Novelty Assessment Report

Paper: The Sample Complexity of Online Reinforcement Learning: A Multi-model Perspective

PDF URL: https://openreview.net/pdf?id=xCelVyUVO2

Venue: ICLR 2026 Conference Submission

Year: 2026

Report Generated: 2026-01-05

Abstract

We study the sample complexity of online reinforcement learning in the general setting of nonlinear dynamical systems with continuous state and action spaces. Our analysis accommodates a large class of dynamical systems ranging from a finite set of nonlinear candidate models to models with bounded and Lipschitz continuous dynamics, to systems that are parametrized by a compact and real-valued set of parameters. In the most general setting, our algorithm achieves a policy regret of $\mathcal{O}(N \epsilon^2 + \mathrm{ln}(m(\epsilon))/\epsilon^2)$, where N is the time horizon, ϵ is a user-specified discretization width, and $m(\epsilon)$ measures the complexity of the function class under consideration via its packing number. In the special case where the dynamics are parametrized by a compact and real-valued set of parameters (such as neural networks, transformers, etc.), we prove a policy regret of $\mathcal{O}(\sqrt{N p})$, where P denotes the number of parameters, recovering earlier sample-complexity results that were derived for linear time-invariant dynamical systems. While this article focuses on characterizing sample complexity, the proposed algorithms are likely to be useful in practice, due to their simplicity, their ability to incorporate prior knowledge, and their benign transient behaviors.

Core Task Landscape

This paper addresses: **Sample Complexity of Online Reinforcement Learning with Nonlinear Dynamics**

A total of **50 papers** were analyzed and organized into a taxonomy with **17 categories**.

Taxonomy Overview

The research landscape has been organized into the following main categories:

- **Theoretical Foundations and Sample Complexity Bounds**
- **Algorithm Design and Optimization Methods**
- **Application Domains and Empirical Studies**

Complete Taxonomy Tree

- Sample Complexity of Online Reinforcement Learning with Nonlinear Dynamics Survey Taxonomy
- Theoretical Foundations and Sample Complexity Bounds
  - General Nonlinear Dynamics and Multi-Model Settings ★ (4 papers)
  - [0] The Sample Complexity of Online Reinforcement Learning: A Multi-model Perspective (Anon et al., 2026) View paper
  - [5] Sample Complexity for Nonlinear Dynamics (Chen Yong-xin, 2022) View paper
  - [28] Sample efficient reinforcement learning in continuous state spaces: A perspective beyond linearity (Malik, 2021) View paper
  - [34] The Sample Complexity of Learning Dynamical Systems (Sattar, 2023) View paper
  - Structured Representations and Operator-Based Methods (3 papers)
  - [2] Rich-observation reinforcement learning with continuous latent dynamics (Song Yu-da, 2024) View paper
  - [37] Near-Optimal Sample Complexity in Reward-Free Kernel-Based Reinforcement Learning (Vakili, 2025) View paper
  - [40] Sparse Learning of Kernel Transfer Operators (Boya Hou, 2021) View paper
  - Robustness and Distributional Shift (3 papers)
  - [1] Sample Complexity of Distributionally Robust Off-Dynamics Reinforcement Learning with Online Interaction (He Yiting, 2025) View paper
  - [17] Hybrid Transfer Reinforcement Learning: Provable Sample Efficiency from Shifted-Dynamics Data (QU Chengrui, 2024) View paper
  - [41] Adversarially Robust Stability Certificates can be Sample-Efficient (Zhang, 2022) View paper
  - Stability-Constrained and Safe Learning (3 papers)
  - [11] On the sample complexity of stability constrained imitation learning (Tu, 2022) View paper
  - [26] Non-asymptotic and Accurate Learning of Nonlinear Dynamical Systems (Yahya Sattar, 2022) View paper
  - [27] Exact Recovery Guarantees for Parameterized Non-linear System Identification Problem under Adversarial Attacks (Zhang Hai-Xiang, 2024) View paper
  - Reward-Free and Exploration-Focused Settings (3 papers)
  - [10] Active Learning for Nonlinear System Identification with Guarantees (Mania, 2022) View paper
  - [18] Optimal Exploration for Model-Based RL in Nonlinear Systems (Wagenmaker, 2023) View paper
  - [23] On the statistical efficiency of reward-free exploration in non-linear rl (Chen, 2022) View paper
  - Linear and Low-Complexity Systems (1 papers)
  - [3] Convergence and sample complexity of policy gradient methods for stabilizing linear systems (Feiran Zhao, 2024) View paper
- Algorithm Design and Optimization Methods
  - Model-Based Policy Optimization (4 papers)
  - [4] Neural network dynamics for model-based deep reinforcement learning with model-free fine-tuning (Nagabandi, 2018) View paper

- [21] SOMBRL: Scalable and Optimistic Model-Based RL (Bhavya Sukhija, 2025) View paper
- [22] Adaptive Nonlinear Model Predictive Horizon Using Deep Reinforcement Learning for Optimal Trajectory Planning (Younes Al Younes, 2022) View paper
- [42] Controllable Flow Matching for Online Reinforcement Learning (Bin Wang, 2025) View paper
- Model-Free and Hybrid Approaches (3 papers)
- [20] Two-step reinforcement learning for model-free redesign of nonlinear optimal regulator (Mei Minami, 2023) View paper
- [30] Non-linear reinforcement learning in large action spaces: Structural conditions and sample-efficiency of posterior sampling (Agarwal, 2022) View paper
- [36] A trust region approach for few-shot sim-to-real reinforcement learning (Paul Daoudi, 2023) View paper
- Algorithm Design for Safety-Critical Systems (4 papers)
- [6] KCRL: Krasovskii-Constrained Reinforcement Learning with Guaranteed Stability in Nonlinear Dynamical Systems (Lale, 2022) View paper
- [8] KCRL: Krasovskii-Constrained Reinforcement Learning with Guaranteed Stability in Nonlinear Discrete-Time Systems (Sahin Lale, 2023) View paper
- [13] Sample-efficient Safe Learning for Online Nonlinear Control with Control Barrier Functions (Wenhao Luo, 2022) View paper
- [19] Fixedâ□□Time Stable Gradient Flows for Optimal Adaptive Control of Continuousâ□□Time Nonlinear Systems (Mahdi Niroomand, 2024) View paper
- Operator-Based and Kernel Methods (3 papers)
- [9] Data-efficient reinforcement learning for complex nonlinear systems (Vrushabh S. Donge, 2023) View paper
- [31] Sample-Efficient Online Control Policy Learning with Real-Time Recursive Model Updates (Zhang Zi-xin, 2025) View paper
- [45] From Embedding to Control: Representations for Stochastic Multi-Object Systems (Cheng Xiaoyuan, 2025) View paper
- Deep Learning and Neural Architectures (3 papers)
- [24] Sample-efficient diffusion-based control of complex nonlinear systems (Chen Hongyi, 2025) View paper
- [38] Kolmogorov-Arnold inspired convolutional networks for enhancing PPO-based online reinforcement learning (N. Islam, 2025) V iew paper
- [47] Deep reinforcement learning with symmetric data augmentation applied for aircraft lateral attitude tracking control (Li Yifei, 2024) View paper
- Preference-Based and Reward Learning (1 papers)
- [7] Sample-Efficient Preference-based Reinforcement Learning with Dynamics Aware Rewards (Metcalf, 2024) View paper
- Adaptive and Online Learning Mechanisms (3 papers)
- [29] Data-Based Feedback Relearning Control for Uncertain Nonlinear Systems With Actuator Faults (Chaoxu Mu, 2022) View paper
- [43] Multi-Kernel Enhanced Receding-Horizon Reinforcement Learning for Steering Control of Intelligent Vehicles (Changxin Zhang, 2025) View paper
- [50] Adaptive Self-Organizing Clustering Dual-Buffer Safe Reinforcement Learning for Nonlinear Optimal Control (Roya Khalili Amirabadi, n.d.) View paper

• Application Domains and Empirical Studies
- Robotics and Autonomous Systems (5 papers)
- [32] A sample efficient model-based deep reinforcement learning algorithm with experience replay for robot manipulation (Cheng Zhang, 2020) View paper
- [44] On the computation of Control Lyapunov Functions for building safe and stabilizing controllers (Guzman, 2024) View paper
- [46] A hierarchical reinforcement learning approach for optimal path tracking of wheeled mobile robots (Lei Zuo, 2013) View paper
- [48] Scalable Supervision for Safe and Efficient Online Robot Learning (Balakrishna, 2022) View paper
- [49] Model-Based Design for Legged Robots: Predictive Control and Reinforcement Learning (Agrawal, 2022) View paper
- PDE-Governed Systems and Continuous Spatial Dynamics (2 papers)
- [14] Learning a model is paramount for sample efficiency in reinforcement learning control of PDEs (Werner, 2023) View paper
- [15] Numerical Evidence for Sample Efficiency of Model-Based Over Model-Free Reinforcement Learning Control of Partial Differential Equations (Stefan Werner, 2024) View paper
- Multi-Agent Systems and Flocking Control (1 papers)
- [16] Improving Sample Efficiency of Multiagent Reinforcement Learning With Nonexpert Policy for Flocking Control (Yunbo Qiu, 2023) View paper
- General Empirical Studies and Benchmarks (5 papers)
- [12] Sample-Efficient Reinforcement Learning and Its Applications (Hua Zheng, 2024) View paper
- [25] Safe reinforcement learning for power system control: A review (Peipei Yu, 2024) View paper
- [33] Learning and Control of Dynamical Systems (Lale, 2023) View paper
- [35] LEADS: Learning Dynamical Systems that Generalize Across Environments (Yin Yuan, 2022) View paper
- [39] Reinforcement Learning vs Optimal Control: Sparse Nonlinear Dynamical Systems Between Theory and Practice (D Maran, 2025) View paper

Narrative

Core task: sample complexity of online reinforcement learning with nonlinear dynamics. The field addresses how many interactions an agent needs to learn effective policies when the underlying system evolves according to complex, nonlinear rules. The taxonomy organizes research into three main branches: Theoretical Foundations and Sample Complexity Bounds, which develops rigorous guarantees for learning efficiency under various structural assumptions; Algorithm Design and Optimization Methods, which proposes concrete techniques—ranging from model-based planning to policy gradient stabilization—that exploit or adapt to nonlinear structure; and Application Domains and Empirical Studies, which demonstrate these ideas in robotics, power systems, and other real-world settings. Representative works span from early neural dynamics models like Neural Dynamics Model-Based[4] to recent advances in robust off-dynamics transfer (Robust Off-Dynamics RL[1]) and reward-free exploration (Reward-free Nonlinear Efficiency[23]), illustrating both the maturation of theoretical tools and the growing diversity of practical deployment scenarios.

Several active lines of work highlight key trade-offs and open questions. One thread examines how structural properties—such as low-dimensional latent representations (Rich-observation Latent Dynamics[2]) or specific function classes—can be leveraged to improve sample efficiency, while another explores robustness when the learned model or policy must transfer across different dynamics (Robust Off-Dynamics RL[1], Hybrid Transfer RL[17]). Safety-aware methods (Safe Learning Barrier Functions[13], Krasovskii-Constrained RL[6]) balance exploration with constraint satisfaction, a concern especially pressing in physical systems. Within this landscape, Multi-model Online RL[0] sits naturally among works addressing general nonlinear settings and multi-model scenarios, sharing thematic ground with Nonlinear Dynamics Complexity[5] and Learning Dynamical Systems Complexity[34]. Its emphasis on handling multiple candidate models distinguishes it from single-model analyses, offering a complementary perspective on how uncertainty over system structure affects sample complexity and algorithmic design.

Related Works in Same Category

The following **3 sibling papers** share the same taxonomy leaf node with the original paper:

1. Sample Complexity for Nonlinear Dynamics

Authors: Chen Yong-xin, Yongxin Chen, Vaidya Umesh, Umesh Vaidya, U. Vaidya | **Year/Venue**: 2022 | **URL**: View paper

Abstract

We consider the identification problems for nonlinear dynamical systems. An explicit sample complexity bound in terms of the number of data points required to recover the models accurately is derived. Our results extend recent sample complexity results for linear dynamics. Our approach for obtaining sample complexity bounds for nonlinear dynamics relies on a linear, albeit infinite dimensional, representation of nonlinear dynamics provided by Koopman and Perron-Frobenius operator. We exploit the...

Relationship Analysis

Both papers belong to the General Nonlinear Dynamics and Multi-Model Settings category, establishing sample complexity bounds for nonlinear systems. They overlap in addressing sample complexity for nonlinear dynamics with continuous state spaces, but differ fundamentally in approach: the original paper focuses on online reinforcement learning with policy regret guarantees using posterior sampling and multi-model adaptive control, while the candidate paper focuses on system identification using Koopman operator theory to lift nonlinear dynamics to infinite-dimensional linear representations, without addressing the exploration-exploitation tradeoff or control policy optimization.

2. Sample efficient reinforcement learning in continuous state spaces: A perspective beyond linearity

Authors: Malik, Dhruv, Dhruv Malik, Pacchiano, Aldo, et al. (12 authors total) | **Year/Venue**: 2021 | **URL**: View paper

Abstract

Reinforcement learning (RL) is empirically successful in complex nonlinear Markov decision processes (MDPs) with continuous state spaces. By contrast, the majority of theoretical RL literature requires the MDP to satisfy some form of linear structure, in order to guarantee sample efficient RL. Such efforts typically assume the transition dynamics or value function of the MDP are described by linear functions of the state features. To resolve this discrepancy between theory and practice, we intro...

Relationship Analysis

Both papers belong to the General Nonlinear Dynamics and Multi-Model Settings category, addressing sample complexity for broad classes of nonlinear systems. They overlap in studying online RL with nonlinear dynamics and continuous state spaces, but differ fundamentally in their approaches: the original paper uses a multi-model framework with Hedge-type updates and posterior sampling across finite/parametric model classes, while the candidate paper introduces the Effective Planning Window (EPW) condition that requires no linearity assumptions and focuses on gaming benchmarks like Atari, using a different structural condition based on planning horizons rather than model selection.

3. The Sample Complexity of Learning Dynamical Systems

Authors: Sattar, Yahya | **Year/Venue**: 2023 | **URL**: View paper

Abstract

Machine learning has emerged as a leading force in revolutionizing technology, education, and almost every aspect of our lives. Reinforcement learning is a sub-field of machine learning that deals with the effects of dynamic feedback and systems that interact with the environment. In these settings, classic statistical and algorithmic guarantees often do not hold because of non i.i.d. data, dynamic feedback, and distribution shift.We develop a framework for single trajectory learning of nonlinea...

Relationship Analysis

Both papers belong to the General Nonlinear Dynamics and Multi-Model Settings category, addressing sample complexity for broad classes of nonlinear systems. They overlap in studying online reinforcement learning with nonlinear dynamics and establishing sample complexity bounds using single-trajectory learning. The original paper focuses on multi-model posterior sampling with Hedge-type updates achieving O(ln(m)) regret for finite model sets and O(√Np) for parametric systems, while the candidate paper emphasizes mixing arguments and empirical risk minimization landscape analysis for learning nonlinear dynamical systems, bilinear systems, and Markov jump systems with optimal sample complexity rates.

Contributions Analysis

Overall novelty summary. The paper develops online reinforcement learning algorithms with nonasymptotic policy-regret guarantees for nonlinear continuous-state-action systems, accommodating finite candidate model sets, Lipschitz-bounded dynamics, and compact parameter spaces. It resides in the 'General Nonlinear Dynamics and Multi-Model Settings' leaf, which contains four papers total—a moderately populated niche within the broader 'Theoretical Foundations and Sample Complexity Bounds' branch. This leaf explicitly focuses on broad nonlinear classes and multi-model scenarios, distinguishing it from specialized structural assumptions like Koopman operators or kernel embeddings found in sibling leaves.

The taxonomy reveals neighboring leaves addressing structured representations (Koopman/kernel methods, three papers), robustness under distributional shift (three papers), stability-constrained learning (three papers), and reward-free exploration (three papers). The paper's multi-model framing and general function-class treatment position it at the intersection of model-selection uncertainty and nonlinear complexity, diverging from operator-based linearization strategies and from purely adversarial robustness analyses. Its scope note explicitly excludes specialized structural assumptions, aligning with the leaf's mandate to handle parametric and function-approximation settings without restrictive linearity or embedding constraints.

Among 26 candidates examined across three contributions, none yielded clear refutations. The first contribution (suite of algorithms with regret guarantees) examined 10 candidates with zero refutable overlaps; the second (frequentist guarantees under persistence of excitation) examined 6 with zero refutations; the third (separation principle for model identification and control) examined 10 with zero refutations. This limited search scope—top-K semantic matches plus citation expansion—suggests that within the examined neighborhood, the specific combination of multi-model handling, general nonlinear function classes, and nonasymptotic regret bounds appears less directly addressed by prior work.

Based on the 26-candidate search, the work's novelty appears to stem from its unified treatment of diverse nonlinear model classes and its explicit multi-model perspective, rather than from introducing entirely new algorithmic primitives. The taxonomy context indicates a moderately active research direction with established sibling papers, yet the contribution-level statistics show no immediate prior work overlap within the examined set. Acknowledging the search's limited scope, a more exhaustive review might uncover closer antecedents, particularly in adjacent leaves addressing parametric or function-approximation settings.

This paper presents **3 main contributions**, each analyzed against relevant prior work:

Contribution 1: Suite of online RL algorithms with nonasymptotic policy-regret guarantees for nonlinear continuous systems

Description: The authors introduce multiple algorithms that achieve provable policy-regret bounds in the online non-episodic setting with continuous state-action spaces and nonlinear dynamics. These algorithms build on Hedge-type updates and posterior sampling while incorporating exploration to ensure rapid convergence of the posterior distribution over models.

This contribution was assessed against **10 related papers** from the literature. Papers with potential prior art are analyzed in detail with textual evidence; others receive brief assessments.

1. Regret guarantees for online deep control

URL: View paper

Brief Assessment

Online Deep Control Regret[67] focuses on online episodic control with neural network policies over linear time-varying dynamics, not the general non-episodic setting with nonlinear dynamics that the original paper addresses.

2. Local Linearity: the Key for No-regret Reinforcement Learning in Continuous MDPs

URL: View paper

Brief Assessment

Local Linearity No-Regret[71] focuses on a different structural assumption (local linearity) and introduces a novel MDP representation class. The original paper's contribution centers on Hedge-type updates and posterior sampling with exploration, which is a distinct algorithmic approach not addressed in the candidate.

3. Sample-efficient and Scalable Exploration in Continuous-Time RL

URL: View paper

Brief Assessment

Continuous-Time Exploration[69] focuses on continuous-time dynamics using ODEs with probabilistic models (Gaussian processes, Bayesian neural networks) for uncertainty-aware learning. The original paper addresses discrete-time nonlinear dynamical systems with a multi-model perspective and Hedge-type updates. These represent fundamentally different technical approaches to continuous control.

4. Sublinear regret for an actor-critic algorithm in continuous-time linear-quadratic reinforcement learning

URL: View paper

Brief Assessment

Actor-Critic Continuous-Time LQR[70] focuses on continuous-time linear-quadratic systems with actor-critic methods, not the general nonlinear dynamics with Hedge-type updates and posterior sampling addressed in the original paper.

5. No-regret reinforcement learning in smooth mdps

URL: View paper

Brief Assessment

No-Regret Smooth MDPs[72] focuses on smooth MDPs with orthogonal feature representations (Legendre polynomials) and achieves regret bounds under smoothness assumptions. The original paper addresses general nonlinear dynamics with multi-model perspectives and hedge-type updates, which is a fundamentally different algorithmic approach and problem formulation.

6. Reinforcement learning in near-continuous time for continuous state-action spaces

URL: View paper

Brief Assessment

Near-Continuous Time RL[73] focuses on near-continuous time systems with Poisson-clock interactions and diffusive approximations for planning, whereas the original paper addresses discrete-time online RL with hedge-type updates and posterior sampling. The technical approaches and problem formulations differ substantially.

7. Online Off-Policy Reinforcement Learning for Optimal Control of Unknown Nonlinear Systems Using Neural Networks

URL: View paper

Brief Assessment

Online Off-Policy Neural[74] focuses on optimal control using temporal difference learning and model-free HJB equations for unknown nonlinear systems, without providing policy-regret bounds or addressing the exploration-exploitation tradeoff that is central to the original paper's contribution.

8. Fixedâ□□Time Stable Gradient Flows for Optimal Adaptive Control of Continuousâ□□Time Nonlinear Systems

URL: View paper

Brief Assessment

Fixed-Time Adaptive Control[19] focuses on fixed-time convergence for optimal control using gradient flows and single network adaptive critic, not on policy-regret bounds in online non-episodic settings with exploration-exploitation tradeoffs.

9. Information theoretic regret bounds for online nonlinear control

URL: View paper

Brief Assessment

Information Theoretic Nonlinear[68] focuses on episodic settings with known RKHS structure and bounded second moments, whereas the original paper addresses non-episodic online learning with general nonlinear dynamics under persistence of excitation assumptions.

10. Efficient exploration in continuous-time model-based reinforcement learning

URL: View paper

Brief Assessment

Efficient Continuous-Time Exploration[75] focuses on continuous-time model-based RL with ODE dynamics and measurement selection strategies, whereas the original paper addresses online non-episodic RL with discrete-time updates and posterior sampling. The technical approaches and problem formulations differ substantially.

Contribution 2: Frequentist policy-regret guarantees via additional exploration under persistence of excitation

Description: Unlike prior posterior sampling reinforcement learning works that provide Bayesian guarantees, the authors' algorithms add deliberate exploration (Gaussian noise) and establish frequentist policy-regret bounds. The analysis relies on persistence of excitation assumptions from system identification rather than mixing assumptions, making it applicable near stability boundaries.

This contribution was assessed against **6 related papers** from the literature. Papers with potential prior art are analyzed in detail with textual evidence; others receive brief assessments.

1. Logarithmic Regret Bound in Partially Observable Linear Dynamical Systems

URL: View paper

Brief Assessment

Logarithmic Regret POMDP[51] focuses on partially observable linear dynamical systems with specific identifiability conditions, whereas the original paper addresses general nonlinear dynamics with continuous state-action spaces using posterior sampling and hedge-type updates.

2. Thompson Sampling Achieves Regret in Linear Quadratic Control

URL: View paper

Brief Assessment

Thompson Sampling LQR[52] focuses on linear quadratic control with Thompson sampling, not general nonlinear RL frameworks. The candidate uses persistence of excitation assumptions specific to LQR systems, while the original work addresses broader nonlinear dynamical systems with different identifiability conditions.

3. Agnostic system identification for model-based reinforcement learning

URL: View paper

Brief Assessment

Agnostic System Identification[53] focuses on agnostic system identification with no-regret online learning guarantees, not on frequentist regret bounds under persistence of excitation assumptions. The candidate addresses train-test distribution mismatch in model-based RL through iterative data collection, which is a different technical approach from the original's deliberate Gaussian noise exploration with persistence of excitation analysis.

4. Approximate Bayesian Reinforcement Learning for System Identification

URL: View paper

Brief Assessment

Approximate Bayesian RL[56] focuses on system identification through Bayesian model variance reduction for exploration, not on establishing frequentist regret bounds or persistence of excitation analysis for reinforcement learning.

5. The Reward Biased Method: An Optimism based Approach for Reinforcement Learning

URL: View paper

Brief Assessment

Reward Biased Optimism[55] focuses on reward-biased maximum likelihood estimation for exploration-exploitation trade-offs, not on frequentist regret guarantees with deliberate Gaussian noise exploration under persistence of excitation assumptions.

6. No-Regret Methods for Learning Sequential Predictions

URL: View paper

Brief Assessment

No-Regret Sequential Predictions[54] focuses on imitation learning and structured prediction using online learning reductions, not on reinforcement learning with persistence of excitation assumptions for system identification. The candidate does not address frequentist regret bounds in RL settings with deliberate Gaussian exploration.

Contribution 3: Separation principle decoupling model identification and certainty-equivalent control for nonlinear dynamics

Description: The authors establish a separation principle where the algorithm decouples best model identification (via posterior sampling over candidate models) from certainty-equivalent control (applying policies corresponding to sampled models). This simplifies policy evaluation and enables explicit characterization of policy regret via packing numbers rather than more complex measures like eluder dimension.

This contribution was assessed against **10 related papers** from the literature. Papers with potential prior art are analyzed in detail with textual evidence; others receive brief assessments.

1. Adaptive control of nonlinear non-minimum phase systems using actorâcritic reinforcement learning

URL: View paper

Brief Assessment

Actor-Critic Nonlinear Adaptive[58] focuses on non-minimum phase systems using cascade control with actor-critic RL for reference signal generation, not on separating model identification from certainty-equivalent control in the context of posterior sampling over candidate models.

2. Structured Control Nets for Deep Reinforcement Learning

URL: View paper

Brief Assessment

Structured Control Nets[66] focuses on neural network architecture for policy representation in deep RL, splitting control into nonlinear and linear modules for improved performance. This is architecturally distinct from the original paper's algorithmic separation of model identification via posterior sampling and certainty-equivalent control for sample complexity analysis.

3. Deep-reinforcement-learning-based separation control in a two-dimensional airfoil

URL: View paper

Brief Assessment

Deep RL Airfoil Separation[59] focuses on active flow control in airfoil aerodynamics using deep reinforcement learning for jet control strategies, not on theoretical separation principles for model identification and certainty-equivalent control in general nonlinear dynamical systems.

4. Fault-Tolerant Control for Topside Separation Systems via Output-Feedback Reinforcement Learning

URL: View paper

Brief Assessment

Fault-Tolerant Output-Feedback RL[57] focuses on fault-tolerant control for topside separation systems using output-feedback reinforcement learning, not on establishing separation principles for model identification and certainty-equivalent control in general nonlinear dynamics.

5. Discovering Flow Separation Control Strategies in 3D Wings via Deep Reinforcement Learning

URL: View paper

Brief Assessment

3D Wing Separation Control[65] focuses on applying deep reinforcement learning to active flow control for aerodynamic applications, not on theoretical separation principles for model identification and control in general nonlinear dynamical systems.

6. Scalable autonomous separation assurance with heterogeneous multi-agent reinforcement learning

URL: View paper

Brief Assessment

Heterogeneous Separation Assurance[60] focuses on multi-agent reinforcement learning for aircraft separation assurance in airspace sectors, not on separation principles for model identification and control in general nonlinear dynamical systems.

7. Autonomous separation assurance with deep multi-agent reinforcement learning

URL: View paper

Brief Assessment

Deep Multi-Agent Separation[62] focuses on multi-agent conflict resolution in air traffic control using deep reinforcement learning, not on establishing separation principles for model identification and control in general nonlinear dynamical systems.

8. Improved Robot Path Planning Method Based on Deep Reinforcement Learning

URL: View paper

Brief Assessment

Deep RL Path Planning[61] focuses on robot path planning using DDQN with dimensionality reduction and dual-branch networks for navigation/obstacle avoidance. It does not address model identification, certainty-equivalent control, or separation principles for nonlinear dynamical systems in reinforcement learning.

9. Observer-Feedback-Feedforward Controller Structures in Reinforcement Learning

URL: View paper

Brief Assessment

Observer-Feedback Controller Structures[64] addresses observer-feedback separation in neural network architectures for partially observable systems, not the separation of model identification from certainty-equivalent control in reinforcement learning as proposed in the original paper.

10. Multiairport Departure Scheduling via Multiagent Reinforcement Learning

URL: View paper

Brief Assessment

Multiairport Departure Scheduling[63] focuses on air traffic scheduling using multiagent reinforcement learning for combinatorial optimization, not on separation principles for model identification and control in nonlinear dynamical systems.

Appendix: Text Similarity Detection

No high-similarity text segments were detected across any compared papers.

References

- [0] The Sample Complexity of Online Reinforcement Learning: A Multi-model Perspective View paper
- [1] Sample Complexity of Distributionally Robust Off-Dynamics Reinforcement Learning with Online Interaction View paper
- [2] Rich-observation reinforcement learning with continuous latent dynamics View paper
- [3] Convergence and sample complexity of policy gradient methods for stabilizing linear systems View paper
- [4] Neural network dynamics for model-based deep reinforcement learning with model-free fine-tuning View paper
- [5] Sample Complexity for Nonlinear Dynamics View paper
- [6] KCRL: Krasovskii-Constrained Reinforcement Learning with Guaranteed Stability in Nonlinear Dynamical Systems View paper
- [7] Sample-Efficient Preference-based Reinforcement Learning with Dynamics Aware Rewards View paper
- [8] KCRL: Krasovskii-Constrained Reinforcement Learning with Guaranteed Stability in Nonlinear Discrete-Time Systems View paper
- [9] Data-efficient reinforcement learning for complex nonlinear systems View paper
- [10] Active Learning for Nonlinear System Identification with Guarantees View paper
- [11] On the sample complexity of stability constrained imitation learning View paper
- [12] Sample-Efficient Reinforcement Learning and Its Applications View paper
- [13] Sample-efficient Safe Learning for Online Nonlinear Control with Control Barrier Functions View paper
- [14] Learning a model is paramount for sample efficiency in reinforcement learning control of PDEs View paper
- [15] Numerical Evidence for Sample Efficiency of Model-Based Over Model-Free Reinforcement Learning Control of Partial Differential Equations View paper
- [16] Improving Sample Efficiency of Multiagent Reinforcement Learning With Nonexpert Policy for Flocking Control View paper
- [17] Hybrid Transfer Reinforcement Learning: Provable Sample Efficiency from Shifted-Dynamics Data View paper
- [18] Optimal Exploration for Model-Based RL in Nonlinear Systems View paper
- [19] Fixedâ□□Time Stable Gradient Flows for Optimal Adaptive Control of Continuousâ□□Time Nonlinear Systems View paper
- [20] Two-step reinforcement learning for model-free redesign of nonlinear optimal regulator View paper

- [21] SOMBRL: Scalable and Optimistic Model-Based RL View paper
- [22] Adaptive Nonlinear Model Predictive Horizon Using Deep Reinforcement Learning for Optimal Trajectory Planning View paper
- [23] On the statistical efficiency of reward-free exploration in non-linear rl View paper
- [24] Sample-efficient diffusion-based control of complex nonlinear systems View paper
- [25] Safe reinforcement learning for power system control: A review View paper
- [26] Non-asymptotic and Accurate Learning of Nonlinear Dynamical Systems View paper
- [27] Exact Recovery Guarantees for Parameterized Non-linear System Identification Problem under Adversarial Attacks View paper
- [28] Sample efficient reinforcement learning in continuous state spaces: A perspective beyond linearity View paper
- [29] Data-Based Feedback Relearning Control for Uncertain Nonlinear Systems With Actuator Faults View paper
- [30] Non-linear reinforcement learning in large action spaces: Structural conditions and sample-efficiency of posterior sampling View paper
- [31] Sample-Efficient Online Control Policy Learning with Real-Time Recursive Model Updates View paper
- [32] A sample efficient model-based deep reinforcement learning algorithm with experience replay for robot manipulation View paper
- [33] Learning and Control of Dynamical Systems View paper
- [34] The Sample Complexity of Learning Dynamical Systems View paper
- [35] LEADS: Learning Dynamical Systems that Generalize Across Environments View paper
- [36] A trust region approach for few-shot sim-to-real reinforcement learning View paper
- [37] Near-Optimal Sample Complexity in Reward-Free Kernel-Based Reinforcement Learning View paper
- [38] Kolmogorov-Arnold inspired convolutional networks for enhancing PPO-based online reinforcement learning View paper
- [39] Reinforcement Learning vs Optimal Control: Sparse Nonlinear Dynamical Systems Between Theory and Practice View paper
- [40] Sparse Learning of Kernel Transfer Operators View paper
- [41] Adversarially Robust Stability Certificates can be Sample-Efficient View paper
- [42] Controllable Flow Matching for Online Reinforcement Learning View paper
- [43] Multi-Kernel Enhanced Receding-Horizon Reinforcement Learning for Steering Control of Intelligent Vehicles View paper
- [44] On the computation of Control Lyapunov Functions for building safe and stabilizing controllers View paper
- [45] From Embedding to Control: Representations for Stochastic Multi-Object Systems View paper
- [46] A hierarchical reinforcement learning approach for optimal path tracking of wheeled mobile robots View paper
- [47] Deep reinforcement learning with symmetric data augmentation applied for aircraft lateral attitude tracking control View paper
- [48] Scalable Supervision for Safe and Efficient Online Robot Learning View paper
- [49] Model-Based Design for Legged Robots: Predictive Control and Reinforcement Learning View paper
- [50] Adaptive Self-Organizing Clustering Dual-Buffer Safe Reinforcement Learning for Nonlinear Optimal Control View paper
- [51] Logarithmic Regret Bound in Partially Observable Linear Dynamical Systems View paper
- [52] Thompson Sampling Achieves Regret in Linear Quadratic Control View paper
- [53] Agnostic system identification for model-based reinforcement learning View paper
- [54] No-Regret Methods for Learning Sequential Predictions View paper
- [55] The Reward Biased Method: An Optimism based Approach for Reinforcement Learning View paper
- [56] Approximate Bayesian Reinforcement Learning for System Identification View paper
- [57] Fault-Tolerant Control for Topside Separation Systems via Output-Feedback Reinforcement Learning View paper
- [58] Adaptive control of nonlinear non-minimum phase systems using actorâcritic reinforcement learning View paper
- [59] Deep-reinforcement-learning-based separation control in a two-dimensional airfoil View paper
- [60] Scalable autonomous separation assurance with heterogeneous multi-agent reinforcement learning View paper
- [61] Improved Robot Path Planning Method Based on Deep Reinforcement Learning View paper
- [62] Autonomous separation assurance with deep multi-agent reinforcement learning View paper
- [63] Multiairport Departure Scheduling via Multiagent Reinforcement Learning View paper
- [64] Observer-Feedback-Feedforward Controller Structures in Reinforcement Learning View paper
- [65] Discovering Flow Separation Control Strategies in 3D Wings via Deep Reinforcement Learning View paper
- [66] Structured Control Nets for Deep Reinforcement Learning View paper
- [67] Regret guarantees for online deep control View paper
- [68] Information theoretic regret bounds for online nonlinear control View paper
- [69] Sample-efficient and Scalable Exploration in Continuous-Time RL View paper
- [70] Sublinear regret for an actor-critic algorithm in continuous-time linear-quadratic reinforcement learning View paper
- [71] Local Linearity: the Key for No-regret Reinforcement Learning in Continuous MDPs View paper
- [72] No-regret reinforcement learning in smooth mdps View paper
- [73] Reinforcement learning in near-continuous time for continuous state-action spaces View paper
- [74] Online Off-Policy Reinforcement Learning for Optimal Control of Unknown Nonlinear Systems Using Neural Networks View paper
- [75] Efficient exploration in continuous-time model-based reinforcement learning View paper