Journal of Robotics and Control (JRC)
Volume 5, Issue 5, 2024
ISSN: 2715-5072, DOI: 10.18196/jrc.v5i5.22148

Reinforcement Learning-Based Trajectory Control for Mecanum Robot with Mass Eccentricity Considerations

Minh Dong Nguyen [1], Manh Tien Ngo [2*], Hiep Do Quang [3] , Nam Dao Phuong [4]
[1,2] Institute of Physics, Vietnam Academy of Science and Technology, Hanoi, Vietnam
[3] Faculty of Electrical, University of Economics Technology for Industries, Hanoi, Vietnam
[4] School of Electrical and Electronic Engineering, Hanoi University of Science and Technology, Hanoi, Vietnam
Email: [1] nmdong20@gmail.com, [2] tieniop@gmail.com,
[3] dqhiep@uneti.edu.vn, [4] nam.daophuong@hust.edu.vn
*Corresponding Author

***Abstract*—This article presents a robust optimal tracking control approach for a Four Mecanum Wheeled Robot (FMWR) using an online actor-critic reinforcement learning (RL) algorithm to address the challenge of precise trajectory tracking problem in the presence of mass eccentricity and friction uncertainty. In order to handle these obstacles, a detailed dynamics model is derived using Lagrange's equation, and the Hamilton–Jacobi–Bellman (HJB) equation is solved by iteration algorithm with policy evaluation and improvement. The training laws of optimal control law and value function are proposed after minimizing the modified Hamiltonian function. Moreover, to handle the time-varying property of tracking error model, a transform is given with the addition of time derivative term. Simulation Studies demonstrate the approach's effectiveness, significantly improving trajectory tracking accuracy and robustness against disturbances. This research contributes to mobile robotics by enhancing control precision and reliability in dynamic environments.**

***Keywords*—Optimal Control, Reinforcement Learning, Trajectory Tracking, Friction Uncertainty**

I. INTRODUCTION

In the field of robotics, the control of mobile robots is becoming increasingly important for both industrial automation and research activities. High precision, real-time performance, and robustness to external disturbances are essential requirements for the design of control systems for these robots. Among various mobile robot designs, the Mecanum wheel stands out because it can move in any direction while supporting high loads [1]. This study addresses a specific problem: developing a robust tracking control system for a Four Mecanum Wheeled Robot (FMWR) that accounts for mass eccentricity using an online actor-critic reinforcement learning algorithm.

Extensive research has been conducted on both system modeling and tracking control algorithms to ensure that robots can accurately and swiftly follow desired trajectories. The dynamic model of a mobile robot equipped with four Mecanum wheels is detailed in [2], [3], while its kinematic model and the relationships between kinematics and platform dynamics are presented in [4], [5]. A comprehensive dynamics model that accounts for mass eccentricity and friction uncertainty is provided in [6]. Additionally, a mathematical model integrating both dynamic and kinematic aspects is analyzed and evaluated in [7].

Several studies have focused on control design for mobile robots. Enhancements in control performance for forward and inverse motion have been achieved using an optimal PID controller based on a differential evolution algorithm [7], as well as a PID controller designed for more accurate point tracking of the Mecanum robot [8]. Additionally, a nonlinear Backstepping control law, validated for its efficiency, has been applied based on the dynamic and kinematic model [9], [10]. A robot dynamics-based sliding mode controller for trajectory tracking is presented in [11], [52]–[54], [56].

Reinforcement learning (RL) offers distinct advantages by not requiring precise system models, as demonstrated in [12]. For instance, a reinforcement learning controller in [13] achieves optimal morality for wheeled mobile robots (WMR), with control PD parameters determined within the action space. However, continuous control remains a challenge due to the decomposition of the action space into multiple subspaces.

Studies such as [14], [57]–[60] have designed reinforcement learning algorithms for systemless MIMO routing uncertainty with prioritized control of output signals. The steering control algorithm for WMR, based on the proposed RL algorithm, was examined considering system time delay and slip effects [15]–[21], [63]–[70]. The RL algorithm for solving the optimal control solution for the Hamilton–Jacobi–Bellman (HJB) equation facilitates policy evaluation and improvement [25]–[28], [71]–[80]. Optimal regulation and tracking of nonlinear systems have been addressed using an actor-critic architecture incorporating both actor and critic neural networks (NNs) [20], [49]–[51].

However, these methods typically solve the optimal control problem offline, requiring multiple learning trials before the controller is trained, without ensuring system stability during the learning phase [22]–[24].

An intelligent actor-critic learning control method augmented by a fuzzy broad learning system with output recurrent feedback has been proposed for obstacle-avoiding trajectory tracking of heterogeneous Mecanum-wheeled omnidirectional mobile robots, accounting for unmodeled errors and varying parameters in response to the working environment [35]–[48]. Furthermore, the optimal tracking control for a three-omni-wheeled mobile robot under external disturbances, using an online actor-critic synchronous learning algorithm, was investigated in [34]. However, this study did not consider the impact of the robot's center of gravity deviation and changes in mass.

Despite these advancements, a significant gap persists in the literature concerning the integration of mass eccentricity into the control scheme for FMWR. Most existing studies either neglect mass eccentricity or address it inadequately, resulting in suboptimal control performance. This research bridges this gap by developing a robust tracking control system that incorporates mass eccentricity through an online actor-critic reinforcement learning algorithm. Section II presents the kinematic and dynamic modeling of the FMWR. Section III details the design of the controller using the online actor-critic algorithm. Section IV provides the results of simulation experiments, followed by a discussion in Section V. Finally, Section VI concludes the paper, summarizing the key findings and contributions.

In this paper, we present a novel approach to developing a robust tracking control scheme for a Fast Mobile Wheeled Robot (FMWR), utilizing an online actor-critic algorithm to specifically tackle the challenge posed by mass eccentricity. Our primary aim is to ensure precise adherence to a predefined trajectory within the shortest possible timeframe and with minimal deviation. This research significantly advances the field of mobile robot control by seamlessly integrating mass eccentricity into the control scheme, effectively addressing a notable gap in the existing literature.

Our experimental results underscore the efficacy of the proposed control system, demonstrating its ability to guide the robot along the desired trajectory with remarkable accuracy and efficiency, thus bolstering the overall performance and reliability of FMWR in practical applications. Section II elucidates the kinematic and dynamic modeling of the FMWR, laying the foundation for our subsequent controller design discussed in Section III, which leverages the online actor-critic algorithm. Section IV presents the outcomes of simulation experiments, providing valuable insights discussed further in Section V. Finally, Section VI concludes the paper, summarizing the key findings and contributions made throughout our research endeavor.

II. The Four Mecanum Wheel Mobile Robot

A. Kinematic of FMWR

The mecanum wheel is designed with passive rollers placed around the main wheel, ensuring the robot can move in all directions. The FMWR model uses four active wheels guided independently by four special motors. Robot moving, the rollers convert part of the translation force into horizontal sliding force, helping the robot move flexibly in many different directions. Let's consider the wheel frame in the robot coordinate [5] being shown in Fig. 1.

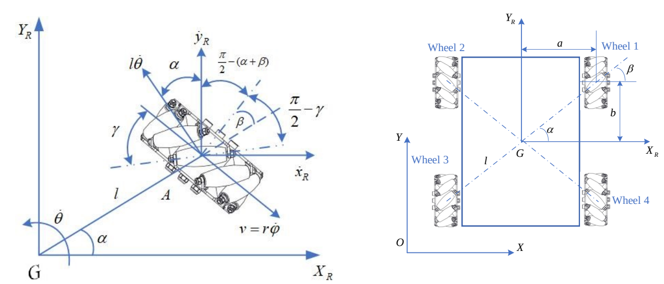


Fig. 1. The four mecanum wheel robot in the world frame

In the case of no slipping occurs along the contact of roller's axis, the same velocity can also be computed from the wheel's rotation speed as:

$$[\sin(\alpha+\beta+\gamma) - \cos(\alpha+\beta+\gamma) - l\cos(\beta+\gamma)]R(\theta)\dot{q} = r\dot{\varphi}\cos\gamma \tag{1}$$

where α is angle of GA with horizontal axis, β is the angle between the vector GA and the main wheel axis, and $\gamma = 45^0$ is the deflection angle of the passive roller. The geometric center G to the wheel center A is l, and r is the main wheel's radius. The orientation of the inertia frame with respect to the robot frame can be expressed as matrix

$$R(\theta) = \begin{bmatrix} \cos\theta & \sin\theta & 0 \\ -\sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{bmatrix} \tag{2}$$

where θ is the angle between axes X_R and X. In the next section, We will construct the kinematic and dynamic equations for FMWR.

In the considered FMWR, assuming that ach mecanum wheel is controlled respectively by an independent motor as arranged in Fig. 1 with wheel has equal radius and distances. The parameters α, β and γ for each wheel are presented in Table I. and substituted into equation (1)

TABLE I. The Parameters of Each Mecanum Wheel

Wheels	α_i	β_i	γ_i
1	$\tan^{-1}(b/a)$	$-\tan^{-1}(b/a)$	$(\pi/2+\pi/4)$
2	$\pi-\tan^{-1}(b/a)$	$\tan^{-1}(b/a)$	$-(\pi/2+\pi/4)$
3	$\pi+\tan^{-1}(b/a)$	$-\tan^{-1}(b/a)$	$(\pi/2+\pi/4)$
4	$2\pi+\tan^{-1}(b/a)$	$\tan^{-1}(b/a)$	$-(\pi/2+\pi/4)$

After computation, the Jacobian matrix is identified as follows:

$$J=\begin{bmatrix}\sqrt{2}/2 & \sqrt{2}/2 & l\sin(\pi/4-\alpha)\\ \sqrt{2}/2 & -\sqrt{2}/2 & l\sin(\pi/4-\alpha)\\ -\sqrt{2}/2 & -\sqrt{2}/2 & l\sin(\pi/4-\alpha)\\ -\sqrt{2}/2 & \sqrt{2}/2 & l\sin(\pi/4-\alpha)\end{bmatrix}\begin{bmatrix}\cos\theta & \sin\theta & 0\\ -\sin\theta & \cos\theta & 0\\ 0 & 0 & 1\end{bmatrix} \tag{3}$$

The forward kinematic equation of the FMWR can be obtained as follows:

$$\begin{bmatrix}\dot{x}\\ \dot{y}\\ \dot{\theta}\end{bmatrix}=-(\sqrt{2}/2)rJ^{+}\begin{bmatrix}\dot{\varphi}_1\\ \dot{\varphi}_2\\ \dot{\varphi}_3\\ \dot{\varphi}_4\end{bmatrix} \tag{4}$$

where $J^{+}=\left(J^{T}J\right)^{-1}J^{T}$ is the pseudoinverse of J. The velocities of wheels and angular velocities are determined by

$$\begin{bmatrix}\dot{x}_R\\ \dot{y}_R\\ \dot{\theta}\end{bmatrix}=\begin{bmatrix}\cos\theta & \sin\theta & 0\\ -\sin\theta & \cos\theta & 0\\ 0 & 0 & 1\end{bmatrix}\begin{bmatrix}\dot{x}\\ \dot{y}\\ \dot{\theta}\end{bmatrix} \tag{5}$$

B. Dynamic of FMWR

Consider the FMWR to be show in Fig. 2, where G, G' are the geometric and mass center, respectively [5].

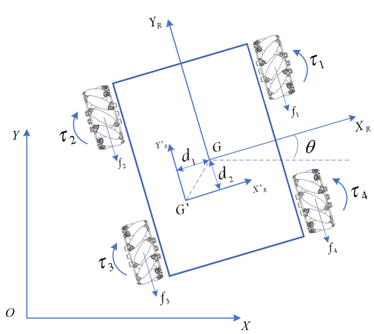


Fig. 2. Schematic of the FMWR

The total kinetic energy E of the mobile robot including those of the platform and four Mecanum wheels can be computed as below:

$$E=\frac{1}{2}\left[m_b v_{G'}^{T}v_{G'}+I_b\dot{\theta}^2+\sum_{i=1}^{4}m_{wi}(r\dot{\varphi}_i)^2+\sum_{i=1}^{4}I_i\dot{\varphi}_i^2\right] \tag{6}$$

where m_b is the mass of the platform, and m_{wi} is the mass of the ith wheel; I_b is the moment of inertia of the platform, and I_i is the moment inertia of the ith wheel about its main axis.

The system moves on the ground, the gravitational potential energy of the system $T=0$, The sum of the kinetic energies of the body and wheels is: $L=E+T$

$$L=\frac{1}{2}m_b v_{G'}^{T}v_{G'}+\frac{1}{2}I_b\dot{\theta}^2+\frac{1}{2}\sum_{i=1}^{4}m_{wi}(r\dot{\varphi}_i)^2+\frac{1}{2}\sum_{i=1}^{4}I_i\dot{\varphi}_i^2 \tag{7}$$

Using the Lagrange's equations [30]:

$$\frac{d}{dt}\left(\frac{\partial L}{\partial \dot{q}_i}-\frac{\partial L}{\partial q_i}\right)=Q_i \tag{8}$$

where $Q_i,(i=1,2,3)$ is torque generalized, $q_i=[x,y,\theta]^T$ is generalized coordinate vector, f_i is the contact friction force of the wheel with the floor, can be derived as follows [29]:
The dynamic equations for FMWR:

$$M(q)\ddot{q}+C(q,\dot{q})\dot{q}+B\xi=B\tau \tag{9}$$

where

$$\tau=\begin{bmatrix}\tau_1 & \tau_2 & \tau_3 & \tau_4\end{bmatrix}^T, f=\begin{bmatrix}f_1 & f_2 & f_3 & f_4\end{bmatrix}^T,$$

$$\xi=r.f.\text{diag}\begin{bmatrix}\text{sgn}(\dot{\varphi}_1) & \text{sgn}(\dot{\varphi}_2) & \text{sgn}(\dot{\varphi}_3) & \text{sgn}(\dot{\varphi}_4)\end{bmatrix}$$

$$M=[m_{ij}]_{3\times 3}, m_11=m_b+4\left(m_w+\tfrac{I}{r^2}\right);$$

$$m_{22}=m_b+4\left(m_w+\tfrac{I}{r^2}\right); m_{12}=m_{21}=0;$$

$$m_{13}=m_{31}=m_b\left(d_1\sin\theta+d_2\cos\theta\right);$$

$$m_{23}=m_{32}=m_b\left(-d_1\cos\theta+d_2\sin\theta\right)$$

$$m_{33}=m_b\left(d_1^2+d_2^2\right)+I_b+8\left(m_w+\tfrac{I}{r^2}\right)l^2\sin^2(\pi/4-\alpha)$$

$$C=\begin{bmatrix}0 & 0 & 0\\ 0 & 0 & 0\\ m_b\dot{\theta}\left(d_1\cos\theta-d_2\sin\theta\right) & m_b\dot{\theta}\left(d_1\sin\theta+d_2\cos\theta\right) & 0\end{bmatrix}^T$$

$$B=\frac{1}{r}\begin{bmatrix}-(\cos\theta-\sin\theta) & -(\sin\theta+\cos\theta) & -\sqrt{2}l\sin(\pi/4-\alpha)\\ -(\cos\theta+\sin\theta) & -(\sin\theta-\cos\theta) & -\sqrt{2}l\sin(\pi/4-\alpha)\\ \cos\theta-\sin\theta & \sin\theta+\cos\theta & -\sqrt{2}l\sin(\pi/4-\alpha)\\ \cos\theta+\sin\theta & \sin\theta-\cos\theta & -\sqrt{2}l\sin(\pi/4-\alpha)\end{bmatrix}^T \tag{10}$$

According to equation (5), it obtains the kinematic equation of FMWR as follows:

$$v_q=R(\theta)\dot{q} \tag{11}$$

Taking the time derivative of equation (11), it implies that:

$$\dot{v}_q=R(\theta)\ddot{q}+\dot{\theta}\dot{R}(\theta)\dot{q} \tag{12}$$

Equation (9) implies the following equality:

$$\ddot{q}=M^{-1}B^T\tau-M^{-1}B^T\xi-M^{-1}C\dot{q} \tag{13}$$

Substituting (13) into equation (12), it obtains that:

$$\dot{v}_q=R(\theta)M^{-1}1B^T\tau-R(\theta)M^{-1}1B^T\xi-R(\theta)R(\theta)^{-1}1$$
$$M^{-1}1Cv_q+\dot{\theta}\dot{R}(\theta)R(\theta)^{-1}1v_q \tag{14}$$

Substituting, can obtain the following dynamic equation of FMWR as below:

$$\dot{v}_q=\bar{f}(q)v_q+\bar{g}(q)\tau+\bar{g}(q)d(q,v_q) \tag{15}$$

where

$$\bar{f}(q)=\left[\dot{\theta}\dot{R}(\theta)R(\theta)^{-1}-M^{-1}C\right]$$

$$\bar{g}(q)=R(\theta)M^{-1}B^T$$

$$fd(q,v_q)=-R(\theta)M^{-1}B^T\xi$$

The problem of controlling a robot to follow a set trajectory $q_r(t)$ in a short time with a low cost function. The dynamic model (15) is rewritten as:

$$\dot{x} = f(x) + g(x)\tau + g(x)d\left(q, v_q\right) \tag{16}$$

where

$$\begin{array}{l} x = (q^T, v_q^T)^T, \\ f(x) = (v_q^T, (f(q)v_q)^T)^T, \\ g(x) = (\mathbf{0}_{3\times3}, g(q)^T)^T \end{array} \tag{17}$$

To develop RL technique for time-invariant systems, we reference system expressed by

$$\dot{x}_r = h_r(x_r) \tag{18}$$

We defining the tracking error as $e = x - x_r$

$$\begin{array}{l} \dot{e} = f(x) + g(x)\tau - h_r(x_r) \\ \quad = f(x) - h_r(x_r) + g(x)\tau + g(x)u \end{array} \tag{19}$$

where $u = \tau - \tau_r$,

$$\tau_r(x_r) = g^+(x_r)(h_r(x_r) - f(x_r)) \tag{20}$$

By employing a new concatenated state as $z = (e^T, x_r^T)^T$, then the dynamics of z is formulated by an invariant-time system as

$$\dot{z} = F(z) + G(z)u \tag{21}$$

where

$$F(z) = \begin{pmatrix} f(x) - h_r(x_r) + g(x)\tau_r \\ h_r(x_r) \end{pmatrix}, G(z) = \begin{pmatrix} g(x) \\ \mathbf{0}_{6\times4} \end{pmatrix}$$

III. Controller Design on the Online Actor-Critic Algorithm

A. Controller design

To control the robot with equation (21) to follow the trajectory, we use the HJB function with the following the Hamiltonian function [34] is given by

$$\begin{array}{l} H(z, u, \bigtriangledown V) = \lambda_M^2(z) + z^T\bar{Q}z + u^T(z)Ru(z) \\ -\gamma V(z) + \bigtriangledown V^T(z)(F(z) + G(z)u(z)) \end{array} \tag{22}$$

With the optimal function clearly defined as follows:

$$V^*(z) = \min_{u\in\pi(\Omega)} \int_t^\infty e^{-\gamma(s-t)}(\lambda_M^2(z) + z^T\bar{Q}z + u^TRu)ds \tag{23}$$

The optimal control signals is

$$u^*(z) = \arg\min_{u\in\pi(\Omega)}[H(z, u, \bigtriangledown V^*(z)] = -\frac{1}{2}R^{-1}G^T(z)\bigtriangledown V^*(z) \tag{24}$$

The control system follows the structure of Online Actor - Critic. The signal control from the Actor will be continuously updated and included in the system size when the signal control is optimal as shown in Fig. 3.

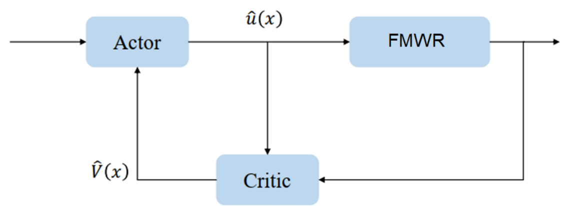


Fig. 3. Control diagram online AC

For optimal control, we use the online actor–critic algorithm to solve the HJB equation using a single-layer NN as

$$V^*(z) = W^T\phi(z) + \varepsilon_v(z) \tag{25}$$

where W is an ideal constant weight vector, the number of neurons is N, $\varepsilon_v(z)$ is the approximation error.

Based on formulas (25) and optimal control, the Actor NNs and Critic NNs approximation for the optimal policy are given by:

$$\hat{u}(z, \hat{W}_a) = -\frac{1}{2}R^{-1}G^T(z)\bigtriangledown\phi^T(z)\hat{W}_a \tag{26}$$

where $\hat{W}_c$ and $\hat{W}_a$ a are the estimates.

The signal control is determined as follows: (20), the tracking controller τ of FWMR can be obtained as

$$\tau = -\frac{1}{2}R^{-1}G^T(z)\bigtriangledown\phi^T(z)\hat{W}_a + g^+(x_r)(h_r(x_r) - f(x_r)) \tag{27}$$

Using the approximations $\hat{u}$ and $\hat{V}$:

$$\begin{array}{l} \hat{H}(z, \hat{W}_c, \hat{W}_a) = \lambda_M^2(z) + z^T\bar{Q}z + \frac{1}{4}\hat{W}_a^TD_1\hat{W}_a \\ -\gamma\hat{W}_c^T\phi + \hat{W}_c^T\bigtriangledown\phi(F - \frac{1}{2}GR^{-1}G^T\bigtriangledown\phi^T\hat{W}_a) \end{array} \tag{28}$$

Define the Bellman error as $\delta = \hat{H} - H^*$,

$$\begin{array}{l} \delta(z, \hat{W}_c, \hat{W}_a) = \lambda_M^2(z) + z^T\bar{Q}z + \frac{1}{4}\hat{W}_a^TD_1\hat{W}_a \\ +\hat{W}_c^T(\bigtriangledown\phi(F + G\hat{u}) - \gamma\phi) \end{array} \tag{29}$$

Training critic NN to minimize the integral error, The tuning law for the actor NN is developed as

$$\dot{\hat{W}}_a = -\eta_{a1}(\hat{W}_a - \hat{W}_c) - \eta_{a2}\hat{W}_a + \frac{\eta_a}{4}D_1\hat{W}_a\frac{\bar{\sigma}^T}{m_\sigma}\hat{W}_c \tag{30}$$

where $\eta_{a1} > 0$, $\eta_{a2} > 0$ and $\eta_a > 0$ are tuning parameters

B. Stability Analysis

Considering the candidate Lyapunov function for the stability of the system

$$V(t) = V^*(t) + \frac{1}{2\eta_c}\tilde{W}_c^T(t)\Gamma^{-1}(t)\tilde{W}_c(t) + \frac{1}{2\eta_a}\tilde{W}_a^T(t)\tilde{W}_a(t) \tag{31}$$

In which the optimal cost function has the following derivative:

$$\dot{V}^*t = W^T \bigtriangledown \phi^F - \frac{1}{2} W^T D_1 \hat{W}_a + \bigtriangledown \varepsilon_v^T (F \\ - \frac{1}{2} G R^- 1 G^T \bigtriangledown \phi^T \hat{W}_a) \tag{32}$$

From the HJB equation (III-A), we can rewrite the optimal cost function determined can be rewritten as:

$$\begin{aligned} \dot{V}^*(t) = & -\lambda_M^2(z) - z^T \bar{Q} z + \gamma W^T \phi + \frac{1}{4} W^T D_1 W \\ & + \frac{1}{2} \tilde{W}_a^T D_1 W + \frac{1}{2} \tilde{W}_a^T \bigtriangledown \phi D_2 \bigtriangledown \varepsilon_v \\ & + \frac{1}{4} \bigtriangledown \varepsilon_v^T D_2 \bigtriangledown \varepsilon_v + \gamma \varepsilon_v(z) \end{aligned} \tag{33}$$

into the derivative of $V(t)$ yields

$$\begin{aligned} \dot{V}(t) = & -\lambda_M^2(z) - e^T Q e - \frac{1}{4} W^T D_1 W - \frac{\tilde{W}_c^T \bar{\sigma}}{m_\sigma} \varepsilon_H \\ & -\tilde{W}_c^T A_1 \tilde{W}_c - \tilde{W}_a^T A_2 \tilde{W}_a + \tilde{W}_a^T A_3 \tilde{W}_c + \tilde{W}_a^T B_1 + B_2 \end{aligned} \tag{34}$$

where

$$A_1 = \frac{1}{2} \bar{\sigma} \bar{\sigma}^T + \frac{\beta}{2} \Gamma^{-1}, A_2 = \frac{(\eta_{a1} + \eta_{a2})}{\eta_a} I - \frac{1}{4} D_1 \frac{\bar{\sigma}^T}{m_\sigma} W$$

$$A_3 = \frac{1}{4} D_1 W \frac{\bar{\sigma}^T}{m_\sigma} + \frac{\eta_{a1}}{\eta_a} I$$

$$B_1 = \frac{1}{2} D_1 W + \frac{1}{2} \bigtriangledown \phi D_2 \bigtriangledown \varepsilon_v - \frac{1}{4} D_1 W \frac{\bar{\sigma}^T}{m_\sigma} W + \frac{\eta_{a2}}{\eta_a} W$$

$$B_2 = \gamma W^T \phi + \frac{1}{4} \bigtriangledown \varepsilon_v^T D_2 \bigtriangledown \varepsilon_v + \gamma \varepsilon_v(z)$$

The boundness ofε_v and $\bigtriangledown \varepsilon_v$, positive constants $\kappa_1, \kappa_2, \kappa_3$ such that $||B_1|| \leq \kappa_1, ||B_2|| \leq \kappa_2, ||(1/4) D_1 (\bar{\sigma}^T / m_\sigma) W|| \leq \kappa_3$. Using the Young's inequality, (34) becomes

$$\begin{aligned} \dot{V}(t) \leq & -\underline{q} ||e||^2 - \frac{\beta}{4\alpha_2} ||\tilde{W}_c||^2 - \frac{\eta_{a1} + \eta_{a2}}{2\eta_a} ||\tilde{W}_a||^2 \\ & - \left(\frac{\beta}{4\alpha_2} - \frac{1}{2\epsilon} \left(\frac{\eta_a 1}{\eta_a} + \kappa_3 \right) \right) ||\tilde{W}_c||^2 \\ & - \left(\frac{\eta_{a1} + \eta_{a2}}{2\eta_a} - \kappa_3 - \frac{\epsilon}{2} \left(\frac{\eta_{a1}}{\eta_a} + \kappa_3 \right) \right) \\ & \times ||\tilde{W}_a||^2 + \frac{\varepsilon_h}{\sqrt{\upsilon \alpha_1}} ||\tilde{W}_c|| + \kappa_1 ||\tilde{W}_a|| + \kappa_2 \end{aligned} \tag{35}$$

where $\underline{q} = \lambda_{min}(Q)$. Choose $(\beta / 4\alpha_2) \geq (1/2\epsilon)((\eta_{a1}/\eta_a) + \kappa_3)$ and $((\eta_{a1} + \eta_{a2}/2\eta_a) \geq \kappa_3 + (\epsilon/2)((\eta_{a1}/\eta_a) + \kappa_3)$ (35) becomes

$$\dot{V}(t) \leq -\bar{\omega}_1 ||\zeta||^2 + \bar{\omega}_2 ||\zeta|| + \kappa_2$$

Applying (23) and (26) gives

$$u^* - \hat{u} = -\frac{1}{2} R^{-1} G^T (\bigtriangledown \phi^T(z) \tilde{W}_a + \bigtriangledown \varepsilon_v)$$

According one can get

$$||u^* - \hat{u}|| \leq \frac{1}{2\lambda_{min}(R)} b_g (b_{\phi z} b_\zeta + b_{\varepsilon z}) = b_u \tag{36}$$

where b_u is a positive constant Therefore, the tracking error dynamics of the FMWR (21) is eventually bounded uniformly, which further shows that the state $q(t)$ can track the desired trajectory $qr(t)$.

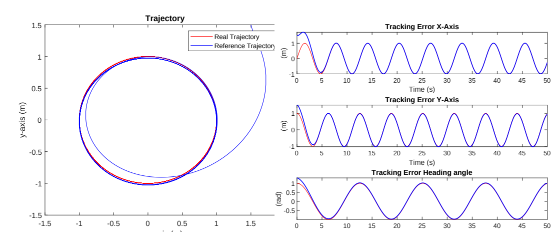
IV. Simulation and Results

In this section, we present a simulation for the Four Mecanum robot model based on MATLAB Simulink environment. The parameters of FMWR: m_b = 12 kg, m_w=0.313kg, a=0.2 m, b=0.3 m, r = 0.0508 m, I = 0.5 $kg.m^2$, I_w = 4.0378 x 10^{-4} $kg.m^2$, g=9.8 m/s^2 And the controller parameters are chosen as

$$x(0) = \begin{pmatrix} 1.5 & 1.5 & 1.4 & 0 & 0 & 0 \end{pmatrix}^T$$
$$\hat{W}_c(0) = 600 \times \mathbf{1}_{51x1}, \hat{W}_a(0) = 6 \times \mathbf{1}_{51\times1}, \Gamma(0) = 10 \times \mathbf{I}_{51}.$$

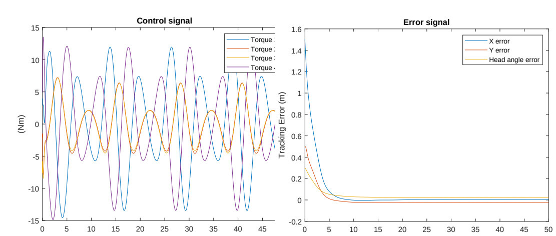
we choose $Q = \mathbf{I}_6$ and $R = \mathbf{I}_3$ the control gains are $\eta_c = 2, \eta_{a1} = 2, \eta_{a2} = 0.001, \eta_a = 0.001, \beta = 0.001, \gamma = 0.0001, \nu = 0.002$. The reference trajectory $q_r = [\sin(t) \cos(t) \cos(0.5t)]^T$.

Simulation results are shown in Figs. 4-6. The system state continues to approach reference trajectory at 5s. We can see simulation results, the neural network weights W_a and W_c converge. The robot's real trajectory also tends to follow the set trajectory at 5s. The corresponding control torques of the four Mecanum wheels are shown in Fig. 5(b) maximum is τ_i = [11.6 6.15 6.15 11.6] Nm, In Fig. 6, the first ten critic and actor weights finally converge to W_a=(-226.6 560.3 578.7 973.9 328.5 81.9 666 501.1 422.9 647.3) and W_c= (-226.5 560.1 578.5 973.6 328.4 81.8 665.8 501 422.8 647) It can be seen from Fig. 5(b) The robot's moving trajectory followed the trajectory set with the reinforcement learning algorithm, deviating from the trajectory set at 16 seconds is within the error range [0.00262 -0.0274 0.0218].



(a) (b)

Fig. 4. (a) Tracking trajectory. (b) Tracking performance of x, y, and θ



(a) (b)

Fig. 5. (a) Control signal. (b) the tracking error

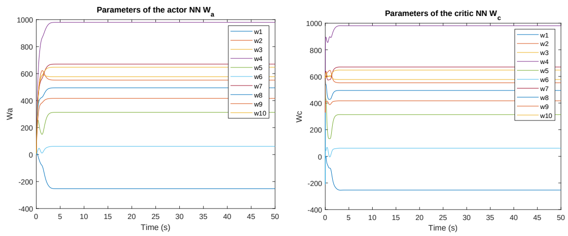
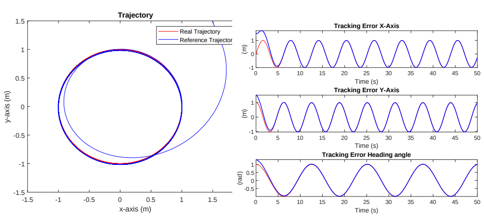


Fig. 6. Weights of the critic and the actor networks

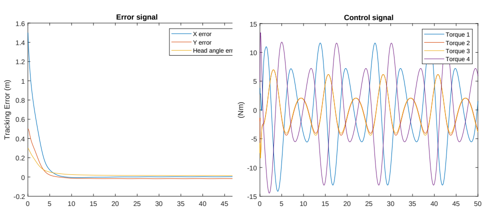
In the simulation, we consider the platform having eccentricity with $d_1 = d_2 = 0.02m$, and the mass has variation $\Delta m_b = 3kg$. Second simulation results are shown in Figs. 7–9.

From Fig. 8(a) we know that the undesired displacement along the x-axes and y-axes are can be kept within 0.0041 0.0052 m and -0.00385 -0.00371 m, respectively, and the orientation θ error is within 0.0312 0.0321 rad. In Fig. 9, the first ten critic and actor weights finally converge to W_a=(-223.1 558.6 577.8 967.1 343.4 81.4 665.7 506.7 423.1 646.6) and W_c=(-223.2 558.8 578 967.4 343.5 81.45 665.98 506.85 423.3 646.8).



(a) (b)

Fig. 7. (a) Tracking trajectory. (b) Tracking performance of x, y, and θ



(a) (b)

Fig. 8. (a) Control signal. (b) the tracking error $q(t)$ - $q_r(t)$

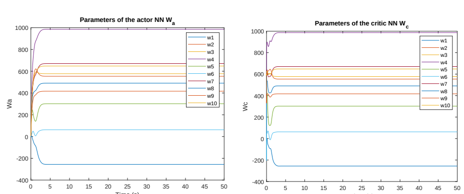


Fig. 9. Weights of the critic and the actor networks

The corresponding control torques shown in Fig. 8(b) maximum is τ_i = [12.06 6.45 6.45 12.06]^T Nm, the control torque is lower compared to the results from the [61], [62] study, this reflects the optimality of the designed method. Indicating that the algorithm is practical and energy saving in real-world applications. Control system for FMWR ensures minimal deviation from the preset trajectory, even in the presence of mass eccentricity and dynamic uncertainties.

V. Conclusion

This study proposes a robust optimal tracking control approach for a Four Mecanum Wheeled Robot (FMWR) in the presence of mass eccentricity and friction uncertainty using an online actor-critic synchronous RL algorithm. The completed mathematical model of mecanum robot is established through Lagrange's equation. Moreover, the modified Hamiltonian function is generated under the discount factor influence. By minimizing its square, the training laws of actor and critic NNs are proposed to obtain RL algorithm. Moreover, the tracking problem and the convergence of learning algorithm are guaranteed by Lyapunov stability theory. Simulation experiments demonstrated the efficacy of the proposed method, significantly improving trajectory tracking accuracy and robustness against disturbances. The results revealed that our control system ensures minimal deviation from the preset trajectory, even in the presence of mass eccentricity and dynamic uncertainties. Future investigation should involve the extension to the experiment systems.

Acknowledgment

- Funding: This work was funded by the scientific project of the Vietnam Academy of Science and Technology: "Research and development the intelligent humanoid robot IVASTBot in front desk application", code: CT0000.01/24-25
- Conflict of interest/Competing interests: The authors declare no potential conflict of interests.

References

[1] I. Zeidis and K. Zimmermann, "Dynamics of a four-wheeled mobile robot with Mecanum wheels," *ZAMM-Journal of Applied Mathematics and Mechanics/Zeitschrift für Angewandte Mathematik und Mechanik*, vol. 99, no. 12, 2019, doi: 10.1002/zamm.201900173.

[2] H. Taheri, B. Qiao, and N. Ghaeminazhad, "Kinematic model of a four mecanum wheeled mobile robot," *International journal of computer applications*, vol. 113, no. 3, pp. 6-9, 2015, doi: 10.5120/19804-1586.

[3] M. Abdelrahman *et al.*, "A description of the dynamics of a four-wheel Mecanum mobile system as a basis for a platform concept for special purpose vehicles for disabled persons," *Ilmenau Scientific Colloquium*, pp. 1–10, 2014.

[4] N. Tlale and M. de Villiers, "Kinematics and Dynamics Modelling of a Mecanum Wheeled Mobile Platform," *2008 15th International Conference on Mechatronics and Machine Vision in Practice, Auckland, New Zealand*, pp. 657-662, 2008, doi: 10.1109/MMVIP.2008.4749608.

[5] L. C. Lin, and H. Y. Shih, "Modeling and adaptive control of an omni-mecanum-wheeled robot," *Intelligent Control and Automation*, vol. 4, no. 2, 2013, doi: 10.4236/ica.2013.42021.

[6] F. Becker *et al.*, "An approach to the kinematics and dynamics of a four-wheel Mecanum vehicle," *Special Issue of Scientific Journal of Iftomm Problems of Mechanics*, pp. 27–37, 2014.

[7] P. Wu, K. Wang, J. Zhang, and Q. Zhang, “Optimal design for pid controller based on de algorithm in omnidirectional mobile robot,” In *MATEC Web of Conferences*, vol. 95, 2017, doi: 10.1051/matecconf/20179508014.

[8] C. S. Shijin and K. Udayakumar, “Speed control of wheeled mobile robots using PID with dynamic and kinematic modelling,” *2017 International Conference on Innovations in Information, Embedded and Communication Systems (ICIIECS)*, pp. 1-7, 2017, doi: 10.1109/ICIIECS.2017.8275962.

[9] A. Katpatal, A. Parwekar, and A. K. Jha, “Model-Based synchronized control of a robotic dual-arm manipulator,” In *Advances in Mechanical Engineering*, pp. 645-654, 2021.

[10] B. Dumitrascu, A. Filipescu, V. Minzu and A. Filipescu, “Backstepping control of wheeled mobile robots,” *15th International Conference on System Theory, Control and Computing*, pp. 1-6, 2011.

[11] R. Solea, A. Filipescu and U. Nunes, “Sliding-mode control for trajectory-tracking of a Wheeled Mobile Robot in presence of uncertainties,” *2009 7th Asian Control Conference*, pp. 1701-1706, 2009.

[12] J. Kober, J. A. Bagnell, and J. Peters, “Reinforcement learning in robotics: A survey,” *The International Journal of Robotics Research*, vol. 32, no. 11, pp. 1238-1274, 2023, doi: 10.1177/0278364913495721.

[13] L. Zuo, X. Xu, C. Liu, and Z. Huang, “A hierarchical reinforcement learning approach for optimal path tracking of wheeled mobile robots,” *Neural Computing and Applications*, vol. 23, pp. 1873–1883, 2013, doi: 10.1007/s00521-012-1243-4.

[14] Y. -J. Liu, L. Tang, S. Tong, C. L. P. Chen and D. -J. Li, “Reinforcement Learning Design-Based Adaptive Tracking Control With Less Learning Parameters for Nonlinear Discrete-Time MIMO Systems,” in *IEEE Transactions on Neural Networks and Learning Systems*, vol. 26, no. 1, pp. 165-176, 2015, doi: 10.1109/TNNLS.2014.2360724.

[15] S. Li, L. Ding, H. Gao, Y. -J. Liu, N. Li and Z. Deng, “Reinforcement Learning Neural Network-Based Adaptive Control for State and Input Time-Delayed Wheeled Mobile Robots,” in *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 50, no. 11, pp. 4171-4182, 2020, doi: 10.1109/TSMC.2018.2870724.

[16] K. G. Vamvoudakis and F. L. Lewis, “Online actor critic algorithm to solve the continuous-time infinite horizon optimal control problem,” *2009 International Joint Conference on Neural Networks*, pp. 3180-3187, 2009, doi: 10.1109/IJCNN.2009.5178586.

[17] R. Kamalapurkar *et al.*, “Approximate optimal trajectory tracking for continuous-time nonlinear systems,” *Automatica*, vol. 51, pp. 40-48, 2015, doi: 10.1016/j.automatica.2014.10.103.

[18] H. Zhang, Q. Wei and Y. Luo, “A Novel Infinite-Time Optimal Tracking Control Scheme for a Class of Discrete-Time Nonlinear Systems via the Greedy HDP Iteration Algorithm,” in *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, vol. 38, no. 4, pp. 937-942, 2008, doi: 10.1109/TSMCB.2008.920269.

[19] Z. Hendzel, “Robust neural networks control of omni-mecanum wheeled robot with hamilton-jacobi inequality,” *Journal of Theoretical and Applied Mechanics*, vol. 56, no. 4, pp. 1193-1204, 2018, doi: 10.15632/jtam-pl.56.4.1193.

[20] K. G. Vamvoudakis, and F. L. Lewis, “Online actor–critic algorithm to solve the continuous-time infinite horizon optimal control problem,” *Automatica*, vol. 46, no. 5, pp. 878-888, 2010, doi: 10.1016/j.automatica.2010.02.018.

[21] R. Kamalapurkar, P. Walters, J. Rosenfeld, and W. E. Dixon, “Model-based reinforcement learning for approximate optimal regulation,” *Reinforcement Learning for Optimal Feedback Control*, pp. 99-148, 2018.

[22] S. P. Singh, “Reinforcement learning with a hierarchy of abstract models,” In *Proceedings of the National Conference on Artificial Intelligence*, vol. 10, 1992.

[23] D. Mitrovic, S. Klanke, and S. Vijayakumar, “Adaptive optimal feedback control with learned internal dynamics models,” *From motor learning to interaction learning in robots*, pp. 65-84, 2010, doi: 10.1007/978-3-642-05181-4_4.

[24] P. Abbeel, M. Quigley, and A. Y. Ng, “Using inaccurate models in reinforcement learning,” In *Proceedings of the 23rd international conference on Machine learning*, pp. 1-8. 2006.

[25] P. Mehta and S. Meyn, “Q-learning and Pontryagin’s Minimum Principle,” *Proceedings of the 48h IEEE Conference on Decision and Control (CDC) held jointly with 2009 28th Chinese Control Conference*, pp. 3598-3605, 2009.

[26] K. G. Vamvoudakis, and F. L. Lewis, “Online actor–critic algorithm to solve the continuous-time infinite horizon optimal control problem,” *Automatica*, vol. 46, no. 5, pp. 878-888, 2010, doi: 10.1016/j.automatica.2010.02.018.

[27] H. Zhang, L. Cui, X. Zhang and Y. Luo, “Data-Driven Robust Approximate Optimal Tracking Control for Unknown General Nonlinear Systems Using Adaptive Dynamic Programming Method,” in *IEEE Transactions on Neural Networks*, vol. 22, no. 12, pp. 2226-2236, 2011, doi: 10.1109/TNN.2011.2168538.

[28] S. Bhasin *et al.*, “A novel actor–critic–identifier architecture for approximate optimal control of uncertain nonlinear systems,” *Automatica*, vol. 49, no. 1, pp. 82-92, 2013, doi: 10.1016/j.automatica.2012.09.019.

[29] S. Bhasin *et al.*, “A novel actor–critic–identifier architecture for approximate optimal control of uncertain nonlinear systems,” *Automatica*, vol. 49, no. 1, pp. 82-92, 2013, doi: 10.1016/j.automatica.2012.09.019.

[30] R. Ortega PhD *et al.*, “Euler-Lagrange systems,” *Passivity-based Control of Euler-Lagrange Systems*, pp. 15-37, 1998, doi: 10.1007/978-1-4471-3603-3_2.

[31] B. A. Finlayson, “The method of weighted residuals and variational principles,” *Classics in Applied Mathematics*, 2013, doi: 10.1137/1.9781611973242.

[32] R. Kamalapurkar *et al.*, “Approximate optimal trajectory tracking for continuous-time nonlinear systems,” *Automatica*, vol. 51, pp. 40-48, 2015, doi: 10.1016/j.automatica.2014.10.103.

[33] 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, doi: 10.1016/j.automatica.2014.05.011.

[34] D. Zhang, G. Wang and Z. Wu, “Reinforcement Learning-Based Tracking Control for a Three Mecanum Wheeled Mobile Robot,” in *IEEE Transactions on Neural Networks and Learning Systems*, vol. 35, no. 1, pp. 1445-1452, 2024, doi: 10.1109/TNNLS.2022.3185055.

[35] C. -C. Tsai, Y. -S. Chen and F. -C. Tai, “Intelligent adaptive distributed consensus formation control for uncertain networked heterogeneous swedish-wheeled omnidirectional multi-robots,” *2016 55th Annual Conference of the Society of Instrument and Control Engineers of Japan (SICE)*, pp. 154-159, 2016, doi: 10.1109/SICE.2016.7749219.

[36] C. -C. Tsai, Y. -S. Chen and F. -C. Tai, “Intelligent adaptive distributed consensus formation control for uncertain networked heterogeneous swedish-wheeled omnidirectional multi-robots,” *2016 55th Annual Conference of the Society of Instrument and Control Engineers of Japan (SICE)*, pp. 154-159, 2016.

[37] X. Yu and Z. Man, “Model reference adaptive control systems with terminal sliding modes,” *International Journal of Control*, vol. 64, no. 6, pp. 1165-1176, 2007, doi: 10.1080/00207179608921680.

[38] H. Y. Zheng, *Fuzzy cooperative EKF localization and adaptive integral terminal sliding-mode formation control using fuzzy broad learning system and artificial potential function for uncertain networking heterogeneous omnidirectional multirobots*, Doctoral dissertation, Master Thesis, Department of Electrical Engineering, National Chung Hsing University, 2020.

[39] C. L. P. Chen, “Broad Learning System and its Structural Variations,” *2018 IEEE 16th International Symposium on Intelligent Systems and Informatics (SISY)*, pp. 000011-00012, 2018, doi: 10.1109/SISY.2018.8524681.

[40] S. Feng and C. L. P. Chen, “Fuzzy Broad Learning System: A Novel Neuro-Fuzzy Model for Regression and Classification,” in *IEEE Transactions on Cybernetics*, vol. 50, no. 2, pp. 414-424, 2020, doi: 10.1109/TCYB.2018.2857815.

[41] C. C. Tsai *et al.*, “Adaptive Reinforcement Learning Formation Control Using ORFBLS for Omnidirectional Mobile Multi-Robots,” *International Journal of Fuzzy Systems*, vol. 25, pp. 1756–1769, 2023.

[42] Chuan-Kai Lin, “Adaptive critic autopilot design of Bank-to-turn missiles using fuzzy basis function networks,” in *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, vol. 35, no. 2, pp. 197-207, 2005.

[43] C. -K. Lin, “H_∞ reinforcement learning control of robot manipulators

using fuzzy wavelet networks," *Fuzzy Sets and Systems*, vol. 160, no. 12, pp. 1765-1786, 2009.
[44] Y. Hu, W. Wang, H. Liu and L. Liu, "Reinforcement Learning Tracking Control for Robotic Manipulator With Kernel-Based Dynamic Model," in *IEEE Transactions on Neural Networks and Learning Systems*, vol. 31, no. 9, pp. 3570-3578, 2020, doi: 10.1109/TNNLS.2019.2945019.
[45] P. Zhu, W. Dai, W. Yao, J. Ma, Z. Zeng and H. Lu, "Multi-Robot Flocking Control Based on Deep Reinforcement Learning," in *IEEE Access*, vol. 8, pp. 150397-150406, 2020, doi: 10.1109/ACCESS.2020.3016951.
[46] X. Gao, R. Gao, P. Liang, Q. Zhang, R. Deng and W. Zhu, "A Hybrid Tracking Control Strategy for Nonholonomic Wheeled Mobile Robot Incorporating Deep Reinforcement Learning Approach," in *IEEE Access*, vol. 9, pp. 15592-15602, 2021, doi: 10.1109/ACCESS.2021.3053396.
[47] C. C. Tsai *et al*., "Adaptive Reinforcement Learning Formation Control Using ORFBLS for Omnidirectional Mobile Multi-Robots," *International Journal of Fuzzy Systems*, vol. 25, pp. 1756–1769, 2023, doi: 10.1007/s40815-023-01491-4.
[48] C. C. Tsai, C. -C. Chan, Y. -C. Li, and F. -C. Tai, "Intelligent adaptive PID control using fuzzy broad learning system: an application to tool-grinding servo control systems," *International Journal of Fuzzy Systems*, vol. 22, pp. 2149–2162, 2020, doi: 10.1007/s40815-020-00913-x.
[49] D. Wang, M. Ha, and M. Zhao, *Advanced Optimal Control and Applications Involving Critic Intelligence*, Springer, Singapore, 2023, doi: 10.1007/978-981-19-7291-1.
[50] M. G. Kumar *et al*., "A nonlinear hidden layer enables actor–critic agents to learn multiple paired association navigation," *Cerebral Cortex*, vol. 32, no. 18, pp. 3917–3936, doi: 10.1093/cercor/bhab456.
[51] Y. Zhang, L. Guo, B. Gao, T. Qu and H. Chen, "Deterministic Promotion Reinforcement Learning Applied to Longitudinal Velocity Control for Automated Vehicles," in *IEEE Transactions on Vehicular Technology*, vol. 69, no. 1, pp. 338-348, 2020, doi: 10.1109/TVT.2019.2955959..
[52] Z. Feng and J. Fei, "Super-Twisting Sliding Mode Control for Micro Gyroscope Based on RBF Neural Network," in *IEEE Access,* vol. 6, pp. 64993-65001, 2018, doi: 10.1109/ACCESS.2018.2877398.
[53] W. Lin, X. Huo, Z. Jin, B. Wu and Z. Liu, "Sliding Mode Control of Manipulator Based on Nominal Model and Nonlinear Disturbance Observer," *IECON 2018 - 44th Annual Conference of the IEEE Industrial Electronics Society*, pp. 5519-5524, 2018, doi: 10.1109/IECON.2018.8592926.
[54] Y. Pan, X. Li, H. Wang, and H. Yu, "Continuous sliding mode control of compliant robot arms: A singularly perturbed approach," *Mechatronics*, vol. 52, pp. 127-134, 2018, doi: 10.1016/j.mechatronics.2018.04.005.
[55] K. Bai, X. Gong, S. Chen, Y. Wang, Z. Liu, "Sliding mode nonlinear disturbance observer-based adaptive back-stepping control of a humanoid robotic dual manipulator," *Robotica*, vol. 36, no. 11, pp. 1728–1742, 2018, doi: 10.1017/S026357471800067X.
[56] S. Khorashadizadeh and M. Sadeghijaleh, "Adaptive fuzzy tracking control of robot manipulators actuated by permanent magnet synchronous motors," *Computers and Electrical Engineering*, vol. 72, pp. 100-111, 2018, doi: 10.1016/j.compeleceng.2018.09.010.
[57] M. Chen, S. S. Ge and B. V. E. How, "Robust Adaptive Neural Network Control for a Class of Uncertain MIMO Nonlinear Systems With Input Nonlinearities," in *IEEE Transactions on Neural Networks*, vol. 21, no. 5, pp. 796-812, 2010, doi: 10.1109/TNN.2010.2042611.
[58] H. Han and J. Qiao, "A Self-Organizing Fuzzy Neural Network Based on a Growing-and-Pruning Algorithm," in *IEEE Transactions on Fuzzy Systems*, vol. 18, no. 6, pp. 1129-1143, 2010, doi: 10.1109/TFUZZ.2010.2070841.
[59] C. L. P. Chen, Y. -J. Liu and G. -X. Wen, "Fuzzy Neural Network-Based Adaptive Control for a Class of Uncertain Nonlinear Stochastic Systems," in *IEEE Transactions on Cybernetics*, vol. 44, no. 5, pp. 583-593, 2014, doi: 10.1109/TCYB.2013.2262935.
[60] Q. Yang, Z. Yang and Y. Sun, "Universal Neural Network Control of MIMO Uncertain Nonlinear Systems," in *IEEE Transactions on Neural Networks and Learning Systems*, vol. 23, no. 7, pp. 1163-1169, 2012.
[61] D. N. Minh *et al*., "An Adaptive Fuzzy Dynamic Surface Control Tracking Algorithm for Mecanum Wheeled Mobile Robots," *International Journal of Mechanical Engineering and Robotics Research*, vol. 12, no. 6, pp. 354-361, 2023.
[62] Z. Yuan *et al*., "Trajectory tracking control of a four mecanum wheeled mobile platform: an extended state observer-based sliding mode approach," *IET Control Theory and Applications*, vol. 14, no. 3, pp. 415-426, 2023.
[63] M. A. -Khalaf, and F. L. Lewis, "Nearly optimal control laws for nonlinear systems with saturating actuators using a neural network HJB approach," *Automatica*, vol. 41, no. 5, 2005, pp. 779-791, 2005.
[64] D. Vrabie and F. Lewis, "Neural network approach to continuous-time direct adaptive optimal control for partially unknown nonlinear systems," *Neural Networks*, vol. 22, no. 3, pp. 237-246, doi: 10.1016/j.neunet.2009.03.008.
[65] S. Bhasin *et al*., "A novel actor–critic–identifier architecture for approximate optimal control of uncertain nonlinear systems," *Automatica*, vol. 49, no. 1, pp. 82-92, 2013, doi: 10.1016/j.automatica.2012.09.019.
[66] H. Modares, F. L. Lewis and M. -B. Naghibi-Sistani, "Adaptive Optimal Control of Unknown Constrained-Input Systems Using Policy Iteration and Neural Networks," in *IEEE Transactions on Neural Networks and Learning Systems*, vol. 24, no. 10, pp. 1513-1525, 2013, doi: 10.1109/TNNLS.2013.2276571.
[67] K. G. Vamvoudakis, and F. L. Lewis, "Online actor–critic algorithm to solve the continuous-time infinite horizon optimal control problem," *Automatica*, vol. 46, no. 5, pp. 878-888, 2010, doi: 10.1016/j.automatica.2010.02.018.
[68] K. G. Vamvoudakis, D. Vrabie and F. L. Lewis, "Online adaptive learning of optimal control solutions using integral reinforcement learning," *2011 IEEE Symposium on Adaptive Dynamic Programming and Reinforcement Learning (ADPRL)*, pp. 250-257, 2011, doi: 10.1109/ADPRL.2011.5967359.
[69] H. -N. Wu and B. Luo, "Neural Network Based Online Simultaneous Policy Update Algorithm for Solving the HJI Equation in Nonlinear H_∞ Control," in *IEEE Transactions on Neural Networks and Learning Systems*, vol. 23, no. 12, pp. 1884-1895, 2012, doi: 10.1109/TNNLS.2012.2217349.
[70] T. Basar and P. Bernhard, "H/sup /spl infin//-0ptimal Control and Related Minimax Design Problems: A Dynamic Game Approach," in *IEEE Transactions on Automatic Control*, vol. 41, no. 9, pp. 1397-, 1996, doi: 10.1109/TAC.1996.536519.
[71] J. Kim and I. Yang, "Hamilton-Jacobi-Bellman equations for Q-learning in continuous time," *Proceedings of Machine Learning Research*, vol. 120, pp. 1–10, 2020.
[72] M. A. Bucci, "Nonlinear optimal control using deep reinforcement learning," In *IUTAM Laminar-Turbulent Transition: 9th IUTAM Symposium*, pp. 279-290, 2022, doi: 10.1007/978-3-030-67902-6_24.
[73] B. Luo, H. -N. Wu, T. Huang, and D. Liu. "Reinforcement learning solution for HJB equation arising in constrained optimal control problem," *Neural Networks*, vol. 71, pp. 150-158, 2015.
[74] H. Wiltzer, D. Meger, and M. G. Bellemare, "Distributional hamilton-jacobi-bellman equations for continuous-time reinforcement learning," In I*nternational Conference on Machine Learning*, pp. 23832-23856, 2022.
[75] J. W. Kim *et al*., "A model-based deep reinforcement learning method applied to finite-horizon optimal control of nonlinear control-affine system," *Journal of Process Control* vol. 87, pp. 166-178, 2020.
[76] W. Xiao, Q. Zhou, Y. Liu, H. Li and R. Lu, "Distributed Reinforcement Learning Containment Control for Multiple Nonholonomic Mobile Robots," in *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 69, no. 2, pp. 896-907, 2022.
[77] N. T. Luy, N. T. Thanh, and H. M. Tri, "Reinforcement learning-based intelligent tracking control for wheeled mobile robot," *Transactions of the Institute of Measurement and Control*, vol. 36, no. 7, pp. 868-877, 2014.
[78] L. Ding, M. Zheng, S. Li, H. Yang, H. Gao, and Z. Deng, "Neural-based online finite-time optimal tracking control for wheeled mobile robotic system with inequality constraints," *Asian Journal of Control*, vol. 26, no. 1, pp. 297-311, 2024, doi: 10.1002/asjc.3203.
[79] N. D. Dien, N. T. Luy, L. K. Lai, and T. T. Hai, "Optimal tracking control for robot manipulators with input constraint based reinforcement learning," *Journal of Computer Science and Cybernetics*, vol. 39, no. 2, pp. 175-189, 2023, doi: 10.15625/1813-9663/18099.
[80] H. V. Doan and N. T. -T. Vu, "Robust optimal control for uncertain wheeled mobile robot based on reinforcement learning: ADP approach," *Bulletin of Electrical Engineering and Informatics*, vol. 13, no. 3, pp. 1524-1534, 2024, doi: 10.11591/eei.v13i3.7054.