Resource Constrained Safe Reinforcement Learning for Lane Following of Wheeled Mobile Robots

Mahtab Noor Shaan[1], Vignesh Narayanan[2], and Avimanyu Sahoo[3]

Abstract— **This paper proposes a safe and resource-efficient near optimal lane following control scheme for wheeled mobile robots (WMRs). The method performs online co-optimization of both the tracking control inputs and their execution instants, while rigorously enforcing lane-keeping safety and actuator constraints. A differential game–inspired event-based adaptive dynamic programming (ADP) and reinforcement learning (RL)-based scheme is developed to synthesize a constrained control policy and adaptively schedule control execution instants. A non-quadratic performance index is formulated that simultaneously penalizes the system states, the constrained control effort, the input error induced by aperiodic (event-like) execution, and a reciprocal control barrier function (CBF) for lane keeping, enforcing safety. The associated Hamilton–Jacobi–Isaacs (HJI) equation is solved approximately using a single critic neural network. Simulation results for state and input safety constraints demonstrate the effectiveness of the proposed approach, yielding up to a 61% reduction in control updates compared with continuous-feedback baselines.**

I. INTRODUCTION

Mobile robots (MRs) are increasingly deployed in various domains such as warehouse automation, visual inspection, autonomous delivery, and transportation systems. In these applications, it is imperative that MRs reliably follow prescribed trajectories while ensuring lane-keeping and collision avoidance, despite limitations in sensing, computation, and actuation [1], [2].

Classical optimal tracking controllers, particularly those based on Hamilton–Jacobi–Bellman (HJB) theory, provide a principled framework for synthesizing feedback policies. However, closed-form solutions to the HJB equation are generally intractable for nonlinear systems. Approximate dynamic programming (ADP) [3] and reinforcement learning (RL) [4] techniques address this challenge by using neural networks (NNs) to approximate value functions, enabling near-optimal control for nonlinear systems with unknown dynamics [5]. These methods have been successfully applied to tracking problems [6], [7]. Despite their success, conventional ADP/RL controllers are computationally expensive [6]–[8]. To mitigate these issues, event-triggered control (ETC) [9] based ADP/RL schemes that reduce resource usage by aperiodically updating control and sensing for tracking control [10], [11] also proposed.

Beyond optimality, safety is crucial in robotic applications, particularly under actuator constraints and in tasks such as lane keeping and obstacle avoidance. Control barrier functions (CBFs), which enforce forward invariance of safe sets, have been proposed for real-time safety certification [12]. Classical optimal control with CBFs incorporates safety constraints into a quadratic program [12]. CBFs have also been integrated with ADP/RL methods. In these approaches, actuator constraints and CBF-based collision-avoidance terms for goal-reaching problems are embedded in the performance index or objective function [13], [14]. However, these approximation-based solution methods are computationally complex and are primarily designed for point-to-point navigation with obstacle avoidance rather than trajectory tracking.

To address these challenges, this paper proposes an event-triggered ADP/RL framework that jointly optimizes trajectory-tracking performance and control execution while ensuring safety. The **main contributions** are twofold. First, an event-triggered optimal control formulation that incorporates actuator constraints and a CBF for lane keeping via a non-quadratic performance index [13], together with the error induced by aperiodic control execution is developed, thereby jointly addressing tracking, safety, and aperiodic implementation. Second, an approximate solution to the resulting event-triggered differential game using a single-critic ADP/RL scheme is proposed, producing a safe, near-optimal control policy and an adaptive event-triggering mechanism. Third, a new CBF that guarantees lane safety with forward invariant guarantee of the safe set is introduced.

II. BACKGROUND AND PROBLEM FORMULATION

In this section, a brief background on the wheeled MR (WMR) kinematics, system reformulation for optimal tracking control, and CBF for safety is presented.

A. Tracking Control of WMR

Consider a WMR represented by a nonholonomic unicycle kinematic model given by

$$\dot{x} = v\cos\theta, \quad \dot{y} = v\sin\theta, \quad \dot{\theta} = \omega, \tag{1}$$

where x and y specify the robot's position in an inertial reference frame, and θ denotes its heading angle measured from the x-axis. The linear and angular velocities are denoted by v and ω, respectively.

This work was supported by National Science Foundation under grants 2337998 and 2337999 and Air Force Office of Scientific Research under grant FA9550-24-1-0228.

[1]Mahtab Noor Shaan and [3]Avimanyu Sahoo are with the Department of Electrical Engineering, University of Alabama in Huntsville (UAH), AL, USA (ms0355@uah.edu and avimanyu.sahoo@uah.edu). [2]Vignesh Narayanan is with the AI Institute, University of South Carolina, Columbia, SC, USA (deyr@email.sc.edu and vignar@sc.edu). .

In a compact form, the model can be represented as an affine nonlinear system as

$$\dot{\zeta} = f(\zeta) + g(\zeta)u, \tag{2}$$

where the vector function $f(\zeta) = \mathbf{0} \in \mathbb{R}^3$ is the internal dynamics and matrix function $g(\zeta) = \begin{bmatrix} \cos\theta & 0 \\ \sin\theta & 0 \\ 0 & 1 \end{bmatrix} \in \mathbb{R}^{3\times 2}$ is the control coefficient with $\zeta = \begin{bmatrix} x & y & \theta \end{bmatrix}^T \in \mathbb{R}^3$ is the state vector and $u = [v \;\; \omega]^T \in \mathcal{U}_c \subset \mathbb{R}^2$ is the control input vector with $\mathcal{U}_c \subset \mathbb{R}^2$ is an admissible set of control inputs. The control coefficient satisfies $g^\top(\zeta)g(\zeta) = I_2$, and thus $g^\dagger(\zeta) = g^\top(\zeta)$ and $\|g^\dagger(\zeta)\|_2 = 1$ for all ζ.

The control goal is to track a bounded, continuously differentiable reference trajectory $\zeta_d(t)$ generated by

$$\dot{\zeta}_d = h(\zeta_d), \quad \zeta_d(0) = \zeta_{d0}. \tag{3}$$

Assumption 1 (Reference feasibility): The reference dynamics (3) are assumed to be feasible for the unicycle kinematics, i.e., $h(\zeta_d) \in \mathrm{Range}(g(\zeta_d))$ for all ζ_d. Equivalently, there exist scalar functions $v_d(\zeta_d)$ and $\omega_d(\zeta_d)$ such that $h(\zeta_d) = \begin{bmatrix} v_d(\zeta_d)\cos\theta_d & v_d(\zeta_d)\sin\theta_d & \omega_d(\zeta_d) \end{bmatrix}^\top$.

Defining the tracking error $e = \zeta - \zeta_d$, and using (2) and (3), the error dynamics can be expressed as

$$\dot{e} = f(e + \zeta_d) + g(e + \zeta_d)u - h(\zeta_d). \tag{4}$$

The steady-state input corresponding to ζ_d [10] is

$$u_{\text{steady}} = g^\dagger(\zeta_d)\big(h(\zeta_d) - f(\zeta_d)\big). \tag{5}$$

Defining the augmented state $z = \begin{bmatrix} e^T, \zeta_d^T \end{bmatrix}^T$, the dynamics can be expressed as [6]

$$\dot{z} = F(z) + G(z)u_{\text{mm}}, \tag{6}$$

where $F(z) = \begin{bmatrix} f(e + \zeta_d) + g(e + \zeta_d)u_{\text{steady}} - h(\zeta_d) \\ h(\zeta_d) \end{bmatrix}$, $G(z) = \begin{bmatrix} g(e + \zeta_d) \\ 0 \end{bmatrix}$, and $u_{\text{mm}} = u - u_{\text{steady}}$ is the mismatch input. The augmented system (6) transform the tracking problem to a regulation problem, where the goal is to design u_{mm} to compute the control input $u = u_{\text{mm}} + u_{\text{steady}} \in \mathbb{R}^2$, such that the tracking error $e \to 0$ as $t \to \infty$.

B. Event-based Tracking Control (ETC)

In an event-based control [10], [11], feedback information is available only at a discrete set of triggering instants $\{t_k\}_{k\in\{0,\mathbb{N}\}} \subset \mathbb{R}_{\geq 0}$, with $t_0 = 0$ and $t_k < t_{k+1}$ for all $k \in \{0, \mathbb{N}\}$. At each triggering instant t_k, the system state and reference trajectory are sampled and transmitted to the controller. Between consecutive triggering instants, the sampled signals are held constant, yielding

$$\zeta_{es}(t) = \zeta(t_k), \quad \zeta_{ds}(t) = \zeta_d(t_k), \quad \forall t \in [t_k, t_{k+1}), \tag{7}$$

and the controller applies a zero-order-hold (ZOH) input of the form $u_{es}(t) = \mu(\zeta_{es}(t))$ to the system.

Define the event-sampled augmented state as

$$z_{es}(t) = z(t_k) = \begin{bmatrix} e^\top(t_k), \, \zeta_{ds}^\top(t) \end{bmatrix}^\top, \; \forall t \in [t_k, t_{k+1}). \tag{8}$$

The corresponding error due to aperiodic sampling is given by $e_s(t) = z_{es}(t) - z(t)$. Similarly, the aperiodic control execution error is defined as $e_u(t) = u_{es}(t) - u(t)$.

Under event-triggered implementation, the system dynamics can be expressed as

$$\dot{\zeta} = f(\zeta) + g(\zeta)u + g(\zeta)e_u. \tag{9}$$

Consequently, the augmented dynamics in (6) in the event-triggered representation can be expressed as

$$\dot{z} = F(z) + G(z)u_{\text{mm}} + G(z)e_u, \tag{10}$$

where the additional term $G(z)e_u$ captures the effect of sampling-induced control mismatch.

Remark 1: The event-triggered implementation introduces a matched disturbance through the input channel, represented by the control sampling error e_u in the augmented dynamics. A triggering mechanism is therefore designed to determine the sampling instants while preserving the controller's performance requirements.

C. Safety Constraints via Control Barrier Functions (CBF)

To formalize safety for the WMR system in (2), define the safe set

$$\mathcal{C} = \{\zeta \in \mathbb{R}^3 \mid h_{\text{cbf}}(\zeta) \geq 0\}, \tag{11}$$

for a continuously differentiable function $h_{\text{cbf}} : \mathbb{R}^3 \to \mathbb{R}$. The boundary and interior of the safe set are given by $\partial\mathcal{C} = \{\zeta \mid h_{\text{cbf}}(\zeta) = 0\}$ and $\mathrm{Int}(\mathcal{C}) = \{\zeta \mid h_{\text{cbf}}(\zeta) > 0\}$ respectively.

Definition 1 (Zeroing Control Barrier Function): [15] For the control-affine system (2), h_{cbf} is a zeroing CBF if there exists an extended class-$\mathcal{K}$ function η such that

$$L_f h_{\text{cbf}}(\zeta) + L_g h_{\text{cbf}}(\zeta)u \geq -\eta(h_{\text{cbf}}(\zeta)) \tag{12}$$

for all ζ in $\mathcal{D}$ and for all admissible inputs u. The $L_f h_{\text{cbf}}$ and $L_g h_{\text{cbf}}$ are the Lie derivatives of h_{cbf} along f and g.

Any Lipschitz continuous controller satisfying (12) renders $\mathcal{C}$ forward invariant, thereby certifying safety, and is demonstrated [15].

D. Problem Statement

In practical applications, a WMR is subject to hard actuator limits that constrain the admissible control inputs, and it must satisfy safety requirements, such as lane keeping and obstacle avoidance, throughout its motion. At the same time, the control strategy must account for limited computational and communication resources, motivating the need to reduce the frequency of control execution and state feedback without compromising closed-loop performance.

From a control perspective, the robot must accurately track the desired trajectory while maintaining safety at all times. This task is further complicated by actuator constraints, which limit the corrective actions available during transient phases, and by event-triggered execution, which introduces sampling-induced disturbances from aperiodic control updates. Moreover, achieving near-optimal tracking performance in the presence of unknown or nonlinear dynamics necessitates learning-based control mechanisms, whose convergence and stability must be guaranteed under both safety and resource constraints. A solution to the above problems is presented in the next sections.

III. Performance Index Design Under Actuator and Safety Constraints

This section presents the formulation of the objective function for the co-optimization problem with safety constraints.

A. Actuator Constraints and Feasibility

In practical robotic systems, the control inputs $u = [v \;\; \omega]^\top$ are subject to actuator limitations and must therefore remain within prescribed bounds. To ensure physical realizability of the control law while preserving stability and safety guarantees, the applied control input is required to lie in an admissible set that encodes these actuator constraints. Accordingly, define the admissible control input set as

$$\mathcal{U}_c := \Big\{ u = [v \;\; \omega]^\top \in \mathbb{R}^2 \;\Big|\; v_{\min} \le v \le v_{\max}, \\ \omega_{\min} \le \omega \le \omega_{\max} \Big\}, \tag{13}$$

where $v_{\min}$ and $v_{\max}$ denote the minimum and maximum allowable linear velocities, and $\omega_{\min}$ and $\omega_{\max}$ denote the minimum and maximum allowable angular velocities imposed by the actuators. All control inputs synthesized in this work are constrained to lie in $\mathcal{U}_c$.

To guarantee the control input is attainable under actuator limitations, the applied control $u = u_{\text{steady}} + u_{\text{mm}}$ must satisfy the admissible input constraint $u(t) \in \mathcal{U}_c$ for all $t \ge 0$. In the ideal tracking regime, the mismatch input $u_{\text{mm}} \to 0$, implying $u \to u_{\text{steady}}$. Consequently, a necessary condition for feasibility is that the steady-state input remains admissible, i.e.,

$$u_{\text{steady}}(t) \in \mathcal{U}_c, \quad \forall t \ge 0.$$

Let $\bar{u}_{\text{ss}} := [\bar{v}_{\text{ss}} \quad \bar{\omega}_{\text{ss}}]^\top$ denote known bounds on the steady-state input components such that $|v_{\text{ss}}(t)| \le \bar{v}_{\text{ss}}$, $|\omega_{\text{ss}}(t)| \le \bar{\omega}_{\text{ss}}$, $\forall t \ge 0$, with $\bar{v}_{\text{ss}} < \min\big(|v_{\min}|, |v_{\max}|\big)$, $\bar{\omega}_{\text{ss}} < \min\big(|\omega_{\min}|, |\omega_{\max}|\big)$.

To preserve feasibility under the control decomposition, the mismatch input is restricted to the compact set

$$\mathcal{U}_{\text{cmm}} := \big\{ u_{\text{mm}} = [v_{\text{mm}} \;\; \omega_{\text{mm}}]^\top \in \mathbb{R}^2 \;\big|\; |v_{\text{mm}}| \le \bar{v}, \\ |\omega_{\text{mm}}| \le \bar{\omega} \big\}, \tag{14}$$

where $\bar{v} := \min\big(|v_{\min}|, |v_{\max}|\big) - \bar{v}_{\text{ss}}$ is the bound and $\bar{\omega} := \min\big(|\omega_{\min}|, |\omega_{\max}|\big) - \bar{\omega}_{\text{ss}}$.

Under these bounds, any control input of the form $u = u_{\text{steady}} + u_{\text{mm}}$ with $u_{\text{steady}} \in \mathcal{U}_c$ and $u_{\text{mm}} \in \mathcal{U}_{\text{cmm}}$ is guaranteed to satisfy the actuator constraints.

B. Control Barrier Function for Missmatch Control Input

Given the augmented system dynamics in (6), we redefine the safe set in the augmented state space as

$$\mathcal{C}_{\text{aug}} := \big\{ z \in \mathbb{R}^6 \;\big|\; h_{\text{cbf}}(z) \ge 0 \big\}, \qquad z(0) \in \mathcal{C}_{\text{aug}}. \tag{15}$$

To incorporate the actuator and safety considerations into the optimal tracking objective, the instantaneous cost can be defined as

$$r_{\text{safe}}(z, u_{\text{mm}}) = z^\top \bar{Q} z + R_u(u_{\text{mm}}) + B(z), \tag{16}$$

where $\bar{Q} = \begin{bmatrix} Q & 0 \\ 0 & 0 \end{bmatrix}$, with $Q \succ 0$ penalizing the tracking error. The actuator constraint is give by

$$R_u(u_{\text{mm}}) = 2 \sum_{i=1}^{2} \int_0^{u_i} \bar{u}_i r_i \tanh^{-1}\Big(\frac{\zeta_i}{\bar{u}_i}\Big)\, d\zeta_i, \tag{17}$$

where u_i denotes the i-th component of u_{mm}, $r_i \in \mathbb{R}_{>0}$ are diagonal entries of the positive-definite matrix $R = \operatorname{diag}(r_1, r_2)$, and $\bar{u} = [\bar{v}, \;\; \bar{\omega}]^\top$ represents the symmetric mismatch input bounds defined in (14). continuously differentiable function $B : \operatorname{Int}(\mathcal{C}_{\text{aug}}) \to \mathbb{R}_{\ge 0}$ [13] satisfy

$$\inf_{z \in \operatorname{Int}(\mathcal{C}_{\text{aug}})} B(z) \ge 0, \qquad \lim_{z \to \partial \mathcal{C}_{\text{aug}}} B(z) = \infty, \tag{18}$$

with $B\Big(\big[0^\top \quad \zeta_d^\top\big]^\top\Big) = 0$, which ensures that the barrier penalty does not affect the nominal tracking objective when the tracking error vanishes. A commonly adopted choice inspired by [13] is $B(z) = \frac{s(z)}{h_{\text{cbf}}(z)}$, where $s : \mathbb{R}^6 \to [0, 1]$ is a smooth scheduling function that activates the barrier penalty only in a neighborhood of $\partial \mathcal{C}_{\text{aug}}$, thereby preventing unnecessary penalization deep inside the safe set [16].

The performance index can be expressed as

$$J_{\text{safe}}(z, u_{\text{mm}}) = \int_0^\infty r_{\text{safe}}\big(z(\tau), u_{\text{mm}}(\tau)\big)\, d\tau. \tag{19}$$

C. Performance Index for Aperiodic Control Execution

To simultaneously (i) minimize the tracking performance index while enforcing actuator feasibility and safety constraints, and (ii) promote aperiodic control execution by maximizing admissible inter-event times, the problem can be reformulated as a two-player zero-sum differential game [11]. In this formulation, the controller seeks to minimize the cost with respect to the mismatch input u_{mm}, while an adversarial player selects the worst-case admissible bound on the sampling-induced control error.

Accordingly, the safe performance index for the event-triggered augmented system (10) can be redefined as

$$J_{\text{safe}}\big(z, u_{\text{mm}}, \hat{e}_u\big) = \int_0^\infty \Big(r_{\text{safe}}\big(z(\tau), u_{\text{mm}}(\tau)\big) \\ - \alpha^2\, \hat{e}_u^\top(\tau)\, \hat{e}_u(\tau) \Big)\, d\tau, \tag{20}$$

where $\alpha > \alpha^*$ is the H_∞ attenuation level that penalizes the admissible sampling-error bound $\hat{e}_u$ associated with the control sampling error e_u [17], with $\alpha^* > 0$ denoting the minimal achievable attenuation. The solution to the co-optimization problem is presented next.

IV. Safe Event-triggered Reinforcement Learning Control Design

A. Exact Solution of the Co-optimization Problem

The aperiodic control problem induced by event-triggered implementation can be cast as a zero-sum differential game. In this formulation, the mismatch input u_{mm} acts as the minimizing player, while the admissible sampling-error threshold $\hat{e}_u$ represents a maximizing player that captures worst-case effects of aperiodic control. The interaction between these players leads to a saddle-point value function [17] given by

$$V^*(z) = \min_{u_{\mathrm{mm}}(\cdot)} \max_{\hat{e}_u(\cdot)} \int_0^\infty \Big(z^\top(\tau)\bar{Q}z(\tau) + R_u\big(u_{\mathrm{mm}}(\tau)\big) + B\big(z(\tau)\big) - \alpha^2\, \hat{e}_u^\top(\tau)\hat{e}_u(\tau)\Big)\, d\tau. \tag{21}$$

Based on the event-triggered augmented dynamics in (10), the Hamiltonian for the differential game is given by

$$\mathcal{H}\big(z, u_{\mathrm{mm}}, \hat{e}_u, \nabla V^*(z)\big) := z^\top \bar{Q} z + R_u(u_{\mathrm{mm}}) + B(z) - \alpha^2\, \hat{e}_u^\top \hat{e}_u + \big(\nabla V^*(z)\big)^\top \big(F(z) + G(z)u_{\mathrm{mm}} + G(z)\hat{e}_u\big). \tag{22}$$

The corresponding HJI equation is obtained as $\min_{u_{\mathrm{mm}}} \max_{\hat{e}_u} \mathcal{H}\big(z, u_{\mathrm{mm}}, \hat{e}_u, \nabla V^*(z)\big) = 0$. Stationary saddle-point strategies follow from first-order optimality conditions. The minimizing policy admits a saturation-compatible representation of the form

$$u^*_{\mathrm{mm}}(z) = -\bar{u}\,\mathrm{Tanh}\bigg(\frac{1}{2}\,\mathrm{diag}(\bar{u})^{-1}\, R^{-1}\, G^\top(z)\, \nabla V^*(z)\bigg), \tag{23}$$

while the maximizing policy associated with the sampling-error bound is given by

$$\hat{e}^*_u(z) = \frac{1}{2\alpha^2}\; G^\top(z)\, \nabla V^*(z). \tag{24}$$

where $\mathrm{Tanh}(\beta) \triangleq [\tanh(\beta_1), \tanh(\beta_2)]^\top$ for $\beta \in \mathbb{R}^2$.

Combining the optimal mismatch input with the feedforward steady-state term yields safe optimal control law

$$\begin{aligned} u^*(z) &= u^*_{\mathrm{mm}}(z) + g^\dagger(\zeta_d)\big(h(\zeta_d) - f(\zeta_d)\big) \\ &= -\bar{u}\,\mathrm{Tanh}\bigg(\frac{1}{2}\,\mathrm{diag}(\bar{u})^{-1}\, R^{-1}\, G^\top(z)\, \nabla V^*(z)\bigg) \\ &\quad + g^\dagger(\zeta_d)\big(h(\zeta_d) - f(\zeta_d)\big), \end{aligned} \tag{25}$$

with the corresponding event-sampled implementation

$$\begin{aligned} u^*_{es}(t) &= -\bar{u}\,\mathrm{Tanh}\Big(\frac{1}{2}\,\mathrm{diag}(\bar{u})^{-1}\, R^{-1}\, G^\top(z_{es}(t)) \\ &\nabla V^*(z_{es}(t))\Big) + g^\dagger(\zeta_{ds}(t))\big(h(\zeta_{ds}(t)) - f(\zeta_{ds}(t))\big), \end{aligned} \tag{26}$$

where $\nabla V^*(z_{es}) = \frac{\partial V^*}{\partial z}\big|_{z=z_{es}}$. By using (23) and (24), it holds that $\mathcal{H}\big(z, u^*_{\mathrm{mm}}, \hat{e}^*_u, \nabla V^*(z)\big) = 0$.

Treating $\hat{e}^*_u(z)$ as a state-dependent worst-case bound, the sampling instants are determined according to

$$\begin{aligned} t_{k+1} = \inf\Big\{ t > t_k \;\Big|\; & e_u^\top(t) e_u(t) = \max\Big\{\gamma^2, \\ & \frac{1}{4\alpha^4}\, \nabla V^*(z)^\top G(z) G^\top(z)\, \nabla V^*(z)\Big\}\Big\}, \end{aligned} \tag{27}$$

with $t_0 = 0$, where $\gamma > 0$ guarantees a strictly positive lower bound on inter-event times, thereby excluding Zeno behavior.

B. Approximate Solution to the Co-optimization Problem

The optimal safe value function generally lacks a closed-form expression and must be approximated, but directly approximating it is numerically unstable due to large gradients from reciprocal barrier terms near the safe-set boundary. To address this, a structured approximation is used that splits the value function into a known bounded and an unknown residual parts. A bounded barrier-like function $B_{\mathrm{bound}}(z) = \frac{s(z)}{h_{\mathrm{cbf}}(z)+c}$ with $c > 0$ captures the safety contribution in a numerically well-conditioned way, while the remaining value function component $V_r(z)$ is smooth over the augmented safe set and represents tracking performance, constrained control effort, and sampling-induced effects.

Accordingly, the optimal value function is approximated using the structured representation

$$V^*(z) \;\approx\; V_r(z) + B_{\mathrm{bound}}(z), \tag{28}$$

where only the residual term $V_r(z)$ is learned, while $B_{\mathrm{bound}}(z)$ is incorporated analytically. This decomposition does not alter the optimal control problem, but constrains the function class used for value function approximation.

The residual component is approximated using a functional-link (critic) neural network as

$$V_r(z) = W^\top \Phi(z) + \varepsilon(z), \tag{29}$$

where $W \in \mathbb{R}^{l_N}$ denotes the ideal weight vector, $\Phi : \mathbb{R}^6 \to \mathbb{R}^{l_N}$ is an activation map with $\Phi(0) = 0$, l_N is the number of basis functions, and $\varepsilon(z)$ represents the approximation error.

The corresponding estimated value function, evaluated using event-sampled information, is defined as

$$\hat{V}(z_{es}) = \hat{W}^\top \Phi(z_{es}) + B_{\mathrm{bound}}(z_{es}), \forall t \in [t_k, t_{k+1}), \tag{30}$$

where $\hat{W}$ is the adaptive critic weight vector.

The gradient of the estimated safe value function then follows as

$$\nabla \hat{V}(z_{es}) = \nabla \Phi^\top(z_{es})\, \hat{W} + \nabla B_{\mathrm{bound}}(z_{es}), \tag{31}$$

for all $t \in [t_k, t_{k+1})$, where $\nabla B_{\mathrm{bound}}(z_{es}) := \frac{\partial B_{\mathrm{bound}}}{\partial z}\big|_{z=z_{es}}$.

Using (31), the event-based safe control input is given by

$$\begin{aligned} u_{es}(t) &= -\bar{u}\,\mathrm{Tanh}\Big(\frac{1}{2}\,\mathrm{diag}(\bar{u})^{-1} R^{-1} G^\top(z_{es})\big(\nabla \Phi^\top(z_{es})\hat{W} \\ &+ \nabla B_{\mathrm{bound}}(z_{es})\big)\Big) + g^\dagger(\zeta_{ds})\big(h(\zeta_{ds}) - f(\zeta_{ds})\big), \end{aligned} \tag{32}$$

for all $t \in [t_k, t_{k+1})$, and the corresponding estimate of the worst-case sampling-error bound is

$$\hat{e}_u(z) = \frac{1}{2\alpha^2}\, G^\top(z)\big(\nabla \Phi^\top(z_{es})\hat{W} + \nabla B_{\mathrm{bound}}(z_{es})\big). \tag{33}$$

Over each inter-event interval, the Bellman residual is computed as $\delta_k = \int_{t_{k-1}}^{t_k} \Big(z^\top \bar{Q} z + R_u(u_{\mathrm{mm}}) + B(z) - \alpha^2 \hat{e}_u^\top \hat{e}_u \Big) d\tau + \hat{W}^\top \Delta\Phi(\tau_{k-1}) + \Delta B_{\mathrm{bound}}(\tau_{k-1})$, where $\Delta\Phi(\tau_{k-1}) = \Phi(z(t_k)) - \Phi(z(t_{k-1}))$ and $\Delta B_{\mathrm{bound}}(\tau_{k-1}) = B_{\mathrm{bound}}(z(t_k)) - B_{\mathrm{bound}}(z(t_{k-1}))$.

To handle the availability of the augmented state only at the event instants, the critic weights are updated via a discrete jump at each trigger, given by

$$\hat{W}(t_k^+) = \hat{W}(t_k^-) - \frac{\alpha_2 \Delta\Phi(\tau_{k-1})\delta_k}{\big(1 + \Delta\Phi^\top(\tau_{k-1})\Delta\Phi(\tau_{k-1})\big)^2}, \tag{34}$$

and continuously evolved between events according to

$$\dot{\hat{W}} = \frac{-\alpha_1 \Delta\Phi(\tau_{k-1})\delta_k,}{\big(1 + \Delta\Phi^\top(\tau_{k-1})\Delta\Phi(\tau_{k-1})\big)^2}, t \in (t_k, t_{k+1}), \tag{35}$$

where $\alpha_1, \alpha_2 > 0$ are learning gains.

The event-triggering condition in (27) is reformulated using the estimated gradient as

$$t_{k+1} = \inf\Big\{t > t_k \;\Big|\; e_u^\top e_u = \max\Big\{\gamma^2,\ \frac{1}{4\alpha^4}\big(\nabla\Phi^\top(z_{es}) \hat{W} + \nabla B_{\text{bound}}(z_{es})\big)^\top G(z)G^\top(z)\big(\nabla\Phi^\top(z_{es})\hat{W} + \nabla B_{\text{bound}}(z_{es})\big)\Big\}\Big\}, \quad t_0 = 0. \tag{36}$$

C. Design of CBF for Lane Keeping

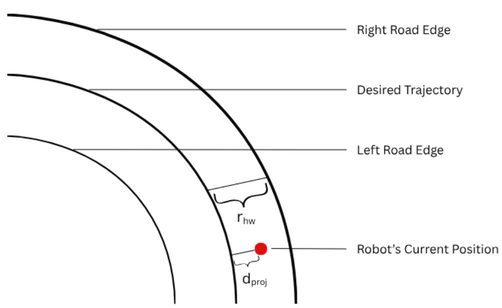


Fig. 1. Figure demonstrating the lane-keeping scenario for CBF design.

Motivated by the barrier-based safety formulations in [13], a reciprocal CBF penalty is introduced as $B(z) = \frac{k_p\, s(z)}{h_{\text{cbf}}(z)}$, where $k_p \in \mathbb{R}_{>0}$ is a constant gain.

To ensure lane-keeping safety, the wheeled mobile robot (WMR) must remain within a predefined lane. The lane is modeled as a tubular neighborhood around the desired trajectory, which serves as the *centerline*, with half-width r_{hw}, as illustrated in Fig. 1. The reciprocal CBF penalty $B(z)$ penalizes deviations of the WMR's position from this prescribed centerline.

Let $\zeta_r = [x_r\ \ y_r\ \ \theta_r]^\top$ denote the current state of the WMR, and let $p = [x_r\ y_r]^\top$ represent its planar position. The reference trajectory is given by $\zeta_d(t) = [x_d(t)\ y_d(t)\ \theta_d(t)]^\top$, with associated planar coordinates $p_d(t) = [x_d(t)\ \ y_d(t)]^\top$. Geometrically, the closest point on the reference trajectory to the current robot position is defined as the orthogonal projection of p onto the trajectory curve.

The corresponding shortest lateral distance between the WMR and the reference trajectory is expressed as $d_{proj} = \big|\, n^\top(\zeta_r - \zeta_d)\,\big|$, where n denotes the unit normal vector to the reference trajectory evaluated at the projection point, as depicted in Fig. 1. Using this distance measure, the lane-keeping CBF is defined by $h_{\text{cbf}}(z) = r_{hw} - d_{proj}$.

By construction, $h_{\text{cbf}}(z) > 0$ indicates that the WMR lies strictly within the lane, $h_{\text{cbf}}(z) = 0$ corresponds to the lane boundary, and $h_{\text{cbf}}(z) < 0$ signifies violation of the safety constraint. Consequently, the CBF satisfies Definition 1 and constitutes a valid CBF for the lane-keeping task.

To smoothly modulate the influence of the barrier penalty, the scheduling function is chosen as $s(z) = \frac{1}{2} + \frac{1}{2}\cos\Big(\frac{\pi\, h_{\text{cbf}}(z)}{r_{hw}}\Big)$. This choice ensures that $s(z) \to 0$ when the WMR is close to the centerline and $s(z) \to 1$ as the state approaches the boundary of the admissible lane, thereby activating the safety penalty primarily near $\partial\mathcal{C}_{aug}$.

V. Simulation Results

This section presents simulation results used to evaluate the effectiveness of the proposed control framework. The reference trajectory is chosen as a closed elliptical path defined by $\zeta_d(t) = [x_d(t),\, y_d(t),\, \theta_d(t)]^\top$, where $x_d(t) = r_1\cos(\omega_d t)$, $y_d(t) = r_2\sin(\omega_d t)$, and $\theta_d(t) = \tan^{-1}\big(\dot{y}_d(t)/\dot{x}_d(t)\big)$. The critic neural network employs $l_N = 23$ hidden-layer neurons, with weights initialized using a uniform distribution over $[0, 1]$. To ensure persistence of excitation, a zero-mean Gaussian probing signal bounded in $[0, 1]$ is injected into the regressor for the first $10\,\text{s}$ of the simulation. The lane is modeled as a road segment whose centerline coincides with the reference trajectory, with a prescribed half-width of $r_{hw} = 0.2\,\text{m}$. The barrier penalty gain is selected as $k_p = 3.5$, and the bounded barrier term $B_{\text{bound}}(z) = \frac{k_p s(z)}{h_{\text{cbf}}(z)+c}$ is implemented with $c = 1$ to improve numerical robustness. The remaining simulation parameters are summarized in Table I.

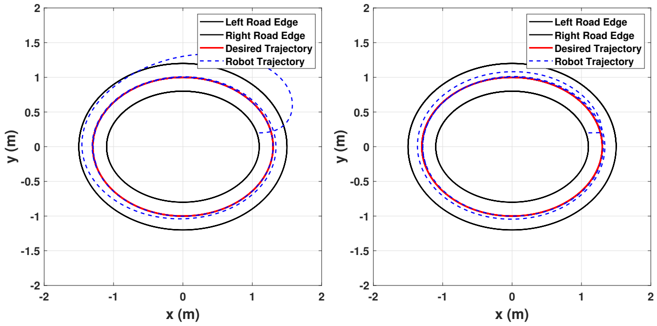
Two simulation scenarios are considered. In the first scenario, the controller is evaluated under actuator constraints only. In the second scenario, both actuator constraints and the proposed CBF-based safety mechanism are enforced. All simulations are performed over a $20\,\text{s}$ horizon, with the initial condition $\zeta_0 = [1.1,\ 0.2,\ 0]^\top$.

TABLE I
Parameter Values.

Parameter	Value	Parameter	Value
r_1	1.3 m	γ	0.001
r_2	1.0 m	$\bar{v}$	5 ms^{-1}
ω_d	0.8 rad	$\bar{\omega}$	5 rads^{-1}
α_1	0.004	Q	$I_{3\times3}$
α_2	0.08	R	$I_{2\times2}$
α	5.0	l_N	23

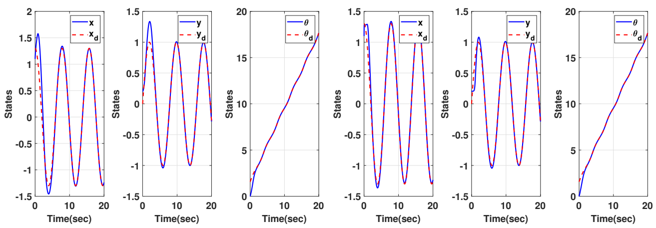
The trajectory tracking performance of the WMR is illustrated in Fig. 2(a)–(b). The desired trajectory is shown as the centerline, with the surrounding black curves indicating the road boundaries corresponding to a half-width of $0.2\,\text{m}$. A relatively large initial tracking error is observed due to the use of an approximate controller with randomly initialized neural network weights. As the critic weights converge toward their optimal values, the WMR trajectory approaches the desired path. However, when only actuator constraints are enforced, the large initial error causes the WMR to violate the lane boundaries, as seen in Fig. 2(a). In contrast, when both actuator constraints and the proposed CBF are applied, the WMR remains within the lane, as seen in Fig. 2(b), whenever the initial condition lies inside the safe set, demonstrating the effectiveness of the safety-enforced control strategy.

The time evolution of the WMR states and the corresponding reference states is depicted in Fig. 3(a)–(b). In both scenarios, the system states ultimately follow the desired trajectory, with the tracking error converging toward zero. However, when only actuator constraints are imposed, the transient tracking error remains relatively large, as shown in Fig. 3(a). In contrast, enforcing the CBF in conjunction with actuator constraints significantly reduces the tracking error



(a) (b)

Fig. 2. Trajectory of the WMR (a) under actuator constraints and (b) both actuator and CBF.

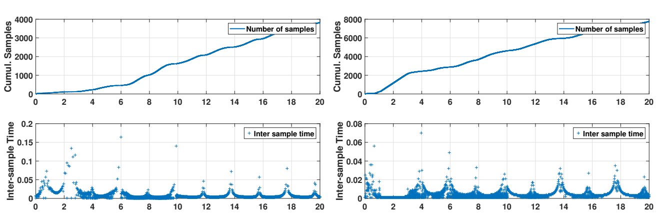


(a) (b)

Fig. 3. Time evolution of states (a) under actuator constraint and (b) both actuator and CBF constraints.

throughout the learning process, as illustrated in Fig. 3(b). This improvement arises because the CBF actively shapes the learning dynamics, preventing large deviations and maintaining the robot within the lane even during the initial learning.

The event-triggered tracking performance of the WMR under actuator constraints is illustrated in Fig. 4(a)–(b). The cumulative number of control executions and the corresponding inter-sampling intervals are shown in the top and bottom panels, respectively. When only actuator constraints are enforced, the controller is updated 3826 times over the $20\,\mathrm{s}$ simulation horizon, compared to 20,000 periodic samples, resulting in an 80.9% reduction in feedback communication and controller execution. When the CBF-based safety mechanism is additionally activated, the number of control executions increases to 7763, corresponding to a 61.2% reduction in feedback communication. The reduced communication savings in the safety-enforced case are expected, as maintaining lane-keeping safety requires more frequent control updates.



(a) (b)

Fig. 4. Event-triggered system performance under actuator constraints. (a) Cumulative number of samples. (b) Inter-sampling instants.

VI. Conclusion

This paper presents an event-triggered ADP/RL framework for trajectory tracking of wheeled mobile robots with actuator limits, with and without safety enforcement. Without safety constraints, the controller achieves accurate tracking with significantly fewer feedback updates, improving communication efficiency. To ensure lane-keeping safety, a CBF is embedded into the event-triggered differential game and the Bellman residual, coupled with a state-dependent triggering rule that balances tracking, safety, and control efficiency. Simulations show the robot stays within lane boundaries with safety enforced, maintains good tracking, and reduces sensing and control executions by up to 61.2% compared to periodic feedback, demonstrating that CBF-based real-time safety certification can be effectively integrated with event-triggered ADP/RL.

Acknowledgments

The authors acknowledge that generative AI tools are used to correct grammar and polish their own text.

References

[1] S. Al Mahmud, A. Kamarulariffin, A. M. Ibrahim, and A. J. H. Mohideen, "Advancements and challenges in mobile robot navigation: A comprehensive review of algorithms and potential for self-learning approaches," *Journal of Intelligent & Robotic Systems*, vol. 110, no. 3, p. 120, 2024.

[2] J. E. Solanes and L. Gracia, "Mobile robots: Trajectory analysis, positioning and control," *Applied Sciences*, vol. 15, no. 1, 2025.

[3] H. Zhang, D. Liu, Y. Luo, and D. Wang, *Adaptive dynamic programming for control: algorithms and stability*. Springer Science & Business Media, 2012.

[4] R. S. Sutton, A. G. Barto, *et al.*, *Reinforcement learning: An introduction*, vol. 1. MIT press Cambridge, 1998.

[5] F. L. Lewis and D. Liu, *Reinforcement learning and approximate dynamic programming for feedback control*. John Wiley & Sons, 2013.

[6] R. Kamalapurkar, H. Dinh, S. Bhasin, and W. E. Dixon, "Approximate optimal trajectory tracking for continuous-time nonlinear systems," *Automatica*, vol. 51, pp. 40–48, 2015.

[7] H. Modares and F. L. Lewis, "Optimal tracking control of nonlinear partially-unknown constrained-input systems using integral reinforcement learning," *Automatica*, vol. 50, no. 7, pp. 1780–1792, 2014.

[8] X. Yang and H. He, "Self-learning robust optimal control for continuous-time nonlinear systems with mismatched disturbances," *Neural Networks*, vol. 99, pp. 19–30, 2018.

[9] P. Tabuada, "Event-triggered real-time scheduling of stabilizing control tasks," *IEEE Transactions on Automatic control*, vol. 52, no. 9, pp. 1680–1685, 2007.

[10] A. Sahoo and V. Narayanan, "Event-based near optimal sampling and tracking control of nonlinear systems," in *2018 IEEE Conference on Decision and Control (CDC)*, pp. 55–60, IEEE, 2018.

[11] A. Sahoo and V. Narayanan, "Optimization of sampling intervals for tracking control of nonlinear systems: A game theoretic approach," *Neural Networks*, vol. 114, pp. 78–90, 2019.

[12] A. D. Ames, J. W. Grizzle, and P. Tabuada, "Control barrier function based quadratic programs with application to adaptive cruise control," in *53rd IEEE conference on decision and control*, pp. 6271–6278, IEEE, 2014.

[13] M. H. Cohen and C. Belta, "Approximate optimal control for safety-critical systems with control barrier functions," in *2020 59th IEEE conference on decision and control (CDC)*, pp. 2062–2067, IEEE, 2020.

[14] Z. Marvi and B. Kiumarsi, "Safe reinforcement learning: A control barrier function optimization approach," *International Journal of Robust and Nonlinear Control*, vol. 31, no. 6, pp. 1923–1940, 2021.

[15] A. D. Ames, S. Coogan, M. Egerstedt, G. Notomista, K. Sreenath, and P. Tabuada, "Control barrier functions: Theory and applications," in *2019 18th European control conference (ECC)*, pp. 3420–3431, Ieee, 2019.

[16] P. Deptula, H.-Y. Chen, R. A. Licitra, J. A. Rosenfeld, and W. E. Dixon, "Approximate optimal motion planning to avoid unknown moving avoidance regions," *IEEE Transactions on Robotics*, vol. 36, no. 2, pp. 414–430, 2019.

[17] T. Başar and P. Bernhard, *H-infinity optimal control and related minimax design problems: A dynamic game approach*. Springer Science & Business Media, 2008.