

Novelty Assessment Report

Paper: Target Drift in Multi-Constraint Lagrangian RL: Theory and Practice

PDF URL: <https://openreview.net/pdf?id=1MHZGpvSDJ>

Venue: ICLR 2026 Conference Submission

Year: 2026

Report Generated: 2026-01-05

Abstract

Lagrangian-based methods are one of the dominant approaches for safe reinforcement learning (RL) in constrained Markov decision processes, commonly used across domains with multiple constraints. While some implementations combine all constraints into a mixed penalty term and others use one estimator per constraint, the fundamental question of which design is theoretically sound has received little scrutiny. We provide the first theoretical analysis showing that the mixed-critic architecture induces a persistent bias due to target drift from evolving Lagrange multipliers. In contrast, dedicated-critic design—separate critics for reward and each constraint—avoids this issue. We also validate our findings in a simulated but realistic energy system with multiple physical constraints, where the dedicated-critic method achieves stable learning and consistent constraint satisfaction, while the mixed-critic method fails. Our results offer a principled argument for preferring dedicated-critic architectures in multi-constraint safe RL problems.

Disclaimer

This report is **AI-GENERATED** using Large Language Models and WisPaper (a scholar search engine). It analyzes academic papers' tasks and contributions against retrieved prior work. While this system identifies **POTENTIAL** overlaps and novel directions, **ITS COVERAGE IS NOT EXHAUSTIVE AND JUDGMENTS ARE APPROXIMATE**. These results are intended to assist human reviewers and **SHOULD NOT** be relied upon as a definitive verdict on novelty.

Note that some papers exist in multiple, slightly different versions (e.g., with different titles or URLs). The system may retrieve several versions of the same underlying work. The current automated pipeline does not reliably align or distinguish these cases, so human reviewers will need to disambiguate them manually.

If you have any questions, please contact: mingzhang23@m.fudan.edu.cn

Core Task Landscape

This paper addresses: **Lagrangian reinforcement learning with multiple constraints**

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:

- **Lagrangian Method Design and Optimization**
- **Safety Formulations and Constraint Types**
- **Application Domains**
- **Multi-Agent and Cooperative Settings**
- **Learning Paradigms and Data Efficiency**
- **Constraint Learning and Inverse Problems**
- **Constraint Handling in Non-RL Optimization**
- **Surveys and Methodological Reviews**
- **Budgeted and Risk-Adaptive RL**

Complete Taxonomy Tree

- Lagrangian reinforcement learning with multiple constraints Survey Taxonomy
- Lagrangian Method Design and Optimization
 - Multiplier Update and Control-Theoretic Enhancements (4 papers)
 - [1] Responsive Safety in Reinforcement Learning by PID Lagrangian Methods (Adam Stooke, 2022) [View paper](#)
 - [24] Predictive Lagrangian Optimization for Constrained Reinforcement Learning (Zhang Tian-qi, 2025) [View paper](#)
 - [29] State-Dependent Lagrange Multipliers for State-Wise Safety in Constrained Reinforcement Learning (Minseok Seo, 2025) [View paper](#)
 - [50] Model-Based Chance-Constrained Reinforcement Learning via Separated Proportional-Integral Lagrangian (Peng Bai-yu, 2021) [View paper](#)
 - Primal-Dual and Distributional Approaches (5 papers)
 - [3] Off-Policy Conservative Distributional Reinforcement Learning With Safety Constraints (Heng-Rui Zhang, 2025) [View paper](#)
 - [7] Conservative Distributional Reinforcement Learning with Safety Constraints (Zhang Hengrui, 2022) [View paper](#)
 - [34] A Stochastic Composite Augmented Lagrangian Method For Reinforcement Learning (Li Yongfeng, 2023) [View paper](#)
 - [39] Last-iterate convergent policy gradient primal-dual methods for constrained mdps (Ding Dong-sheng, 2023) [View paper](#)
 - [43] Constrained differential dynamic programming: A primal-dual augmented lagrangian approach (Jallet, 2022) [View paper](#)
 - Multi-Constraint Architecture and Target Drift ★ (4 papers)
 - [0] Target Drift in Multi-Constraint Lagrangian RL: Theory and Practice (Anon et al., 2026) [View paper](#)
 - [13] State augmented constrained reinforcement learning: Overcoming the limitations of learning with rewards (Calvo-Fullana, 2023) [View paper](#)
 - [18] Gradient shaping for multi-constraint safe reinforcement learning (Yao, 2024) [View paper](#)
 - [32] Multi-Constraint Safe RL with Objective Suppression for Safety-Critical Applications (Z Zhou, 2024) [View paper](#)
 - Gradient Manipulation and Multi-Objective Optimization (2 papers)
 - [4] Constraints as Rewards: Reinforcement Learning for Robots without Reward Functions (Ishihara, 2025) [View paper](#)
 - [12] Multi-Constraint Optimization for Real-Time Bidding: A Reinforcement Learning Approach (H Weng, 2024) [View paper](#)
- Safety Formulations and Constraint Types
 - Risk-Sensitive and Distributional Safety (3 papers)
 - [5] Risk-Constrained Reinforcement Learning With Augmented Lagrangian Multiplier for Portfolio Optimization (Bayaraa Enkhsaikhan, 2025) [View paper](#)

- [28] CVaR-Constrained Policy Optimization for Safe Reinforcement Learning (Qiyuan Zhang, 2024) [View paper](#)
- [42] Augmented Lagrangian Risk-constrained Reinforcement Learning for Portfolio Optimization (Student Abstract) (Jo, 2025) [View paper](#)
- Temporal Logic and Formal Specifications (1 papers)
- [23] Deep reinforcement learning under signal temporal logic constraints using Lagrangian relaxation (Junya Ikemoto, 2022) [View paper](#)
- Barrier and Lyapunov Function Methods (1 papers)
- [6] Stable and Safe Reinforcement Learning via a Barrier-Lyapunov Actor-Critic Approach (Liqun Zhao, 2023) [View paper](#)
- Robust and Adversarial Constraints (1 papers)
- [30] Robust lagrangian and adversarial policy gradient for robust constrained markov decision processes (David M. Bossens, 2024) [View paper](#)
- Application Domains
 - Energy Systems and Power Grids (5 papers)
 - [8] A deep reinforcement learning-based charging scheduling approach with augmented Lagrangian for electric vehicles (Guibin Chen, 2025) [View paper](#)
 - [15] A Reinforcement Learning Embedded Surrogate Lagrangian Relaxation Method for Fast Solving Unit Commitment Problems (Yu-Hang Zhu, 2025) [View paper](#)
 - [21] An Augmented Lagrangian-Based Safe Reinforcement Learning Algorithm for Carbon-Oriented Optimal Scheduling of EV Aggregators (Xiaoying Shi, 2024) [View paper](#)
 - [33] Real-Time Optimal Power Flow: A Lagrangian Based Deep Reinforcement Learning Approach (Ziming Yan, 2020) [View paper](#)
 - [44] Distributed Reinforcement Learning for Real-Time Batteries Control Using Lagrangian Decomposition (Eleni Stai, 2022) [View paper](#)
 - Autonomous Driving and Transportation (3 papers)
 - [14] Constrained Reinforcement Learning-Enabled Policies With Augmented Lagrangian for Cooperative Intersection Management (Zhenhai Gao, 2024) [View paper](#)
 - [16] Safe Reinforcement Learning in Autonomous Driving With Epistemic Uncertainty Estimation (Zheng Zhang, 2024) [View paper](#)
 - [38] Constraints Driven Safe Reinforcement Learning for Autonomous Driving Decision-Making (Fei Gao, 2024) [View paper](#)
 - Network Management and Resource Allocation (4 papers)
 - [9] Augmented lagrangian-based reinforcement learning for network slicing in IIoT (QI Qi, 2022) [View paper](#)
 - [27] RL-mpca: A reinforcement learning based multi-phase computation allocation approach for recommender systems (Zhou Jiahong, 2023) [View paper](#)
 - [35] Causal CSI-Based Trajectory Design and Power Allocation for UAV-Enabled Wireless Networks Under Average Rate Constraints: A Constrained Reinforcement Learning Approach (Hyemin Yu, 2025) [View paper](#)
 - [46] Constraint-Aware Deep Reinforcement Learning for End-to-End Resource Orchestration in Mobile Networks (Liu Qiang, 2021) [View paper](#)
 - Combinatorial Optimization and Scheduling (1 papers)
 - [45] Learning to solve soft-constrained vehicle routing problems with lagrangian relaxation (Tang, 2022) [View paper](#)
 - Communication and Remote Estimation (1 papers)
 - [19] Semantic-Aware Remote Estimation of Multiple Markov Sources Under Constraints (Jiping Luo, 2024) [View paper](#)
- Multi-Agent and Cooperative Settings (2 papers)
 - [36] Scalable Constrained Policy Optimization for Safe Multi-agent Reinforcement Learning (Lin Li, 2024) [View paper](#)
 - [48] Provably efficient generalized lagrangian policy optimization for safe multi-agent reinforcement learning (Ding Dong-sheng, 2023) [View paper](#)
- Learning Paradigms and Data Efficiency
 - Off-Policy and Offline Constrained RL (1 papers)
 - [41] Poce: Primal policy optimization with conservative estimation for multi-constraint offline reinforcement learning (Guan Jiayi, 2024) [View paper](#)
 - Model-Based and Predictive Approaches (1 papers)
 - [11] Model-based safe deep reinforcement learning via a constrained proximal policy optimization algorithm (Jayant, 2022) [View paper](#)
 - Meta-Learning and Hyperparameter Adaptation (1 papers)
 - [17] Meta SAC-Lag: Towards Deployable Safe Reinforcement Learning via MetaGradient-based Hyperparameter Tuning (Homayoun Honari, 2024) [View paper](#)
 - Intervention-Based and Human-in-the-Loop Learning (2 papers)
 - [2] Safe RLHF: Safe Reinforcement Learning from Human Feedback (Josef Dai, 2023) [View paper](#)
 - [40] Real-world Reinforcement Learning from Suboptimal Interventions (Yinuo Zhao, 2025) [View paper](#)
- Constraint Learning and Inverse Problems (2 papers)
 - [47] Uncertainty-aware constraint inference in inverse constrained reinforcement learning (S Xu, 2023) [View paper](#)
 - [49] Concurrent Learning of Policy and Unknown Safety Constraints in Reinforcement Learning (Yifru, 2024) [View paper](#)
- Constraint Handling in Non-RL Optimization (4 papers)
 - [20] Conditionally adaptive augmented Lagrangian method for physics-informed learning of forward and inverse problems using artificial neural networks (Hu Qi-Feng, 2025) [View paper](#)
 - [25] An adaptive augmented Lagrangian method for training physics and equality constrained artificial neural networks (Basir, 2023) [View paper](#)
 - [31] Lagrangian duality for constrained deep learning (Ferdinando Fioretto, 2020) [View paper](#)
 - [37] Composition and Alignment of Diffusion Models using Constrained Learning (Hounie, 2025) [View paper](#)
- Surveys and Methodological Reviews (2 papers)
 - [10] A Review of Safe Reinforcement Learning Methods for Modern Power Systems (Tong Su, 2024) [View paper](#)
 - [22] Safe Reinforcement Learning with Constraints: A Survey (Zhengyu Chen, 2025) [View paper](#)
- Budgeted and Risk-Adaptive RL (1 papers)
 - [26] Scaling up budgeted reinforcement learning (Carrara Nicolas, 2022) [View paper](#)

Narrative

Core task: Lagrangian reinforcement learning with multiple constraints. The field addresses how agents can optimize rewards while satisfying several simultaneous safety or resource limits, typically by adjusting Lagrange multipliers during training. The taxonomy

reveals a rich structure spanning methodological innovations in Lagrangian optimization (including multi-constraint architectures and convergence guarantees), diverse safety formulations (from hard constraints to distributional risk measures), and a wide array of application domains such as power systems, autonomous driving, and network resource management. Several branches focus on learning paradigms that improve data efficiency or handle offline settings, while others explore multi-agent coordination, constraint learning from demonstrations, and budgeted or risk-adaptive frameworks. Representative works like PID Lagrangian Safety[1] and Safe RLHF[2] illustrate how controller-inspired updates and human-feedback integration can stabilize constraint satisfaction, whereas studies such as Conservative Distributional Safety[3] and Constraints as Rewards[4] show alternative formulations that blend distributional robustness with reward shaping.

A particularly active line of research examines architectural and algorithmic refinements for handling multiple constraints simultaneously, balancing the need for stable multiplier updates against the risk of objective drift or oscillation. Target Drift Lagrangian[0] sits squarely in this branch, proposing mechanisms to prevent the primary reward objective from being overshadowed when many constraints compete for attention. Nearby works like Gradient Shaping Multi Constraint[18] and Objective Suppression Safety[32] tackle related trade-offs by reshaping policy gradients or dynamically adjusting constraint priorities, highlighting ongoing debates about how to maintain reward progress without violating safety bounds. Meanwhile, methods such as State Augmented Constrained[13] and Conditionally Adaptive Lagrangian[20] explore state-dependent or adaptive multiplier schedules, offering complementary perspectives on when and how aggressively to enforce each constraint. Collectively, these efforts underscore a central tension: achieving fast, stable convergence in multi-constraint settings while preserving the agent's ability to optimize its core objective.

Related Works in Same Category

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

1. State augmented constrained reinforcement learning: Overcoming the limitations of learning with rewards

Authors: Calvo-Fullana, Miguel, Paternain, Santiago, Chamon, et al. (8 authors total) | **Year/Venue:** 2023 | **URL:** [View paper](#)

Abstract

A common formulation of constrained reinforcement learning involves multiple rewards that must individually accumulate to given thresholds. In this class of problems, we show a simple example in which the desired optimal policy cannot be induced by any weighted linear combination of rewards. Hence, there exist constrained reinforcement learning problems for which neither regularized nor classical primal-dual methods yield optimal policies. This work addresses this shortcoming by augmenting the s...

Relationship Analysis

Both papers belong to the Multi-Constraint Architecture and Target Drift category, addressing architectural design choices for handling multiple constraints in Lagrangian reinforcement learning. The original paper focuses on the target drift phenomenon caused by evolving Lagrange multipliers in mixed-critic architectures and proposes dedicated critics as a solution, while the candidate paper addresses the fundamental limitation that no fixed weighted combination of rewards can induce optimal policies for certain constrained problems, proposing state augmentation with Lagrange multipliers as part of the MDP state space. The key difference is that the original paper analyzes critic architecture bias within standard Lagrangian methods, whereas the candidate paper fundamentally redefines the problem formulation by augmenting the state space to overcome inherent limitations of regularized approaches.

2. Gradient shaping for multi-constraint safe reinforcement learning

Authors: Yao, Yihang, Yihang Yao, Liu Zu-xin, Zuxin Liu, et al. (20 authors total) | **Year/Venue:** 2024 | **URL:** [View paper](#)

Abstract

Online safe reinforcement learning (RL) involves training a policy that maximizes task efficiency while satisfying constraints via interacting with the environments. In this paper, our focus lies in addressing the complex challenges associated with solving multi-constraint (MC) safe RL problems. We approach the safe RL problem from the perspective of Multi-Objective Optimization (MOO) and propose a unified framework designed for MC safe RL algorithms. This framework highlights the manipulation o...

Relationship Analysis

Both papers belong to the Multi-Constraint Architecture and Target Drift category, addressing architectural design choices for handling multiple constraints in Lagrangian reinforcement learning. While the original paper provides theoretical analysis of target drift phenomena caused by evolving Lagrange multipliers in mixed-critic versus dedicated-critic architectures, this candidate paper approaches multi-constraint safe RL from a Multi-Objective Optimization perspective and proposes a gradient shaping method to handle redundant and conflicting constraints. The key difference is that the original paper focuses on critic architecture design and target drift theory, whereas this paper focuses on gradient manipulation strategies for constraint handling without explicitly analyzing the target drift issue.

3. Multi-Constraint Safe RL with Objective Suppression for Safety-Critical Applications

Authors: Z Zhou, J Booher, W Liu, A Petiushko, A Garg | **Year/Venue:** 2024 | **URL:** [View paper](#)

Abstract

RL into an unconstrained one. [6] demonstrates Lagrangian relaxation can be adopted to safe RL. We observe that previous safe RL algorithms struggle in multi-constraint scenarios.

Relationship Analysis

Both papers belong to the Multi-Constraint Architecture and Target Drift category, addressing architectural design choices for handling multiple constraints in Lagrangian RL. The original paper provides a theoretical analysis of target drift in mixed-critic versus dedicated-critic architectures, proving that mixed critics suffer from dual-induced bias while dedicated critics eliminate this drift. The candidate paper focuses on a practical method called Objective Suppression that adaptively suppresses task reward objectives based on safety critics, combining this with hierarchical methods like Recovery RL rather than analyzing the fundamental critic architecture design.

Contributions Analysis

Overall novelty summary. The paper contributes a theoretical analysis distinguishing mixed-critic from dedicated-critic architectures in multi-constraint Lagrangian RL, arguing that mixed critics induce persistent bias through target drift from evolving multipliers. It resides in the 'Multi-Constraint Architecture and Target Drift' leaf, which contains only four papers total, making this a relatively sparse research direction within the broader taxonomy of 50 papers. This leaf focuses specifically on architectural choices for handling multiple constraints simultaneously, excluding single-constraint methods and gradient manipulation techniques that belong elsewhere in the taxonomy.

The taxonomy reveals that neighboring leaves address related but distinct concerns: 'Gradient Manipulation and Multi-Objective Optimization' (2 papers) explores constraint aggregation and gradient shaping, while 'Multiplier Update and Control-Theoretic Enhancements' (4 papers) focuses on adaptive update mechanisms. The sibling papers in the same leaf examine target drift phenomena and architectural trade-offs, but the taxonomy structure suggests limited prior work explicitly comparing mixed versus dedicated critic

designs. The broader 'Lagrangian Method Design and Optimization' branch contains 15 papers across four leaves, indicating moderate activity in foundational method development compared to application-focused branches.

Among 16 candidates examined across three contributions, no clearly refuting prior work was identified. The theoretical analysis of mixed-critic bias examined zero candidates, suggesting this specific framing may be novel or that semantic search did not surface relevant comparisons. The dedicated-critic design contribution examined six candidates with none refuting, while empirical validation examined ten candidates with none refuting. This limited search scope—16 total candidates rather than hundreds—means the analysis captures top semantic matches and immediate citations but cannot claim exhaustive coverage of all potentially overlapping work in multi-constraint architectures or target drift phenomena.

Based on the limited search scope, the work appears to occupy a relatively underexplored niche within multi-constraint Lagrangian RL, specifically addressing architectural design choices that have received less systematic theoretical treatment than multiplier update mechanisms or gradient manipulation techniques. The absence of refuting candidates among 16 examined suggests potential novelty, though the small search scale and sparse taxonomy leaf (4 papers) leave open the possibility of relevant work outside the top semantic matches or in adjacent research communities not fully captured by this taxonomy.

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

Contribution 1: Theoretical analysis of mixed-critic bias in multi-constraint Lagrangian RL

Description: The authors formally prove that training a single mixed critic on aggregated constraint signals introduces structural bias in actor updates. This bias arises because the critic's target drifts as Lagrange multipliers evolve during training, violating the stationarity assumption required by temporal-difference learning and leading to persistent error in policy gradient estimation.

This contribution was assessed against **0 related papers** from the literature. Papers with potential prior art are analyzed in detail with textual evidence; others receive brief assessments.

Contribution 2: Dedicated-critic design eliminates dual-induced drift

Description: The authors prove that maintaining separate critics for reward and each individual constraint eliminates the dual-driven drift problem entirely. This design yields stationary targets that depend only on the policy, not on evolving multipliers, enabling stable policy gradient estimation in multi-constraint settings.

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. CAFL-L: Constraint-Aware Federated Learning with Lagrangian Dual Optimization for On-Device Language Models

URL: [View paper](#)

Brief Assessment

Federated Learning Lagrangian[63] applies Lagrangian dual optimization to federated learning resource constraints, not to multi-constraint safe RL critic architectures. The candidate does not address critic design for policy gradient methods or target drift in value function estimation.

2. Hybrid Actor-Critic Based Low-Overhead Scheduling Using MDP for Large-Scale Edge Computing Networks

URL: [View paper](#)

Brief Assessment

Hybrid Actor Critic Scheduling[64] focuses on edge computing resource scheduling using a hybrid actor-critic approach for mixed discrete-continuous action spaces, not on multi-constraint Lagrangian RL with separate critics per constraint to eliminate dual-driven drift.

3. Learning Constrained Optimization with Deep Augmented Lagrangian Methods

URL: [View paper](#)

Brief Assessment

Deep Augmented Lagrangian[61] focuses on learning dual solutions for constrained optimization problems, not on critic architectures in multi-constraint reinforcement learning. The candidate addresses optimization solver training, while the original contribution concerns RL policy gradient estimation.

4. Multi-Objective Lagrangian Inverse Function Stratified Monte Carlo Method for Quantifying Instability Risks in Compressor Aerodynamic Systems

URL: [View paper](#)

Brief Assessment

Lagrangian Inverse Stratified[66] focuses on Monte Carlo methods for quantifying instability risks in compressor aerodynamic systems, not on reinforcement learning critic architectures or Lagrangian methods for constrained MDPs.

5. Dual-Critic Multi-Agent Deep Reinforcement Learning for Multi-Zone HVAC Safety Control

URL: [View paper](#)

Brief Assessment

Dual Critic HVAC[65] applies a dual-critic architecture to HVAC control but does not address or analyze the theoretical problem of dual-induced drift in Lagrangian methods. The candidate focuses on practical HVAC temperature control rather than the fundamental bias analysis in multi-constraint RL.

6. A Twin Primal-Dual DDPG Algorithm for Safety-Constrained Reinforcement Learning

URL: [View paper](#)

Brief Assessment

Twin Primal Dual DDPG[62] focuses on using twin critics to prevent overestimation in value functions via clipped double Q-learning, not on eliminating dual-induced drift through separate reward/constraint critics. The architectural motivation differs fundamentally from the original paper's drift-elimination claim.

Contribution 3: Empirical validation in constrained bandit and power system environments

Description: The authors validate their theoretical results through experiments in both a constrained bandit problem and a complex energy control task with multiple interacting constraints. The dedicated-critic approach demonstrates stable training, lower constraint violations, and better Pareto frontiers compared to the mixed-critic baseline.

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. Load shedding control strategy in power grid emergency state based on deep reinforcement learning

URL: [View paper](#)

Brief Assessment

Load Shedding DRL[52] focuses on load shedding control in power grid emergency states using deep reinforcement learning, but does not address multi-constraint Lagrangian RL with dedicated vs. mixed critic architectures or constraint satisfaction stability.

2. Adaptive critic nonlinear robust control: A survey

URL: [View paper](#)

Brief Assessment

Adaptive Critic Survey[51] is a survey paper on adaptive-critic-based robust control for continuous-time nonlinear systems, focusing on theoretical methods rather than empirical validation of dedicated-critic architectures in multi-constraint Lagrangian RL settings.

3. Adaptive Critic-Based Control of Voltage Source Converters in Microgrid Systems

URL: [View paper](#)

Brief Assessment

Adaptive Critic Converters[60] focuses on voltage source converter control in microgrids using adaptive critic methods with PI and neuro-fuzzy controllers, not on constrained bandit problems or multi-constraint Lagrangian RL frameworks with dedicated vs. mixed critics.

4. DHP Adaptive Critic based control of STATCOM in power system

URL: [View paper](#)

Brief Assessment

DHP STATCOM Control[58] focuses on voltage control in power systems using adaptive critic designs for a specific hardware component (STATCOM), not on general multi-constraint Lagrangian RL frameworks or dedicated-critic architectures for constraint satisfaction.

5. Safe Multi-Critic Reinforcement Learning-Based Energy Management and Volt-Var Control in Active Distribution Networks

URL: [View paper](#)

Brief Assessment

Multi Critic Energy Management[56] focuses on volt-var control in distribution networks with multiple operational objectives (voltage regulation, loss minimization, peak reduction), not on comparing dedicated vs. mixed critic architectures in Lagrangian RL or analyzing target drift from evolving multipliers.

6. Multi-agent Double Time Scale Two Critic Deep Reinforcement Learning for Voltage Control of Active Distribution System

URL: [View paper](#)

Brief Assessment

Double Time Scale Voltage[59] focuses on voltage control in active distribution systems using a multi-agent double time scale framework, not on general constrained bandit problems or the dedicated-critic vs. mixed-critic comparison central to the original paper's empirical validation.

7. Dual heuristic dynamic programming control of grid-connected synchronverters

URL: [View paper](#)

Brief Assessment

Dual Heuristic Programming[53] focuses on grid-connected synchronverter control using neural networks for power system optimization, not on constrained bandit problems or multi-constraint Lagrangian RL frameworks with dedicated critics.

8. A Soft Actor-Critic-Based Deep Reinforcement Learning for Nanogrid Energy Management

URL: [View paper](#)

Brief Assessment

Nanogrid Energy Management[54] addresses energy management in off-grid PV-battery nanogrids using SAC-based DRL with safe action projection for physical constraints. This differs from the original paper's focus on dedicated-critic architectures for multi-constraint Lagrangian RL in power systems with multiple interacting constraints and Pareto frontier analysis.

9. Physics-Aware Reinforcement Learning for Flexibility Management in PV-Based Multi-Energy Microgrids Under Integrated Operational Constraints

URL: [View paper](#)

Brief Assessment

Physics Aware Microgrid[57] focuses on physics-informed RL for multi-energy microgrids with PV systems, not on comparing dedicated vs. mixed critic architectures in Lagrangian safe RL. The candidate addresses constraint satisfaction through physics-based projection mechanisms rather than through critic architecture design.

10. Adaptive fuzzy critic based control design for AGC of power system connected via AC/DC tie-lines

URL: [View paper](#)

Brief Assessment

Adaptive Fuzzy AGC[55] addresses automatic generation control in power systems using fuzzy critic-based emotional learning, not dedicated-critic architectures for multi-constraint Lagrangian RL. The technical approaches are fundamentally different.

Appendix: Text Similarity Detection

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

References

- [0] Target Drift in Multi-Constraint Lagrangian RL: Theory and Practice [View paper](#)

- [1] Responsive Safety in Reinforcement Learning by PID Lagrangian Methods [View paper](#)
- [2] Safe RLHF: Safe Reinforcement Learning from Human Feedback [View paper](#)
- [3] Off-Policy Conservative Distributional Reinforcement Learning With Safety Constraints [View paper](#)
- [4] Constraints as Rewards: Reinforcement Learning for Robots without Reward Functions [View paper](#)
- [5] Risk-Constrained Reinforcement Learning With Augmented Lagrangian Multiplier for Portfolio Optimization [View paper](#)
- [6] Stable and Safe Reinforcement Learning via a Barrier-Lyapunov Actor-Critic Approach [View paper](#)
- [7] Conservative Distributional Reinforcement Learning with Safety Constraints [View paper](#)
- [8] A deep reinforcement learning-based charging scheduling approach with augmented Lagrangian for electric vehicles [View paper](#)
- [9] Augmented lagrangian-based reinforcement learning for network slicing in IIoT [View paper](#)
- [10] A Review of Safe Reinforcement Learning Methods for Modern Power Systems [View paper](#)
- [11] Model-based safe deep reinforcement learning via a constrained proximal policy optimization algorithm [View paper](#)
- [12] Multi-Constraint Optimization for Real-Time Bidding: A Reinforcement Learning Approach [View paper](#)
- [13] State augmented constrained reinforcement learning: Overcoming the limitations of learning with rewards [View paper](#)
- [14] Constrained Reinforcement Learning-Enabled Policies With Augmented Lagrangian for Cooperative Intersection Management [View paper](#)
- [15] A Reinforcement Learning Embedded Surrogate Lagrangian Relaxation Method for Fast Solving Unit Commitment Problems [View paper](#)
- [16] Safe Reinforcement Learning in Autonomous Driving With Epistemic Uncertainty Estimation [View paper](#)
- [17] Meta SAC-Lag: Towards Deployable Safe Reinforcement Learning via MetaGradient-based Hyperparameter Tuning [View paper](#)
- [18] Gradient shaping for multi-constraint safe reinforcement learning [View paper](#)
- [19] Semantic-Aware Remote Estimation of Multiple Markov Sources Under Constraints [View paper](#)
- [20] Conditionally adaptive augmented Lagrangian method for physics-informed learning of forward and inverse problems using artificial neural networks [View paper](#)
- [21] An Augmented Lagrangian-Based Safe Reinforcement Learning Algorithm for Carbon-Oriented Optimal Scheduling of EV Aggregators [View paper](#)
- [22] Safe Reinforcement Learning with Constraints: A Survey [View paper](#)
- [23] Deep reinforcement learning under signal temporal logic constraints using Lagrangian relaxation [View paper](#)
- [24] Predictive Lagrangian Optimization for Constrained Reinforcement Learning [View paper](#)
- [25] An adaptive augmented Lagrangian method for training physics and equality constrained artificial neural networks [View paper](#)
- [26] Scaling up budgeted reinforcement learning [View paper](#)
- [27] RL-mpca: A reinforcement learning based multi-phase computation allocation approach for recommender systems [View paper](#)
- [28] CVaR-Constrained Policy Optimization for Safe Reinforcement Learning [View paper](#)
- [29] State-Dependent Lagrange Multipliers for State-Wise Safety in Constrained Reinforcement Learning [View paper](#)
- [30] Robust lagrangian and adversarial policy gradient for robust constrained markov decision processes [View paper](#)
- [31] Lagrangian duality for constrained deep learning [View paper](#)
- [32] Multi-Constraint Safe RL with Objective Suppression for Safety-Critical Applications [View paper](#)
- [33] Real-Time Optimal Power Flow: A Lagrangian Based Deep Reinforcement Learning Approach [View paper](#)
- [34] A Stochastic Composite Augmented Lagrangian Method For Reinforcement Learning [View paper](#)
- [35] Causal CSI-Based Trajectory Design and Power Allocation for UAV-Enabled Wireless Networks Under Average Rate Constraints: A Constrained Reinforcement Learning Approach [View paper](#)
- [36] Scalable Constrained Policy Optimization for Safe Multi-agent Reinforcement Learning [View paper](#)
- [37] Composition and Alignment of Diffusion Models using Constrained Learning [View paper](#)
- [38] Constraints Driven Safe Reinforcement Learning for Autonomous Driving Decision-Making [View paper](#)
- [39] Last-iterate convergent policy gradient primal-dual methods for constrained mdps [View paper](#)
- [40] Real-world Reinforcement Learning from Suboptimal Interventions [View paper](#)
- [41] Poce: Primal policy optimization with conservative estimation for multi-constraint offline reinforcement learning [View paper](#)
- [42] Augmented Lagrangian Risk-constrained Reinforcement Learning for Portfolio Optimization (Student Abstract) [View paper](#)
- [43] Constrained differential dynamic programming: A primal-dual augmented lagrangian approach [View paper](#)
- [44] Distributed Reinforcement Learning for Real-Time Batteries Control Using Lagrangian Decomposition [View paper](#)
- [45] Learning to solve soft-constrained vehicle routing problems with lagrangian relaxation [View paper](#)
- [46] Constraint-Aware Deep Reinforcement Learning for End-to-End Resource Orchestration in Mobile Networks [View paper](#)
- [47] Uncertainty-aware constraint inference in inverse constrained reinforcement learning [View paper](#)
- [48] Provably efficient generalized lagrangian policy optimization for safe multi-agent reinforcement learning [View paper](#)
- [49] Concurrent Learning of Policy and Unknown Safety Constraints in Reinforcement Learning [View paper](#)
- [50] Model-Based Chance-Constrained Reinforcement Learning via Separated Proportional-Integral Lagrangian [View paper](#)
- [51] Adaptive critic nonlinear robust control: A survey [View paper](#)
- [52] Load shedding control strategy in power grid emergency state based on deep reinforcement learning [View paper](#)
- [53] Dual heuristic dynamic programming control of grid-connected synchronverters [View paper](#)
- [54] A Soft Actor-Critic-Based Deep Reinforcement Learning for Nanogrid Energy Management [View paper](#)
- [55] Adaptive fuzzy critic based control design for AGC of power system connected via AC/DC tie-lines [View paper](#)
- [56] Safe Multi-Critic Reinforcement Learning-Based Energy Management and Volt-Var Control in Active Distribution Networks [View paper](#)
- [57] Physics-Aware Reinforcement Learning for Flexibility Management in PV-Based Multi-Energy Microgrids Under Integrated Operational Constraints [View paper](#)
- [58] DHP Adaptive Critic based control of STATCOM in power system [View paper](#)
- [59] Multi-agent Double Time Scale Two Critic Deep Reinforcement Learning for Voltage Control of Active Distribution System [View paper](#)
- [60] Adaptive Critic-Based Control of Voltage Source Converters in Microgrid Systems [View paper](#)
- [61] Learning Constrained Optimization with Deep Augmented Lagrangian Methods [View paper](#)
- [62] A Twin Primal-Dual DDPG Algorithm for Safety-Constrained Reinforcement Learning [View paper](#)
- [63] CAFL-L: Constraint-Aware Federated Learning with Lagrangian Dual Optimization for On-Device Language Models [View paper](#)
- [64] Hybrid Actor-Critic Based Low-Overhead Scheduling Using MDP for Large-Scale Edge Computing Networks [View paper](#)
- [65] Dual-Critic Multi-Agent Deep Reinforcement Learning for Multi-Zone HVAC Safety Control [View paper](#)

- [66] Multi-Objective Lagrangian Inverse Function Stratified Monte Carlo Method for Quantifying Instability Risks in Compressor Aerodynamic Systems [View paper](#)