Dynamic Price Incentive for EV Decarbonization Management in Distribution Network

Hongrong Yang, hy2864
Dongbing Han, dh3071

May 10, 2025

1 Introduction

The purpose of the paper is to optimize dynamic price incentives for electric vehicle (EV) charging to simultaneously maximize social welfare, maintain grid safety, and manage carbon emissions. Building on a DRL-based pricing framework that models the price–demand relationship through EV user behavior at charging stations, this paper extends the model by adding system carbon intensity as a new state variable and introducing penalty terms for carbon limit violations. The oringal framework that formulates the pricing problem as a limited Markov Decision Process (CMDP), capturing the implicit relationship between price and demand for charging of EVs through user decision modeling [1]. To solve this, a safe deep reinforcement learning framework is proposed and the Adaptive Model Based Safe DRL (AMSDRL) algorithm [1] is developed. The state variables in original framework only include the distribution of electric vehicles in the charging stations and the operational status of the grid.

Deep reinforcement learning (DRL), the primary method employed in this paper, is a data-driven control technique that learns and generates near-optimal policies rather than exact solutions. This makes DRL especially suitable for solving real-time optimal control problems under uncertainty. In [2], a DRL-based bi-level optimization framework that coordinates power and transportation networks through EV charging pricing. In [3], a DRL-based approach is used to coordinate EV charging while ensuring user satisfaction and fairness. In [4] a optimal DRL-based EV driving model for scheduling EV charging within integrated power and transportation systems. In [5], a bi-level DRL framework for PEV decision-making is proposed to guide EVs to charging stations while optimizing traffic flow and grid load, effectively balancing user costs and infrastructure constraints. In [6], a DRL-based framework that integrates offline/online algorithms to minimize carbon emission costs from the community through dynamic scheduling of EVs.

Most existing EV charging optimization methods focus on social welfare and grid safety but overlook carbon emissions. To address emerging environmental concerns, this paper extends the original CMDP-based dynamic pricing model by incorporating carbon constraints into the optimization model. Specifically, a new state variable is introduced, system carbon intensity, and includes additional penalties for violating carbon intensity limits. The model is validated through simulation using a modified IEEE 33-bus distribution network integrated with a realistic urban transportation scenario in Shenzhen, China.

2 Mathematical Model

2.1 Carbon Intensity Formulation

The system-wide carbon intensity is calculated as a weighted average of the carbon intensities of all power sources, weighted by their power output contributions. The total injected power is:

$$P_{\text{total}} = \sum_{i \in G} P_{G_i} + \sum_{i \in S} P_{S_i} + \sum_{i \in E} P_{E_i} \tag{1}$$

where P_{G_i}, P_{S_i}, P_{E_i} are the active power output (MW) of generator , static generator and external grid connection, respectively.
The contribution ratio of each source is:

$$R_{G_i} = \frac{P_{G_i}}{P_{\text{total}}}, \quad R_{S_i} = \frac{P_{S_i}}{P_{\text{total}}}, \quad R_{E_i} = \frac{P_{E_i}}{P_{\text{total}}} \tag{2}$$

Then the system-wide carbon intensity is:

$$CI = \sum_{i \in G} (R_{G_i} \cdot C_{G_i}) + \sum_{i \in S} (R_{S_i} \cdot C_{S_i}) + \sum_{i \in E} (R_{E_i} \cdot C_{E_i}) \tag{3}$$

where C_{G_i}, C_{S_i},C_{E_i} are the carbon intensity (kg CO_2/MWh) of generator, static generator and external grid connection. If we assume that power from all sources is uniformly distributed across the network, this value can be assigned to all nodes:

$$CI_{\text{node}_j} = CI \quad \forall j \in B \tag{4}$$

This approach assumes that power from all sources is uniformly distributed across the network, ignoring topology and power flow paths.

The weighted system carbon intensity (CI_{sw}) is a metric that reflects the average carbon intensity across all nodes in the power distribution network, weighted by the load at each node. It explicitly incorporates the node-specific allocation factors $\alpha_{j,i}$, which represent the fraction of node j's load supplied by source i. The formula is given by:

$$CI_{\text{sw}} = \sum_{j \in \text{nodes}} \left(\frac{P_{\text{load}_j}}{\sum_{k \in \text{nodes}} P_{\text{load}_k}} \cdot CI_{\text{node}_j} \right)$$

$$CI_{\text{node}_j} = \sum_{i \in \{G,S,E\}} (\alpha_{j,i} \cdot C_i)$$

where P_{load_j} is the active power load at node j (in MW), $\sum_{k \in \text{nodes}} P_{\text{load}_k}$ indicates the total active power load across all nodes in the network, $\alpha_{j,i}$ is the fraction of node j's load supplied by source i (e.g., generator G, static generator S, or external grid E), and C_i is the carbon intensity of source i.

2.2 Problem Formulation

The structure of the proposed method for real-time carbon intensity management in distribution networks is illustrated in Fig. 1. The Distribution System Operator (DSO) first observes real-time information, including the number of EVs at charging stations (CSs), carbon intensity, grid status, and social welfare metrics. Subsequently, the pre-trained policy is loaded to generate a pricing strategy that enables rapid response for real-time carbon intensity management. This policy is derived from a model-based safe deep reinforcement learning (DRL) controller, trained using limited historical data on EV characteristics alongside power system and transportation network parameters. Notably, the EV data includes only non-private attributes, such as arrival and departure state of charge (SOC), timestamps, and battery capacities, without sensitive personal information like trip chains.

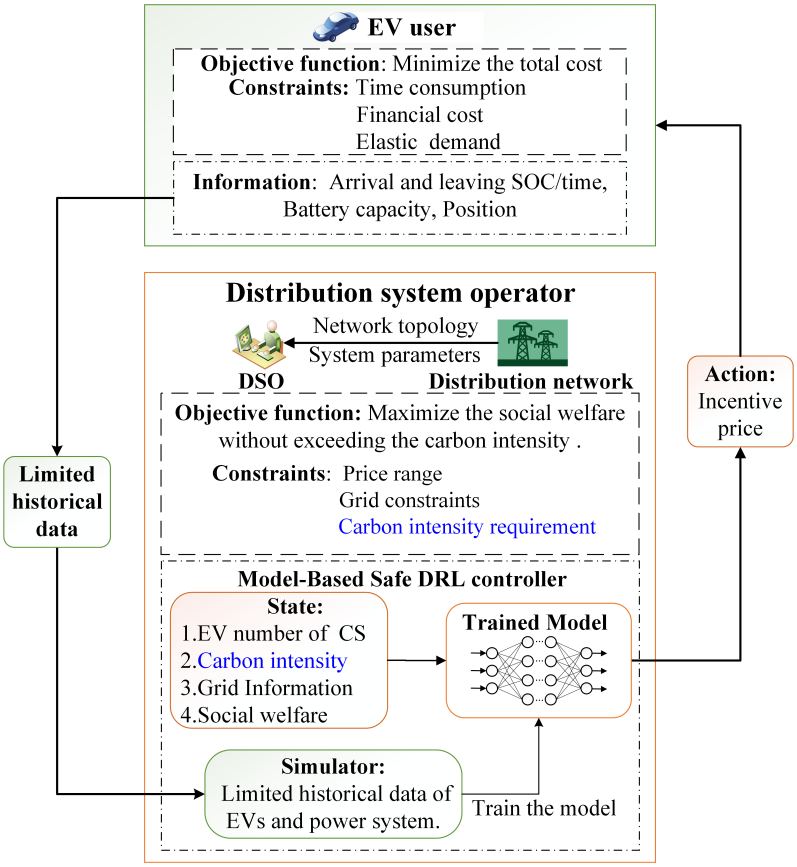

Figure 1: Your descriptive figure title goes here.

Drawn on the bi-level models in [1], the objective function of DSO is to maximize social welfare, without exceeding nodes carbon intensity, is given by:

$$\max \Pi_t^S = \sum_{t \in T} \left(\sum_{n \in B} \sum_{j \in H_n} \Pi_{n,j,t}^{\text{CS}} + \Pi_t^{\text{DSO}} \right), \tag{5}$$

$$\Pi_{n,i,j}^{CS,t} = \Pi_{n,i,j}^{CS,C} - C_{n,i,j}^{CS,P}, \tag{6}$$

$$\Pi_t^{\text{DSO}} = \sum_{n \in B} \sum_{j \in H} C_{n,i,j}^{CS,P} - C_{n,i,j}^{\text{ISO},M} - C_{n,i,j}^{\text{ISO},D}, \tag{7}$$

where $\Pi_{n,j,t}^{\text{CS}}$ denotes the CS profits, which consist of the EV charging profits $\Pi_{n,i,j}^{CS,C}$ and the CS electricity purchase cost $C_{n,i,j}^{CS,P}$; Π_t^{DSO} denotes the total profits of the DSO; and $C_{n,i,j}^{\text{ISO},M}$ and $C_{n,i,j}^{\text{ISO},D}$ are the costs to the DSO of purchasing power from the main grid and distributed generators to sell to CSs, respectively.

The objective of the EV user decision strategy is to choose the CS j to minimize the total cost, which is formulated as:

$$S_t^{EV} = v\sigma S_t^T + (1-v) S_t^F, \tag{8}$$

where v is the weight coefficient reflecting the varying inclinations of EV users toward the two types of costs, and σ is the coefficient that converts time consumption into monetary cost. Specifically, σS_t^T represents the time cost, which includes driving time, waiting time, and charging time, while S_t^F denotes the financial cost.

3 Methodology

3.1 Model-Based Safe DRL Method

The infinite-horizon CMDP in safe reinforcement learning is defined as a tuple $(\mathcal{S}, \mathcal{A}, T, r, c, \gamma, \mu_0)$, where $\mathcal{S}$ and $\mathcal{A}$ are the state and action spaces, respectively, $T : \mathcal{S} \times \mathcal{A} \to \mathcal{D}(\mathcal{S})$ is the transition distribution, and $\mu_0 \in \mathcal{D}(\mathcal{S})$ is the initial state distribution. The reward function is denoted by $r : \mathcal{S} \times \mathcal{A} \to \mathbb{R}$, and the cost function by $c : \mathcal{S} \times \mathcal{A} \to \mathbb{R}$. The discount factor is $\gamma \in (0, 1)$. A policy π is a mapping from states to distributions over actions. To guarantee that the policy π^* is feasible under transition dynamics T, we apply the method in [1, 7] and write the safety model-based framework using the strict and adaptive cost function $J_s(\pi)$ as follows:

$$\begin{aligned} \max_{\pi} \quad & J(\pi) := \mathbb{E}\left[\sum_{t=0}^{\infty} \gamma^t r(s_t, a_t)\right] \\ \text{s.t.} \quad & J_s(\pi) := \mathbb{E}\left[\sum_{t=0}^{\infty} \gamma^t c(s_t, a_t) + \beta u_t(s_t, a_t)\right] \leq C \\ & d_f\big(\hat{T}(s, a), T(s, a)\big) \leq u_t(s, a) \end{aligned} \tag{9}$$

where $\hat{u}_T$ is a heuristic cost penalty derived from the statistics of the transition model. To model the transition, we employ a neural ensemble that outputs a Gaussian distribution $\hat{T}_\theta = \mathcal{N}(\mu_\theta(s_t, a_t), \Sigma_\theta(s_t, a_t))$ and train an ensemble of N models $\{\hat{T}_\theta^i = \mathcal{N}(\mu_\theta^i, \Sigma_\theta^i)\}_{i=1}^N$. Specifically, we use a neural network parameterized by θ to learn and predict transitions in the constrained Markov decision process (CMDP). We then define $\hat{u}_T$ as the maximum Frobenius norm of the standard deviations across the learned ensemble models, which is applied in offline reinforcement learning as follows:

$$u_t(s, a) = \max_i \left\| \sum_e^i (s, a) \right\|_k . \tag{10}$$

However, we observe that directly applying the cost penalty $\hat{u}_T$ is inflexible and results in suboptimal performance in practice. To address this, we introduce an adaptive scalability factor β to balance safety and exploration, as detailed in Equation (10). We employ a proportional-integral (PI) control method to update β with learning rate α as follows:

$$\beta_{i+1} = \beta_i + \alpha\big(J_c(\pi_i) - C\big) \tag{11}$$

where α is the learning rate. When $J_C(\pi) > C$, the value of β is increased to tighten the cost constraint; conversely, when $J_C(\pi) \leq C$, β is decreased to relax the constraint.

3.2 CMDP Formulation

The state includes the number of EVs at charging stations, carbon intensity, grid status, and social welfare metrics. The actions are the charging price of three chargings stations, the reward and cost are as follows:

$$r_t = \lambda_1 \Pi_t^S, \tag{12}$$

$$c_t = \lambda_2 c_t^{ci} + \lambda_3 c_t^g, \tag{13}$$

$$c_t^{ci} = \begin{cases} \sum_{n \in B} CI_{\text{sw},t} - CI_{\text{max},t}, & CI_{\text{sw},t} > CI_{\text{max},t} \\ 0, & \text{else} \end{cases} \tag{14}$$

$$c_t^g = \begin{cases} \sum_{n \in B} \phi_{mn,t} - \phi_{mn,t}^{\max}, & \phi_{mn,t} > \phi_{mn,t}^{\max} \\ 0, & \text{else} \end{cases} \tag{15}$$

where r_t is the reward, and c_t is the total cost, which consists of the carbon intensity cost c_t^{ci} and the congestion cost c_t^g. The parameters $\lambda_1, \lambda_{2,3}$ are scale factors, while $CI_{\max,t}$ and $\phi_{mn,t}^{\max}$ represent the maximum allowable values for carbon intensity and line load percentage, respectively. An additional congestion cost is introduced because the reallocation of EV distribution, performed to satisfy the carbon intensity constraint, may inadvertently cause new congestion in the network.

4 Case Study

All experiments are conducted on a computer equipped with a 5.60 GHz CPU, an NVIDIA RTX 4090 graphics card, and 64 GB of RAM. We develop the simulation environment by integrating a transportation network, a power distribution system, and electric vehicle (EV) charging demand. The entire framework is implemented using PyTorch for the model-based safe DRL algorithm and Pandapower for the power system simulation.The transportation network is constructed using a real map from Shenzhen, China. As shown in Figure 2, the road system includes main roads (blue), secondary roads (red), and urban expressways (green). Three charging stations (CS1, CS2, and CS3) are placed at traffic nodes 2, 9, and 10, marked in yellow.

We adopt a modified IEEE 33-bus distribution network for power system. To better focus on carbon emission dynamics, we simplify congestion-related complexity by modifying certain bus connections and eliminating severely congested line segments (e.g., 1–18, 9–10, and 27–28 in the original setup). CS1-CS3 are connected to buses 13, 14, and 19 respectively, and a distributed generator (DG) is added at bus 5 to simulate renewable integration. More detailed information regarding the transportation network, charging stations (CSs), and EV users can be found in [1].

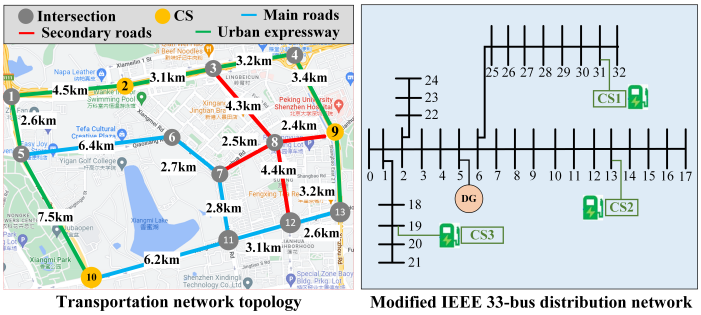

Figure 2: Illustration of the coupled transportation and power network

Figure 3 presents the training process of the proposed model-based safe DRL framework. The top chart illustrates the evolution of social welfare over 500 training iterations, while the bottom chart captures the system constraint violations, including grid congestion and carbon intensity.

In the top plot, the light blue line shows the raw social welfare at each step, while the dark blue line represents the smoothed average over every 50 iterations. Early in training, social welfare fluctuates due to policy exploration. After around 300 iterations, it stabilizes and gradually improves, indicating the agent is learning a better pricing strategy that balances rewards and constraint satisfaction.

The lower chart reveals that constraint violations are frequent and severe in the early phase, particularly within the first 200 iterations. As training progresses, these errors steadily decline and eventually converge. This indicates that the model effectively learns to reduce both grid congestion and carbon intensity violations, leading to a balance of social welfare, grid safety, and carbon emissions.

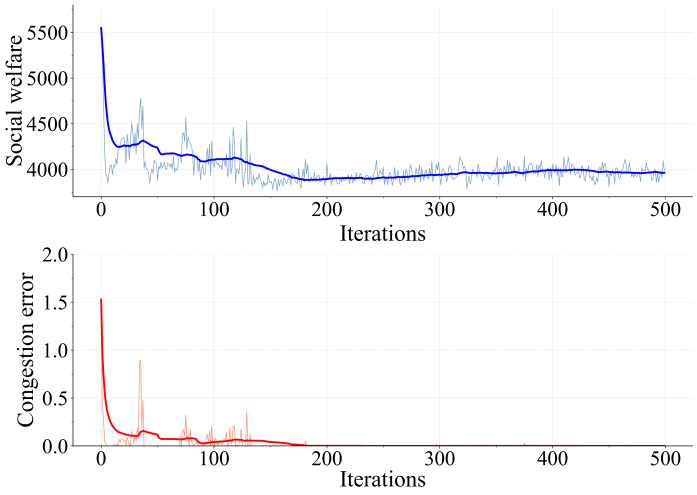

Figure 3: Training results of the model based safe DRL method

Figures 4 and 5 illustrate the effectiveness of our dynamic pricing strategy (continuous data) across the three charging stations (CS1, CS2, and CS3). Figure 4 displays the real-time price trajectories over the course of a full afternoon. While CS2 maintains the highest price throughout, CS1 and CS3 exhibit more dynamic adjustments. Notably, CS3 significantly lowers its price during the final time window (4:45–5:00 PM), setting up a price incentive signal for EV redistribution.

Figure 5 presents the EV behavioral response. For each charging station, we compare the number of EVs arriving (Arrival Number) and the total number present (Total Number) at each time step. CS3 shows the lowest arrival rate for most of the day until the last interval, where the reduced price drives a sharp surge in new arrivals (71 vehicles). This shift confirms that the incentive mechanism effectively influences EV routing decisions.

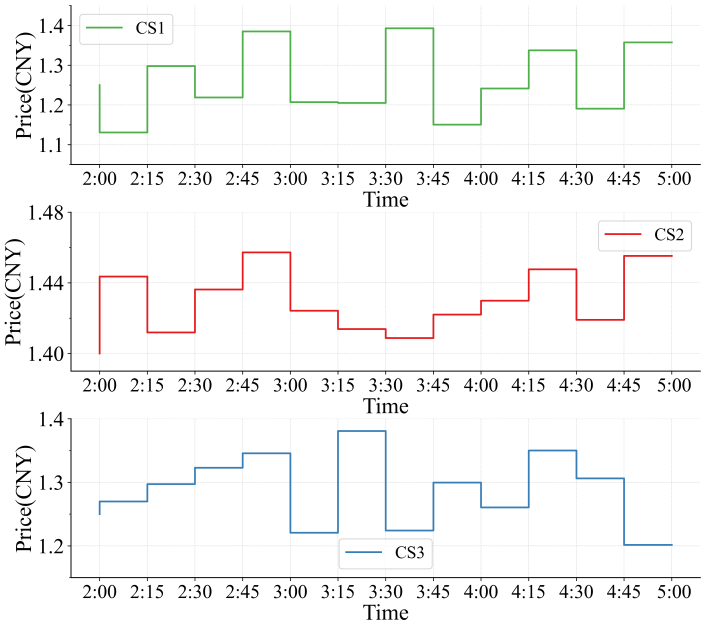

Figure 4: Real-time price of three CSs

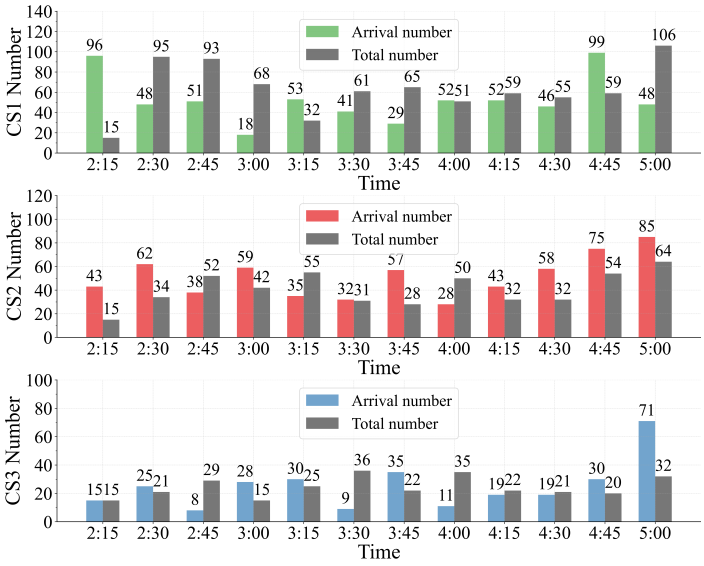

Figure 5: Number of EVs arriving in each time stage

Figure 6 compares the carbon intensity trajectories of the proposed method and the control group over 12 time points. The red dashed line indicates the carbon intensity limit of 790 gCO_2/kWh. The control group, which does not incorporate carbon constraints, frequently exceeds this threshold—especially at time points 6 and 7. In contrast, the proposed method maintains carbon intensity below the limit for nearly all time steps, demonstrating the effectiveness of carbon-aware pricing in constraining emissions. This result confirms that our method successfully guides EV charging behavior toward supporting decarbonization objectives.

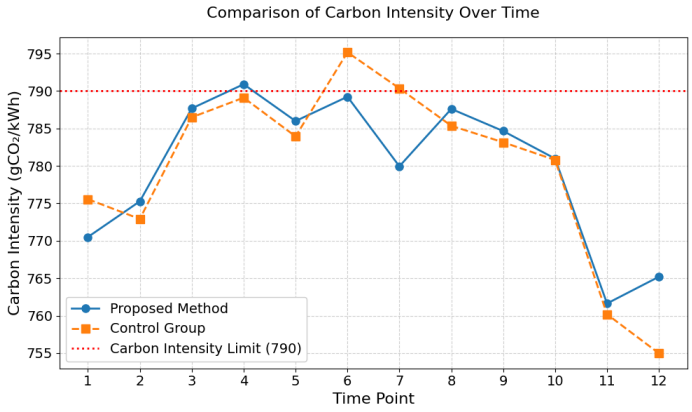

Figure 6: Carbon intensity of the distribution networks

5 Conclusion

This paper extends a model-based safe DRL framework for dynamic EV pricing by incorporating carbon intensity as a new optimization constraint. By adding system-wide carbon metrics to the CMDP formulation, the updated AMSDRL algorithm learns incentive pricing policies that balance social welfare, grid safety, and carbon emissions. Simulation results on a coupled transportation and modified IEEE 33-bus network demonstrate that the proposed method effectively prevents carbon intensity violations, maintains grid stability, and enhances EV load distribution. Compared with the baseline without carbon constraints, our approach maintains carbon intensity below the specified threshold while achieving stable training convergence and improved environmental performance.

References

[1] H. Yang, Y. Xu, and Q. Guo, "Dynamic incentive pricing on charging stations for real-time congestion management in distribution network: an adaptive model-based safe deep reinforcement learning method," *IEEE Transactions on Sustainable Energy*, vol. 15, no. 2, pp. 1100–1113, 2023.

[2] T. Qian, C. Shao, X. Li, X. Wang, and M. Shahidehpour, "Enhanced coordinated operations of electric power and transportation networks via ev charging services," *IEEE Transactions on Smart Grid*, vol. 11, no. 4, pp. 3019–3030, 2020.

[3] J. Fan, H. Wang, and A. Liebman, "Marl for decentralized electric vehicle charging coordination with v2v energy exchange," in *IECON 2023- 49th Annual Conference of the IEEE Industrial Electronics Society*, 2023, pp. 1–6.

[4] H. Cha, M. Chae, M. A. Zamee, and D. Won, "Operation strategy of ev aggregators considering ev driving model and distribution system operation in integrated power and transportation systems," *IEEE Access*, vol. 11, pp. 25 386–25 400, 2023.

[5] Q. Xing, Z. Chen, R. Wang, and Z. Zhang, "Bi-level deep reinforcement learning for pev decision-making guidance by coordinating transportation-electrification coupled systems," *Frontiers in Energy Research*, vol. 10, 01 2023.

[6] V. R. Chifu, T. Cioara, C. B. Pop, H. Rusu, and I. Anghel, "A deep q-learning based smart scheduling of evs for demand response in smart grids," 2024. [Online]. Available: https://arxiv.org/abs/2401.02653

[7] Y. J. Ma, A. Shen, O. Bastani, and J. Dinesh, "Conservative and adaptive penalty for model-based safe reinforcement learning," in *Proceedings of the 36th AAAI Conference on Artificial Intelligence*, vol. 36, no. 5, June 2022, pp. 5404–5412.