Novelty Assessment Report

Paper: Minimax Optimal Adversarial Reinforcement Learning
PDF URL: https://openreview.net/pdf?id=QEcSLhfOoQ
Venue: ICLR 2026 Conference Submission
Year: 2026
Report Generated: 2026-01-01

Abstract

Consider episodic Markov decision processes (MDPs) with adversarially chosen transition kernels, where the transition kernel is adversarially chosen at each episode. Prior works have established regret upper bounds of $\widetilde{\mathcal{O}}(\sqrt{T} + C^P)$, where T is the number of episodes and C^P quantifies the degree of adversarial change in the transition dynamics. This regret bound may scale as large as $\mathcal{O}(T)$, leading to a linear regret. This raises a fundamental question: Can sublinear regret be achieved under fully adversarial transition kernels? We answer this question affirmatively. First, we show that the optimal policy for MDPs with adversarial transition kernels must be history-dependent. We then design an algorithm of Adversarial Dynamics Follow-the-Regularized-Leader (AD-FTRL), and prove that it achieves a sublinear regret of $\mathcal{O}(\sqrt{(|\mathcal{S}||\mathcal{A}|)^K T})$, where K is the horizon length, $|\mathcal{S}|$ is the number of states, and $|\mathcal{A}|$ is the number of actions. Such a regret cannot be achieved by simply solving this problem as a contextual bandit. We further construct a hard MDP instance and prove a matching lower bound on the regret, which thereby demonstrates the **minimax optimality** of our algorithm.

Core Task Landscape

This paper addresses: **reinforcement learning with adversarially chosen transition kernels**

A total of **50 papers** were analyzed and organized into a taxonomy with **22 categories**.

Taxonomy Overview

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

- **Theoretical Foundations and Regret Analysis**
- **Robustness Enhancement Methods**
- **Multi-Agent and Cooperative Settings**
- **Adversarial Attack Methods and Vulnerability Analysis**
- **Domain-Specific Applications**
- **Generalization and Transfer Learning**
- **Imitation Learning and Inverse Reinforcement Learning**
- **Specialized Environments and Testbeds**

Complete Taxonomy Tree

- reinforcement learning with adversarially chosen transition kernels Survey Taxonomy
- Theoretical Foundations and Regret Analysis
  - Adversarial MDPs with Unknown Transitions ★ (4 papers)
  - [0] Minimax Optimal Adversarial Reinforcement Learning (Anon et al., 2026) View paper
  - [6] Dynamic regret of adversarial MDPs with unknown transition and linear function approximation (Long Fei Li, 2024) View paper
  - [26] Learning adversarial markov decision processes with bandit feedback and unknown transition (Chi Jin, 2020) View paper
  - [41] No-regret online reinforcement learning with adversarial losses and transitions (Jin, 2023) View paper
  - Adversarial Restless Multi-Armed Bandits (1 papers)
  - [13] Provably Efficient Reinforcement Learning for Adversarial Restless Multi-Armed Bandits with Unknown Transitions and Bandit Feedback (Xiong Guojun, 2024) View paper
  - Maximum Entropy and Robust RL Theory (1 papers)
  - [50] Maximum entropy RL (provably) solves some robust RL problems (Eysenbach, 2021) View paper
- Robustness Enhancement Methods
  - Adversarial Training for Observation Robustness (5 papers)
  - [1] Robust deep reinforcement learning against adversarial perturbations on state observations (Zhang Huan, 2020) View paper
  - [3] Robust deep reinforcement learning through adversarial attacks and training: A survey (Schott, 2024) View paper
  - [5] Robust deep reinforcement learning through adversarial loss (Tuomas Oikarinen, 2021) View paper
  - [7] Robust deep reinforcement learning with adversarial attacks (Anay Pattanaik, 2017) View paper
  - [24] Robust lane change decision making for autonomous vehicles: An observation adversarial reinforcement learning approach (Xiangkun He, 2022) View paper
  - Adversarial Dynamics and Action-Space Training (5 papers)
  - [4] Robust reinforcement learning using adversarial populations (Vinitsky, 2020) View paper
  - [17] Active Robust Adversarial Reinforcement Learning Under Temporally Coupled Perturbations (Jiacheng Yang, 2025) View paper
  - [18] Robustifying reinforcement learning agents via action space adversarial training (Kai Tan, 2020) View paper
  - [21] Robust adversarial reinforcement learning (Pinto Lerrel, 2017) View paper
  - [22] Robust adversarial reinforcement learning with dissipation inequation constraint (Peng Zhai, 2022) View paper
  - Model-Based Robustness Methods (4 papers)

- [10] Robust Model-Based Reinforcement Learning with an Adversarial Auxiliary Model (Anwar Ali, 2024) View paper
- [14] Rambo-rl: Robust adversarial model-based offline reinforcement learning (Rigter, 2022) View paper
- [28] Towards Robust Model-Based Reinforcement Learning Against Adversarial Corruption (Ye, 2024) View paper
- [31] Model-Based Offline Reinforcement Learning with Adversarial Data Augmentation (Cao Hongye, 2025) View paper
- Distributionally Robust and Risk-Averse Methods (2 papers)
- [8] Distributionally robust policy learning via adversarial environment generation (Allen Z. Ren, 2022) View paper
- [30] Risk averse robust adversarial reinforcement learning (Xinlei Pan, 2019) View paper

- Multi-Agent and Cooperative Settings
  - Multi-Agent Adversarial Robustness (2 papers)
  - [9] Multi-Agent Reinforcement Learning in Adversarial Game Environments: Personalized Anti-Interference Strategies for Heterogeneous UAV Communication (Yeguang Qin, 2025) View paper
  - [11] Robust multi-agent reinforcement learning via adversarial regularization: Theoretical foundation and stable algorithms (Bukharin, 2023) View paper
  - Hierarchical and Distributed Multi-Agent Control (1 papers)
  - [46] Distributed hierarchical reinforcement learning in multi-agent adversarial environments (Navid Naderializadeh, 2022) View paper
- Adversarial Attack Methods and Vulnerability Analysis
  - Observation and State Perturbation Attacks (3 papers)
  - [27] Tactics of adversarial attack on deep reinforcement learning agents (Yen-Chen Lin, 2017) View paper
  - [37] Real-time Adversarial Image Perturbations for Autonomous Vehicles Using Reinforcement Learning (Hyung Jin Yoon, 2025) View paper
  - [49] Vulnerability Analysis for Safe Reinforcement Learning in Cyber-Physical Systems (Shixiong Jiang, 2024) View paper
  - Training-Time and Environment Poisoning Attacks (2 papers)
  - [15] Policy Disruption in Reinforcement Learning:Adversarial Attack with Large Language Models and Critical State Identification (Jiang, 2025) View paper
  - [20] Policy teaching via environment poisoning: Training-time adversarial attacks against reinforcement learning (Rakhsha, 2020) View paper
  - Adversarial Policy and Multi-Agent Attacks (1 papers)
  - [43] Rethinking Adversarial Policies: A Generalized Attack Formulation and Provable Defense in RL (Liu Xiangyu, 2023) View paper
  - Attack Detection and Defense Mechanisms (2 papers)
  - [16] Challenges and countermeasures for adversarial attacks on deep reinforcement learning (Ilahi, 2021) View paper
  - [44] Clustering-based attack detection for adversarial reinforcement learning (RubÃ©n Majadas, 2024) View paper
- Domain-Specific Applications
  - Autonomous Systems and Robotics (2 papers)
  - [29] Multiagent modeling of pedestrian-vehicle conflicts using Adversarial Inverse Reinforcement Learning (Payam Nasernejad, 2023) View paper
  - [33] Coordinated Control of Urban Expressway Integrating Adjacent Signalized Intersections Using Adversarial Network Based Reinforcement Learning Method (Gengyue Han, 2024) View paper
  - Network Security and Intrusion Detection (4 papers)
  - [19] Adversarial environment reinforcement learning algorithm for intrusion detection (G. Caminero, 2019) View paper
  - [40] Adversarial robustness of deep reinforcement learning-based intrusion detection (Mohamed Amine Merzouk, 2024) View paper
  - [45] Reinforcement-learning-based Adversarial Attacks Against Vulnerability Detection Models (Siran Chen, 2024) View paper
  - [47] Adversarial RL-based IDS for evolving data environment in 6LoWPAN (Aryan Mohammadi Pasikhani, 2022) View paper
  - Recommendation Systems and Market Modeling (2 papers)
  - [12] A model-based reinforcement learning with adversarial training for online recommendation (Xueying Bai, 2019) View paper
  - [36] Improving Generalization in Reinforcement Learningâ□□Based Trading by Using a Generative Adversarial Market Model (Chia-Hsuan Kuo, 2021) View paper
  - Blockchain and Privacy-Preserving Systems (3 papers)
  - [32] Adaptive consensus optimization in blockchain using reinforcement learning and validation in adversarial environments. (Rommel Gutierrez, 2025) View paper
  - [38] Privacy and Security for Trustworthy AI/ML in Multi-Agent Critical Infrastructures: An Analysis of Adversarial Dynamics and Protective Strategies (Hossain, 2024) View paper
  - [42] Adversarial Reinforcement Learning Against Statistic Inference on Agent Identity (Yue Tian, 2024) View paper
  - Fault Detection and Anomaly Detection (1 papers)
  - [25] Fault detection method based on adversarial reinforcement learning (Junhuai Li, 2023) View paper
- Generalization and Transfer Learning
  - Domain Adaptation and Visual Generalization (2 papers)
  - [34] Sample-Efficient Reinforcement Learning via Adversarial Self-Loop Dynamics Modeling (Yuchen Liang, 2025) View paper
  - [35] Domain adversarial reinforcement learning (FranÃ§ois-Lavet, 2021) View paper
  - Meta-Learning and Algorithm Discovery (1 papers)
  - [39] Discovering General Reinforcement Learning Algorithms with Adversarial Environment Design (Jackson, 2023) View paper
- Imitation Learning and Inverse Reinforcement Learning (1 papers)
  - [23] Provably efficient adversarial imitation learning with unknown transitions (Xu Tian, 2023) View paper
- Specialized Environments and Testbeds (2 papers)
  - [2] Reinforcement Learning for Adversarial Environments (Christian Carrizales, 2025) View paper
  - [48] Reinforcement Learning Environment with LLM-Controlled Adversary in D&D 5th Edition Combat (Joseph Emmanuel DL Dayo, 2025) View paper

Narrative

Core task: reinforcement learning with adversarially chosen transition kernels. This field examines how agents can learn effective policies when the environment's dynamics are selected by an adversary, rather than being fixed or stochastic in a benign sense. The taxonomy reveals a rich structure spanning eight main branches. Theoretical Foundations and Regret Analysis investigates minimax optimality and regret bounds in adversarial MDPs, often under unknown transitions, as seen in works like Minimax Optimal Adversarial RL[0] and Dynamic Regret Adversarial MDPs[6]. Robustness Enhancement Methods focuses on training techniques that harden policies

against perturbations, including adversarial regularization and distributionally robust approaches such as Distributionally Robust Policy[8]. Adversarial Attack Methods and Vulnerability Analysis explores how to craft effective attacks on trained agents, with studies like Robust DRL Adversarial Perturbations[1] and Adversarial Attacks Training Survey[3] characterizing threat models. Meanwhile, Domain-Specific Applications and Specialized Environments branches demonstrate how adversarial RL principles apply to cybersecurity, autonomous driving, and other real-world testbeds.

Several active lines of work highlight contrasting emphases and open questions. One thread pursues tight regret guarantees for online learning in adversarial MDPs, balancing computational efficiency with statistical optimality; another thread emphasizes practical robustness via adversarial training or auxiliary models that anticipate worst-case perturbations. Minimax Optimal Adversarial RL[0] sits squarely within the theoretical branch on adversarial MDPs with unknown transitions, aiming to establish minimax optimal rates. It shares conceptual ground with Dynamic Regret Adversarial MDPs[6], which also tackles regret minimization under adversarial dynamics, and with No-Regret Online RL[41], which explores no-regret guarantees in online settings. Compared to these neighbors, Minimax Optimal Adversarial RL[0] appears to emphasize achieving the tightest possible bounds in the unknown-transition regime, whereas Dynamic Regret Adversarial MDPs[6] may focus more on time-varying adversaries. This positioning underscores ongoing debates about the trade-offs between sample complexity, computational tractability, and the strength of adversarial assumptions.

Related Works in Same Category

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

1. Dynamic regret of adversarial MDPs with unknown transition and linear function approximation

Authors: Long Fei Li, Longfei Li, Peng Zhao, Long-Fei Li, Zhi-Hua Zhou, et al. (6 authors total) | **Year/Venue**: 2024 | **URL**: View paper

Abstract

We study reinforcement learning (RL) in episodic MDPs with adversarial full-information losses and the unknown transition. Instead of the classical static regret, we adopt dynamic regret as the performance measure which benchmarks the learner's performance with changing policies, making it more suitable for non-stationary environments. The primary challenge is to handle the uncertainties of unknown transition and unknown non-stationarity of environments simultaneously. We propose a general frame...

Relationship Analysis

Both papers belong to the category of Adversarial MDPs with Unknown Transitions, analyzing regret bounds and sample complexity for MDPs with adversarial losses and unknown transition dynamics. The original paper focuses on achieving minimax optimal sublinear regret under fully adversarial transition kernels using history-dependent policies and trajectory-level occupancy measures, while the candidate paper studies dynamic regret (comparing against changing policies rather than a fixed optimal policy) with linear function approximation in tabular, linear, and linear mixture MDPs. The key difference is that the original paper addresses static regret with fully adversarial transitions and proves minimax optimality, whereas the candidate paper tackles dynamic regret in non-stationary environments with structured function approximation.

2. Learning adversarial markov decision processes with bandit feedback and unknown transition

Authors: Chi Jin, Tiancheng Jin, Haipeng Luo, Suvrit Sra, Tiancheng Yu | **Year/Venue**: 2020 | **URL**: View paper

Abstract

â¦ The environment dynamics are usually modeled as a â¦ transition function, and apply Online Mirror Descent over the space of occupancy measures (see Section 2.1) to handle adversarial â¦

Relationship Analysis

Both papers belong to the same taxonomy category focusing on adversarial MDPs with unknown transitions and bandit feedback, analyzing regret bounds and sample complexity. They overlap in addressing the challenge of learning optimal policies under adversarially chosen transition kernels without full knowledge of the environment dynamics. The key difference is that the original paper achieves Õ(√(|S||A|)^K T) regret using history-dependent policies and trajectory-level occupancy measures with a novel regularization approach, while the candidate paper achieves Õ(L|X|√|A|T) regret using state-action occupancy measures with upper occupancy bounds and a tighter confidence set construction for the transition function.

3. No-regret online reinforcement learning with adversarial losses and transitions

Authors: Jin, Tiancheng, Tiancheng Jin, Liu Junyan, Junyan Liu, et al. (18 authors total) | **Year/Venue**: 2023 | **URL**: View paper

Abstract

Existing online learning algorithms for adversarial Markov Decision Processes achieve ${O}(\sqrt{T})$ regret after T rounds of interactions even if the loss functions are chosen arbitrarily by an adversary, with the caveat that the transition function has to be fixed. This is because it has been shown that adversarial transition functions make no-regret learning impossible. Despite such impossibility results, in this work, we develop algorithms that can handle both adversarial losses and adver...

Relationship Analysis

Both papers belong to the same taxonomy category analyzing regret bounds for adversarial MDPs with unknown transitions. They share the core challenge of achieving sublinear regret under adversarially chosen transition kernels and bandit feedback. The key difference is that the original paper focuses on history-dependent policies and achieves Õ(√(|S||A|)^K T) regret with minimax optimality proofs, while the candidate paper uses Markov policies with occupancy measures and achieves Õ(√T + C_P) regret where C_P quantifies transition corruption, additionally providing adaptive guarantees for stochastically constrained losses.

Contributions Analysis

Overall novelty summary. The paper establishes sublinear regret bounds for episodic MDPs with fully adversarial transition kernels, introducing the AD-FTRL algorithm and proving history-dependent optimal policies are necessary. It resides in the 'Adversarial MDPs with Unknown Transitions' leaf, which contains only four papers total, indicating a relatively sparse research direction within the broader taxonomy of 50 papers. This leaf focuses specifically on regret analysis under adversarial losses and unknown dynamics, distinguishing it from known-transition settings and bandit formulations that populate neighboring branches.

The taxonomy reveals this work sits within 'Theoretical Foundations and Regret Analysis,' one of eight major branches addressing adversarial RL. Neighboring leaves include 'Adversarial Restless Multi-Armed Bandits' (bandit feedback without full MDP structure) and 'Maximum Entropy and Robust RL Theory' (formal robustness proofs via MaxEnt). The sibling papers in the same leaf examine dynamic regret bounds and sample complexity under adversarial losses, but the taxonomy's scope notes explicitly exclude known-transition settings, positioning this work at the intersection of unknown dynamics and adversarial control where theoretical guarantees remain challenging.

Among 30 candidates examined, the first contribution (characterizing history-dependent optimal policies) shows one refutable candidate from 10 examined, suggesting some prior theoretical characterization exists in the limited search scope. The second contribution (AD-FTRL algorithm) and third contribution (minimax optimal bounds with matching lower bound) each examined 10 candidates with zero refutations, indicating these algorithmic and optimality results appear more novel within the searched literature. The analysis explicitly

notes this reflects top-K semantic search plus citation expansion, not exhaustive coverage, so additional related work may exist beyond the examined set.

Given the sparse four-paper leaf and limited overlap detected across 30 candidates, the work appears to advance a relatively underexplored theoretical direction. The history-dependent policy characterization shows modest prior overlap, while the algorithmic and optimality contributions exhibit stronger novelty signals within the examined scope. However, the small search scale and narrow leaf population mean this assessment captures local novelty rather than field-wide positioning, and broader literature may contain additional relevant precedents not surfaced by semantic search.

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

Contribution 1: Characterization of optimal policy under adversarial transitions

Description: The authors establish that when transition kernels are chosen adversarially at each episode, the optimal policy must depend on the full history of observations rather than only the current state. This contrasts with standard MDPs where Markov policies are known to be optimal.

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. Robust reinforcement learning on state observations with learned optimal adversary

URL: View paper

Brief Assessment

Learned Optimal Adversary[74] addresses adversarial perturbations on state observations in a fixed MDP, not adversarially changing transition kernels across episodes. The setting and problem formulation differ fundamentally from the original paper's focus on history-dependent policies under adversarial transition dynamics.

---

2. On the foundation of distributionally robust reinforcement learning

URL: View paper

Prior Art Analysis

Foundation Distributionally Robust[75] demonstrates that history-dependent policies are necessary for optimal control in robust MDPs with adversarial transitions, predating the original paper's claim. The candidate paper establishes that when transition kernels vary adversarially, the optimal policy must depend on full history rather than only current state, which directly addresses the same problem setting as the original contribution.

Evidence

Evidence 1 - **Rationale**: Both papers address the fundamental question of optimal policy structure under adversarial transition dynamics, establishing the necessity of history-dependence. - **Original**: we first prove that the optimal policy that minimizes cumulative loss with adversarially chosen transition kernel must be a history-dependent policy, instead of a markov one. - **Candidate**: the controller seeks to maximize the expected cumulative rewards, while the adversary, modeling the possible yet unknown environmental shifts, selects a stochastic environment to hinder performance. this adversarial perspective offers a powerful lens for designing reliable controllers under uncertai...

---

3. A PDE Approach to the Prediction of a Binary Sequence with Advice from Two Historyâ□□Dependent Experts

URL: View paper

Brief Assessment

PDE Binary Sequence Prediction[79] addresses a fundamentally different problem: prediction of binary sequences with history-dependent experts in an adversarial market setting, not reinforcement learning with adversarial transition kernels in MDPs. The paper focuses on stock prediction using expert advice rather than optimal policies in episodic MDPs.

---

4. Online Prediction with Historyâ□□Dependent Experts: The General Case

URL: View paper

Brief Assessment

History-Dependent Experts[80] studies online prediction with expert advice in a stock prediction framework, not reinforcement learning with adversarial MDPs. The candidate focuses on combining expert predictions for binary sequences, while the original addresses optimal policies in episodic MDPs with adversarially chosen transition kernels.

---

5. Relgan: Relational generative adversarial networks for text generation

URL: View paper

Brief Assessment

RelGAN[72] addresses text generation using GANs with relational memory and Gumbel-softmax relaxation, not adversarial reinforcement learning with history-dependent policies under adversarial transition kernels.

---

6. Evading model poisoning attacks in federated learning by a long-short-term-memory-based approach

URL: View paper

Brief Assessment

Model Poisoning LSTM[71] addresses federated learning security through LSTM-based detection of malicious model updates, not reinforcement learning with adversarial transition kernels or history-dependent policies.

---

7. A bayesian learning algorithm for unknown zero-sum stochastic games with an arbitrary opponent

URL: View paper

Brief Assessment

Bayesian Learning Stochastic Games[76] addresses zero-sum stochastic games with an arbitrary opponent using stationary randomized policies, not adversarial transition kernels that change per episode. The candidate's setting involves a fixed but unknown transition kernel, fundamentally different from the original's adversarially-chosen transitions requiring history-dependent policies.

---

8. Two-phase real-time task offloading framework for edge-IoT systems using spiking neuromorphic coordination and holographic memory reuse

URL: View paper

Brief Assessment

Spiking Neuromorphic Coordination[73] focuses on edge-IoT task offloading using neuromorphic coordination and holographic memory, not on adversarial reinforcement learning with history-dependent policies in MDPs.

---

9. Anomaly detection for wind turbines using long short-term memory-based variational autoencoder wasserstein generation adversarial network under semi â□¦

URL: View paper

Brief Assessment

Wind Turbine Anomaly Detection[78] focuses on anomaly detection for wind turbines using deep learning methods (LSTM-VAE-WGAN), not on reinforcement learning with adversarial transition kernels or optimal policy characterization in MDPs.

---

10. Risk-sensitive safety analysis using conditional value-at-risk

URL: View paper

Brief Assessment

Risk-Sensitive Safety Analysis[77] addresses risk-sensitive safety analysis using CVaR for stochastic systems with constraint violations, not adversarial transition kernels in reinforcement learning. The history-dependence mentioned relates to CVaR temporal decomposition for safety analysis, not optimal policy structure under adversarial MDPs.

Contribution 2: AD-FTRL algorithm with sublinear regret guarantee

Description: The authors design a Follow-the-Regularized-Leader algorithm that operates with bandit feedback and unknown adversarial transitions. The algorithm uses trajectory-level occupancy measures and importance sampling with a carefully designed regularization term to achieve sublinear regret without requiring knowledge of transition kernels.

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. Simultaneously learning stochastic and adversarial bandits under the position-based model

URL: View paper

Brief Assessment

Position-Based Model[64] addresses online learning to rank under position-based models with stochastic/adversarial bandits, not general RL with adversarial transition kernels. The candidate uses FTRL for ranking problems with fixed transition structures, while the original tackles episodic MDPs with fully adversarial, unknown transition dynamics.

---

2. A Simple and Adaptive Learning Rate for FTRL in Online Learning with Minimax Regret of and its Application to Best-of-Both-Worlds

URL: View paper

Brief Assessment

Adaptive Learning Rate FTRL[62] focuses on online learning with minimax regret of θ(t^(2/3)) for problems like partial monitoring and graph bandits, not adversarial RL with unknown transition kernels. The technical settings and problem domains are fundamentally different.

---

3. A blackbox approach to best of both worlds in bandits and beyond

URL: View paper

Brief Assessment

Blackbox Best Both Worlds[63] focuses on multi-armed bandits and general online learning settings with adversarial losses but fixed transitions, not on MDPs with adversarially changing transition kernels. The candidate's FTRL framework addresses a fundamentally different problem structure than the original paper's adversarial dynamics setting.

---

4. On the rate of convergence of regularized learning in games: From bandits and uncertainty to optimism and beyond

URL: View paper

Brief Assessment

Convergence Regularized Learning Games[66] focuses on convergence rates to Nash equilibria in games with full/bandit feedback, not on adversarial MDPs with unknown transition kernels. The candidate addresses game-theoretic learning convergence, while the original addresses adversarial RL with trajectory-level occupancy measures.

---

5. Towards best-of-all-worlds online learning with feedback graphs

URL: View paper

Brief Assessment

Feedback Graphs[65] addresses online learning with feedback graphs using FTRL with a novel Tsallis-Shannon entropy regularizer. The original paper focuses on adversarial RL with unknown transition kernels using trajectory-level occupancy measures, which is a fundamentally different problem setting and technical approach.

---

6. The best of both worlds: stochastic and adversarial episodic mdps with unknown transition

URL: View paper

Brief Assessment

Best Both Worlds Episodic[61] addresses episodic MDPs with stochastic/adversarial losses but assumes fixed (though unknown) transitions, not fully adversarial transitions. The candidate's FTRL framework targets best-of-both-worlds guarantees under a different problem setting than the original's fully adversarial dynamics.

---

7. Faster Convergence for Unknown-Game Bandits

URL: View paper

Brief Assessment

Unknown-Game Bandits[68] addresses multi-agent game settings with swap regret minimization, while the original paper focuses on single-agent adversarial MDPs with unknown transitions. The candidate's OFTRL algorithm targets game-theoretic equilibria, not sequential decision-making under adversarial dynamics.

---

8. Self-Concordant Perturbations for Linear Bandits

URL: View paper

Brief Assessment

Self-Concordant Perturbations[69] addresses adversarial linear bandits with known action sets, not adversarial MDPs with unknown transitions. The candidate focuses on FTRL/FTPL for linear optimization problems, not trajectory-level occupancy measures in reinforcement learning settings.

---

9. Best-of-Both Worlds for linear contextual bandits with paid observations

URL: View paper

Brief Assessment

Linear Contextual Paid Observations[67] addresses linear contextual bandits with paid observations, not adversarial MDPs with unknown transitions. The settings are fundamentally different: one involves contextual bandits with optional observation costs, while the original addresses episodic MDPs with adversarially changing transition kernels.

---

10. Adapting to Stochastic and Adversarial Losses in Episodic MDPs with Aggregate Bandit Feedback

URL: View paper

Brief Assessment

Aggregate Bandit Feedback[70] focuses on episodic MDPs with aggregate bandit feedback (observing only cumulative episode loss), not adversarial transitions. The original paper addresses adversarial transition kernels with trajectory-level occupancy measures, which is a fundamentally different problem setting.

Contribution 3: Minimax optimal regret bound with matching lower bound

Description: The authors construct a hard MDP instance and prove a matching lower bound that demonstrates their algorithm achieves the minimax optimal regret. Their proof introduces a new analytical approach using composite hypothesis testing for handling adversarial transitions, providing a complete characterization of the fundamental difficulty of this problem.

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. Online convex optimization in adversarial markov decision processes

URL: View paper

Brief Assessment

Online Convex Optimization[55] addresses a different problem setting (adversarial losses with fixed but unknown transitions) and does not establish minimax optimality for the fully adversarial transition setting studied in the original paper.

---

2. Narrowing the gap between adversarial and stochastic MDPs via policy optimization

URL: View paper

Brief Assessment

Narrowing the Gap[59] addresses adversarial MDPs with oblivious adversaries and full information feedback, achieving Õ(√(h^7sat)) regret. The original paper considers fully adversarial transitions with bandit feedback, achieving Õ(√((|s||a|)^k t)) regret where k is horizon length. These are fundamentally different problem settings with incomparable regret bounds due to different assumptions about adversary knowledge and feedback structure.

---

3. Achieving Near Instance-Optimality and Minimax-Optimality in Stochastic and Adversarial Linear Bandits Simultaneously

URL: View paper

Brief Assessment

Instance-Optimality Minimax-Optimality[60] focuses on linear bandits with stochastic/adversarial rewards, not MDPs with adversarial transitions. The settings are fundamentally different: linear function approximation in bandits versus tabular MDPs with time-varying dynamics.

---

4. Learning adversarial mdps with stochastic hard constraints

URL: View paper

Brief Assessment

Stochastic Hard Constraints[58] addresses constrained MDPs with adversarial losses and stochastic constraints, focusing on constraint violation bounds rather than minimax optimal regret for adversarial transitions. The settings and objectives differ fundamentally from the original paper's focus on adversarial transition kernels.

---

5. Refining minimax regret for unsupervised environment design

URL: View paper

Brief Assessment

Refining Minimax Regret[51] addresses minimax regret in unsupervised environment design (UED) for curriculum learning, not adversarial RL with time-varying transitions. The candidate focuses on refining minimax regret policies to avoid regret stagnation in partially observable settings, whereas the original paper constructs hard MDP instances and proves minimax optimal regret bounds for adversarial transition kernels using composite hypothesis testing.

---

6. Differentially private no-regret exploration in adversarial markov decision processes

URL: View paper

Brief Assessment

Differentially Private Exploration[53] focuses on private learning in adversarial MDPs with differential privacy constraints, not on establishing minimax optimal regret bounds for general adversarial MDPs. The technical approaches differ fundamentally—the candidate addresses privacy-preserving mechanisms while the original develops composite hypothesis testing for adversarial transitions.

---

7. Near-optimal regret for adversarial mdp with delayed bandit feedback

URL: View paper

Brief Assessment

Delayed Bandit Feedback[57] addresses adversarial MDPs with delayed bandit feedback, not adversarial transition kernels. The technical setting and problem formulation differ fundamentally from the original paper's focus on adversarially changing transitions.

---

8. Learning adversarial linear mixture markov decision processes with bandit feedback and unknown transition

URL: View paper

Brief Assessment

Linear Mixture Bandit[52] studies adversarial linear mixture MDPs but focuses on bandit feedback with unknown transitions in a different problem formulation. The candidate's lower bound Ω(dh√k + √hsak) addresses linear mixture MDPs, while the original paper's Ω(√(|S||A|)kt) addresses fully adversarial tabular MDPs with time-varying transitions—these are distinct problem settings with different structural assumptions and complexity measures.

---

9. Dynamic regret of adversarial linear mixture MDPs

URL: View paper

Brief Assessment

Dynamic Regret Linear Mixture[54] focuses on adversarial linear mixture MDPs with dynamic regret analysis, while the original paper addresses adversarial transition kernels in tabular MDPs with history-dependent policies. The candidate's lower bound techniques and problem settings differ fundamentally from the original's composite hypothesis testing approach for adversarial transitions.

---

10. Optimistic regret bounds for online learning in adversarial Markov decision processes

URL: View paper

Brief Assessment

Optimistic Regret Bounds[56] addresses a different problem setting (optimistic online learning with cost predictors in AMDPs) and does not establish minimax lower bounds for adversarial transitions. Their focus is on optimistic regret that scales with predictor quality, not on proving fundamental limits via composite hypothesis testing for adversarial MDPs.

Appendix: Text Similarity Detection

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

References

- [0] Minimax Optimal Adversarial Reinforcement Learning View paper
- [1] Robust deep reinforcement learning against adversarial perturbations on state observations View paper
- [2] Reinforcement Learning for Adversarial Environments View paper
- [3] Robust deep reinforcement learning through adversarial attacks and training: A survey View paper
- [4] Robust reinforcement learning using adversarial populations View paper
- [5] Robust deep reinforcement learning through adversarial loss View paper
- [6] Dynamic regret of adversarial MDPs with unknown transition and linear function approximation View paper
- [7] Robust deep reinforcement learning with adversarial attacks View paper
- [8] Distributionally robust policy learning via adversarial environment generation View paper
- [9] Multi-Agent Reinforcement Learning in Adversarial Game Environments: Personalized Anti-Interference Strategies for Heterogeneous UAV Communication View paper
- [10] Robust Model-Based Reinforcement Learning with an Adversarial Auxiliary Model View paper
- [11] Robust multi-agent reinforcement learning via adversarial regularization: Theoretical foundation and stable algorithms View paper
- [12] A model-based reinforcement learning with adversarial training for online recommendation View paper
- [13] Provably Efficient Reinforcement Learning for Adversarial Restless Multi-Armed Bandits with Unknown Transitions and Bandit Feedback View paper
- [14] Rambo-rl: Robust adversarial model-based offline reinforcement learning View paper
- [15] Policy Disruption in Reinforcement Learning:Adversarial Attack with Large Language Models and Critical State Identification View paper
- [16] Challenges and countermeasures for adversarial attacks on deep reinforcement learning View paper
- [17] Active Robust Adversarial Reinforcement Learning Under Temporally Coupled Perturbations View paper
- [18] Robustifying reinforcement learning agents via action space adversarial training View paper
- [19] Adversarial environment reinforcement learning algorithm for intrusion detection View paper
- [20] Policy teaching via environment poisoning: Training-time adversarial attacks against reinforcement learning View paper
- [21] Robust adversarial reinforcement learning View paper
- [22] Robust adversarial reinforcement learning with dissipation inequation constraint View paper
- [23] Provably efficient adversarial imitation learning with unknown transitions View paper
- [24] Robust lane change decision making for autonomous vehicles: An observation adversarial reinforcement learning approach View paper
- [25] Fault detection method based on adversarial reinforcement learning View paper
- [26] Learning adversarial markov decision processes with bandit feedback and unknown transition View paper
- [27] Tactics of adversarial attack on deep reinforcement learning agents View paper
- [28] Towards Robust Model-Based Reinforcement Learning Against Adversarial Corruption View paper
- [29] Multiagent modeling of pedestrian-vehicle conflicts using Adversarial Inverse Reinforcement Learning View paper
- [30] Risk averse robust adversarial reinforcement learning View paper
- [31] Model-Based Offline Reinforcement Learning with Adversarial Data Augmentation View paper
- [32] Adaptive consensus optimization in blockchain using reinforcement learning and validation in adversarial environments. View paper
- [33] Coordinated Control of Urban Expressway Integrating Adjacent Signalized Intersections Using Adversarial Network Based Reinforcement Learning Method View paper
- [34] Sample-Efficient Reinforcement Learning via Adversarial Self-Loop Dynamics Modeling View paper
- [35] Domain adversarial reinforcement learning View paper
- [36] Improving Generalization in Reinforcement Learningâ□□Based Trading by Using a Generative Adversarial Market Model View paper
- [37] Real-time Adversarial Image Perturbations for Autonomous Vehicles Using Reinforcement Learning View paper
- [38] Privacy and Security for Trustworthy AI/ML in Multi-Agent Critical Infrastructures: An Analysis of Adversarial Dynamics and Protective Strategies View paper
- [39] Discovering General Reinforcement Learning Algorithms with Adversarial Environment Design View paper
- [40] Adversarial robustness of deep reinforcement learning-based intrusion detection View paper

- [41] No-regret online reinforcement learning with adversarial losses and transitions View paper
- [42] Adversarial Reinforcement Learning Against Statistic Inference on Agent Identity View paper
- [43] Rethinking Adversarial Policies: A Generalized Attack Formulation and Provable Defense in RL View paper
- [44] Clustering-based attack detection for adversarial reinforcement learning View paper
- [45] Reinforcement-learning-based Adversarial Attacks Against Vulnerability Detection Models View paper
- [46] Distributed hierarchical reinforcement learning in multi-agent adversarial environments View paper
- [47] Adversarial RL-based IDS for evolving data environment in 6LoWPAN View paper
- [48] Reinforcement Learning Environment with LLM-Controlled Adversary in D&D 5th Edition Combat View paper
- [49] Vulnerability Analysis for Safe Reinforcement Learning in Cyber-Physical Systems View paper
- [50] Maximum entropy RL (provably) solves some robust RL problems View paper
- [51] Refining minimax regret for unsupervised environment design View paper
- [52] Learning adversarial linear mixture markov decision processes with bandit feedback and unknown transition View paper
- [53] Differentially private no-regret exploration in adversarial markov decision processes View paper
- [54] Dynamic regret of adversarial linear mixture MDPs View paper
- [55] Online convex optimization in adversarial markov decision processes View paper
- [56] Optimistic regret bounds for online learning in adversarial Markov decision processes View paper
- [57] Near-optimal regret for adversarial mdp with delayed bandit feedback View paper
- [58] Learning adversarial mdps with stochastic hard constraints View paper
- [59] Narrowing the gap between adversarial and stochastic MDPs via policy optimization View paper
- [60] Achieving Near Instance-Optimality and Minimax-Optimality in Stochastic and Adversarial Linear Bandits Simultaneously View paper
- [61] The best of both worlds: stochastic and adversarial episodic mdps with unknown transition View paper
- [62] A Simple and Adaptive Learning Rate for FTRL in Online Learning with Minimax Regret of and its Application to Best-of-Both-Worlds View paper
- [63] A blackbox approach to best of both worlds in bandits and beyond View paper
- [64] Simultaneously learning stochastic and adversarial bandits under the position-based model View paper
- [65] Towards best-of-all-worlds online learning with feedback graphs View paper
- [66] On the rate of convergence of regularized learning in games: From bandits and uncertainty to optimism and beyond View paper
- [67] Best-of-Both Worlds for linear contextual bandits with paid observations View paper
- [68] Faster Convergence for Unknown-Game Bandits View paper
- [69] Self-Concordant Perturbations for Linear Bandits View paper
- [70] Adapting to Stochastic and Adversarial Losses in Episodic MDPs with Aggregate Bandit Feedback View paper
- [71] Evading model poisoning attacks in federated learning by a long-short-term-memory-based approach View paper
- [72] Relgan: Relational generative adversarial networks for text generation View paper
- [73] Two-phase real-time task offloading framework for edge-IoT systems using spiking neuromorphic coordination and holographic memory reuse View paper
- [74] Robust reinforcement learning on state observations with learned optimal adversary View paper
- [75] On the foundation of distributionally robust reinforcement learning View paper
- [76] A bayesian learning algorithm for unknown zero-sum stochastic games with an arbitrary opponent View paper
- [77] Risk-sensitive safety analysis using conditional value-at-risk View paper
- [78] Anomaly detection for wind turbines using long short-term memory-based variational autoencoder wasserstein generation adversarial network under semi â¦ View paper
- [79] A PDE Approach to the Prediction of a Binary Sequence with Advice from Two Historyâ□□Dependent Experts View paper
- [80] Online Prediction with Historyâ□□Dependent Experts: The General Case View paper