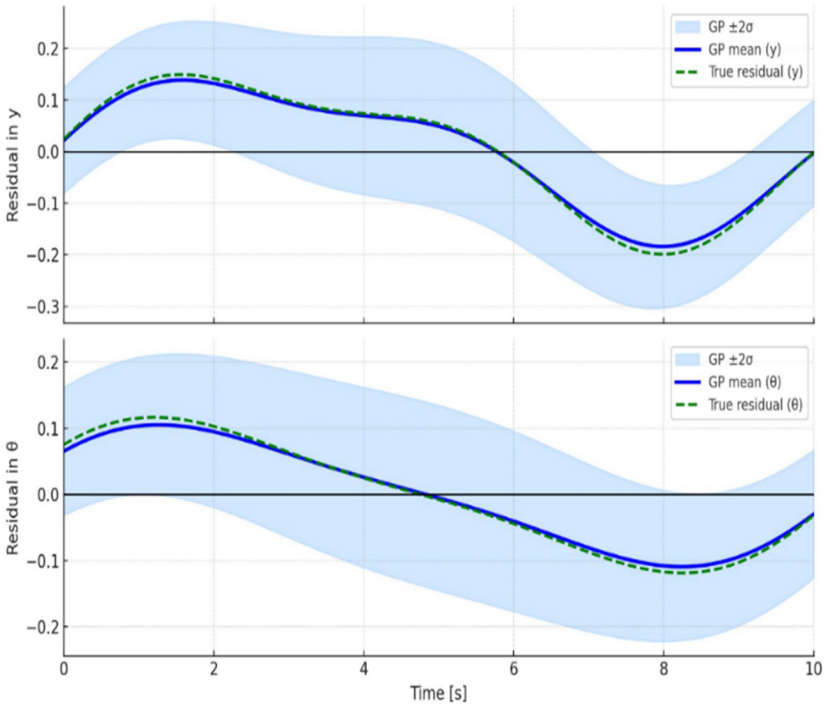

Figure 13. GP confidence over residual dynamics.

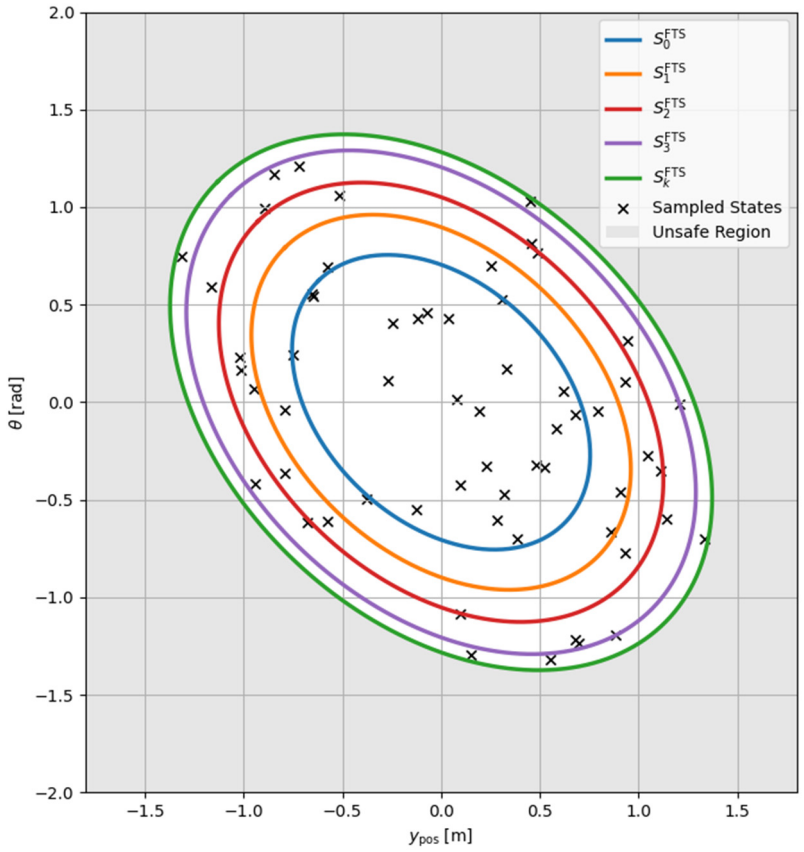

Figure 14. Safe set monotonic expansion.

Figure 14 demonstrates the monotonic expansion of the certified safe set for the vehicle dynamic model under the proposed FxTS framework. The initial safe set S_0^{FxTS} is expanded to S_n^{FxTS} through certified exploration, ensuring that all newly added states satisfy the fixed-time Lyapunov decrease condition. All sampled states remain inside the certified boundary at each iteration, verifying the safety guarantees during exploration, as depicted in Figure 14.

To evaluate the control performance of lateral vehicle dynamics model, the quadratic cost function defined in the safe policy optimization section is used. The instantaneous reward is defined as a negative of this control cost, penalizing the deviations of the states from the equilibrium together with the excessive steering effort. Specifically, the cost is dependent on lateral motion states such as the lateral position, heading angle, and steering control input. The reward function is designed to penalize large lateral deviations and heading errors, with strong penalties near the safe set boundary, while encouraging smooth control inputs. This structured reward framework directs the policy toward achieving improved tracking performance without compromising safety. Consequently, the performance indicators presented in the simulation results, such as states convergence, Lyapunov function evolution, and cumulative reward, are aligned with the cost function formulation and represent the optimization objective while satisfying the discrete-time fixed-time Lyapunov safety condition.

The simulation results conclusively demonstrate the effectiveness of the proposed framework for the nonlinear systems with uncertain dynamics. In both the benchmarks considered for the simulations, that is, the inverted pendulum and vehicle dynamic model, the proposed controller design guarantees the convergence of the state trajectories to the equilibrium point within a uniform, prescribed time, thereby ensuring the FxTS guarantees presented in the theoretical framework. The decreasing evolution of the Lyapunov function and approaching zero within a fixed-time horizon clearly represent that each candidate policy satisfies the fixed-time Lyapunov decrease condition and ensure the fixed-time convergence guarantees through the entire learning process. The safe set evolution clearly demonstrates the monotonic expansion of the certified region of attraction, highlighting the proposed framework's ability to progressively enlarge the admissible region while ensuring the FxTS guarantees. The presented results for both benchmarks clearly verify that the proposed framework achieves fixed-time convergence guarantees, safe exploration, and reinforcement-learning-based policy optimization.

5.3. Qualitative comparative analysis

To further illustrate the behavior of the proposed framework, a qualitative comparative analysis is presented. Two scenarios are considered. First, the performance of FxTS-based safe RL control is compared with an asymptotic safe RL formulation in order to illustrate the differences in convergence behavior and Lyapunov evolution. Second, the behavior of safe RL and unsafe RL analyzed by initializing the system states inside and outside the certified safe region.

The qualitative results further highlight the effectiveness of the proposed framework. In the FxTS, safe RL case, the Lyapunov function exhibits a more rapid decrease, and the system states converge within a bounded time interval relative to the asymptotic formulation as shown in Figures 15 and 16, respectively. This behavior is consistent with the fixed-time stability guarantees ensured by the Lyapunov-based control design. The normalized Lyapunov decay comparison further illustrates the improved convergence behavior of the proposed controller, as shown in Figure 17.

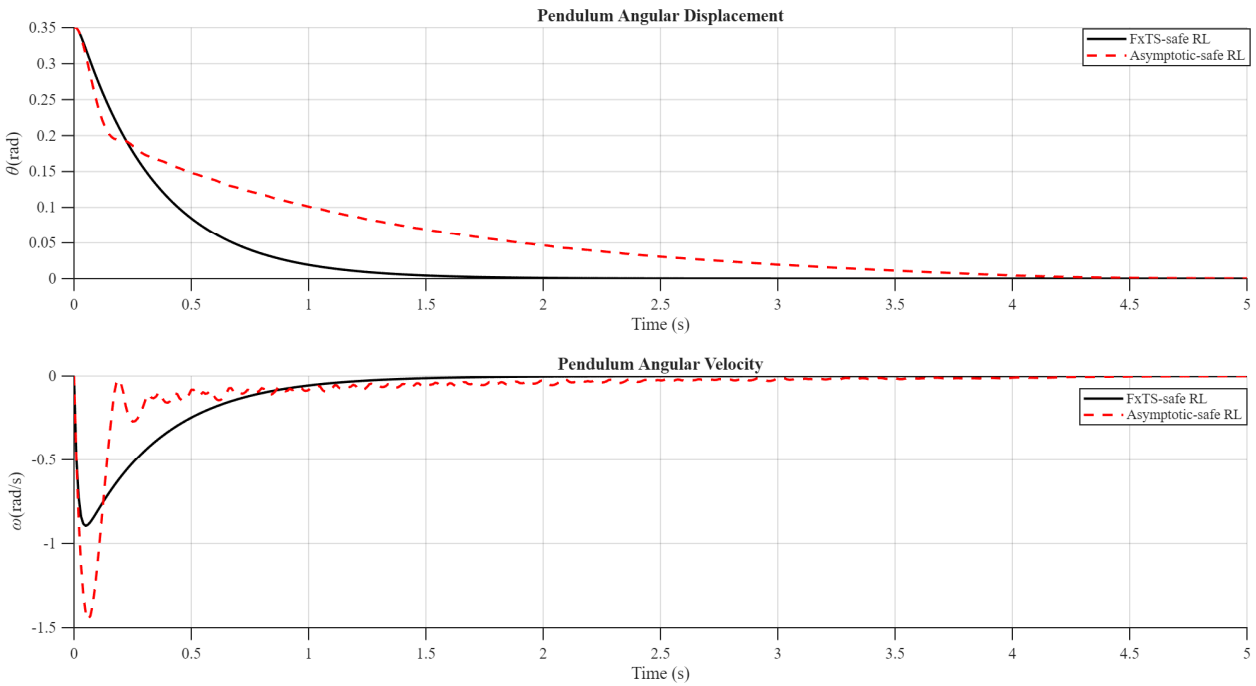

Figure 15. States convergence (FxTS safe RL vs. asymptotic safe RL).

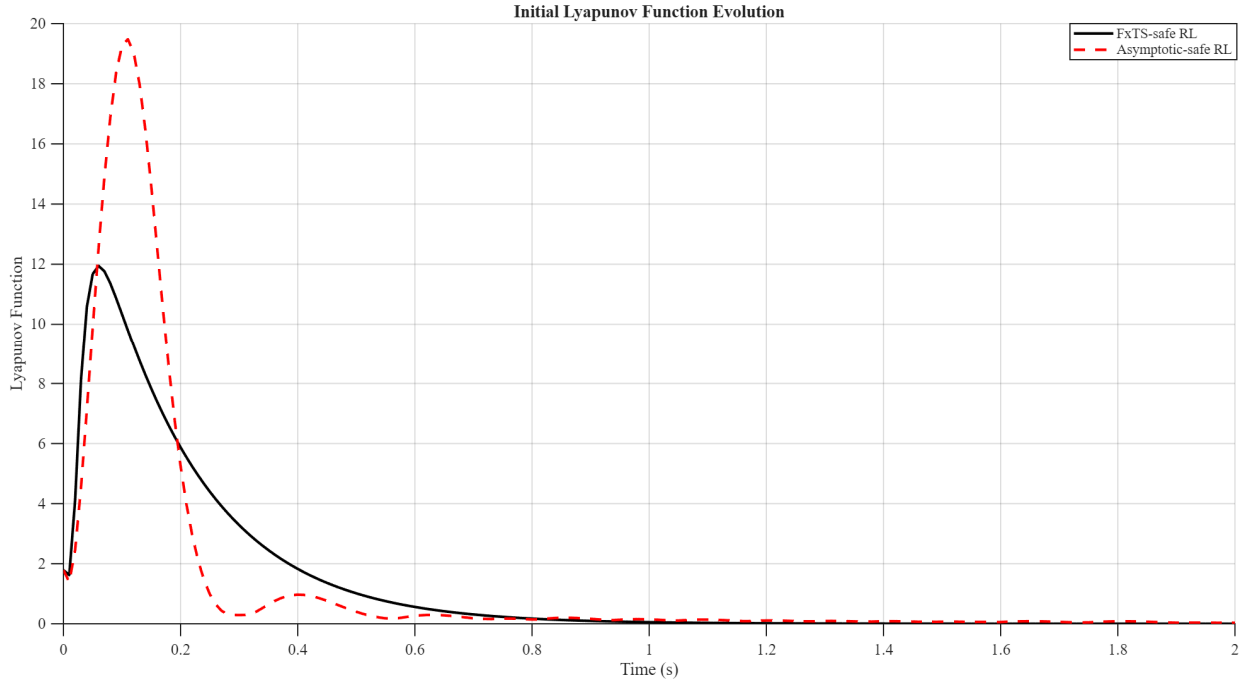

Figure 16. Lyapunov function evolution (FxTS safe RL vs asymptotic safe RL).

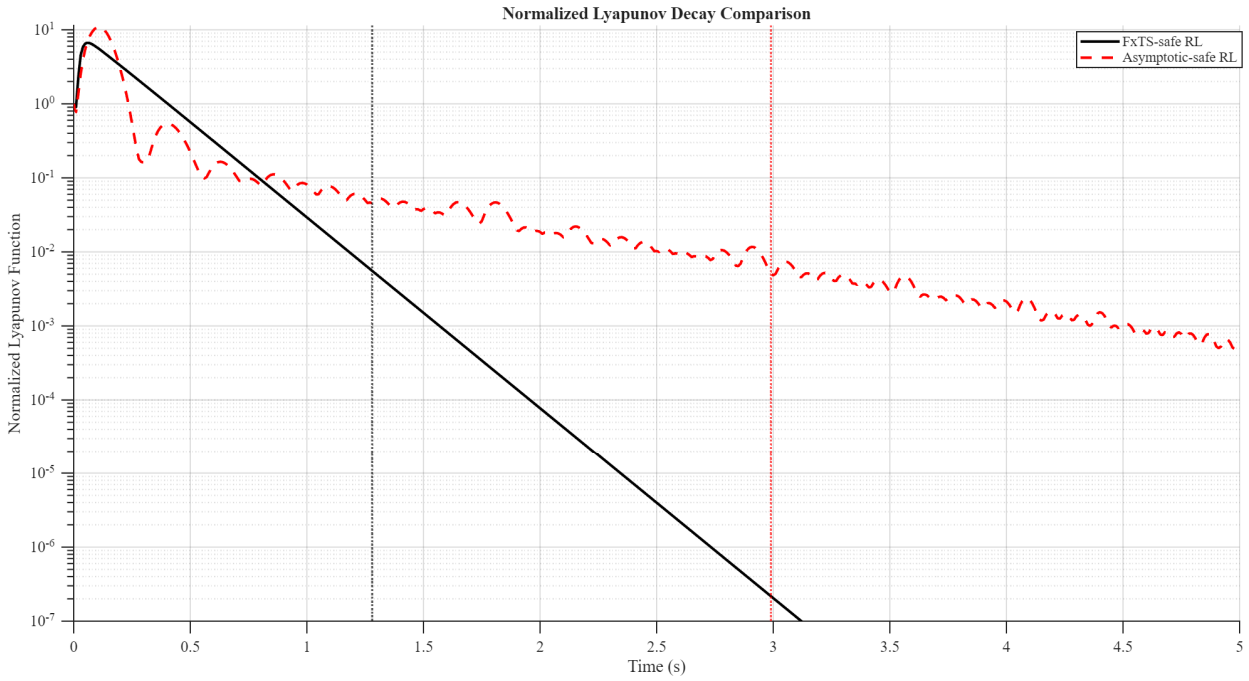

Figure 17. Normalized Lyapunov decay comparison (FxTS safe RL vs. asymptotic safe RL).

The comparison between the safe RL and unsafe RL illustrates the significance of the certified safe region. When the initial condition is selected inside the safe region, the Lyapunov decrease condition is satisfied, and the system states converges to the equilibrium. However, when the initial point is selected outside the certified safe region, the stability guarantees cannot be ensured, and the system exhibits an oscillatory response, as shown in Figures 18 and 19, respectively. These comparisons further demonstrate the stability and safety guarantees ensured by the proposed framework.

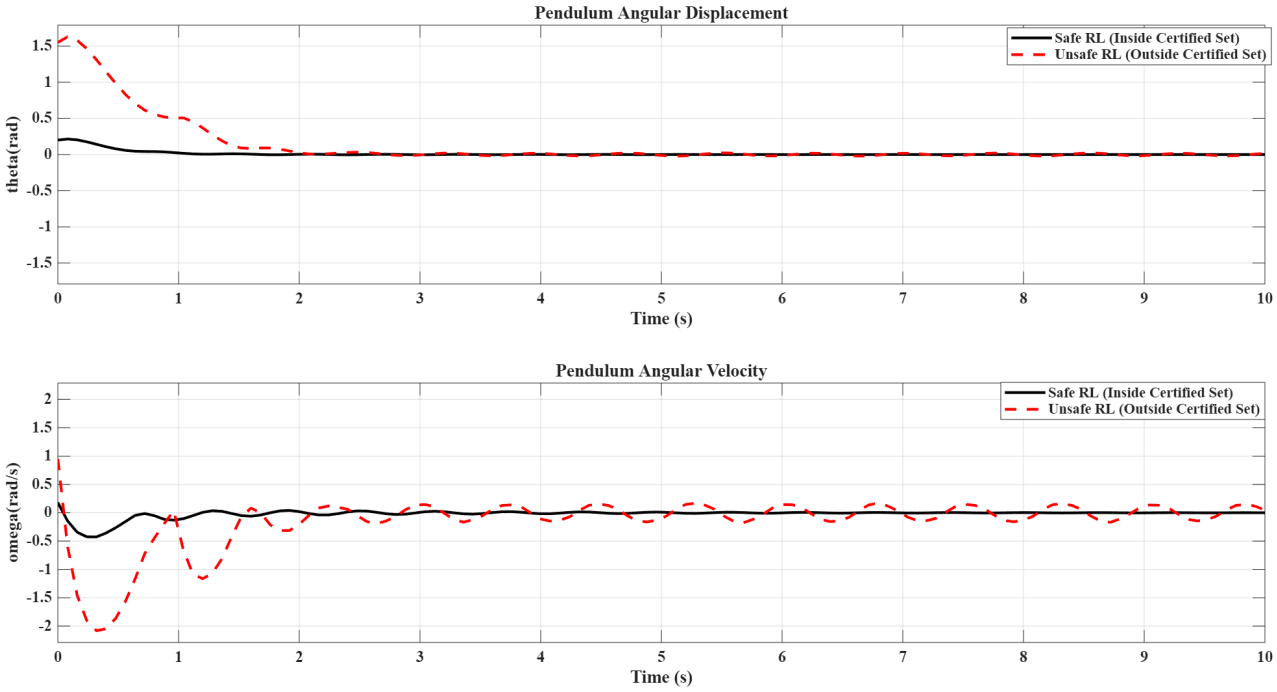

Figure 18. States convergence (safe RL vs. unsafe RL).

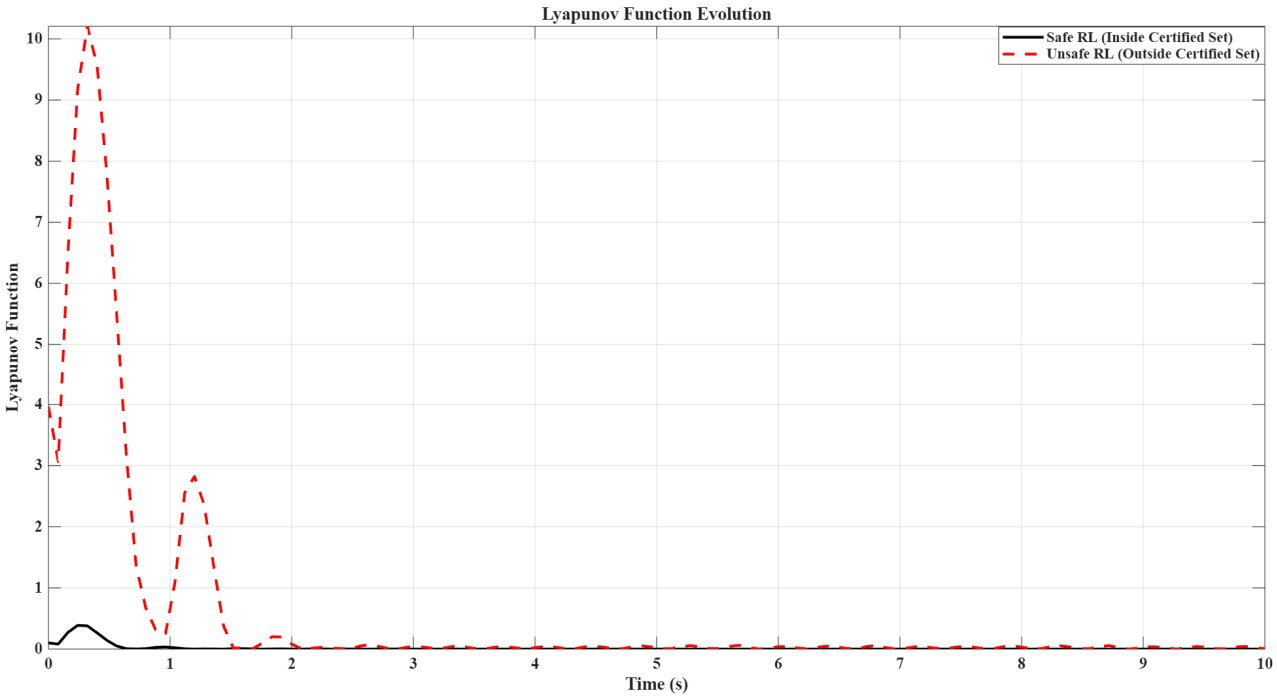

Figure 19. Lyapunov function evolution (safe RL vs. unsafe RL).

6. Conclusions

This study presents a FxTS-driven, safe reinforcement learning framework tailored for nonlinear systems with uncertainty. By integrating a nonlinear quadratic Lyapunov function with GP-based modeling, guarantees both safety and fixed-time convergence irrespective of initial conditions. The neural-network-based policy refined consistently drives the system toward equilibrium within the certified safe region. Simulation results validate the effectiveness of the proposed method to expand the certified safe region and improve policy performance in the presence of uncertainty. As compared to standard approaches, the FxTS-based control framework guarantees convergence within a fixed-time, which is crucial for safety-critical systems.

An important future direction follows from this proposed framework is to implement the methodology on high-fidelity vehicle and robotic systems, including collision avoidance and lane-keeping problems in which fixed-time stability guarantees play a crucial role to provide reliable and predictable transient behavior in safety-critical environments. From the stability perspective, another promising direction will be the combination of robust and sliding-mode control into fixed-time safe reinforcement learning framework. This provides the ability to handle both matched and unmatched disturbance while maintaining fixed-time convergence. Finally, the hardware-based experimental validation of the proposed framework will be a critical step to address. This will be useful to examine the robustness against noise and disturbances, system-unmodeled dynamics, and implementation delays. Moreover, it would be helpful to evaluate the effectiveness of fixed-time, safe reinforcement learning in further real-world applications.

Use of AI tools declaration

The authors declare they have not used artificial intelligence (AI) tools in the creation of this article.

Conflict of interest

The authors declare there is no conflict of interest.

References

1. S. P. Bhat, D. S. Bernstein, Finite-time stability of continuous autonomous systems, *SIAM J. Control Optim.*, **38** (2000), 751–766. https://doi.org/10.1137/S036301299732126X
2. C. Zhang, L. Chang, L. Xing, X. Zhang, Fixed-time stabilization of a class of strict-feedback nonlinear systems via dynamic gain feedback control, *IEEE/CAA J. Autom. Sin.*, **10** (2023), 403–410. https://doi.org/10.1109/JAS.2023.123408
3. W. M. Haddad, K. Verma, V. Chellaboina, Fixed-time stability, uniform strong dissipativity, and stability of nonlinear feedback systems, *Mathematics*, **13** (2025), 1377. https://doi.org/10.3390/math13091377
4. K. Garg, D. Panagou, Robust control barrier and control Lyapunov functions with fixed-time convergence guarantees, in *2021 American Control Conference (ACC)*, (2021), 2292–2297. https://doi.org/10.23919/ACC50511.2021.9482751
5. A. Polyakov, Nonlinear feedback design for fixed-time stabilization of linear control systems, *IEEE Ttrans. Autom. Control*, **57** (2011), 2106–2110. https://doi.org/10.1109/TAC.2011.2179869
6. F. Tatari, H. Modares, Deterministic and stochastic fixed-time stability of discrete-time autonomous systems, *IEEE/CAA J. Autom. Sin.*, **10** (2023), 945–956. https://doi.org/10.1109/JAS.2023.123405
7. J. Lee, W. M. Haddad, Fixed-time stability and optimal stabilisation of discrete autonomous systems, *Int. J. Control*, **96** (2023), 2341–2355. https://doi.org/10.1080/00207179.2022.2092557
8. J. Gu, Y. Wang, A constrained reinforcement learning based approach for cooperative control of multi-UAV in dense obstacle environments, *Sci. China Technol. Sci.*, **69** (2026), 1120601. https://doi.org/10.1007/s11431-025-3076-2
9. X. Liu, L. Zhao, J. Jin, A noise-tolerant fuzzy-type zeroing neural network for robust synchronization of chaotic systems, *Concurrency Comput. Pract. Exp.*, **36** (2024), e8218. https://doi.org/10.1002/cpe.8218s
10. Q. Lu, X. Wu, J. She, F. Guo, L. Yu, Disturbance rejection for systems with uncertainties based on fixed-time equivalent-input-disturbance approach, *IEEE/CAA J. Autom. Sin.*, **11** (2024), 2384–2395. https://doi.org/10.1109/JAS.2024.124650
11. M. T. Vu, S. H. Kim, D. H. Pham, H. L. N. N. Thanh, V. H. Pham, M. Roohi, Adaptive dynamic programming-based intelligent finite-time flexible SMC for stabilizing fractional-order four-wing chaotic systems, *Mathematics*, **13** (2025), 2078. https://doi.org/10.3390/math13132078
12. M. T. Vu, V. T. Nguyen, Q. T. Do, W. Youn, T. H. Nguyen, Robust non-integer predictive control for wind turbine pitch angle regulation in full load regions using deep on-policy learning, *Eng. Appl. Artif. Intell.*, **156** (2025), 111156. https://doi.org/10.1016/j.engappai.2025.111156
13. M. T. Vu, D. H. Pham, V. T. Nguyen, Q. T. Do, A. K. Alanazi, T. H. Nguyen, Adaptive nonlinear integral-backstepping control for frequency stabilization in cyber-physical shipboard microgrids using double deep Q-learning, *Eng. Appl. Artif. Intell.*, **160** (2025), 111943. https://doi.org/10.1016/j.engappai.2025.111943
14. F. Xu, S. Feng, Y. Wang, J. Chang, C. Zhou, Efficient deep reinforcement learning with expert demonstrations for human-machine shared steering control under emergency obstacle avoidance conditions, *IEEE Trans. Veh. Technol.*, **2025** (2025). https://doi.org/10.1109/TVT.2025.3629701

15. W. Yuan, J. Chen, S. Chen, D. Feng, Z. Hu, P. Li, et al., Transformer in reinforcement learning for decision-making: A survey, *Front. Inf. Technol. Electronic Eng.*, **25** (2024), 763–790. https://doi.org/10.1631/FITEE.2300548
16. A. Dong, A. Starr, Y. Zhao, Neural network-based parametric system identification: A review, *Int. J. Syst. Sci.*, **54** (2023), 2676–2688. https://doi.org/10.1080/00207721.2023.2241957
17. M. Forgione, D. Piga, Continuous-time system identification with neural networks: Model structures and fitting criteria, *Eur. J. Control*, **59** (2021), 69–81. https://doi.org/10.1016/j.ejcon.2021.01.008
18. M. Forgione, D. Piga, DynoNet: A neural network architecture for learning dynamical systems, *Int. J. Adapt. Control Signal Process.*, **35** (2021), 612–626. https://doi.org/10.1002/acs.3216
19. C. Qin, X. Ran, D. Zhang, Unsupervised image stitching based on generative adversarial networks and feature frequency awareness algorithm, *Appl. Soft Comput.*, **2025** (2025), 113466. https://doi.org/10.1016/j.asoc.2025.113466
20. D. Zhang, X. Hao, L. Liang, W. Liu, C. Qin, A novel deep convolutional neural network algorithm for surface defect detection, *J. Comput. Design Eng.*, **9** (2022), 1616–1632. https://doi.org/10.1093/jcde/qwac071
21. D. Zhang, X. Hao, D. Wang, C. Qin, B. Zhao, et al., An efficient lightweight convolutional neural network for industrial surface defect detection, *Artif. Intell. Rev*, **56** (2023), 10651–10677. https://doi.org/10.1007/s10462-023-10438-y
22. D. Zhang, C. Yu, Z. Li, C. Qin, R. Xia, A lightweight network enhanced by attention-guided cross-scale interaction for underwater object detection, *Appl. Soft Comput.*, **2025** (2025), 113811. https://doi.org/10.1016/j.asoc.2025.113811
23. X. Song, D. Zheng, S. Song, V. Stojanovic, I. Tejado, Robust anti-disturbance interval type-2 fuzzy control for interconnected nonlinear PDE systems via conjunct observer, *Math. Comput. Simul.*, **227** (2025), 149–167. https://doi.org/10.1016/j.matcom.2024.07.039
24. L. Gao, Z. Zhuang, H. Tao, Y. Chen, V. Stojanovic, Non-lifted norm optimal iterative learning control for networked dynamical systems: A computationally efficient approach, *J. Franklin Inst.*, **361** (2024), 107112. https://doi.org/10.1016/j.jfranklin.2024.107112
25. J. Fang, C. Ren, H. Wang, V. Stojanovic, S. He, Finite-region asynchronous H∞ filtering for 2-D Markov jump systems in Roesser model, *Appl. Math. Comput*, **470** (2024), 128573. https://doi.org/10.1016/j.amc.2024.128573
26. C. Qin, S. Hou, M. Pang, Z. Wang, D. Zhang, Reinforcement learning-based secure tracking control for nonlinear interconnected systems: An event-triggered solution approach, *Eng. Appl. Artif. Intell.*, **161** (2025), 112243. https://doi.org/10.1016/j.engappai.2025.112243
27. C. Qin, M. Pang, Z. Wang, S. Hou, D. Zhang, Observer based fault tolerant control design for saturated nonlinear systems with full state constraints via a novel event-triggered mechanism, *Eng. Appl. Artif. Intell.*, **161** (2025), 112221. https://doi.org/10.1016/j.engappai.2025.112221
28. C. Qin, X. Qiao, J. Wang, D. Zhang, Y. Hou, et al., Barrier-critic adaptive robust control of nonzero-sum differential games for uncertain nonlinear systems with state constraints, *IEEE Trans. Syst. Man Cybern. Syst.*, **54** (2023), 50–63. https://doi.org/10.1109/TSMC.2023.3302656
29. D. Zhang, Y. Wang, L. Meng, J. Yan, C. Qin, Adaptive critic design for safety-optimal FTC of unknown nonlinear systems with asymmetric constrained-input, *ISA Trans.*, **155** (2024), 309–318. https://doi.org/10.1016/j.isatra.2024.09.018

30. D. Zhang, Q. Yuan, L. Meng, R. Xia, W. Liu, C. Qin, Reinforcement learning for single-agent to multi-agent systems: From basic theory to industrial application progress, a survey, *Artif. Intell. Rev.*, **2025** (2025). https://doi.org/10.1007/s10462-025-11439-9
31. G. Dalal, K. Dvijotham, M. Vecerik, T. Hester, C. Paduraru, Y. Tassa, Safe exploration in continuous action spaces, preprint, arXiv:180108757. https://doi.org/10.48550/arXiv.1801.08757
32. S. Gu, L. Yang, Y. Du, G. Chen, F. Walter, J. Wang, et al., A review of safe reinforcement learning: Methods, theories and applications, *IEEE Trans. Pattern Anal. Mach. Intell.*, **46** (2024), 11216–11235. https://doi.org/10.1109/TPAMI.2024.3457538
33. I. Salehi, T. Taplin, A. P. Dani, Learning discrete-time uncertain nonlinear systems with probabilistic safety and stability constraints, *IEEE Open J. Control Syst.*, **1** (2022), 354–365. https://doi.org/10.1109/OJCSYS.2022.3216545
34. F. Berkenkamp, R. Moriconi, A. P. Schoellig, A. Krause, Safe learning of regions of attraction for uncertain, nonlinear systems with gaussian processes, in *2016 IEEE 55th Conference on Decision and Control (CDC)*, (2016), 4661–4666. https://doi.org/10.1109/CDC.2016.7798979
35. J. Schreiter, D. Nguyen-Tuong, M. Eberts, B. Bischoff, H. Markert, M. Toussaint, Safe exploration for active learning with Gaussian processes, in *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*, (2015), 133–149. https://doi.org/10.1007/978-3-319-23461-8_9
36. J. Garcıa, F. Fernández, A comprehensive survey on safe reinforcement learning, *J. Mach. Learn. Res.*, **16** (2015), 1437–1480.
37. J. Baxter, P. L. Bartlett, Infinite-horizon policy-gradient estimation, *J. Artif. Intell. Res.*, **15** (2001), 319–350. https://doi.org/10.1613/jair.806
38. Y. Wang, Z. Wu, Control lyapunov-barrier function-based safe reinforcement learning for nonlinear optimal control, *AIChE J.*, **70** (2024), e18306. https://doi.org/10.1002/aic.18306
39. A. Wachi, W. Hashimoto, X. Shen, K. Hashimoto, Safe exploration in reinforcement learning: A generalized formulation and algorithms, *Adv. Neural Inf. Process. Syst.*, **36** (2023), 29252–29272. https://doi.org/10.52202/075280-1273
40. H. K. Khalil, J. W. Grizzle, *Nonlinear Systems*, Prentice Hall Upper Saddle River, NJ.
41. F. Berkenkamp, M. Turchetta, A. Schoellig, A. Krause, Safe model-based reinforcement learning withstability guarantees, *Adv. Neural Inf. Process. Syst.*, **30** (2017).
42. Y. Chow, O. Nachum, E. Duenez-Guzman, M. Ghavamzadeh, A lyapunov-based approach to safe reinforcement learning, *Adv. Neural Inf. Process. Syst.*, **31** (2018).
43. L. Manda, S. Chen, M. Fazlyab, Learning performance-oriented control barrier functions under complex safety constraints and limited actuation, *preprint*, arXiv:240105629. https://doi.org/10.48550/arXiv.2401.05629
44. R. Munos, T. Stepleton, A. Harutyunyan, M. Bellemare, Safe and efficient off-policy reinforcement learning, *Adv. Neural Inf. Process. Syst.*, **29** (2016).
45. D. Bertsekas, *Dynamic Programming and Optimal Control: Volume I*, Athena Scientific, 2012.