International Journal of Electrical and Computer Engineering (IJECE)
Vol. 15, No. 5, October 2025, pp. 4542~4554
ISSN: 2088-8708, DOI: 10.11591/ijece.v15i5.pp4542-4554

Discount factor-based data-driven reinforcement learning cascade control structure for unmanned aerial vehicle systems

Ngoc Trung Dang, Quynh Nga Duong
Faculty of Electrical Engineering, Thai Nguyen University of Technology, Thai Nguyen, Vietnam

Article Info

Article history:

Received Oct 25, 2024
Revised Jun 18, 2025
Accepted Jul 12, 2025

Keywords:

Approximate/adaptive dynamic programming
Data reinforcement learning
Model-free based control
Quadrotor
Unmanned aerial vehicles

ABSTRACT

This article investigates the discount factor-based data-driven reinforcement learning control (DDRLC) algorithm for completely uncertain unmanned aerial vehicle (UAV) quadrotors. The proposed cascade control structure of UAV is categorized with two control loops of attitude and position sub-systems, which are established the proposed discount factor-based DDRLC algorithm. Through the analysis of the Bellman function's time derivative from two perspectives, a revised Hamilton-Jacobi-Bellman (HJB) equation including a discount factor is developed. Then, in the view of off-policy consideration, an equation is formulated to simultaneously solve the approximate Bellman function and approximate optimal control law in the proposed DDRLC algorithm with guaranteed convergence. According to the modified state variables vector, the development of the discount factor-based DDRLC algorithm in each control loop is indirectly implemented by transforming the time-varying tracking error model into the time invariant system. Finally, a simulation study on the proposed discount factor-based DDRLC algorithm is provided to validate its effectiveness. To validate the tracking performance of the quadrotor, four performance indices are considered, including $IAE_p = 3.0527$, $IAE_\Omega = 0.1175$, $ITAE_p = 1.8408$, and $ITAE_\Omega = 0.0144$, where the subscript p denotes position tracking error and Ω denotes attitude tracking error.

Corresponding Author:

Ngoc Trung Dang
Faculty of Electrical Engineering, Thai Nguyen University of Technology
3-2 Street, Tich Luong Commune, Thai Nguyen City, Vietnam
Email: trungcsktd@tnut.edu.vn

1. INTRODUCTION

In recent decades, unmanned aerial vehicles (UAVs) have been increasingly used to perform various tasks, such as surveillance, military, air traffic control, agriculture management [1]–[3]. To perform task effectively, it is often necessary to develop the trajectory tracking problem and optimal control performance. In practical application, these two control requirements are necessary to develop to the obstacle of external disturbance and dynamic uncertainties. Due to the complexity of UAV model with a high number of variables, an approach of model separation is considered with rotational and translational sub-systems [1]–[7]. In study [6], the control designs for position sub-system and attitude sub-system were implemented similarly by sliding mode control technique (SMC) and the addition of state observer, neural networks (NNs) were considered to handle the obstacle of external disturbance and dynamic uncertainties. Some extensions were developed for multi-rotor UAV model with unknown bounded time-varying disturbance by augmented disturbance observer (DO) based controller, which was implemented under the appointed-time prescribed performance (ATPP) technique [5]. In [2], an adaptive trajectory tracking control was proposed for UAV experiment systems after estimating the necessary variables based on image, inertia measurement. Moreover,

for general robotics control design studied in [8], output feedback law with state observer was presented for surface vessels (SVs) according to event-triggered rule. In order to further handle Backlash-Like hysteresis and external disturbance, an adaptive fuzzy dynamic memory-event-triggered mechanism was studied for a six-rotor UAV by Backstepping recursive framework with the first-order filtering technique [2]. But as far as we known, it can be found that there is little research attention on the optimal control UAV systems.

With the complexity of UAV model and the diversification of practical tasks, it is difficult to obtain the control objectives of complex purposes only relying on a single UAV agent. Hence, UAV researches put forward the concept of multi-agent systems (MASs), which involves two research hotspots of consensus and formation control problems [1], [3], [7], [9]–[12]. In [9], a consensus control law was developed for multiple UAV systems with time delay and cascade model. However, the Kronecker product and Linear Matrix Inequalities (LMIs) were implemented in [9] due to the simplification of UAV model. The research conducted by [13] is concerned with the consensus controller with the sign function. Hence, the stability consideration requires the Fillipov theory employment. Additionally, the bearing persistence of excitation (PE) based leader-follower formation control strategy was proposed for multiple double-integrators in three dimensional (3D) space using the projection of vector on the plane orthogonal to 2-sphere [14]. When each agent was considered more complicated with Euler-Lagrange systems, the state representation can be used to obtain the event-triggered based consensus controller with Kronecker product [15]. The fault-tolerant consensus control problem for nonstrict-feedback nonlinear MASs with intermittent actuator faults was investigated state observer and backstepping technique [16]. Moreover, the formation control of multiple UAVs was also considered by model predictive control (MPC) with the affine tracking error model [17]–[19]. Despite this, studies [17]–[19] did not examine the stability properties of the closed-loop system when operating under MPC framework. For the formation tracking control problem, addressing the time-varying formation (TVF) is also extremely crucial for meeting application requirement [1], [7], [10], [11]. According to linear UAV model, the TVF tracking control was investigated by Kronecker product consideration and LMIs technique [7]. Although the cost function was mentioned in [7] but the optimal control law has not been studied in this work. On the other hand, extended observer (ESO) based backstepping controller was proposed in the second-order attitude sub-system [1]. Furthermore, the estimation of yaw angle in virtual leader was carried out with the connection to the time-varying communication topology as well as the distributed formation tracking control was addressed in the position sub-system [1]. Based on the linear model of fixed-swing UAVs, the TVF tracking control was discussed by employing the solution of Riccati equation [10]. Notably, [11] tackled the TVF tracking control for multiple linear systems by extending Event-Triggered mechanism. Although there has been some research on the distributed control schemes for MASs especially the consensus and formation systems, most of the recent references have focused on simple UAV model and rarely considered the cascade UAV structure as well as optimization-based control formulation.

Implementing the optimal control law in real-world systems requires the use of iterative algorithms to compute solutions to the Hamilton-Jacobi-Bellman (HJB) equations for nonlinear systems or Riccati equations for linear systems, since analytical solutions are typically not feasible. To advance the implementation of optimal control in robotic systems, it is essential to incorporate reinforcement learning control (RLC) in conjunction with methods from approximate and adaptive dynamic programming (ADP), as highlighted in studies [12], [20]–[27]. In [12], [20]–[22], the actor/critic structure was realized via neural network (NN) approximation methods, with learning algorithms for weight adaptation proposed alongside optimization strategies, enabling the closed-loop system to satisfy both tracking performance and optimality requirements. However, it is necessary to eliminate external disturbance and dynamic uncertainties in the practical model, which are handle by traditional robust control design [12], [20]–[22]. A different approach of handling directly the external disturbance and dynamic uncertainties in optimal control law can be known in zero and non-zero sum game methods [28]–[30]. On the other hand, it is different from the simultaneous learning in actor/critic framework in [12], [20]–[22], authors in [31], [32] developed the sequential learning value iteration (VI) algorithm to obtain the Bellman function and optimal control law. Some researchers focused on using data-driven RL to obtain the optimal control strategies for uncertain systems [6], [22], [28], [30], [33]–[37]. According to the data collection in time interval, the approximate optimal function can be computed from the approximate optimal control input without the knowledge of model. However, to handle the complete uncertainty in the inverse direction, the addition of off-policy technique or Q-learning is necessary to consider [2], [36], [37]. A data-driven reinforcement learning control strategy was recently introduced for quadrotors, demonstrating the capability to achieve optimal control while ensuring trajectory tracking, which is closely related to the focus of this article [37]. However, the data-driven RL approach in [37] was applied solely to the attitude subsystem of a UAV, and the associated cost function did not incorporate a discount factor. On account of the above results, we will further explorer the cascade UAV control structure, which involves two data-driven RL with a discount factor-based performance index, and this is another interest of this study.

This study investigates a cascade control architecture for a fully uncertain quadrotor UAV by employing two data-driven RL algorithms based on a performance index with a discount factor. Through constructing a data set tailored to this general class of affine continuous-time systems and integrating a RL strategy using an off-policy algorithm, a control framework is formulated for UAVs with unknown dynamics. The summary contributions of this study are given in the following:

a. Based on the optimal control scheme with a discount factor-based performance index, we further introduce a RL algorithm for an affine continuous-time system to guarantee the finite value of the integral cost function with infinity terminal.
b. We propose a novel data-driven RL based cascade control structure in both two sub-systems for completely uncertain UAVs by off-policy method. Compared with the current results [37], only considering the RL algorithm for the attitude sub-system without discount factor, a data-driven RL based cascade control structure is first proposed for completely uncertain UAVs with a discount factor-based performance index. Finally, simulation results are presented to validate the effectiveness of the proposed model-free, data-driven RL algorithm.

2. CONTROLLER METHODOLODY FOR QUADROTOR

As shown in Figure 1, the Earth-fixed frame and the body-fixed frame are established to describe the dynamic model of the quadrotor. The movements of this Quadrotor as shown in Figure 1 can be established by changes on four lift forces, which are generated by adjusting the angle velocities of four rotors. It can be seen that a vertical movement can be obtained by the variation of the sum of four lift forces on the four rotors. Due to the difference between the counter-torques achieved by the group of rotors (Rotor 1 and Rotor 3) and the group of rotors (Rotor 2 and Rotor 4), the yaw movement is established. Additionally, the pitch and roll movements can be generated by changing the lift forces of each pair, which result in the longitudinal motion and the lateral motion, as shown in Figure 1. The position of the UAV quadrotor and the quadrotor attitude are given as $r = [r_x, r_y, r_z]^T \in \mathbb{R}^3$and $\Omega = [\phi, \theta, \psi]^T \in \mathbb{R}^3$, respectively. It is worth noting that Euler angles Roll-Pitch-Yaw are satisfied the bound condition as $-\pi/2 < \phi < \pi/2$, $-\pi/2 < \theta < \pi/2$ and $-\pi < \psi < \pi$. Moreover, the UAV quadrotor parameters are expressed in Table 1.

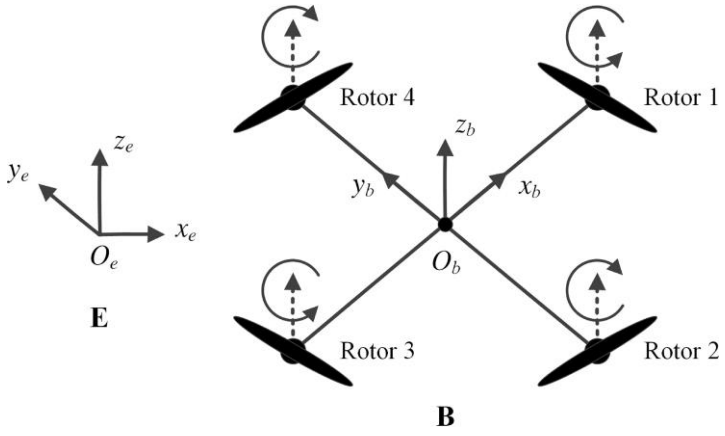

Figure 1. Quadrotor model in North-East-Down (NED) coordinate

Table 1. UAV parameters and variables

UAV parameters	variables
m	Weight of the quadrotor
g	Acceleration of the gravity
$\omega_1, \omega_2, \omega_3, \omega_4$	Angle velocity of each rotor
l	The arm length
$J = diag\{J_\phi, J_\theta, J_\psi\} \in \mathbb{R}^{3\times3}$	The inertia matrix is symmetric and positive definite
k_f, l_c, k_τ	Positive parameters

The rotation matrix $R \in SO(3)$ representing the transformation from the Earth-fixed frame to the body-fixed coordinate system is given as (1):

$$R^T = \begin{bmatrix} c_\theta c_\psi & c_\theta s_\psi & -s_\theta \\ s_\phi s_\theta c_\psi - c_\phi s_\psi & s_\phi s_\theta s_\psi + c_\phi c_\psi & s_\phi c_\theta \\ c_\phi s_\theta c_\psi + s_\phi s_\psi & c_\phi s_\theta s_\psi - s_\phi c_\psi & c_\phi c_\theta \end{bmatrix} \quad (1)$$

where $c(\bullet) = cos(\bullet), s(\bullet) = sin(\bullet)$.

In the view of [1], the complete quadrotor dynamic model can be represented as (2):

$$\begin{aligned} m\ddot{r} &= Rf \\ J\ddot{\Omega} &= C_{(\Omega,\dot{\Omega})}\dot{\Omega} + \tau \end{aligned} \quad (2)$$

where the parameters are given in Table 1 and the Coriolis matrix $C_{(\Omega,\dot{\Omega})} \in \mathbb{R}^{3\times3}$ is described in [2]. Additionally, the force $f \in \mathbb{R}^{3\times1}$ is relative to the body fixed frame of the quadrotor can be obtained as (3):

$$f = \begin{bmatrix} 0 & 0 & \overline{f} \end{bmatrix}^T - R^T \begin{bmatrix} 0 & 0 & mg \end{bmatrix}^T \quad (3)$$

where the lifting force $\overline{f} \in \mathbb{R}$ and the torque $\tau = [\tau_\phi \quad \tau_\theta \quad \tau_\psi] \in \mathbb{R}^3$are given as (4), (5):

$$\overline{f} = k_f(\omega_1^2 + \omega_2^2 + \omega_3^2 + \omega_4^2) \quad (4)$$

$$\tau_\phi = l_c k_f(\omega_2^2 - \omega_4^2), \tau_\theta = l_c k_f(\omega_1^2 - \omega_3^2), \tau_\psi = k_\tau(\omega_1^2 - \omega_2^2 + \omega_3^2 - \omega_4^2) \quad (5)$$

In where, the control signals of the quadrotor (2) are defined as (6):

$$\begin{aligned} u_f &= \omega_1^2 + \omega_2^2 + \omega_3^2 + \omega_4^2, \\ u_\phi &= \omega_2^2 - \omega_4^2, u_\theta = \omega_1^2 - \omega_3^2 \\ u_\psi &= \omega_1^2 - \omega_2^2 + \omega_3^2 - \omega_4^2. \end{aligned} \quad (6)$$

The control objective of this paper is to develop a data-driven RL algorithm based on the optimal control scheme to achieve an optimized tracking control law for a quadrotor, enabling the quadrotor to effectively track the desired trajectory with high accuracy. The optimal control signal ensures trajectory tracking while simultaneously achieving approximate optimality by minimizing the objective function. Additionally, the data-driven RL-based optimal control law is developed for not only the position sub-system but also the attitude sub-system without the knowledge of the UAV model.

Remark 1. Unlike the conventional trajectory tracking control purpose in UAV control systems [1], [3], [6], [7], [11], the control objective in this paper considers both the trajectory tracking performance and the optimal control problem. In addition, both subsystems as shown in Figure 2 achieve a unified framework of optimal control and stability, which is typically difficult to attain due to the time-varying dynamics of the closed-loop systems.

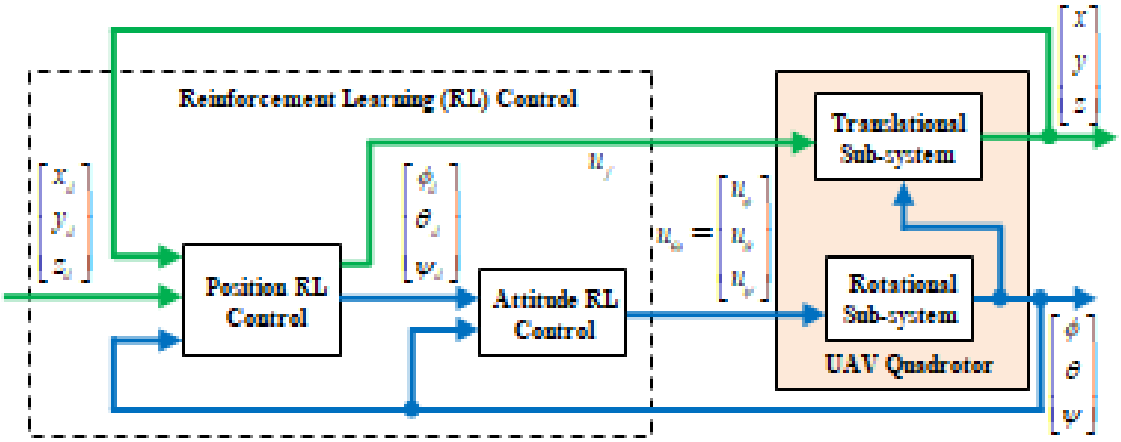

Figure 2. The quadrotor control schematic

In this section, a data-driven reinforcement learning approach is introduced to address the trade-off between tracking performance and optimality within the quadrotor control system. The control architecture illustrated in Figure 2 integrates both position and attitude control strategies under the application of a discount factor. These controllers are updated concurrently using the collected data to handle system uncertainties effectively.

2.1. Discount factor-based RL control design for augmented quadrotor system

First of all, we consider a nonlinear affine system as (7):

$$\frac{d}{dt}\eta(t) = f\big(\eta(t)\big) + g\big(\eta(t)\big)u(t). \tag{7}$$

and the associated cost function is defined by (8):

$$V(\eta(t), u(t)) = \int_t^{\infty}\left[\eta(\tau)^T\overline{Q}\eta(\tau) + u(\tau)^T\overline{R}u(\tau)\right]d\tau. \tag{8}$$

where $\overline{Q} \in \mathbb{R}^{n\times n} > 0, \overline{R} \in \mathbb{R}^{n\times n} > 0$ are both symmetric positive definite matrices. The tracking error model of nonlinear affine systems (7) with the desired trajectory $\eta_d(t)$, which is established by a command generator model $\frac{d}{dt}\eta_d(t) = h\big(\eta_d(t)\big), h(0) = 0$, can be formulated as (9):

$$\frac{d}{dt}e(t) = f\big(\eta(t)\big) - h\big(\eta_d(t)\big) + g\big(\eta(t)\big)u(t). \tag{9}$$

where $e(t) = \eta(t) - \eta_d(t)$, $h\big(\eta_d(t)\big)$is the unknown function. Hence, according to tracking error model (9) and the command generator model $h\big(\eta_d(t)\big)$, we achieve the following augmented system:

$$\frac{d}{dt}\zeta(t) = F\big(\zeta(t)\big) + G\big(\zeta(t)\big)u(t). \tag{10}$$

where

$$\zeta(t) = [e(t) \quad \eta_d(t)]^T, F\big(\zeta(t)\big) = \begin{bmatrix} f\big(e(t)+\eta_d(t)\big) - h\big(\eta_d(t)\big) \\ h\big(\eta_d(t)\big) \end{bmatrix}, G\big(\zeta(t)\big) = \begin{bmatrix} g\big(e(t)+\eta_d(t)\big) \\ 0 \end{bmatrix} \tag{11}$$

The optimal control law $u^*(t)$ is designed to minimize the discounted cost function associated with the augmented system (10).

$$V(\zeta(t), u(t)) = \int_t^{\infty} e^{-\lambda(\tau-t)}\, U(\zeta(\tau), u(\tau))d\tau, \tag{12}$$

where $\lambda > 0$is a discount factor, $U(\zeta(\tau), u(\tau)) \triangleq \zeta(\tau)^T Q\zeta(\tau) + \big(u(\tau)\big)^T Ru(\tau)$, $Q = \begin{bmatrix} \overline{Q} & 0 \\ 0 & 0 \end{bmatrix}$ and $R = \overline{R}$. The addition of the discount factor λ in the cost function (12) is able to guarantee that it will be finite value although the integral terminal is infinity. Therefore, it is unnecessary to explicitly define the admissible control set, as discussed in [2]. The set $\Upsilon(U)$ is defined as the constraint set of control input $u(\zeta)$ such that the cost function (12) is finite. Based on the dynamic programming principle, the tracking Bellman function for the augmented system (10) can be expressed as the following static function:

$$V^*(\zeta(t)) = \min_{u(\zeta(t))\in\Upsilon(U)} V\big(\zeta(t), u(\zeta(t))\big) \tag{13}$$

Based on two approaches for computing the time derivative of the Bellman function $V^*(\zeta(t))$ in (13), the associated Hamiltonian function under a discount factor $\lambda > 0$ is formulated. The first approach involves a direct computation, as detailed:

$$\frac{d}{dt}V^*(\zeta(t)) = \frac{\partial V^*}{\partial \zeta}\frac{d\zeta}{dt} = \frac{\partial V^*}{\partial \zeta}(F(\zeta(t)) + G(\zeta)u^*(t)). \tag{14}$$

where $u^*(t)$ denotes the optimal control input. According to the Bellman principle, a second approach for computing the time derivative of the Bellman function $V^*(\zeta(t))$is formulated by utilizing the static Bellman function in (13):

$$V^*(\zeta(t)) = \int_t^{t+\delta} e^{-\lambda(\tau-t)}\, U\big(\zeta(\tau), u^*(\tau)\big)d\tau + e^{-\lambda\delta}\int_{t+\delta}^{\infty} e^{-\lambda(\tau-(t+\delta))}\, U\big(\zeta(\tau), u^*(\tau)\big)d\tau$$
$$= \int_t^{t+\delta} e^{-\lambda(\tau-t)}\, U\big(\zeta(\tau), u^*(\tau)\big)d\tau + e^{-\lambda\delta}V^*(\zeta(t+\delta)) \quad (15)$$

The representation (15) obtains that:

$$\frac{V^*(\zeta(t))-V^*((\zeta(t+\delta))}{\delta} = \frac{1}{\delta}\int_t^{t+\delta} e^{-\lambda(\tau-t)}\, U\big(\zeta(\tau), u^*(\tau)\big)d\tau + \frac{(e^{-\lambda\delta}-1)}{\delta}V^*(\zeta(t+\delta)). \quad (16)$$

In the view of (16) and (14) as $\delta \to 0$, we achieve that the static Bellman function $V^*(\zeta(t))$ can be soved by the optimal control signal $u^*(t)$ using the following partial derivative equation as (17)

$$U\big(\zeta(t), u^*(t)\big) - \lambda V^*((\zeta(t)) + \frac{\partial V^*}{\partial \zeta}(F(\zeta(t)) + G(\zeta)u^*(t)) = 0. \quad (17)$$

Conversely, to determine the optimal control input $u^*(t)$ using the static Bellman function $V^*((\zeta(t))$ and based on the Bellman principle, the corresponding optimization problem can be formulated as (18):

$$V^*\big(\zeta(t)\big) = \min_{u(s)\in\Upsilon(U)} \Big(\int_t^{t+\delta} U(\zeta(s), u(s))ds + e^{-\lambda\delta}V^*((\zeta(t+\delta))\Big) \quad (18)$$

Since $\delta \to 0^+$, (18) leads to the corresponding optimization problem as (19):

$$\min_{u(t)(\zeta)\in\Upsilon(U)} [U(\zeta(t), u(t)) - \lambda V^*(\zeta(t)) + \frac{\partial V^*}{\partial \zeta}(F(\zeta(t)) + G(\zeta(t))u(t))] = 0. \quad (19)$$

Defining the modified Hamiltonian function in the presence of a discount factor $\lambda > 0$ as (19),

$$H(\zeta, u(t), \nabla V, V) = (\zeta(t))^T Q\zeta(t) + (u(t))^T R\, u(t) - \lambda V(\zeta(t)) + \nabla V^T(\zeta(t))(F(\zeta(t)) + G(\zeta(t))u(t)) \quad (20)$$

where $\nabla V(\zeta) \triangleq \frac{\partial V(\zeta)^T}{\partial \zeta}$, it follows that the optimal control solution is then obtained by (19) as (20),

$$u^*(\zeta(t)) = \underset{u\in\Upsilon(\Omega)}{\operatorname{argmin}}\big[H\big(\zeta, u(t), \nabla V^*(\zeta(t))\big)\big] = -\frac{1}{2}R^{-1}G^T(\zeta(t))\nabla V^*(\zeta(t)) \quad (21)$$

Additionally, substituting the optimal control law $u^*(\zeta(t))$ (21) into (19), it implies the partial derivative equation (PDE) is expressed as (22):

$$H^*\big(\zeta(t), u^*(t), \nabla V^*, V^*(t)\big) = \zeta(t)^T Q\zeta(t) - \frac{1}{4}\nabla V^{*T}\big(\zeta(t)\big)G\big(\zeta(t)\big)R^{-1}G^T\big(\zeta(t)\big)\nabla V^*\big(\zeta(t)\big) - \lambda V^*\big(\zeta(t)\big) + \nabla V^{*T}\big(\zeta(t)\big)F\big(\zeta(t)\big) = 0. \quad (22)$$

Remark 2. Including a positive discount factor $\lambda > 0$ ensures that the cost function in (8) remains finite, even when the state variable $\eta(t)$ does not converge to zero as $t \to \infty$. This consideration leads to the appearance of the term "$\lambda V^*(\zeta)$ " in (19) resulting in necessary adjustments within the discount factor-based RL control framework described in sections 2.2 and 2.3.

2.2. Data-driven proportional-integral position controller

In this section, a cascade control framework for a quadrotor UAV as shown in Figure 2 is formulated following the model separation in (2), where each subsystem applies a discount factor-based optimal control approach. However, due to the inherent uncertainties and nonlinearities present in (22), obtaining a direct analytical solution is infeasible. As a result, a data-driven RL algorithm is employed to estimate the static Bellman function $V^*(\zeta)$ corresponding to the optimal control policy $u^*(\zeta)$ for each subsystem.

The dynamic model of the position sub-system (2) can be modified as (23):

$$\ddot{r} = \frac{1}{m}k_f u_f R[0 \quad 0 \quad 1]^T - g[0 \quad 0 \quad 1]^T = \frac{1}{m}k_f u_r \quad (23)$$

where $u_r = u_f R[0 \quad 0 \quad 1]^T - \frac{mg}{k_f}[0 \quad 0 \quad 1]^T$. For developing the control design of the position sub-system (23), the tracking error model is necessary to made with the time invariant model as shown in (7). Therefore, the state variables vector $x_r = (r_x, \dot{r}_x, r_y, \dot{r}_y, r_z, \dot{r}_z)^T \in \mathbb{R}^6$is applied to reduce the order of (23). Hence, the model (23) can be transformed into the first order system as (24):

$$\dot{x}_r = A_r x_r + B_r u_r \tag{24}$$

where

$$A_r = diag(a_r, a_r, a_r) \in \mathbb{R}^{6\times 6}, a_r = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} \text{ and } B_r = \frac{k_f}{m}\begin{bmatrix} 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}^T.$$

Moreover, due to the time varying of the desired trajectory $r^{ref}(t) = \left[r_x^{ref}(t), r_y^{ref}(t), r_z^{ref}(t)\right]^T \in \mathbb{R}^3$, to transform the tracking error model of the position sub-system (24) into the time invariant model (7), it is necessary to utilize the following assumptions:

Assumption 1. The desired trajectory $r^{ref}(t) = \left[r_x^{ref}(t), r_y^{ref}(t), r_z^{ref}(t)\right]^T \in \mathbb{R}^3$ is bounded and its time derivative $\frac{d}{dt}r^{ref}(t)$ is the Lipschitz function.

Assumption 2. The reference vector $x_r^{ref} = \left[r_x^{ref}, \dot{r}_x^{ref}, r_y^{ref}, \dot{r}_y^{ref}, r_z^{ref}, \dot{r}_z^{ref}\right]^T \in \mathbb{R}^6$ can be completely expressed as (25),

$$\frac{d}{dt}x_r^{ref}(t) = A_{rd}x_r^{ref}(t) \tag{25}$$

Therefore, in the view of (24) and (25), it obtains the time invariant model (7) as:

$$\frac{d}{dt}X_r = \begin{bmatrix} \dot{e}_r \\ \dot{x}_r^{ref} \end{bmatrix} = \begin{bmatrix} A_r & A_r - A_{rd} \\ 0_{6,6} & A_{rd} \end{bmatrix} X_r + \begin{bmatrix} B_r \\ 0_{6,3} \end{bmatrix} u_r$$

where

$$e_r = x_r - x_r^{ref}, X_r = \begin{bmatrix} e_r \\ x_r^{ref} \end{bmatrix} \tag{26}$$

The tracking cost function is modified as (27):

$$V_r\left(X_r(t)\right) = \int_t^\infty e^{-\lambda(s-t)} \times [X_r(s)^T Q_r X_r(s) + u_r(s)^T R_r u_r(s)]ds \tag{27}$$

where $Q_r = \begin{bmatrix} Q_{er} & 0_{6,6} \\ 0_{6,6} & 0_{6,6} \end{bmatrix}$ and $Q_{er} \in \mathbb{R}^{6\times 6}$, $R_r \in \mathbb{R}^{3\times 3}$ are symmetric matrices with positive definiteness. Note that, the term $e^{-\lambda(\tau-t)}$ is added to (27) for ensuring the finite cost function while $X_r = \begin{bmatrix} e_r \\ x_r^{ref} \end{bmatrix}$ does not converge to zero as time approaches infinity. According to (17)-(21) and the off-policy technique [3], the data-driven algorithm is proposed to develop the position controller as follows:

Algorithm 1. Data-driven algorithm for position controller

1. Initialization: Employing the stabilizing policy $u_r^0(X_r)$ and the additional noise $e_r(t)$ to satisfy PE condition. Collecting the input-output data in the quadrotor system and establishing the threshold ϵ_r
2. Policy evaluation: Based on the control input $u_r^i(X_r) = \hat{u}_r^i(X_r) + e_r$ and the control policy $\hat{u}_r^i(X_r)$, we solve the (28) to find simultaneously $V_r^{i+1}(X_r)$ and $u_r^{i+1}(X_r)$:

$$V_r^{i+1}\left(X_r(t+\Delta)\right) - e^{\lambda\Delta}V_r^{i+1}\left(X_r(t)\right) = -\int_t^{t+\Delta} e^{-\lambda(\tau-t-\Delta)}(X_r(\tau)^T Q_r X_r(\tau) + (\hat{u}_r^i)^T R_r \hat{u}_r^i + 2\hat{u}_r^i R_r e_r)\, d\tau; \hat{u}_r^i(t) = u_r^i(t) + e_r(t) \tag{28}$$

3. Policy improvement: Obtain the control policy $u_r^i(X_r) = u_r^{i+1}(X_r), i \to (i+1)$ and go to step 2 until $\|u_r^{i+1} - u_r^i\| < \epsilon_r$.

In the Algorithm 1, the solution of Bellman equation (24) is improved by data collection by the following modification:

$$V_r^{i+1}(X_r(t+\Delta)) - V_r^{i+1}(X_r(t)) = -\int_t^{t+\Delta}\left(X_r{}^T(\tau)Q_rX_r(\tau) + (u_r^i)^T\big(X_r(\tau)\big)R_ru_r^i\big(X_r(\tau)\big)\right)d\tau + \int_t^{t+\Delta}\lambda\, V_r^{i+1}\big(X_r(\tau)\big)d\tau + 2\int_t^{t+\Delta}\left(u_r^{i+1}\big(X_r(\tau)\big)\right)^T R_ru_r^i\big(e_r(\tau)\big)d\tau \quad (29)$$

After achieving the position control signal u_r in the quadrotor control structure as shown in Figure 2, we proceed to compute the reference of attitude control scheme $[\phi_d \quad \theta_d \quad \psi_d]^T$ as follows. According to $u_r = u_f R[0 \quad 0 \quad 1]^T - \frac{mg}{k_f}[0 \quad 0 \quad 1]^T$, it follows that:

$$u_r + \frac{mg}{k_f}[0 \quad 0 \quad 1]^T = u_f\begin{bmatrix}(sin\,\phi)(sin\,\psi) + (cos\,\phi)(cos\,\psi)(sin\,\theta)\\(cos\,\phi)(sin\,\theta)(sin\,\psi) - (cos\,\psi)(sin\,\phi)\\(cos\,\phi)(cos\,\theta)\end{bmatrix} \quad (30)$$

By setting the yaw angle reference $\psi_d(t)$ as a constant number to synchronize in practical applications, based on (30), we can achieve the desired u_f, ϕ_d, θ_d as (31):

$$\begin{aligned} u_f &= \frac{\left(u_{rz} + \frac{mg}{k_f}\right)}{(cos\,\phi)(sin\,\theta)}\\ \phi_d &= arcsin\left(\frac{u_{rx}\,sin\,\psi - u_{ry}cos\psi}{u_f}\right),\\ \theta_d &= arcsin\left(\frac{u_{rx}cos\psi + u_{ry}sin\psi}{u_f\,cos\,\psi}\right). \end{aligned} \quad (31)$$

2.3. Data-driven RL based attitude controller

In this part, a data-driven RL-based attitude control law is similarly designed as above to obtain the input signals u_Ω for satisfying optimal tracking performance with the desired trajectory (31). The attitude dynamic model (2) can be rewritten by (32):

$$\ddot{\Omega} = J^{-1}\tau - J^{-1}C_{(\Omega,\dot{\Omega})}\dot{\Omega} \quad (32)$$

By considering the attitude state vector $x_\Omega = [\phi, \dot{\phi}, \theta, \dot{\theta}, \psi, \dot{\psi}]^T$ and referring to the attitude control structure illustrated in Figure 2, the design approach mirrors the position control strategy described in subsection 2.3. Based on (32), the augmented attitude dynamics can be reformulated as (33):

$$\frac{d}{dt}X_\Omega = \begin{bmatrix}\dot{e}_\Omega\\ \dot{x}_{\Omega d}\end{bmatrix} = \begin{bmatrix}F_\Omega & F_\Omega - F_{\Omega d}\\ 0_{6,6} & F_{\Omega d}\end{bmatrix}X_{\Omega d} + \begin{bmatrix}G_\Omega\\ 0_{6,3}\end{bmatrix}u_\Omega \quad (33)$$

Accordingly, the attitude control strategy is summarized in the Algorithm 2:

Algorithm 2. Data-driven RL based attitude control scheme

1. Initialization: Employing the stabilizing policy $u_s^0(X_s)$ and the additional noise $u_{se}(t)$ to satisfy PE condition. Collecting the input-output data of the quadrotor system.
2. Policy evaluation: Based on the control signal $u_s^i(X_s) = \hat{u}_s^i(X_s) + e_s$ and the control policy $u_s^i(X_s)$, we solve the (34) to find simultaneously $V_s^{i+1}(X_s)$ and $u_s^{i+1}(X_s)$:

$$V_s^{i+1}(X_s(t+\Delta)) - V_s^{i+1}(X_s(t)) = -\int_t^{t+\Delta}\left(X_s{}^T(\tau)Q_sX_s(\tau) + (u_s^i)^T\big(X_s(\tau)\big)R_su_s^i\big(X_s(\tau)\big)\right)d\tau + \int_t^{t+\Delta}\lambda\, V_s^{i+1}\big(X_s(\tau)\big)d\tau + 2\int_t^{t+\Delta}\left(u_s^{i+1}\big(X_s(\tau)\big)\right)^T R_su_s^i\big(e_s(\tau)\big)d\tau \quad (34)$$

3. Policy improvement: Obtain the control policy $u_s^i(X_s) = u_s^{i+1}(X_s), i \to (i+1)$ and go to step 2 until $\|u_s^{i+1} - u_s^i\| < \epsilon_s$

Remark 3. Two data-driven RL algorithms incorporating a discount factor are proposed for the quadrotor, addressing both the attitude and position subsystems. This work extends the study in [37], which focused solely on RL control for the attitude subsystem without considering a discount factor.

3. SIMULATION RESULTS

In this section, we use the example of quadrotor to illustrate the proposed data RL algorithm with the following parameter as follows:

$$m = 2.0(kg), k_w = 1(Ns^2), k_t = 1\left(\frac{Ns^2}{m}\right), g = 9.8\left(\frac{m}{s^2}\right), l_\tau = 0.2(m),$$
$$J = 10^{-3}diag(5.1,5.1,5.2)(kg.m^2).$$

The desired trajectory of the position controller is chosen as: $r_d(t) = [0.5t,\ 0.5t,\ 1.5 + t]^T$, it can be obtained that the (25) is guaranteed with matrix

$$A_{rd} = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

Moreover, the cost function utilizes the weight matrices $Q_e = 100I_6$, $R_p = I_3$, $Q_{\theta e} = 100I_6$, $R_\theta = I_3, \lambda = 0.01, T_{step} = 0.01$, and a discount factor of $\lambda = 0.01$. During the initial data collection phase [17], two proportional-derivative (PD) controllers are applied to the position and attitude loops to gather data for the learning process. To ensure the persistence of excitation (PE) conditions required for the proposed algorithms, noise signals defined as $u_{pe} = \sum_{m=1}^{100} 0.01sin(w_m t)$ and $u_{\theta e} = \sum_{m=1}^{500} 0.002sin(w_m t)$, where each w_m is randomly selected within [−100, 100], are injected into the position and attitude control inputs, respectively. For the critic and actor neural networks, second-order and first-order polynomial activation functions are employed, respectively. It is worth noting that the tracking performance of the proposed data-driven RL-based position and attitude controllers is illustrated in Figures 3 to 7, demonstrating fast convergence with only four iterations required for the algorithm weights to stabilize. Moreover, the position tracking errors converge to zero within 4 seconds, while the attitude tracking errors reach zero in approximately 0.5 seconds, as illustrated in Figures 3 and 5, respectively. Furthermore, Figure 7 demonstrates the quadrotor's trajectory tracking performance relative to a predefined reference path, showing that the quadrotor's position closely follows the reference trajectory with high accuracy. Furthermore, to evaluate the effectiveness of the tracking performance, numerous performance indices, including the integral of absolute error (IAE) and the integral of absolute time-weighted error (IATE), are presented as shown in Table 2.

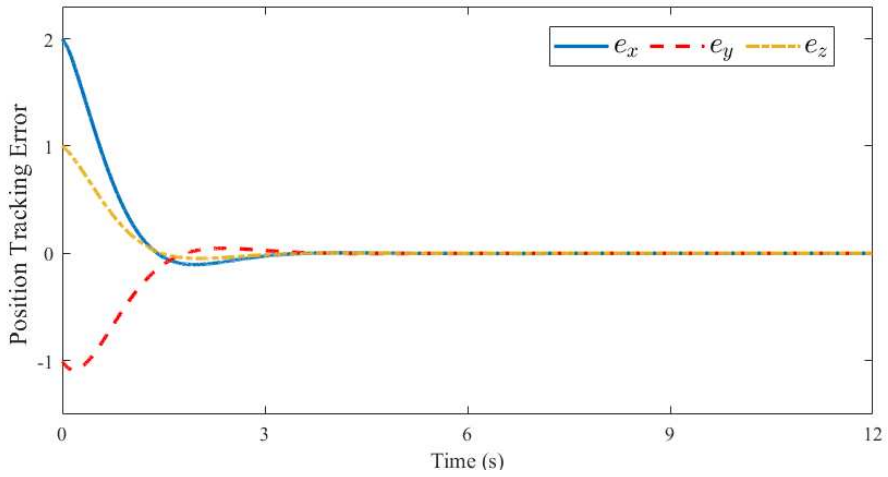

Figure 3. The position tracking error

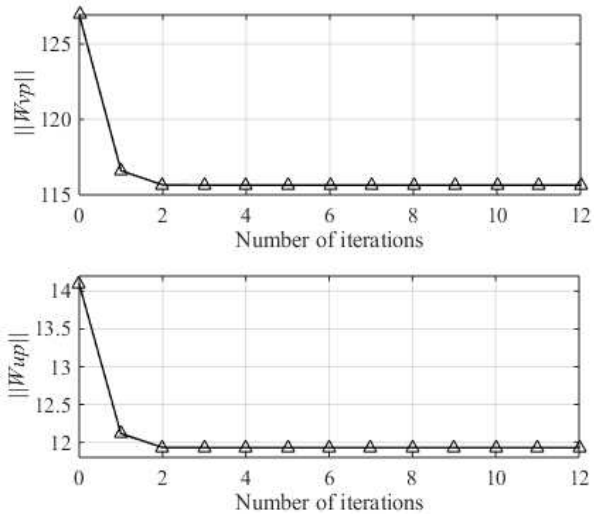

Figure 4. The convergence of training weights in position controller

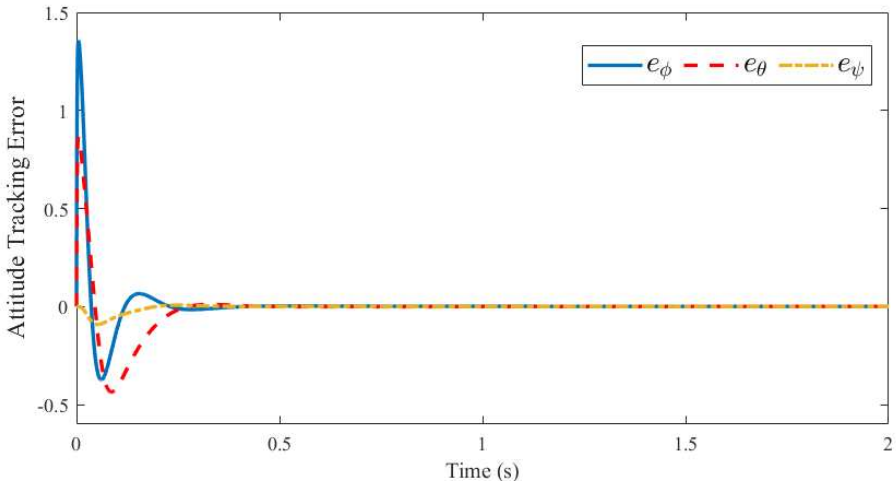

Figure 5. The tracking of orientation angles

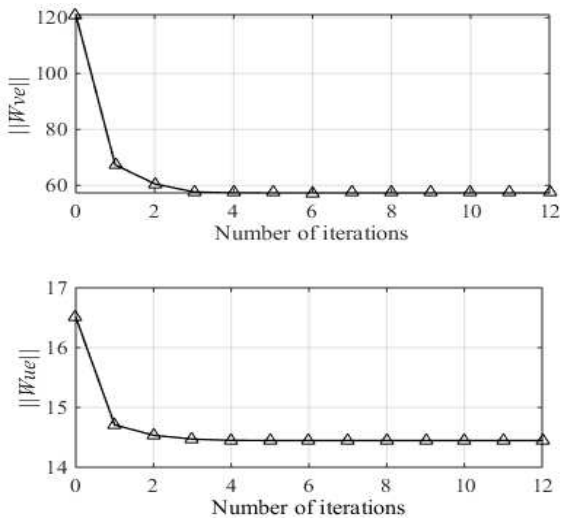

Figure 6. The convergence of training weights in attitude controller

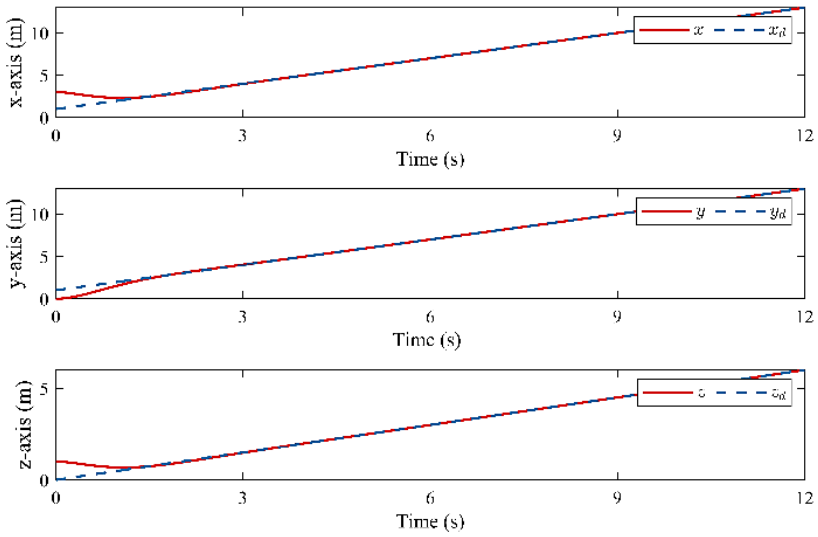

Figure 7. The trajectory tracking of RL control

Table 2. Numerous performance indices

Performance indices	IAE_p	IAE_Ω	$ITAE_p$	$ITAE_\Omega$
Value	3.0527	0.1175	1.8408	0.0144

4. CONCLUSION

A novel data-driven reinforcement learning algorithm incorporating a discount factor was proposed for application in the two subsystems of a UAV quadrotor to address performance challenges in fully uncertain UAV systems. Utilizing the off-policy approach, the model-free cascade control framework was constructed to simultaneously obtain the optimal control law and the corresponding Bellman function. The network weights were adjusted to approximate the solution of the modified Hamilton-Jacobi-Bellman (HJB) equation, with theoretical guarantees of both convergence and stability. A numerical example was provided to demonstrate the effectiveness of the proposed discount factor-based data-driven RL algorithm in the UAV control context.

ACKNOWLEDGEMENTS

This research was supported by Research Foundation funded by Thai Nguyen University of Technology.

REFERENCES

[1] L.-X. Xu, Y.-L. Wang, X. Wang, and C. Peng, "Distributed active disturbance rejection formation tracking control for quadrotor UAVs," *IEEE Transactions on Cybernetics*, vol. 54, no. 8, pp. 4678–4689, Aug. 2024, doi: 10.1109/tcyb.2023.3324752.

[2] S. Wang, H. Chen, J. Liu, and Y. Liu, "Adaptive trajectory tracking of UAV with a cable-suspended load using vision-inertial-based estimation," *Automatica*, vol. 158, p. 111310, Dec. 2023, doi: 10.1016/j.automatica.2023.111310.

[3] X. Shao and D. Ye, "Robust adaptive dynamic memory-event-triggered attitude control for nonlinear multi-UAVs resist actuator hysteresis," *International Journal of Robust and Nonlinear Control*, vol. 33, no. 14, pp. 8315–8335, Jun. 2023, doi: 10.1002/rnc.6821.

[4] Z. Huang and M. Chen, "Augmented disturbance observer-based appointed-time tracking control of UAVs under exogenous disturbance," *IEEE Transactions on Intelligent Vehicles*, vol. 9, no. 1, pp. 2822–2835, Jan. 2024, doi: 10.1109/tiv.2023.3303348.

[5] X. Wu, W. Fei, X. Wu, and R. Zhen, "Fixed-time neuroadaptive practical tracking control based on extended state/disturbance observer for a QUAV with external disturbances and time-varying parameters," *Journal of the Franklin Institute*, vol. 359, no. 8, pp. 3466–3491, May 2022, doi: 10.1016/j.jfranklin.2022.03.018.

[6] J. Li, Z. Xiao, J. Fan, T. Chai, and F. L. Lewis, "Off-policy Q-learning: Solving Nash equilibrium of multi-player games with network-induced delay and unmeasured state," *Automatica*, vol. 136, p. 110076, Feb. 2022, doi: 10.1016/j.automatica.2021.110076.

[7] J. Wang, Z. Zhou, C. Wang, and Z. Ding, "Cascade structure predictive observer design for consensus control with applications to UAVs formation flying," *Automatica*, vol. 121, p. 109200, Nov. 2020, doi: 10.1016/j.automatica.2020.109200.

[8] J. Chen, W. Yang, Z. Shi, and Y. Zhong, "Robust horizontal-plane formation control for small fixed-wing UAVs," *Aerospace Science and Technology*, vol. 131, p. 107958, Dec. 2022, doi: 10.1016/j.ast.2022.107958.

[9] Y. Shen and C. Wei, "Multi-UAV flocking control with individual properties inspired by bird behavior," *Aerospace Science and Technology*, vol. 130, p. 107882, Nov. 2022, doi: 10.1016/j.ast.2022.107882.

[10] Y. Zhao, F. Zhu, and D. Xu, "Event-triggered bipartite time-varying formation control for multiagent systems with unknown inputs," *IEEE Transactions on Cybernetics*, vol. 53, no. 9, pp. 5904–5917, Sep. 2023, doi: 10.1109/tcyb.2022.3208228.
[11] Y. Kang, D. Luo, B. Xin, J. Cheng, T. Yang, and S. Zhou, "Robust leaderless time-varying formation control for nonlinear unmanned aerial vehicle swarm system with communication delays," *IEEE Transactions on Cybernetics*, vol. 53, no. 9, pp. 5692–5705, Sep. 2023, doi: 10.1109/tcyb.2022.3165007.
[12] G. Wen, S. S. Ge, C. L. P. Chen, F. Tu, and S. Wang, "Adaptive tracking control of surface vessel using optimized backstepping technique," *IEEE Transactions on Cybernetics*, vol. 49, no. 9, pp. 3420–3431, Sep. 2019, doi: 10.1109/tcyb.2018.2844177.
[13] G. Wang, B. Luo, and S. Xue, "Integral reinforcement learning-based optimal output feedback control for linear continuous-time systems with input delay," *Neurocomputing*, vol. 460, pp. 31–38, Oct. 2021, doi: 10.1016/j.neucom.2021.06.073.
[14] T. Bian and Z. P. Jiang, "Reinforcement learning and adaptive optimal control for continuous-time nonlinear systems: a value iteration approach," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 33, no. 7, pp. 2781–2790, Jul. 2022, doi: 10.1109/TNNLS.2020.3045087.
[15] H. Zhang, Z. Ming, Y. Yan, and W. Wang, "Data-driven control of agent-based models: An equation/variable-free machine learning approach," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 34, no. 8, pp. 4687–4701, Aug. 2023, doi: 10.1109/tnnls.2021.3116464.
[16] H. Jiang, H. Zhang, Y. Cui, and G. Xiao, "Robust control scheme for a class of uncertain nonlinear systems with completely unknown dynamics using data-driven reinforcement learning method," *Neurocomputing*, vol. 273, pp. 68–77, Jan. 2018, doi: 10.1016/j.neucom.2017.07.058.
[17] H. Zhang, G. Xiao, Y. Liu, and L. Liu, "Value iteration-based H∞ controller design for continuous-time nonlinear systems subject to input constraints," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 50, no. 11, pp. 3986–3995, Nov. 2020, doi: 10.1109/tsmc.2018.2853091.
[18] D. Wang, H. He, and D. Liu, "Intelligent optimal control with critic learning for a nonlinear overhead crane system," *IEEE Transactions on Industrial Informatics*, vol. 14, no. 7, pp. 2932–2940, Jul. 2018, doi: 10.1109/tii.2017.2771256.
[19] J. Li, G. Zhang, C. Liu, and W. Zhang, "COLREGs-constrained adaptive fuzzy event-triggered control for underactuated surface vessels with the actuator failures," *IEEE Transactions on Fuzzy Systems*, vol. 29, no. 12, pp. 3822–3832, Dec. 2021, doi: 10.1109/tfuzz.2020.3028907.
[20] J. Sun and T. Long, "Event-triggered distributed zero-sum differential game for nonlinear multi-agent systems using adaptive dynamic programming," *ISA Transactions*, vol. 110, pp. 39–52, Apr. 2021, doi: 10.1016/j.isatra.2020.10.043.
[21] T. Li, W. Bai, Q. Liu, Y. Long, and C. L. P. Chen, "Distributed fault-tolerant containment control protocols for the discrete-time multiagent systems via reinforcement learning method," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 34, no. 8, pp. 3979–3991, Aug. 2023, doi: 10.1109/tnnls.2021.3121403.
[22] H. Wang, W. Ren, W. Yu, and D. Zhang, "Fully distributed consensus control for a class of disturbed second-order multi-agent systems with directed networks," *Automatica*, vol. 132, p. 109816, Oct. 2021, doi: 10.1016/j.automatica.2021.109816.
[23] Z. Tang, R. Cunha, T. Hamel, and C. Silvestre, "Formation control of a leader–follower structure in three dimensional space using bearing measurements," *Automatica*, vol. 128, p. 109567, Jun. 2021, doi: 10.1016/j.automatica.2021.109567.
[24] S. Li, W. Zou, and Z. Xiang, "Neural-network-based consensus of multiple Euler-Lagrange systems with an event-triggered mechanism," *Journal of the Franklin Institute*, vol. 358, no. 16, pp. 8625–8638, Oct. 2021, doi: 10.1016/j.jfranklin.2021.08.033.
[25] W. Wu, Y. Li, and S. Tong, "Neural network output-feedback consensus fault-tolerant control for nonlinear multiagent systems with intermittent actuator faults," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 34, no. 8, pp. 4728–4740, Aug. 2023, doi: 10.1109/tnnls.2021.3117364.
[26] F. Ghaderi, A. Toloei, and R. Ghasemi, "Formation control and obstacle avoidance of a multi-quadrotor system based on model predictive control and improved artificial potential field," *International Journal of Engineering*, vol. 37, no. 1, pp. 115–126, 2024, doi: 10.5829/ije.2024.37.01a.11.
[27] F. Ghaderi, A. Toloei, and R. Ghasemi, "Quadrotor control for tracking moving target, and dynamic obstacle avoidance based on potential field method," *International Journal of Engineering*, vol. 36, no. 10, pp. 1720–1732, 2023, doi: 10.5829/ije.2023.36.10a.01.
[28] S. K. Bhat and G. D. Deepak, "Predictive modelling and optimization of double ring electrode based cold plasma using artificial neural network," *International Journal of Engineering*, vol. 37, no. 1, pp. 83–93, 2024, doi: 10.5829/ije.2024.37.01a.08.
[29] V. Tu Vu, T. L. Pham, and P. N. Dao, "Disturbance observer-based adaptive reinforcement learning for perturbed uncertain surface vessels," *ISA Transactions*, vol. 130, pp. 277–292, Nov. 2022, doi: 10.1016/j.isatra.2022.03.027.
[30] K. Nguyen, V. T. Dang, D. D. Pham, and P. N. Dao, "Formation control scheme with reinforcement learning strategy for a group of multiple surface vehicles," *International Journal of Robust and Nonlinear Control*, vol. 34, no. 3, pp. 2252–2279, Nov. 2023, doi: 10.1002/rnc.7083.
[31] H. Nguyen, H. B. Dang, and P. N. Dao, "On-policy and off-policy Q-learning strategies for spacecraft systems: An approach for time-varying discrete-time without controllability assumption of augmented system," *Aerospace Science and Technology*, vol. 146, p. 108972, Mar. 2024, doi: 10.1016/j.ast.2024.108972.
[32] W. Zhao, H. Liu, F. L. Lewis, and X. Wang, "Data-driven finite-horizon H∞ tracking control with event-triggered mechanism for the continuous-time nonlinear systems," *IEEE Transactions on Cybernetics*, vol. 52, no. 8, pp. 7889–7898, Aug. 2022, doi: 10.1109/tcyb.2021.3049486.
[33] M. Z. Mohd Tumari, M. A. Ahmad, M. H. Suid, M. R. Ghazali, and M. O. Tokhi, "An improved marine predators algorithm tuned data-driven multiple-node hormone regulation neuroendocrine-PID controller for multi-input–multi-output gantry crane system," *Journal of Low Frequency Noise, Vibration and Active Control*, vol. 42, no. 4, pp. 1666–1698, Jun. 2023, doi: 10.1177/14613484231183938.
[34] R. Strässer, J. Berberich, and F. Allgöwer, "Robust data-driven control for nonlinear systems using the Koopman operator," *IFAC-PapersOnLine*, vol. 56, no. 2, pp. 2257–2262, 2023, doi: 10.1016/j.ifacol.2023.10.1190.
[35] D. G. Patsatzis, L. Russo, I. G. Kevrekidis, and C. Siettos, "Data-driven control of agent-based models: An equation/variable-free machine learning approach," *Journal of Computational Physics*, vol. 478, p. 111953, Apr. 2023, doi: 10.1016/j.jcp.2023.111953.
[36] G. M. Qian, M. R. Bin Ghazali, M. A. Bin Ahmad, and M. S. Bin Mohd Shukri, "Data driven sigmoid proportional-integral-derivative (SPID) controller for twin rotor MIMO system," *Journal Européen des Systèmes Automatisés*, vol. 56, no. 1, pp. 105–113, Feb. 2023, doi: 10.18280/jesa.560114.
[37] X. Wang, J. Berberich, J. Sun, G. Wang, F. Allgöwer, and J. Chen, "Model-based and data-driven control of event- and self-triggered discrete-time linear systems," *IEEE Transactions on Cybernetics*, vol. 53, no. 9, pp. 6066–6079, Sep. 2023, doi: 10.1109/tcyb.2023.3272216.

BIOGRAPHIES OF AUTHORS

Ngoc Trung Dang born in 1984, received his Ph.D. in automation from Thai Nguyen University of Technology in 2017. He is currently the deputy head of the Faculty of Electrical Engineering at Thai Nguyen University of Technology. His research focuses on robotics control, modeling, and simulation. He has contributed to several national research projects in the field of intelligent systems and automation. He actively participates in academic activities and serves as a reviewer for journals in control and robotics. He can be contacted at email: trungcsktd@tnut.edu.vn.

Quynh Nga Duong born in 1985, holds an M.Sc. in electrical engineering from Thai Nguyen University of Technology, Vietnam. She is currently a lecturer at the Faculty of Electrical Engineering, Thai Nguyen University of Technology, Thai Nguyen City, Vietnam. Her main research interests include control robotics, electrical engineering, automation, and solar energy. She is actively engaged in academic and applied research projects focusing on the integration of renewable energy and intelligent control systems. She can be contacted at email: duongquynhngaktd@tnut.edu.vn.