Bulletin of Electrical Engineering and Informatics
Vol. 13, No. 3, June 2024, pp. 1524~1534
ISSN: 2302-9285, DOI: 10.11591/eei.v13i3.7054

Robust optimal control for uncertain wheeled mobile robot based on reinforcement learning: ADP approach

Hoa Van Doan[1], Nga Thi-Thuy Vu[2]
[1]Department of Electrical Engineering, University of Economics-Technology for Industries, Hanoi, Vietnam
[2]School of Electrical-Electronics Engineering, Hanoi University of Science and Technology, Hanoi, Vietnam

Article Info

Article history:

Received Jun 23, 2023
Revised Oct 4, 2023
Accepted Oct 16, 2023

Keywords:

Actor-critic
Adaptive dynamic programming
Disturbances
Reinforcement learning
Wheel slips
Wheeled mobile robot

ABSTRACT

This paper presents a robust optimal control approach for the wheel mobile robot system, which considers the effects of external disturbances, uncertainties, and wheel slipping. The proposed method utilizes an adaptive dynamic programming (ADP) technique in conjunction with a disturbance observer. Initially, the system's state space model is formulated through the utilization of kinematic and dynamic models. Subsequently, the ADP method is employed to establish an online adaptive optimal controller, which solely relies on a single neural network for the purpose of function approximation. The utilization of the disturbance observer in conjunction with the compensation controller serves to alleviate the effects of disturbances. The Lyapunov theorem establishes the stability of the complete closed-loop system and the convergence of the weights of the neural network. The proposed approach has been shown to be effective through simulation under the effect of the disturbances and the change of the desired trajectory.

Corresponding Author:

Nga Thi-Thuy Vu
School of Electrical-Electronics Engineering, Hanoi University of Science and Technology
Hanoi, Vietnam
Email: nga.vuthithuy@hust.edu.vn

1. INTRODUCTION

A wheeled mobile robot (WMR) is a system that can move from one location to another autonomously, without the need for external intervention or assistance [1]. Unlike robotic arms, which can only work in a specific space, WMR can move freely within a predetermined workspace to achieve the desired goal. Proportional integral derivative (PID) controllers [2], feedback linearization controllers [3], backstepping controllers [4], sliding mode controllers [5], [6], adaptive controllers [7], [8], robust controllers [9], [10], fuzzy controllers [11], [12], and neural network-based controllers [13], [14] are just a few of the control methods proposed for WMR. The WMR is assumed to roll without slipping in these studies. In practice, however, due to the presence of nonlinear components such as friction, wheel slip, and so on, some studies have added these components to the mathematical model of the WMR to improve accuracy [15]–[18]. In [15], [16], the friction and wheel slip components are included in the robot's kinematics and dynamics models, and then robust controllers for tracking control are established. Vu *et al.* [17] presents an adaptive control method based on a disturbance estimator that can compensate for the effects of wheel slip and external disturbances acting on both kinematic and dynamic loops. Similar control structures for uncertain WMR with kinematic and dynamic control loops are presented in [18]. However, rather than using two disturbance observers in the inner and outer loops, which complicates the system, the controllers in [18] are designed to deal with disturbances using the adaptive fuzzy type 2 control technique. Because the controller parameters are updated based on the optimal rule to adapt to changes in working conditions, the disturbance observers are removed.

The previously mentioned control techniques have fulfilled the criteria for achieving high-quality trajectory tracking. Nevertheless, the issue of determining the optimal index with respect to tracking error and control energy remains unresolved. Reinforcement learning (RL) [19] and adaptive dynamic programming (ADP) [20] are efficient techniques that leverage optimization rules of dynamic programming to address optimization problems. RL is utilized for the purpose of determining the resolution of the Hamilton-Jacobi-Bellman (HJB) and Hamilton-Jacobi-Isaacs (HJI) equations. This is due to the fact that these equations are nonlinear differential equations, rendering it challenging to obtain solutions through analytical approaches, particularly in the case of nonlinear systems. Prior research has employed the conventional ADP control framework [21], [22], featuring two neural networks referred to as actor-critic (AC), frequently neglecting the impact of disturbances on the system. The proposed approach employs a neural network, specifically an actor network, to approximate the optimal control law. Additionally, a critic network is utilized to evaluate the control law and approximate the optimal cost function. Subsequent to this achievement, a number of algorithms have been developed for nonlinear systems that are subject to disturbance effects [23]–[28]. The algorithms in [23]–[25] employ the ADP structure, which incorporates three neural networks. Notably, an additional neural network has been incorporated into the AC structure to estimate the upper bound of noise. A reinforcement learning based trajectory tracking controller is proposed for the autopilot system of underactuated surface vessel (USVs) influenced by input disturbances and input signal constraints [26]. By using the tracking error conversion technique to handle the error constraint problem, it is ensured that the USVs can accurately follow the set trajectory. However, the updating rule of the weights of the actor and critic neural networks is sequential, which reduces the convergence speed of the parameters. Sun and Liu [27] propose robust optimal control for the rocket autopilot using ADP technique combined with nonlinear disturbance observer and adaptive sliding mode controller. The AC architecture is used to design an adaptive optimal controller using only one critic neural network. However, because the recognition process is additionally combined, the controller has a high computational complexity and is difficult to implement. An algorithm based on online adaptive reinforcement learning method is developed for the optimal control problem of continuous nonlinear systems with model uncertainty [28]. To approximate the solution of the HJB equation, an actor-critic-identity (ACI) structure is used based on three neural networks: actor and critic networks that estimate the optimal control law and the optimal cost function, respectively, and the third network is used for system dynamics identification. The utilization of a control structure that involves two or three neural networks may ensure the good performances for uncertain nonlinear system but can result in a complex calculation process and inefficient use of resources, ultimately resulting in a reduction in the rate of convergence.

This paper develops an adaptive optimal controller for tracking control of a WMR system using the online adaptive dynamic programming technique in cooperation with a disturbance observer. The control scheme consists of two parts: the first part is the optimal component to optimize the cost function and the second component is the compensation component that uses the estimated disturbances to remove the effect of model uncertainty and the system noise. The optimal controller is designed based on the value iteration (VI) method and simultaneously updates the weight matrix. The stability of the whole system using the optimal component and the disturbance observer is demonstrated under the uniformly ultimately bounded (UUB) condition. Finally, some simulations are performed to prove the correctness of the algorithm. The simulation results show that the proposed scheme gives good performance for both the nominal working and when affected by uncertainties and external disturbances.

2. SYSTEM MODELLING

Considering the three-wheel mobile robot, two independent drive wheels at the rear and one rudder at the front, subject to nonholonomic constraints as shown in Figure 1. In Figure 1, G is the WMR's center of mass, $M(x_M, y_M)$ is the center of the axle connecting the two rear wheels, and θ is the WMR's direction angle. Let $\dot{\varphi}_R$ and $\dot{\varphi}_L$ represent the angular velocities of the right and left wheels, respectively. μ_R and μ_L represent longitudinal slip of the right and left wheels, while δ represents the the lateral slip along the wheel shaft. Taking the effect of wheel slip into account, the kinematic equation for WMR is [17], [18]:

$$\begin{cases} \dot{x}_M = \beta \cos\theta - \dot{\delta} \sin\theta \\ \dot{y}_M = \beta \sin\theta + \dot{\delta} \cos\theta \end{cases} \tag{1}$$

where β is the linear velocity perpendicular to the axis joining the two rear wheels and ϖ is angular velocity of the WMR.

The dynamic model of the WMR is as (2) [17], [18]:

$$M\dot{v} + Bv + Q\ddot{\mu} + C\dot{\delta} + G\ddot{\delta} + \tau_d = \tau \tag{2}$$

where τ_d is input disturbance, $v = [\dot{\varphi}_R \quad \dot{\varphi}_L]^T$, $\mu = [\mu_R \quad \mu_L]^T$

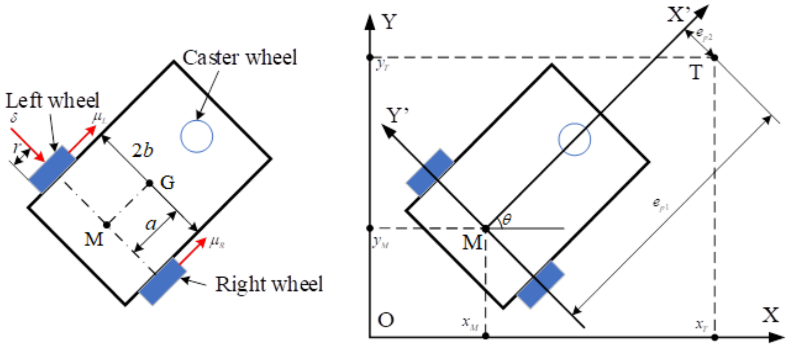


Figure 1. WMR model and coordinate system

The objective of this study is to control the WMR, i.e. point $M(x_M, y_M)$, to track the reference trajectory, represented by point $T(x_T, y_T)$ with mininize consumption of energy. The position error between the center of robot (point M) and the target point (point T) is determined as (3):

$$e_p = \begin{bmatrix} e_{p1} \\ e_{p2} \end{bmatrix} = \begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} x_T - x_M \\ y_T - y_M \end{bmatrix} = H \begin{bmatrix} x_T - x_M \\ y_T - y_M \end{bmatrix} \tag{3}$$

where H is the transform matrix: $H = \begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix}$
The time derivative of (3) is obtainted as (4):

$$\dot{e}_p = \begin{bmatrix} \dot{e}_{p1} \\ \dot{e}_{p2} \end{bmatrix} = \kappa v + \xi_1 \tag{4}$$

where κ and ξ_1 are calculated as in the documents [17], [18]

Define state variables $x_1 = e_p$; $x_2 = \dot{x}_1 + \lambda x_1$ where λ is a positive scalar. Based on (2) and (4), the time derivative of x_1 and x_2 has the following form:

$$\dot{x}_1 = \dot{e}_p = \kappa v + \xi_1 \tag{5}$$

$$\dot{x}_2 = E x_2 - E\lambda x_1 + Z\tau + d \tag{6}$$

where: $E = E_1\kappa^{-1}$, $d = \xi_3 - E_1\kappa^{-1}\xi_1$, $E_1 = -\kappa M^{-1}B$, $Z = \kappa M^{-1}$,
$\xi_2 = -M^{-1}(Q\ddot{\mu} + C\dot{\delta} + G\ddot{\delta} + \tau_d)$,$\xi_3 = \kappa\xi_2 + \dot{\kappa}v + \dot{\xi}_1 + \lambda\kappa v + \lambda\xi_1$

Rewriting (5) and (6) in the state space form, the following is obtained:

$$\dot{x} = f(x) + g_u u + g_d d \tag{7}$$

where: $f(x) = \begin{bmatrix} x_2 - \lambda x_1 \\ E x_2 - \lambda E x_1 \end{bmatrix}$; $g_u = \begin{bmatrix} 0 \\ Z \end{bmatrix}$; $g_d = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$; $\tau = u$
Because system (7) is nonlinear and affected by disturbance d, the controller u is established as follows to achieve optimal performances:

$$u = u_r(x) + u_d(x) \tag{8}$$

where $u_r(x)$ is the optimal control component which will be designed using the ADP method and $u_d(x)$ compensation control component.

3. ROBUST ADAPTIVE OPTIMAL CONTROLLER DESIGN FOR THE WMR

The ADP algorithm can only be applied to nonlinear systems when ignoring impact noise, which reduces the applicability of the algorithm in practice. Therefore, we propose to combine the ADP algorithm with a disturbance observer to design a sustainable optimal tracking controller for WMR containing an uncertain component, affected by disturbances. Figure 2 illustrates the proposed controller's structure diagram. The controller is constructed of two parts: an optimal component $u_r(x)$ and a disturbance compensation component $u_d(x)$. The detailed design for each part is shown.

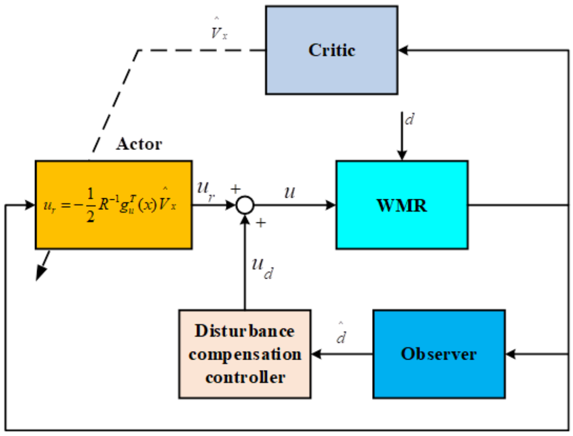


Figure 2. ADP control structure combined with disturbance observer

3.1. The adaptive optimal controller

In the case of $d = 0$, the nonlinear system (7) is rewritten as (9):

$$\dot{x} = f(x) + g_u(x)u_r \tag{9}$$

Assumption 1: $f(x)$ satisfies the lipschitz continuous condition in the set Ω where Ω is a set which consists of all possible solution of (9).
Define the cost function:

$$J(x) = \int_t^{\infty} r(x(\tau), u_r(x(\tau)))\, d\tau \tag{10}$$

where $r(x(\tau), u_r(\tau)) = x^T(\tau)Qx(\tau) + u_r^T(x)Ru_r(x)$ in which $Q \in \mathbb{R}^{2n\times 2n}$, $R \in \mathbb{R}^{n\times n}$ are symmetric positive definite matrices.
The Hamilton function is defined as (11):

$$H(x, u_r, J_x) = \left(\frac{\partial J}{\partial x}\right)^T \dot{x} + r(x, u_r) = \left(\frac{\partial J}{\partial x}\right)^T \dot{x} + x^TQx + u_r^TRu_r \tag{11}$$

For the system (9) to have the optimal solution, there must exist a function $V(x, u_r)$ satisfying the HJB:

$$H(x, u_r, V) = \left(\frac{\partial V}{\partial x}\right)^T \dot{x} + x^TQx + u_r^TRu_r = 0 \tag{12}$$

The optimal control signal u_r is then determined by using the formula that is presented:

$$u_r = \underset{u_r}{\operatorname{argmin}}\{H(x, u_r, V)\} \tag{13}$$

By solving (13) using (12), the following is obtained:

$$u_r = -\frac{1}{2}R^{-1}g_u^T\frac{\partial V}{\partial x} \tag{14}$$

However, the system (11) is nonlinear, so (12) cannot be solved directly, which means the function $V(x, u_r)$ cannot be found by analytical methods. To overcome this difficulty, the function $V(x, u_r)$ is approximated by a neural network as (15):

$$V(x, u_r) = W^T \Phi(x) + \varepsilon(x) \tag{15}$$

where W is the weight matrix of the neural network, $\Phi(x)$ is the active function which is a function of state variable x, and $\varepsilon(x)$ is the approximation error.
Using this approximation, the controller u_r^* becomes:

$$u_r = -\frac{1}{2} R^{-1} g_u^T \frac{\partial V}{\partial x} = -\frac{1}{2} R^{-1} g_u^T W^T \frac{\partial \Phi}{\partial x} \tag{16}$$

Unfortunately, the real value of W in (15) is unknown; therefore, it is replaced by an estimation, and the function $V(x, u_r)$ is also presented by its estimation, as (17):

$$\hat{V}(x, u_r) = \widehat{W}^T \Phi(x) \tag{17}$$

where $\widehat{W}$ is the estimation of W, which is updated by the following law:

$$\dot{\widehat{W}} = -\alpha_1 \frac{\hat{\sigma}}{\left(\hat{\sigma}^T \hat{\sigma} + 1\right)^2} \left(\hat{\sigma}^T \widehat{W} + Q(x) + u_r^T R u_r\right) + \frac{1}{2} \alpha_2 \frac{\partial \Phi}{\partial x} g_u(x) R^{-1} g_u^T(x) x \tag{18}$$

in which $\hat{\sigma} = \frac{\partial \Phi}{\partial x} \dot{x}$
Finally, the optimal controller u_r is employed as (19):

$$u_r = -\frac{1}{2} R^{-1} g_u^T \widehat{W}^T \frac{\partial \Phi}{\partial x} \tag{19}$$

3.2. Observer based adaptive controller design

Consider system (7) which is affected by the disturbance d:

$$\dot{x} = f(x) + g_u(x) u + g_d(x) d \tag{20}$$

As mentioned in section 2, the controller u consists of a compensation control component that compensates for the effects of system uncertainties and disturbances. In this study, this control component is utilized as shown (21):

$$u_d(x) = -Z^{-1} \hat{d} \tag{21}$$

where $\hat{d}$ is the estimation of d, the value of $\hat{d}$ is determined by the following observer [29]–[31]:

$$\begin{cases} \hat{d} = \eta + \rho(x) \\ \dot{\eta} = -h(x)\{g_d(x)[\eta + \rho(x)] + f(x) + g_u(x) u\} \end{cases} \tag{22}$$

in which $\eta \in \mathbb{R}^l$ is the internal state of the observer, $\rho(x) \in \mathbb{R}^l$ is a designed vector, and $h(x) = \frac{\partial \rho(x)}{\partial x}$ is the gain of the observer. The convergence of the observer is presented in detail in [29]–[31].

3.3. Stability of overall system

Replace the control components (8) and (21) into system (20), the dynamic of the closed loop system is express as (23):

$$\dot{x} = f(x) + g_u u_r - g_u Z^{-1} \hat{d} + g_d\left(\tilde{d} + \hat{d}\right) \tag{23}$$

Due to $g_d = Z^{-1} g_u$, the following is obtained:

$$\dot{x} = f(x) + g_u u_r + g_d \tilde{d} \tag{24}$$

In order to demonstrate the stability of the system which consists of the adaptive optimal controller, the disturbance observer, and the WMR, the following Lyapunov function is chosen:

$$L(X,t) = v_1 x^T x + v_2 \int_t^\infty (x^T Qx + u_r^T R u_r)\, d\tau + v_3 \int_t^\infty (\tilde{d}^T \tilde{d})\, d\tau + \frac{v_4}{\alpha_2} \widetilde{W}^T \widetilde{W} \quad (25)$$

where $\widetilde{W} = W - \widehat{W}$, $X = [x \quad u_r \quad \tilde{d} \quad \widetilde{W}]$ and v_i (i=1,2,3,4) is positive scalar.
Due to u_r is the solution of (10), the component $\int_t^\infty (x^T Qx + u_r^{iT} R u_r^i)\, d\tau$ is limited. In addition, the observer in section 3.2 is exponentially convergent, so the integrator $\int_t^\infty (\tilde{d}^T \tilde{d})\, d\tau$ is convergent. These leads to the following result:

$$0 \le L(X,t) \le \eta_1 \|X\| \quad (26)$$

where η_1 is a positive constant.
The time derivative of $L(X,t)$ is determined as (27):

$$\dot{L} = 2v_1 x^T (f + g_u u_r + Z^{-1} g_u \tilde{d}) - v_2 (x^T Qx + u_r^T R u_r) - v_3 \|\tilde{d}\|^2 - \frac{2v_4}{\alpha_2} \widetilde{W}^T W = 2v_1 x^T (f + g_u u_r) + 2v_1 x^T Z^{-1} g_u \tilde{d} - v_2 (x^T Qx + u_r^T R u_r) - v_3 \|\tilde{d}\|^2 - \frac{2v_4}{\alpha_2} \widetilde{W}^T \dot{\widehat{W}} \quad (27)$$

We have:

$$2v_1 x^T \gamma(x) g_u \tilde{d} \le v_1 \|x\|^2 + v_1 \|Z^{-1} g_u\|^2 \|\tilde{d}\|^2 \quad (28)$$

$$x^T Qx + u_r^{iT} R u_r^i \ge \lambda \|x\|^2 \|u_r^i\|^2{}_{min_{min}} \quad (29)$$

or

$$-v_2 (x^T Qx + u_r^{iT} R u_r^i) \le -v_2 \lambda \|x\|^2{}_2 \|u_r^i\|^2{}_{min_{min}} \quad (30)$$

(27) is equivalent to:

$$\dot{L} \le 2v_1 \|x^T\| \|f(x)\| + 2v_1 \|x^T\| \|g_u\| \|u_r\| + v_1 \|x\|^2 + v_1 \|Z^{-1} g_u\|^2 \|\tilde{d}\|^2 - v_2 \lambda \|x\|^2{}_2 \|u_r\|^2{}_3 \|\tilde{d}\|^2 \frac{\widehat{2v_4}^T}{\alpha_2}{}_{min_{min}} \quad (31)$$

Using updating law (18) leads to (32):

$$-\frac{2v_4}{\alpha_2} \widetilde{W}^T \dot{\widehat{W}} = -\frac{2v_4}{\alpha_2} \widetilde{W}^T \dot{\widehat{W}}_1 - \frac{2v_4}{\alpha_2} \widetilde{W}^T W_2 = 2v_4 \frac{\alpha_1}{\alpha_2} \widetilde{W}^T \frac{\hat{\sigma}}{(\hat{\sigma}^T \hat{\sigma} + 1)^2} (\hat{\sigma}^T W - \hat{\sigma}^T \widetilde{W} + Q(x) + u_r^T R u_r) - v_4 \widetilde{W}^T \frac{\partial \Phi}{\partial x} g_u(x) R^{-1} g_u^T(x) x \quad (32)$$

Define $\varepsilon_H = \hat{\sigma}^T W + Q(x) + u_r^T R u_r$
Then:

$$-\frac{2v_4}{\alpha_2} \widetilde{W}^T \dot{\widehat{W}} = -2v_4 \frac{\alpha_1}{\alpha_2} \frac{\widetilde{W}^T \hat{\sigma} \hat{\sigma}^T \widetilde{W}}{(\hat{\sigma}^T \hat{\sigma} + 1)^2} + 2v_4 \frac{\alpha_1}{\alpha_2} \frac{\widetilde{W}^T \hat{\sigma}}{(\hat{\sigma}^T \hat{\sigma} + 1)^2} \varepsilon_H - v_4 \widetilde{W}^T \Phi_x g_u(x) R^{-1} g_u^T(x) x \quad (33)$$

$$-\frac{2v_4}{\alpha_2} \widetilde{W}^T \dot{\widehat{W}} \le -2v_4 \frac{\alpha_1}{\alpha_2} \left\| \frac{\hat{\sigma}}{\hat{\sigma}^T \hat{\sigma} + 1} \right\|^2 \|\widetilde{W}\|^2 + v_4 \frac{\alpha_1}{\alpha_2} \left\| \frac{\hat{\sigma}}{\hat{\sigma}^T \hat{\sigma} + 1} \right\|^2 \|\widetilde{W}\|^2 + v_4 \frac{\alpha_1}{\alpha_2} \left\| \frac{\varepsilon_H}{\hat{\sigma}^T \hat{\sigma} + 1} \right\|^2 - v_4 \widetilde{W}^T \Phi_x g_u(x) R^{-1} g_u^T(x) x = -v_4 \frac{\alpha_1}{\alpha_2} \left\| \frac{\hat{\sigma}}{\hat{\sigma}^T \hat{\sigma} + 1} \right\|^2 \|\widetilde{W}\|^2 + v_4 \frac{\alpha_1}{\alpha_2} \left\| \frac{\varepsilon_H}{\hat{\sigma}^T \hat{\sigma} + 1} \right\|^2 - v_4 \widetilde{W}^T \Phi_x g_u(x) R^{-1} g_u^T(x) x \quad (34)$$

Also, because $f(x)$ is Lipschitz then $f(x) \le k\|x\|$where k is a positive constant. Substituting $u_r = -\frac{1}{2} R^{-1} g_u^T \left(\frac{\partial \Phi}{\partial x}\right)^T \widehat{W}$ and (34) into (31), the following is obtained:

$$\dot{L} \le ((2k+1)v_1 - v_2 \lambda_{min}() \|x\|^2{}_2 \|u_r^i\|^2{}_2 (v_1 \|Z^{-1} g_u\|^2 - v_3) \|\tilde{d}\|^2{}_{min}) - v_4 \frac{\alpha_2}{\alpha_1} \left\| \frac{\hat{\sigma}}{\hat{\sigma}^T \hat{\sigma} + 1} \right\|^2 \|\widetilde{W}\|^2 + v_4 \frac{\alpha_2}{\alpha_1} \left\| \frac{\varepsilon_H}{\hat{\sigma}^T \hat{\sigma} + 1} \right\|^2 - (v_1 \|\widehat{W}^T\| + v_4 \|\widetilde{W}^T\|) \left\| \frac{\partial \Phi}{\partial x} \right\| \|g_u(x)\|^2 \|R^{-1}\| \|x\|) \quad (35)$$

Choose $v_4 = v_1$ and note that $\widetilde{W} + \widehat{W} = W$, then:

$$\dot{L} \leq ((2k+1)v_1 - v_2\lambda_{min}()\|x\|^2{}_2\|u_r\|^2(v_1\|Z^{-1}g_u\|^2 - v_3)\|\tilde{d}\|^2{}_{min}) - v_4\frac{\alpha_2}{\alpha_1}\left\|\frac{\hat{\sigma}}{\hat{\sigma}^T\hat{\sigma}+1}\right\|^2\|\widetilde{W}\|^2 + v_4\frac{\alpha_2}{\alpha_1}\left\|\frac{\varepsilon_H}{\hat{\sigma}^T\hat{\sigma}+1}\right\|^2 - v_1\|W^T\|\left\|\frac{\partial\Phi}{\partial x}\right\|\|g_u(x)\|^2\|R^{-1}\|\|x\|) \quad (36)$$

$$\dot{L} \leq ((2k+1)v_1 - v_2\lambda_{min}()\|x\|^2{}_2\|u_r^i\|^2{}_{min}) + (v_1\|Z^{-1}g_u\|^2 - v_3)\|\tilde{d}\|^2 - v_4\frac{\alpha_2}{\alpha_1}\left\|\frac{\hat{\sigma}}{\hat{\sigma}^T\hat{\sigma}+1}\right\|^2\|\widetilde{W}\|^2 + v_4\frac{\alpha_2}{\alpha_1}\left\|\frac{\varepsilon_H}{\hat{\sigma}^T\hat{\sigma}+1}\right\|^2 + v_1\lambda x) \quad (37)$$

where $\lambda = \left\|W_{max}^T\left\|\left(\frac{\partial\Phi}{\partial x}\right)_{max}\left\|g_u(x)_{max}{}^2\|R^{-1}\|\right\|\right\|\right\|$

$$\dot{L} \leq ((2k+1)v_1 - v_2\lambda_{min}()\|x\|^2{}_2\|u_r^i\|^2(v_1\|Z^{-1}g_u\|^2 - v_3)\|\tilde{d}\|^2{}_{min}) - v_4\frac{\alpha_2}{\alpha_1}\left\|\frac{\hat{\sigma}}{\hat{\sigma}^T\hat{\sigma}+1}\right\|^2\|\widetilde{W}\|^2 + v_4\frac{\alpha_2}{\alpha_1}\left\|\frac{\varepsilon_H}{\hat{\sigma}^T\hat{\sigma}+1}\right\|^2 + \frac{v_1\lambda^2}{2} + \frac{v_1}{2}\|x\|^2) \quad (38)$$

Define $L_{1x} = \left(2k+\frac{3}{2}\right)v_1 - v_2\lambda_{min}$; $L_{1u} = -v_2\lambda_{min}$; $L_{1d} = v_1\|g_u\gamma\|^2 - v_3$

$$L_{1W} = -v_4\frac{\alpha_2}{\alpha_1}\left\|\frac{\hat{\sigma}}{\hat{\sigma}^T\hat{\sigma}+1}\right\|^2;\ L_{1\varepsilon} = v_4\frac{\alpha_2}{\alpha_1}\left\|\frac{\varepsilon_H}{\hat{\sigma}^T\hat{\sigma}+1}\right\|^2 + \frac{v_1\lambda^2}{2}$$

As a result, (38) is rewritten as (39):

$$\dot{L} \leq L_{1x}\|x\|^2 + L_{1u}\|u_r^i\|^2 + L_{1d}\|\tilde{d}\|^2 + L_{1W}\|\widetilde{W}\|^2 + L_{1\varepsilon} \quad (39)$$

If v_i, i=1, ..., 4 satisfies: $v_2 \geq \frac{(2k+2)v_1}{\lambda_{min}}$, $v_3 > v_1\|g_u\gamma\|^2$, $v_4 = v_1$, and $\|x\| \geq \sqrt{\frac{L_{1\varepsilon}}{-L_{1x}}}$ or $\|u_r^i\| \geq \sqrt{\frac{L_{1\varepsilon}}{-L_{1u}}}$ or $\|\tilde{d}\| \geq \sqrt{\frac{L_{1\varepsilon}}{-L_{1d}}}$ or $\|\widetilde{W}\| \geq \sqrt{\frac{L_{1\varepsilon}}{-L_{1W}}}$

Then

$$\dot{L} \leq \gamma\|X\|^2 \quad (40)$$

where $\gamma \leq max\{L_{1x}, L_{1u}, L_{1d}, L_{1W}\}$is a negative number.

4. SIMULATIONS AND DISCUSSION

To verify the correctness of the optimal tracking control algorithm based on ADP algorithm with Actor-Critic structure, we perform numerical simulation on MATLAB/Simulink software with the parameters of the WMR given in Table 1 and the designed parameters as follows:

$$\alpha_1 = 0.25, \alpha_2 = 0.01, \gamma = diag([1000,1000,1000,1,1,1,1,1])$$

Table 1. Wheel mobile robot parameters

Parameters	Value
Weight of the platform (m_G)	30 kg
Weight of each wheel (m_W)	1 kg
Inertial moment of the platform (I_G)	15.625 kgm^2
Inertial moment of each wheel (rotation axis - I_W)	0.1 kgm^2
Inertial moment of each wheel (diameter axis - I_D)	0.0025 kgm^2
Distance between the M and G (a)	0.3 m
Radius of the wheel shaft (b)	0.75 m
Radius of the wheel (r)	0.15 m

The simulation is executed under two kinds of trajectory: straight line and circle line. In each case, the trajectory in *xy*-reference frame, the position error in time domain, and the velocity error in time domain are illustrated. The simulation results are depicted in Figures 3, 4(a), 4(b), 5, 6(a), and 6(b).

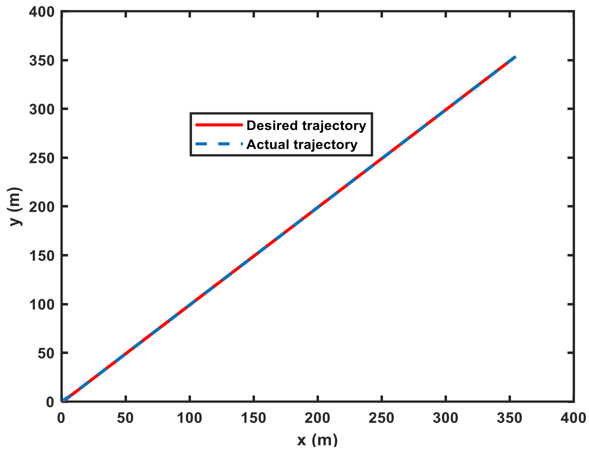
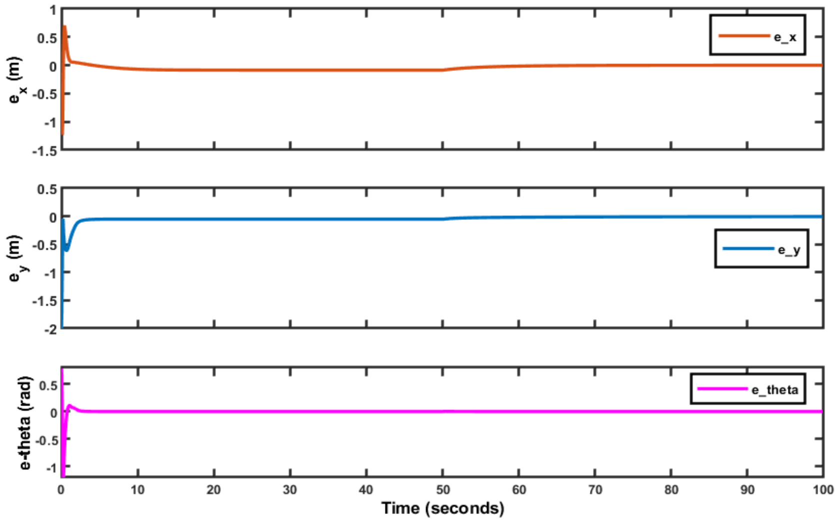
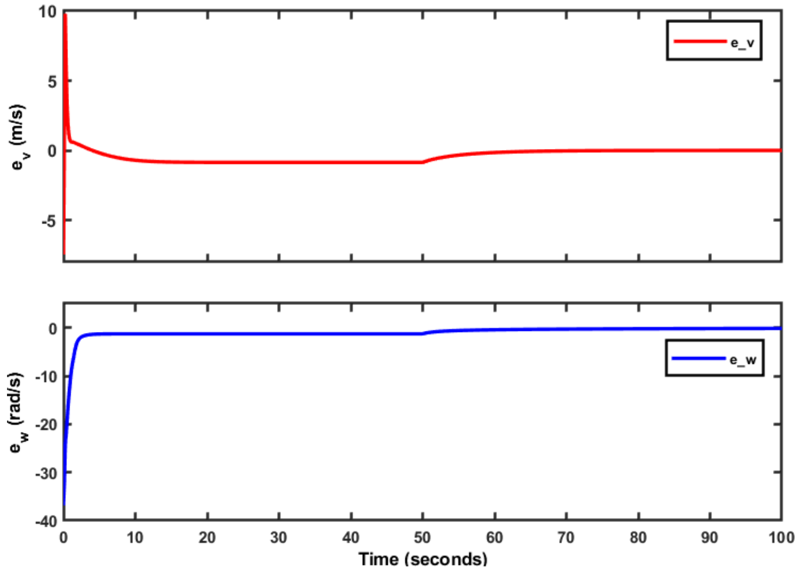


Figure 3. Response of the ADP controller with linear reference trajectory



(a)



(b)

Figure 4. Tracking error: (a) position tracking error and (b) velocity errors with straight-line reference trajectory

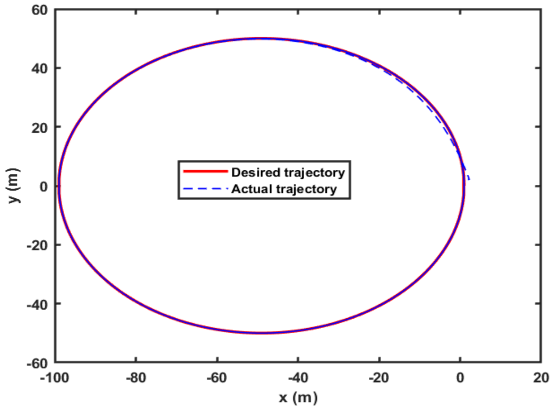
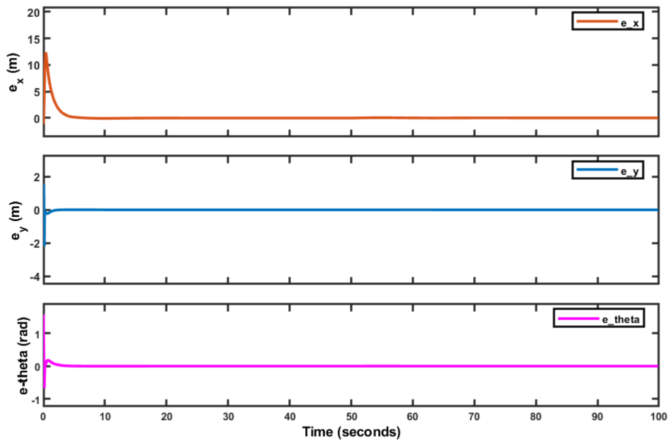
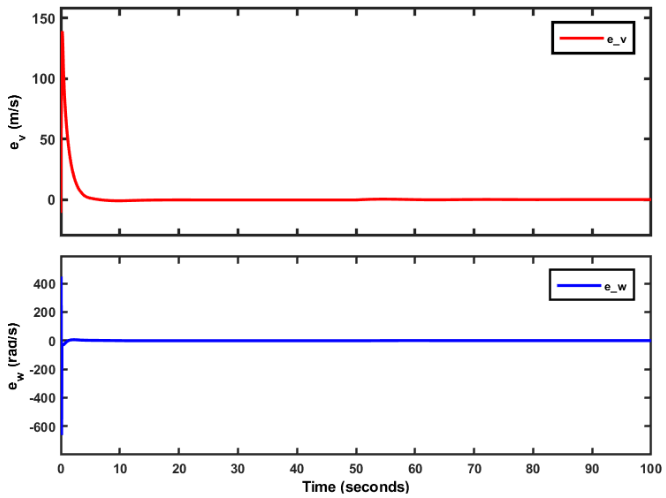


Figure 5. Response of the ADP controller with circle reference trajectory



(a)



(b)

Figure 6. Tracking error: (a) position tracking error and (b) velocity errors with circle reference trajectory

From the simulation results, it can be seen that in the first stage, the critic neural network is in the learning process, so the quality of the tracking is not good. However, after a period of 8 s, the optimal control law finishes the learning process and convergese to the optimal value. This leads to an increased quality of the WMR's tracking and the WMR follows the reference trajectory. The tracking in the *x* and *y* axis and the direction angle θ gives a large error at initial state; however, after the learning period, the tracking error is almost zero for all variables in both case of simulations.

5. CONCLUSION

The present study introduces a novel approach that combines online adaptive dynamic programming with a disturbance observer to address the challenge of robust optimization in the context of nonlinear systems. The proposed approach, featuring a singular neural network, yields superior results in terms of enhanced system quality and reduced computational overhead. The mathematical proof of the stability of the entire system, comprising the optimal controller and disturbance observer components, is established via Lyapunov theory. Ultimately, a simulation was conducted to assess the efficacy of the algorithm that was put forth. Results of the simulation indicate that the observer-based optimal adaptive dynamic programming methodology possesses the capability to yield a favorable response for the wheel mobile robot, even when confronted with instances of system uncertainties and external disturbances.

However, the above method still needs to know the internal dynamic information of the system to be able to update the controller parameters. In the next research direction, we use data about the state of the system to calculate a control algorithm that does not depend on the system's dynamic model.

REFERENCES

[1] S. G. Tzafestas, “Introduction to Mobile Robot Control,” *Introduction to Mobile Robot Control*, pp. 1–691, 2013, doi: 10.1016/C2013-0-01365-5.
[2] J. Meng, A. Liu, Y. Yang, Z. Wu, and Q. Xu, “Two-Wheeled Robot Platform Based on PID Control,” in *Proceedings - 2018 5th International Conference on Information Science and Control Engineering, ICISCE 2018*, IEEE, Jul. 2018, pp. 1011–1014, doi: 10.1109/ICISCE.2018.00208.
[3] T. Taniguchi and M. Sugeno, “Trajectory tracking controls for non-holonomic systems using dynamic feedback linearization based on piecewise multi-linear models,” *IAENG International Journal of Applied Mathematics*, vol. 47, no. 3, pp. 339–351, 2017.
[4] S. Rudra, R. K. Barai, and M. Maitra, “Design and implementation of a block-backstepping based tracking control for nonholonomic wheeled mobile robot,” *International Journal of Robust and Nonlinear Control*, vol. 26, no. 14, pp. 3018–3035, Sep. 2016, doi: 10.1002/rnc.3485.
[5] N. K. Goswami and P. K. Padhy, “Sliding mode controller design for trajectory tracking of a non-holonomic mobile robot with disturbance,” *Computers and Electrical Engineering*, vol. 72, pp. 307–323, Nov. 2018, doi: 10.1016/j.compeleceng.2018.09.021.
[6] A. Mustafa, N. K. Dhar, and N. K. Verma, “Event-Triggered sliding mode control for trajectory tracking of nonlinear systems,” *IEEE/CAA Journal of Automatica Sinica*, vol. 7, no. 1, pp. 307–314, Jan. 2020, doi: 10.1109/JAS.2019.1911654.
[7] O. Mohareri, R. Dhaouadi, and A. B. Rad, “Indirect adaptive tracking control of a nonholonomic mobile robot via neural networks,” *Neurocomputing*, vol. 88, pp. 54–66, Jul. 2012, doi: 10.1016/j.neucom.2011.06.035.
[8] A. Onat and M. Ozkan, “Dynamic adaptive trajectory tracking control of nonholonomic mobile robots using multiple models approach,” *Advanced Robotics*, vol. 29, no. 14, pp. 913–928, Jul. 2015, doi: 10.1080/01691864.2015.1014836.
[9] M. M. Fateh and A. Arab, “Robust control of a wheeled mobile robot by voltage control strategy,” *Nonlinear Dynamics*, vol. 79, no. 1, pp. 335–348, Jan. 2015, doi: 10.1007/s11071-014-1667-8.
[10] P. N. Chandra and S. J. Mija, “Robust controller for trajectory tracking of a Mobile Robot,” in *1st IEEE International Conference on Power Electronics, Intelligent Control and Energy Systems, ICPEICES 2016*, IEEE, Jul. 2017, pp. 1–6, doi: 10.1109/ICPEICES.2016.7853137.
[11] M. Abdelwahab, V. Parque, A. M. R. Fath Elbab, A. A. Abouelsoud, and S. Sugano, “Trajectory tracking of wheeled mobile robots using Z-Number based fuzzy logic,” *IEEE Access*, vol. 8, pp. 18426–18441, 2020, doi: 10.1109/ACCESS.2020.2968421.
[12] A. Štefek, V. T. Pham, V. Krivanek, and K. L. Pham, “Optimization of fuzzy logic controller used for a differential drive wheeled mobile robot,” *Applied Sciences (Switzerland)*, vol. 11, no. 13, p. 6023, Jun. 2021, doi: 10.3390/app11136023.
[13] P. Bozek, Y. L. Karavaev, A. A. Ardentov, and K. S. Yefremov, “Neural network control of a wheeled mobile robot based on optimal trajectories,” *International Journal of Advanced Robotic Systems*, vol. 17, no. 2, pp. 1-10, Mar. 2020, doi: 10.1177/1729881420916077.
[14] Z. Chen, Y. Liu, W. He, H. Qiao, and H. Ji, “Adaptive-Neural-Network-Based Trajectory Tracking Control for a Nonholonomic Wheeled Mobile Robot with Velocity Constraints,” *IEEE Transactions on Industrial Electronics*, vol. 68, no. 6, pp. 5057–5067, Jun. 2021, doi: 10.1109/TIE.2020.2989711.
[15] D. Wang and C. B. Low, “Modeling and analysis of skidding and slipping in wheeled mobile robots: Control design perspective,” *IEEE Transactions on Robotics*, vol. 24, no. 3, pp. 676–687, Jun. 2008, doi: 10.1109/TRO.2008.921563.
[16] C. B. Low and D. Wang, “GPS-based tracking control for a car-like wheeled mobile robot with skidding and slipping,” *IEEE/ASME Transactions on Mechatronics*, vol. 13, no. 4, pp. 480–484, Aug. 2008, doi: 10.1109/TMECH.2008.2000827.
[17] N. T. T. Vu, L. X. Ong, N. H. Trinh, and S. T. H. Pham, “Robust adaptive controller for wheel mobile robot with disturbances and wheel slips,” *International Journal of Electrical and Computer Engineering*, vol. 11, no. 1, pp. 336–346, Feb. 2021, doi: 10.11591/ijece.v11i1.pp336-346.
[18] V. Q. Ha, S. H. T. Pham, and N. T. T. Vu, “Adaptive Fuzzy Type-II Controller for Wheeled Mobile Robot with Disturbances and Wheelslips,” *Journal of Robotics*, vol. 2021, pp. 1–11, Sep. 2021, doi: 10.1155/2021/6946210.
[19] R. S. Sutton and A. G. Barto, “Reinforcement Learning: An Introduction,” *in Cambridge, MA: MIT press, 2018*, 2018.
[20] D. Liu, S. Xue, B. Zhao, B. Luo, and Q. Wei, “Adaptive Dynamic Programming for Control: A Survey and Recent Advances,” *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 51, no. 1, pp. 142–160, 2021, doi: 10.1109/TSMC.2020.3042876.
[21] 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, Apr. 2009, doi: 10.1016/j.neunet.2009.03.008.
[22] 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.
[23] H. He, Z. Ni, and J. Fu, "A three-network architecture for on-line learning and optimization based on adaptive dynamic programming," *Neurocomputing*, vol. 78, no. 1, pp. 3–13, 2012, doi: 10.1016/j.neucom.2011.05.031.
[24] S. Bhasin, R. Kamalapurkar, M. Johnson, K. G. Vamvoudakis, F. L. Lewis, and W. E. Dixon, "A novel actor-critic-identifier architecture for approximate optimal control of uncertain nonlinear systems," *Automatica*, vol. 49, no. 1, pp. 82–92, Jan. 2013, doi: 10.1016/j.automatica.2012.09.019.
[25] H. N. Wu and B. Luo, "Neural network based online simultaneous policy update algorithm for solving the HJI equation in nonlinear H\infty control," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 23, no. 12, pp. 1884–1895, 2012, doi: 10.1109/TNNLS.2012.2217349.
[26] Z. Zheng, L. Ruan, M. Zhu, and X. Guo, "Reinforcement learning control for underactuated surface vessel with output error constraints and uncertainties," *Neurocomputing*, vol. 399, pp. 479–490, 2020, doi: 10.1016/j.neucom.2020.03.021.
[27] J. Sun and C. Liu, "Disturbance observer-based robust missile autopilot design with full-state constraints via adaptive dynamic programming," *Journal of the Franklin Institute*, vol. 355, no. 5, pp. 2344–2368, 2018, doi: 10.1016/j.jfranklin.2018.01.005.
[28] R. Song and F. L. Lewis, "Robust optimal control for a class of nonlinear systems with unknown disturbances based on disturbance observer and policy iteration," *Neurocomputing*, vol. 390, pp. 185–195, 2020, doi: 10.1016/j.neucom.2020.01.082.
[29] W. H. Chen, "Disturbance observer based control for nonlinear systems," *IEEE/ASME Transactions on Mechatronics*, vol. 9, no. 4, pp. 706–710, 2004, doi: 10.1109/TMECH.2004.839034.
[30] J. Yang, W. H. Chen, and S. Li, "Non-linear disturbance observer-based robust control for systems with mismatched disturbances/uncertainties," *IET Control Theory and Applications*, vol. 5, no. 18, pp. 2053–2062, 2011, doi: 10.1049/iet-cta.2010.0616.
[31] K. Dupree, P. M. Patre, Z. D. Wilcox, and W. E. Dixon, "Asymptotic optimal control of uncertain nonlinear EulerLagrange systems," *Automatica*, vol. 47, no. 1, pp. 99–107, 2011, doi: 10.1016/j.automatica.2010.10.007.

BIOGRAPHIES OF AUTHORS

Hoa Van Doan was born in 1984. He graduated Engineering Degree majoring in Automatic Control at the Thai Nguyen University of Technology in 2008. Defense Master Degree in Control Engineering and Automation at Thai Nguyen University of Technology, Vietnam in 2012. Now, he works at the Department of Electrical Engineering, University of Economics-Technology for Industries. The primary research: adaptive control, robust control, fuzzy logic control, mobile robots, neural network, and artificial intelligence. He can be contacted at email: rvhoa@uneti.edu.vn.

Nga Thi-Thuy Vu received the B.Eng. degree in Electrical Engineering and the M.Sc. degree in Electrical Engineering School from Hanoi University of Science and Technology, Vietnam, in 2005 and 2009, respectively, and the Ph.D. degree in Engineering from Dongguk University, Korea, in 1996. She is currently an Associate Professor at Hanoi University of Science and Technology, Vietnam, lecturing undergraduates and postgraduates in control theory and applications at the School of Electrical and Electronics, Hanoi University of Science and Technology, Vietnam. Her research interest focuses on driven control, power electronics control, mechatronics system control, and renewable energy system. She can be contacted at email: nga.vuthithuy@hust.edu.vn.