Formation Control of Multiple Unmanned Aerial Vehicle Systems using Integral Reinforcement Learning

Ngoc Trung Dang [1*], Quynh Nga Duong [2]
[1, 2] Faculty of Electrical Engineering, Thai Nguyen University of Technology, Thai Nguyen, Vietnam
Email: [1] trungcsktd@tnut.edu.vn, [2] duongquynhngaktd@tnut.edu.vn
*Corresponding Author

***Abstract*—Formation control of Unmanned Aerial Vehicles (UAVs), especially quadrotors, has many practical applications in contour mapping, transporting, search and rescue. This article solves the formation tracking requirement of a group of multiple UAVs by formation control design in outer loop and integral Reinforcement Learning (RL) algorithms in position sub-system. First, we present the formation tracking control structure, which uses a cascade description to account for the model separation of each UAV. Second, based on value function of inner model, a modified iteration algorithm is given to obtain the optimal controller in the presence of discount factor, which is necessary to employ due to the finite requirement of infinite horizon based cost function. Third, the integral RL control is developed to handle dynamic uncertainties of attitude sub-systems in formation UAV control scheme with a discount factor to be employed in infinite horizon based cost function. Specifically, the advantage of the proposed control is pointed out in not only formation tracking problem but also in the optimality effectiveness. Finally, the simulation results are conducted to validate the proposed formation tracking control of a group of multiple UAV system.**

***Keywords*—Integral Reinforcement Learning (RL); Unmanned Aerial Vehicles (UAVs); Formation Control; Approximate/Adaptive Dynamic Programming (ADP); Model-Free Based Control.**

I. INTRODUCTION

The formation tracking control of Unmanned Aerial Vehicles (UAVs) has been investigated in recent time with many applications, such as in transportation, agriculture, military, etc. [1]-[8]. In [1], according to conventional sliding mode control (SMC), the disturbance in UAV was handled by the term of sign function to obtain the finite time performance. Moreover, the obstacle of finite time convergence has been carried out by the existence of exponential function in control scheme [1]. The extension of handling finite-time control introduced in [1] was considered with event-triggered mechanism (ETM) [2]. In order to improve the control performance of each UAV, the optimal control law has been presented in [3] with the Model-free solution, which was developed by integral RL technique and data collection method. Additionally, to address the obstacle of input saturation, the back correction technique was discussed by establishing the dynamic equation of error between the actual control input and computational control input to achieve the stability of closed control system. For handling the dynamic uncertainty, Neural Network method has been applied to combine with exponential function to obtain the Fixed-Time Cooperative Control design for uncertain systems [5]. The consideration of formation control was extended for the case of heterogeneous multi-agent systems with UAV and USV [6]. As we have all known, it is difficult to obtain the state variables due to the measurement technique and physical property. The authors in [7] developed disturbance observer (DO) to improve the formation tracking control of multiple UAVs. By employing state observer, output formation tracking control of heterogeneous multi-agent systems was implemented with the optimality [8]. On the other hand, optimal control problem has been discussed in connection to classical consideration of NN and sliding mode control [9]. The constraint problem was mentioned in formation tracking control with an UAV team [10]. However, the above references of formation control mainly focus on implementing the Lyapunov control method, which only guarantee the tracking problem. Moreover, this approach met the difficulties of actuator saturation and expanding the control performance. The optimality consideration with two main directions of optimization and optimal control is able to handle actuator saturation by transforming the constraint in optimal control into optimization, which can be obtained the general solutions. It follows that the optimal control design can be known as the significant approach to satisfy the tracking problem in the presence of constraints in UAV. Moreover, the optimal control is able to achieve not only formation tracking problem in an UAV team but also minimization of a given cost function, which has not been mentioned in the previous UAV researches [22]-[80].

Reinforcement Learning (RL) control (or Approximate/Adaptive Dynamic Programming (ADP)) has been developed with numerous results [11]-[21]. In [11], authors proposed the appropriate constraint set to satisfy tracking problem. In RL control problem, as we have all known, due to the challenge of solving Hamilton-Jacobi-Bellman for nonlinear system and Riccati for linear model, learning methods was developed to solve for finding the roots. In [12] and [14], the disturbance influence was integrated into optimal control problem to solve the HJB, Ricatti equations. Authors in [13] developed the model-based RL control by using two Neural Networks, which are used to approximate optimal value function and control policy [13]. The learning process was described by obtaining the training method of weights based on the minimization of square of error between computed Hamilton function and actual Hamilton function [13]. Additionally, Data-driven technique was improved to handle the actuator saturation and nonlinear model [15]. Nonzero-sum tracking games has been known as the establishment was considered with many cost function [16]. Authors in [16] pointed out the Event-triggered neural experience replay learning, which is able to relax the Persistence of excitation (PE) condition. The work in [17] address the time-varying systems by lifting method and presenting two On-Policy and Off- Policy iterations. Additionally, the Actor/Critic method for multi-agent systems introduced in [18] was the extension of the work in [13]. The work of [19][20] develops the learning technique for linear systems with disturbance to be integrated in [11]. The formation tracking problem has been handled in a group of multiple Surface Vessels with cascade controller [21]. To my best knowledge, the development of optimal control for multi-agent systems with fully dynamic model has not been much done yet [22]-[80]. The research contribution is described as follows. This paper focuses on formation control structure of a team of multiple UAVs with integral RL method and the consideration of dynamic model in each UAV to achieve both the formation tracking control effectiveness and minimization of a given cost function in model-free situation.

The rest of the paper is structured as follows. In Section II, we briefly introduce the algebraic graph theory and attitude models of UAVs. In Section III, we further develop the RL control for nonlinear systems with a discount factor, the formation control of outer loop and a model-free RL algorithms is developed for attitude sub-system. In Section IV, the proposed algorithms are illustrated with a group of UAV control system to demonstrate their performance. This article concludes with a summary of the findings in Section V.

II. PRELIMINARIES AND PROBLEM STATEMENTS

In this section, the basic theories are mentioned to describe the research problem, including graph theory, mathematical model of each quadrotor and control objective.

A. *Algebraic Graph Theory*

In this paper, the multiple UAVs are depicted by a connected graph $G = (V, \mho, A)$, where $V = (\upsilon_1, \upsilon_2, \ldots, \upsilon_n)$ is the set of node, in which each node is used to describe an UAV. Matrix $A = [a_{ij}]$ denotes the adjacency matrix, $\mho \subseteq V \times V$ defines the edge set. If the existence of a connection from node υ_j to node υ_i, then the edge $\underline{\upsilon}_{ij} = (\upsilon_i, \upsilon_j) \in \mho$, and node υ_j is known as a neighbor of node υ_i. Additionally, we obtain the neighbor set of υ_i is defined by $N_i = \{\upsilon_j : (\upsilon_i, \upsilon_j) \in \mho\}$. The adjacency matrix $A = [a_{ij}]$ is defined as $a_{ij} = 1$ if $\underline{\upsilon}_{ij} = (\upsilon_i, \upsilon_j) \in \mho$, and $a_{ij} = 0$ if $\underline{\upsilon}_{ij} = (\upsilon_i, \upsilon_j) \notin \mho$. The graph $G = (V, \mho, A)$is known as undirected if and only if $a_{ij} = a_{ji}$. An undirected graph is called as connected if the existence of a path between an abitrary pair $(\upsilon_i, \upsilon_{i1}), (\upsilon_{i1}, \upsilon_{i2}), \ldots, (\upsilon_{ik}, \upsilon_j)$. Consider the Graph $G = (V, \mho, A)$, it leads to the Laplacian matrix to be known as $L^{n\times n} = D - A$, where $D = diag\{\sum_{k=1}^{n} a_{1k}, \ldots, \sum_{k=1}^{n} a_{nk}\}$. In the formation control design as described in next sections, the neighbor set $N_i = \{\upsilon_j : (\upsilon_i, \upsilon_j) \in \mho\}$ plays an important role.

B. *Mathematical Model of each UAV*

Fig. 1 depicts the kinematics and dynamics of an UAV with two reference frames to be considered as a Body-fixed frame $B = \{x_B, y_B, z_B\}$ and a earth-fixed inertial frame $E = \{x_E, y_E, z_E\}$. In Fig. 1, each UAV is steered via four propeller force F_1, F_2, F_3, F_4.

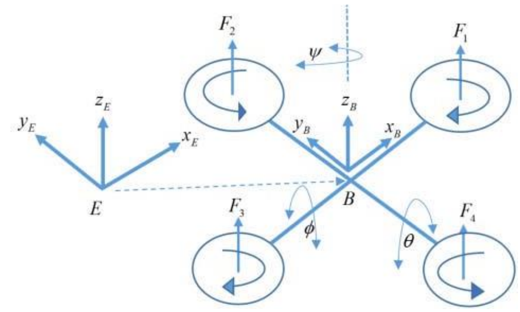

Fig. 1. UAV model and two reference frames [2]

In the view of [1], the positions vector of each UAV and the angels vector are considered under the earth-fixed inertial frame, the Euler angles Roll-Pitch-Yaw, which are bounded as $-\pi/2 < \phi < \pi/2$, $-\pi/2 < \theta < \pi/2$ and $-\pi < \psi < \pi$. This is an important assumption for formation control design. The following rotation matrix is employed to describe UAV model:

$$R = \begin{bmatrix} c_\theta c_\psi & c_\theta s_\psi & -s_\theta \\ s_\phi s_\theta c_\psi - c_\phi s_\psi & s_\phi s_\theta s_\psi + c_\phi c_\psi & s_\phi c_\theta \\ c_\phi s_\theta c_\psi + s_\phi s_\psi & c_\phi s_\theta s_\psi - s_\phi c_\psi & c_\phi c_\theta \end{bmatrix} \tag{1}$$

The relation between the angular velocities vector given by $\overline{\omega} = (p, q, r)^T$, and the Euler angles rate $\dot{\Omega} = (\dot{\phi}, \dot{\theta}, \dot{\psi})^T$ as:

$$\omega = \begin{pmatrix} 1 & 0 & -s_\theta \\ 0 & c_\phi & s_\phi c_\theta \\ 0 & -s_\phi & c_\phi c_\theta \end{pmatrix} \dot{\Omega} \tag{2}$$

In the light of [1], the dynamics model of each quadrotor can be written as:

$$\begin{aligned} m\ddot{r}_i &= R_i f_i \\ J\ddot{\Omega}_i &= C_{(\Omega_i, \dot{\Omega}_i)} \dot{\Omega}_i + \tau_i \end{aligned} \tag{3}$$

The control objective is to design a formation control for multiple quadrotors to guarantee both the tracking problem, which is known as $\underset{t\to\infty}{Lim} p_i = p_i^*$, and minimize the discount factor-based cost function (5). Moreover, it can be seen that in the proposed control system (Fig. 2), the optimal control law is developed for rotational sub-systems in Eqn. (3) by integral reinforcement learning and the position control law plays the important role of outer loop controller.

Remark 1. Unlike the conventional formation tracking control objective only consider the convergence of tracking error, the control objective in this article considers both the trajectory tracking performance and the minimization of the given performance index. Moreover, the formation tracking control problem is also necessary to be guaranteed in outer control loop.

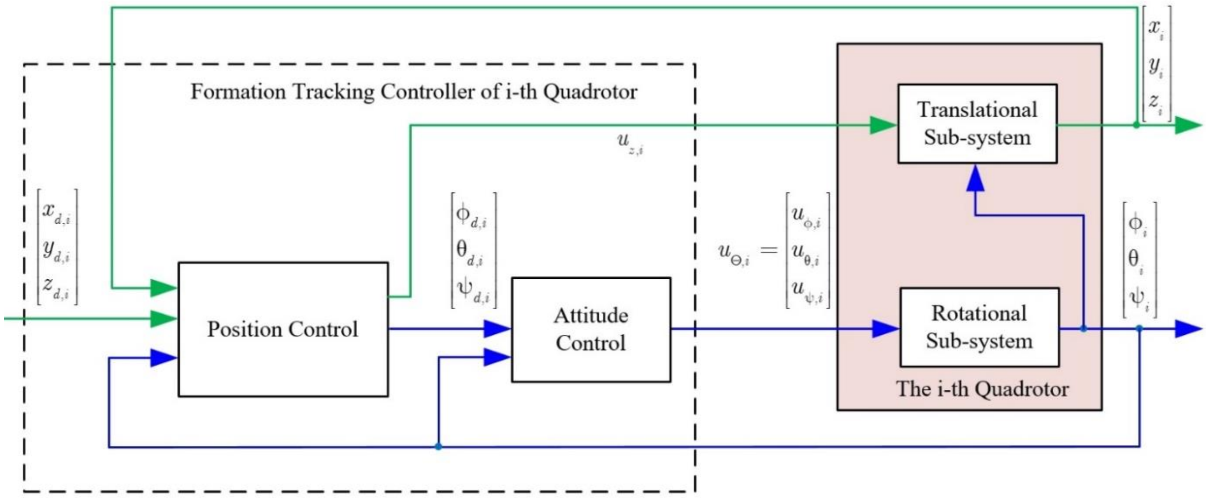

Fig. 2. The formation control schematic of multiple UAV [2]

III. Formation Control Design

In this part, a formation control structure illustrated in Fig. 2 is presented with two control loops, including *Position Controller* and *Attitude Controller* after considering each UAV as two sub-systems (see Eqn. (3)) The *Position Controller* is designed to satisfy the formation tracking problem and its outputs are the desired Euler angles $(\phi^{des}, \theta^{des}, \psi^{des})^T$. The *Attitude Control Design* plays the role of not only tracking problem but also minimizing the cost function. Moreover, it should be noted that the optimal control problem of inner loop control is solved by integral RL strategy to handle the disadvantage of complete uncertainty, which has not been studied in conventional methods.

A. RL Control Law with a Discount Factor

The optimal control design is studied for the nonlinear continuous system:

$$\frac{d}{dt}\zeta = F(\zeta) + G(\zeta)u(t) \tag{4}$$

with the following infinite horizon performance index in the presence of a positive discount factor $\lambda > 0$.

$$V(\zeta(t), u(t)) = \int_t^{\infty} e^{-\lambda(s-t)}\, U(\zeta(s), u(s))ds, \tag{5}$$

where $U(\zeta, u(s)) \triangleq \zeta^T Q\zeta + (u(s))^T Ru(s)$, Q and $R \in \mathbb{R}^{n\times n}$ are two positive-definite symmetric constant matrices. It can be seen that the addition of discount factor in cost function (8) is to guarantee the finite value of the cost function as time converges to infinity. The Bellman function obtained from Dynamic programming is described as follows:

$$V^*(\zeta(t)) = \min_{u(\zeta(t))\in \Upsilon(U)} V\big(\zeta(t), u(\zeta)\big) \tag{6}$$

In the light of [21], it follows that:

$$U\big(\zeta(t), u^*(t)\big) - \lambda V^*((\zeta(t)) + \frac{\partial V^*}{\partial \zeta}(F(\zeta) + G(\zeta)u^*) = 0 \tag{7}$$

and the following optimization problem:

$$V^*(\zeta(t)) = \min_{u(\zeta)\in \Upsilon(U)} \left(\int_t^{t+\Delta} U(\zeta, u(t))ds + e^{-\lambda\Delta}V^*((\zeta(t+\Delta))\right) \tag{8}$$

Therefore, it implies the following modified optimization problem:

$$\min_{u(t)(\zeta)\in \Upsilon(U)} [\, U(\zeta, u(t)) - \lambda V^*(\zeta) + \frac{\partial V^*}{\partial \zeta}(F(\zeta) + G(\zeta)u(t))] = 0 \tag{9}$$

We define the modified Hamiltonian function being associated with a discount factor $\lambda > 0$ as:

$$H(\zeta, u(t), \nabla V, V) = \zeta^T Q\zeta + (u(t))^T R\, u(t) - \lambda V(\zeta) + \nabla V^T(\zeta)(F(\zeta) + G(\zeta)u(\zeta)) \tag{10}$$

Then, it follows that:

$$u^*(\zeta) = \operatorname*{argmin}_{u\in \Upsilon(\Omega)}[H(\zeta, u(t), \nabla V^*(\zeta))] = -\frac{1}{2}R^{-1}G^T(\zeta)\nabla V^*(\zeta) \tag{11}$$

and the partial derivative equation (PDE) is achieved as:

$$H^*(\zeta, u^*, \nabla V^*, V^*) = \zeta^T Q\zeta - \frac{1}{4}\nabla V^{*T}(\zeta)G(\zeta)R^{-1}G^T(\zeta)\nabla V^*(\zeta - \lambda V^*(\zeta) + \nabla V^{*T}(\zeta)F(\zeta) = 0 \tag{12}$$

However, it is impossible to analytically solve the PDE (12) to find the Bellman function from the optimal control signal $u^*(\zeta)$. Therefore, the proposed algorithm is considered to solve in Sections C.

B. Formation Control Scheme

In this part, to develop the formation control design, the outer model of UAV can be described by (13).

$$\begin{cases} \dot{r}_i = v_i \\ \dot{v}_i = u_i + d_{ui} \end{cases} \tag{13}$$

where $r_i \in R^d$ denotes the agent position, v_i, u_i are the velocity and control input of each agent, respectively. The disturbance d_{ui} satisfies that $\|d_{ui}\| \le D_{ui}$.

The formation control objective is to find the control scheme to satisfy that $\lim_{t\to\infty} p_i = p_i^*$. It can be seen that, the following control signal is guaranteed the convergence (14).

$$\dot{p}_i = u_i = k_p(p_i^* - p_i), i = 1, \ldots, n \tag{14}$$

where k_p is a positive number.

Based on graph theory of describing the connectivity in quadrotors team, we can improve the control performance of formation control design with the traditional control input to be modified as:

$$\dot{p}_i = u_i = k_p(p_i^* - p_i) + \sum_{j\in N_i} a_{\mathrm{ij}}((p_j^* - p_j) - (p_i^* - p_i)), i = 1,\ldots,n \quad (15)$$

where $a_{\mathrm{ij}} > 0$ is the weight number to be connected to $(v_j, v_i) \in E(G)$. Therefore, the dynamic of closed system is expressed as:

$$\dot{p} = k_p(p^* - p) - (\mathscr{L} \otimes I_d)(p^* - p) \quad (16)$$

where $p = vec(p_1,\ldots p_n), p^* = vec(p_1^*,\ldots,p_n^*)$ and $e_p \triangleq p^* - p$.

After obtaining the control signal in the control structure (Fig. 1), the reference of the attitude control scheme can be obtained $u_r = (u_{rn}, u_{re}, u_{rd})$. According to [1], the desired yaw angle ψ^{des} can be chosen as zero and ψ^{des}, θ^{des} can be easily solved as follows:

$$\begin{aligned} u_{rp} &= \sqrt{u_{rn}^2 + u_{re}^2 + (u_{rd} + u_{r_0})^2}, \psi^{des} = 0, \\ \phi^{des} &= arcsin\left(\frac{u_{rn}\sin(\psi^{des}) - u_{re}\cos(\psi^{des})}{u_{rp}}\right), \\ \theta^{des} &= arctan\left(\frac{u_{rn}\cos(\psi^{des}) + u_{re}\sin(\psi^{des})}{u_{rd} + u_{r_0}}\right) \end{aligned} \quad (17)$$

C. IRL for Attitude Control Design

After the desired attitudes are obtained in Section B, we continue to design the attitude controller in Fig. 1 for satisfying the optimal control problem. The model in Eq. (3) can be written as:

$$\ddot{\Omega} = J^{-1}T - J^{-1}C_{(\Omega,\dot{\Omega})}\dot{\Omega} \quad (18)$$

According to the states vector $x_\Omega = [\phi, \dot{\phi}, \theta, \dot{\theta}, \psi, \dot{\psi}]^T$, and the model in Eq. (3), we can obtain the following attitude model as:

$$\dot{X}_{\Omega d} = \begin{bmatrix} \dot{e}_\Omega \\ \dot{x}_{\Omega d} \end{bmatrix} = \begin{bmatrix} F_\Omega & F_\Omega - F_{\Omega d} \\ 0_{6,6} & F_{\Omega d} \end{bmatrix} X_{\Omega d} + \begin{bmatrix} G_\Omega \\ 0_{6,3} \end{bmatrix} u_\Omega \quad (19)$$

Therefore, the attitude control design in Fig. 2 can be developed in the following algorithm (Fig. 3).

Algorithm 1: IRL Control scheme

Step 1 (Initialization): Selecting stabilizing control policy and the disturbance term $u_{\Omega e}(t)$, the threshold to satisfy the PE condition, and implementing the data collection.

Step 2 (Policy Evaluation): For each control signal $u_\Omega^i(X_\Omega)$, solve simultaneously the $V_\Omega^{i+1}(X_\Omega)$ and $u_\Omega^{i+1}(X_\Omega)$ by the equation (20).

$$\begin{aligned} &V_\Omega^{i+1}(X_\Omega(t+\delta t)) - V_\Omega^{i+1}(X_\Omega(t)) = \\ &-\int_t^{t+\delta t}\Big(X_\Omega(\tau)^T Q_\Omega X_\Omega(\tau) + \\ &\left(u_\Omega^i(X_\Omega(\tau))\right)^T R_\Omega u_\Omega^i(X_\Omega(\tau))\Big)d\tau + \\ &\int_t^{t+\delta t}\lambda V_\Omega^{i+1}(X_\Omega(\tau))d\tau + \\ &2\int_t^{t+\delta t}\left(u_\Omega^{i+1}(X_\Omega(\tau))\right)^T R_\Omega u_\Omega^i(X_\Omega(\tau))d\tau - \\ &2\int_t^{t+\delta t}\left(u_\Omega^{i+1}(X_\Omega(\tau))\right)^T R_\Omega(u_\Omega^0(\tau) + u_{\Omega e})d\tau \end{aligned} \quad (20)$$

Step 3 (Policy Improvement): Update the control policy $u_\Omega^i(X_\Omega) = u_\Omega^{i+1}(X_\Omega), i \to (i+1)$ and come back to Step 2 until $\|u_p^{i+1} - u_p^i\| < \epsilon_p$.

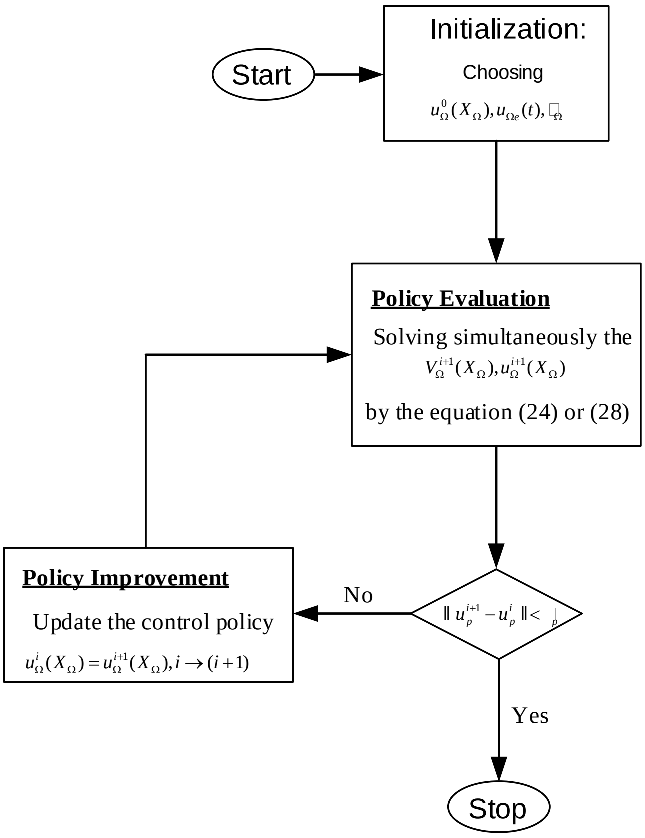

Fig. 3. Flowchart of Algorithm 1

In this part, it can be seen that the learning time is considered from the initial time to the time of completing RL Algorithms 1. Furthermore, under the dynamic uncertainties in models (3), the proposed RL Algorithms 1 are implemented with data collection in the practical system, which are utilized to solve the Eq. (15)–(20). However, it is worth noting that the existence of root in Eq. (15)–(20), (Algorithms 1), requires the Persistence of Excitation (PE) condition as shown in [10][11]. To guarantee the PE condition in Alg.1, it is necessary to insert the disturbance term $u_{\Omega e}(t)$ into the input signal at each step of Alg. 1. The computational efficiency is absolutely depended on collected data, which are implemented in many time periods. Furthermore, the reference of the inner model (Fig. 2) is time varying function implies the difficulties in designing the inner controller. On the other hand, it is emphasizing that the advantage of the proposed integral RL is described in uncertain model, which has been mentioned in the almost researches on robotic control systems [21]-[80].

IV. SIMULATION STUDIES

Consider a group of three perturbed UAV with the parameters to be given as follows:$m = 2.0(kg), k_w = 1(Ns^2), k_t = 1, g = 9.8(m/s^2), l_\tau = 0.2(m)$. The desired velocities and input disturbances are existed in force $\delta = 0.1\,sin(\pi t)\,(Nm)$. The existence of disturbance guarantees the relation between the simulation studies and practical systems. Additionally, thanks to the model-free property of the proposed integral RL method, the given parameters are sufficient to implement the simulation. It follows that the robustness and reliability of the proposed control system is satisfied. The formation control implementation is evaluated as follows. To develop RL control for inner control loop, the activation functions of the critic and actor parts are second-order polynomials and first-order polynomials, respectively. In the light of [13], the selection of activation function can be developed and kept in learning process. The training weights of value function and optimal control are computed as described in Alg. 1. For the purpose of satisfying the existence of solution in Alg. 1, the PE conditions in inner control loop is necessary to guaranteed under the probing noises $u_{\Theta e} = \sum_{m=1}^{500} 0.002 sin(w_m t)$ (w_m is a frequency chosen in range randomly), which are added to the position and attitude controllers respectively.

Based on the above simulation scenario, it is worth noting that the convergence of the weights in algorithms 1 is shown in Fig. 4, which validates the optimal control problem. The fact is that the convergence of learning process points out the Bellman function as well as the optimal control law. Moreover, the control performance is satisfied not only the optimal control problem but also the tracking effectiveness as shown in Fig. 5, Fig. 6. It follows that the unification of tracking problem and optimality is guaranteed to obtain the advantage in comparison with the previous researches [50 – 60]. On the other hand, under the integral RL method for inner sub-system and formation control law for outer model, it can be seen that the formation control is guaranteed as shown in Fig. 7. Finally, the advantage of the proposed integral RL is also shown in Fig. 8 with the better performance index in comparison with the relating research [39].

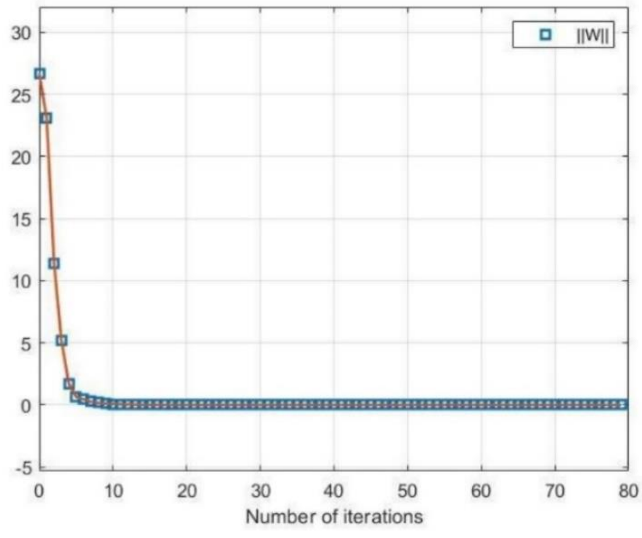

Fig. 4. The weight convergence of the learning stage

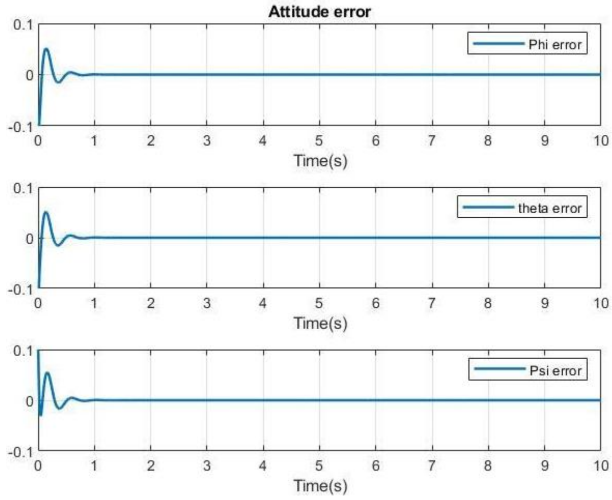

Fig. 5. The tracking of Attitude sub-system

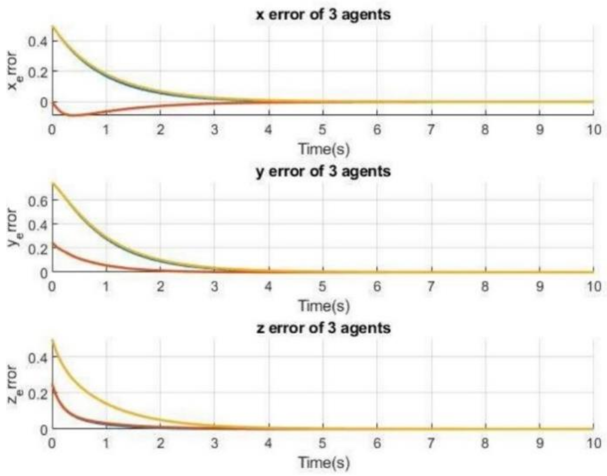

Fig. 6. The tracking of Position sub-system

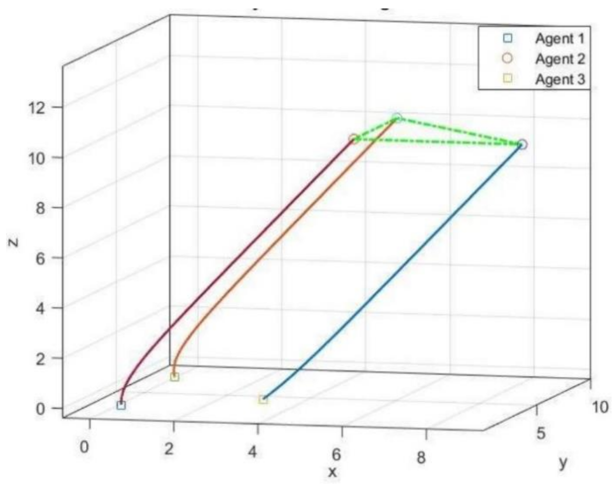

Fig. 7. The formation tracking control

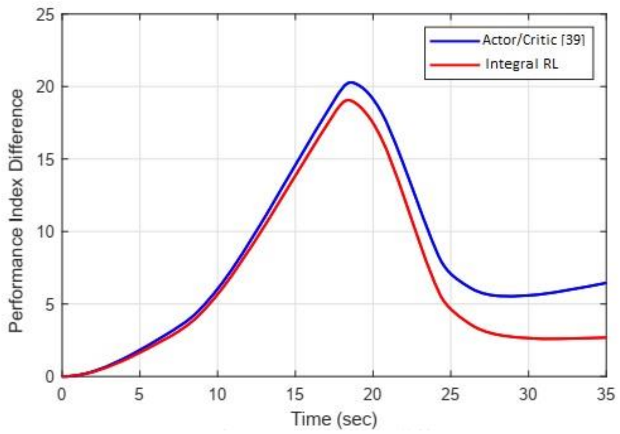

Fig. 8. The comparison of performance index between the Actor/Critic in [39] and the proposed integral RL method

V. CONCLUSION

This paper has proposed the formation control of multiple UAVs with model-free data integral RL strategy to be considered for attitude sub-system in multiple UAVs to effectively address both formation tracking and optimal control, which has not mentioned in the previous researches. The advantage of the proposed integral RL strategy is to implement the control law without the knowledge of UAV model. Therefore, the possibility of extending the proposed method for practical systems is determined. The main approach is to employ the Off-Policy RL algorithm using a discount factor in cost function to guarantee the bound of cost function and solve the model-free disadvantage without UAV model knowledge. Additionally, implementing data collection and considering computation techniques are analyzed to obtain simultaneously the actor and critic parts. Additionally, simulation studies are developed to point out the performance of the proposed Integral RL algorithms in the UAV control system. However, it should be noted that the proposed method requires the computation on many equations, which are established from collected data. In future work, we will consider extending the proposed strategies to the formation control problem of multiple UAVs.

ACKNOWLEDGMENT

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

REFERENCES

[1] X. Wang, Y. Yu, and Z. Li, "Distributed sliding mode control for leader-follower formation flight of fixed-wing unmanned aerial vehicles subject to velocity constraints," *International Journal of Robust and Nonlinear Control*, vol. 31, no. 6, pp. 2110–2125, June 2020, doi: 10.1002/rnc.5030.

[2] J. Wang, C. Bi, D. Wang, Q. Kuang, and C. Wang, "Finite-time distributed event-triggered formation control for quadrotor UAVs with experimentation," *ISA transactions*, vol. 126, pp. 585–596, 2022, doi: 10.1016/j.isatra.2021.07.049.

[3] O. Elhaki and K. Shojaei, "A novel model-free robust saturated reinforcement learning-based controller for quadrotors guaranteeing prescribed transient and steady state performance," *Aerospace Science and Technology*, vol. 119, pp. 107–128, 2021, doi: 10.1016/j.ast.2021.107128.

[4] T. Yang, P. Zhang, and H. Chen, "Distributed prescribed-time leader-follower formation control of surface vehicles with unknowns and input saturation," *ISA Transactions*, vol. 134, pp. 16–27, 2023, doi: 10.1016/j.isatra.2022.07.033.

[5] D. Liu, Z. Liu, C. P. Chen, and Y. Zhang, "Neural Network-based Smooth Fixed-Time Cooperative Control of High-Order Multi-Agent Systems With Time-Varying Failures," *Journal of the Franklin Institute*, no. 16, p. 8553–8578, 2022, doi: 10.1016/j.jfranklin.2022.08.058.

[6] H. Liu, P. Weng, X. Tian, and Q. Mai, "Distributed adaptive fixed-time formation control for UAV-USV heterogeneous multi-agent systems," *Ocean Engineering*, vol. 267, no. 4, pp. 113–240, 2023, doi: 10.1016/j.oceaneng.2022.113240.

[7] H. Xu, G. Cui, Q. Ma, Z. Li, and W. Hao, "Fixed-Time Disturbance Observer-Based Distributed Formation Control for Multiple QUAVs," *IEEE Transactions on Circuits and Systems II: Express Briefs*, vol. 70, no. 6, pp. 2181 – 2185, 2023, doi: 10.1109/TCSII.2022.3233438.

[8] Y. Shi, X. Dong, Y. Hua, J. Yu, and Z. Ren, "Distributed output formation tracking control of heterogeneous multi-agent systems using reinforcement learning," *ISA transactions*, vol. 138, pp. 318–328, 2023, doi: 10.1016/j.isatra.2023.03.003.

[9] T. Sun, C. Liu, and X. Wang, "Distributed anti-windup NN-sliding mode formation control of multi-ships with minimum cost," *ISA transactions*, vol. 138, pp. 49-62, 2023, doi: 10.1016/j.isatra.2023.03.011.

[10] Z. Hu and X. Jin, "Formation control for an UAV team with environment-aware dynamic constraints," *IEEE Transactions on Intelligent Vehicles*, vol. 9, no. 1, pp. 1465 – 1480, 2023, doi: 10.1109/TIV.2023.3295354.

[11] T. H. Nguyen, D. Q. Bui, and P. N. Dao, "An efficient Min/Max Robust Model Predictive Control for nonlinear discrete-time systems with dynamic disturbance," *Chaos, Solitons & Fractals*, vol. 180, p. 114551, March 2024, doi: 10.1016/j.chaos.2024.114551.

[12] J. Wang, J. Liu, Y. Zheng, and D. Zhang, "Data-based L2gain optimal control for discrete-time system with unknown dynamics," *Journal of the Franklin Institute*, vol. 360, no. 6, pp. 4354–4377, 2023, doi: 10.1016/j.jfranklin.2023.02.030.

[13] 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, 2022, doi: 10.1016/j.isatra.2022.03.027.

[14] M. Liu, Q. Cai, D. Li, W. Meng, and M. Fu, "Output feedback Q-learning for discrete-time finite-horizon zero-sum games with application to the H∞ control," *Neurocomputing*, vol. 529, pp. 48–55, 2023, doi: 10.1016/j.neucom.2023.01.050.

[15] L. Chen and F. Hao, "Optimal tracking control for unknown nonlinear systems with uncertain input saturation: A dynamic event-triggered ADP algorithm," *Neurocomputing*, vol. 564, no. 2, 2024, doi: 10.1016/j.neucom.2023.126964.

[16] X. Cui, B. Peng, B. Wang, and L. Wang, "Event-triggered neural experience replay learning for nonzero-sum tracking games of unknown continuous-time nonlinear systems," *International Journal of Robust and Nonlinear Control*, vol. 35, no. 12, pp. 6553–6575, 2023, doi: 10.1002/rnc.6709.

[17] 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, no. 10, p. 108972, 2024, doi: 10.1016/j.ast.2024.108972.

[18] Z. Ming, H. Zhang, J. Zhang, and X. Xie, "A novel actor–critic–identifier architecture for nonlinear multiagent systems with gradient descent method," *Automatica*, vol. 155, pp. 111–128, 2023, doi: 10.1016/j.automatica.2023.111128.

[19] Y. Jiang, W. Gao, J. Wu, T. Chai, and F. L. Lewis, "Reinforcement learning and cooperative H∞ output regulation of linear continuous-time multi-agent systems," *Automatica*, vol. 148, p. 110768, 2023, doi: 10.1016/j.automatica.2022.110768.

[20] B. Lian, W. Xue, Y. Xie, F. L. Lewis, and A. Davoudi, "Off-policy inverse Q-learning for discrete-time antagonistic unknown systems," *Automatica*, vol. 155, pp. 111–171, 2023, doi: 10.1016/j.automatica.2023.111171.

[21] 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, 2024, doi: 10.1002/rnc.7083.

[22] P. N. Dao and H. A. N. Duc, "Nonlinear RISE based integral reinforcement learning algorithms for perturbed Bilateral Teleoperators with variable time delay," *Neurocomputing*, vol. 605, p. 128355, 2024, doi: 10.1016/j.neucom.2024.128355.

[23] T. T. Cao, Q. H. Dao, and P. N. Dao, "Q-Learning Algorithm for F16 Aircraft: An approach to Discrete-Time Systems," *2023 12th International Conference on Control, Automation and Information Sciences (ICCAIS)*, pp. 1-4, 2023, doi: 10.1109/ICCAIS59597.2023.10382319.

[24] X. K. Tran, D. T. Le, D. T. Tran, and P. N. Dao, "A Reinforcement Learning Method for Control Scheme of Permanent Magnet Synchronous Motor," *International Conference on Advances in Information and Communication Technology*, pp. 121-126, 2023, doi: 10.1007/978-3-031-50818-9_15.

[25] D. T. T. Thi, H. L. Thi, P. N. Dao, and T. L. Nguyen, "Online Critic Dynamic Programming Based Optimal Control for Three-Shaft Web Machine Systems," *2023 Asia Meeting on Environment and Electrical Engineering (EEE-AM)*, pp. 1-5, 2023, doi: 10.1109/EEE-AM58328.2023.10394730.

[26] T. T. Cao, V. Q. Nguyen, H. A. N. Duc, Q. P. Nguyen, and P. N. Dao, "Policy Iteration-Output Feedback Adaptive Dynamic Programming Tracking Control for a Two-Wheeled Self Balancing Robot," *The International Conference on Intelligent Systems & Networks*, pp. 603-609, 2023, doi: 10.1007/978-981-99-4725-6_71.

[27] T. T. Cao, M. H. Vu, V. C. Nguyen, T. A. Nguyen, and P. N. Dao, "Formation Control Scheme of Multiple Surface Vessels with Model Predictive Technique," *The International Conference on Intelligent Systems & Networks*, pp. 312-317, 2023, doi: 10.1007/978-981-99-4725-6_39.

[28] Y. Dong and J. Chen "Nonlinear observer-based approach for cooperative control of networked rigid spacecraft systems," *Automatica*, vol. 128, p. 109552, 2021, doi: 10.1016/j.automatica.2021.109552.

[29] Z. Marvi and B. Kiumarsi, "Safe reinforcement learning: A control barrier function optimization approach," *International Journal of Robust and Nonlinear Control*, vol. 31, no. 6, pp. 1923-1940, 2021, doi: 10.1002/rnc.5132.

[30] S. Zhang, B. Zhao, D. Liu, and Y. Zhang, "Observer-based event-triggered control for zero-sum games of input constrained multi-player nonlinear systems," *Neural Networks*, vol. 144, pp. 101-112, 2021, doi: 10.1016/j.neunet.2021.08.012.

[31] T. Wang, Y. Wang, X. Yang, and J. Yang, "Further results on optimal tracking control for nonlinear systems with nonzero equilibrium via adaptive dynamic programming," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 34, no. 4, pp. 1900-1910, 2021, doi: 10.1109/TNNLS.2021.3105646.

[32] Y. Ouyang, L. Xue, L. Dong, and C. Sun, "Neural network-based finite-time distributed formation-containment control of two-layer quadrotor UAVs," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 52, no. 8, pp. 4836-4848, 2021, doi: 10.1109/TSMC.2021.3103013.

[33] D. Yu, J. Long, C. P. Chen, and Z. Wang, "Adaptive swarm control within saturated input based on nonlinear coupling degree," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 52, no. 8, pp. 4900-4911, 2021, doi: 10.1109/TSMC.2021.3102587.

[34] M. Wu, L. Liu, and Z. Yu, "Augmented safety guarantee-based area keeping control for an underactuated USV with environmental disturbances," *ISA transactions*, vol. 127, pp. 415-422, 2022.

[35] Y. Ouyang, C. Sun, and L Dong, "Actor–critic learning based coordinated control for a dual-arm robot with prescribed performance and unknown backlash-like hysteresis," *ISA transactions*, vol. 126, pp. 1-13, 2022.

[36] D. Wang, M. Zhao, M. Ha, and L. Hu, "Adaptive-critic-based hybrid intelligent optimal tracking for a class of nonlinear discrete-time systems," *Engineering Applications of Artificial Intelligence*, vol. 105, p. 104443, 2021.

[37] Y. Zhang, B. Zhao, D. Liu, and S. Zhang, "Event-triggered control of discrete-time zero-sum games via deterministic policy gradient adaptive dynamic programming," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 52, no. 8, pp. 4823-4835, 2021.

[38] K. Shao, R. Tang, F. Xu, X. Wang, and J. Zheng, "Adaptive sliding mode control for uncertain Euler–Lagrange systems with input saturation," *Journal of the Franklin Institute*, vol. 358, no. 16, pp. 8356-8376, 2021, doi: 10.1016/j.jfranklin.2021.08.027.

[39] G. Wen and C. P. Chen, "Optimized backstepping consensus control using reinforcement learning for a class of nonlinear strict-feedback-dynamic multi-agent systems," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 34, no. 3, 1524-1536, 2021.

[40] X. Z. Zhang, J. Lan, Y.-J. Liu, and L. Liu, "Adaptive event-triggered control of multi-agent systems with state constraints and unknown disturbances," *IET Control Theory & Applications*, vol. 15, no. 17, pp. 2171-2182, 2021, doi: 10.1049/cth2.12183.

[41] Y. Xu, T. Li, W. Bai, Q. Shan, L. Yuan, and Y. Wu, "Online event-triggered optimal control for multi-agent systems using simplified ADP and experience replay technique," *Nonlinear Dynamics*, vol. 106, no. 1, pp. 509-522, 2021, doi: 10.1007/s11071-021-06816-2.

[42] Jin, Xin, Yang Shi, Yang Tang, Herbert Werner, and Jürgen Kurths. "Event-triggered fixed-time attitude consensus with fixed and switching topologies." IEEE transactions on automatic control, vol. 67, no. 8, pp. 4138-4145, 2021, DOI: 10.1109/TAC.2021.3108514.

[43] S. Wang, S. Wen, K. Shi, X. Zhou, and T. Huang, "Approximate optimal control for nonlinear systems with periodic event-triggered mechanism," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 34, no. 6, pp. 2722-2731, 2021.

[44] Y. Yang, H. Modares, K. G. Vamvoudakis, W. He, C.-Z. Xu, and D. C. Wunsch, "Hamiltonian-driven adaptive dynamic programming with approximation errors," *IEEE Transactions on Cybernetics*, vol. 52, no. 12, pp. 13762-13773, 2022, doi: 10.1109/TCYB.2021.3108034.

[45] W. Sun, S. Diao, S.-F. Su, and Z.-Y. Sun, "Fixed-time adaptive neural network control for nonlinear systems with input saturation," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 34, no. 4, pp. 1911-1920, 2023, doi: 10.1109/TNNLS.2021.3105664.

[46] Y. Cheng, B. Xu, Z. Lian, Z. Shi, and P. Shi, "Adaptive learning control of switched strict-feedback nonlinear systems with dead zone using NN and DOB," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 34, no. 5, pp. 2503-2512, 2023.

[47] S. A. A. Rizvi and Z. Lin, "An iterative Q-learning scheme for the global stabilization of discrete-time linear systems subject to actuator saturation," *International Journal of Robust and Nonlinear Control*, vol. 29, no. 9, pp. 2660-2672, 2019, doi: 10.1002/rnc.4514.

[48] Y.-Y. Liu, Z-S. Wang, and Z. Shi, "H∞ tracking control for linear discrete-time systems via reinforcement learning," *International Journal of Robust and Nonlinear Control*, vol. 30, no. 1, pp. 282-301, 2020, doi: 10.1002/rnc.4762.

[49] Z. Shi and Z. Wang, "Adaptive output-feedback optimal control for continuous-time linear systems based on adaptive dynamic programming approach," *Neurocomputing*, vol. 438, pp. 334-344, 2021, doi: 10.1016/j.neucom.2021.01.070.

[50] M. Long, S. Housheng, and Z. Zeng, "Output-feedback global consensus of discrete-time multiagent systems subject to input saturation via Q-learning method," *IEEE Transactions on Cybernetics*, vol. 52, no. 3, pp. 1661-1670, 2022.

[51] Y. Liu and H. Su, "General second-order consensus of discrete-time multiagent systems via Q-learning method," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 52, no. 3, pp. 1417-1425, 2022, doi: 10.1109/TSMC.2020.3019519.

[52] Y. Liu, H. Su, and Z. Zeng, "Second-order consensus for multiagent systems with switched dynamics," *IEEE Transactions on Cybernetics*, vol. 52, no. 6, pp. 4105-4114, 2022.

[53] H. Su, Q. Qiu, X. Chen, and Z. Zeng, "Distributed adaptive containment control for coupled reaction-diffusion neural networks with directed topology," *IEEE Transactions on Cybernetics*, vol. 52, no. 7, pp. 6320-6330, 2022, DOI: 10.1109/TCYB.2020.3034634.

[54] X. Wang, G.-P. Jiang, H. Su, and Z. Zeng, "Consensus of continuous-time linear multiagent systems with discrete measurements," *IEEE Transactions on Cybernetics*, vol. 52, no. 5, pp. 3196-3206, 2022.

[55] Y. Yang, Z. Ding, R. Wang, H. Modares, and D. C. Wunsch, "Data-driven human-robot interaction without velocity measurement using off-policy reinforcement learning," *IEEE/CAA Journal of Automatica Sinica*, vol. 9, no. 1, pp. 47-63, 2021.

[56] W. Xue, B. Lian, J. Fan, P. Kolaric, T. Chai, and F. L. Lewis, "Inverse reinforcement Q-learning through expert imitation for discrete-time

systems," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 34, no. 5, pp. 2386-2399, 2023.

[57] Z. Pan, C. Zhang, Y. Xia, H. Xiong, and X. Shao, "An improved artificial potential field method for path planning and formation control of the multi-UAV systems," *IEEE Transactions on Circuits and Systems II: Express Briefs*, vol. 69, no. 3, pp. 1129-1133, 2021.

[58] J. Sun, H. Zhang, Y. Yan, S. Xu, and X. Fan, "Optimal regulation strategy for nonzero-sum games of the immune system using adaptive dynamic programming," *IEEE Transactions on cybernetics*, vol. 53, no. 3, pp. 1475-1484, 2023, doi: 10.1109/TCYB.2021.3103820.

[59] G. Wen, W. Hao, W. Feng, and K. Gao, "Optimized backstepping tracking control using reinforcement learning for quadrotor unmanned aerial vehicle system," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 52, no. 8, pp. 5004-5015, 2022.

[60] Z. Zhang, C. Wen, L. Xing, and Y. Song, "Adaptive event-triggered control of uncertain nonlinear systems using intermittent output only," *IEEE Transactions on Automatic Control*, vol. 67, no. 8, pp. 4218-4225, 2022, doi: 10.1109/TAC.2021.3115435.

[61] J. Song, D. Shi, Y. Shi, and J. Wang, "Proportional-integral event-triggered control of networked systems with unmatched uncertainties," *IEEE Transactions on Industrial Electronics*, vol. 69, no. 9, pp. 9320-9330, 2022, doi: 10.1109/TIE.2021.3112997.

[62] S. A. Arogeti and F. L. Lewis, "Static output-feedback H∞ control design procedures for continuous-time systems with different levels of model knowledge," *IEEE Transactions on Cybernetics*, vol. 53, no. 3, pp. 1432-1446, 2023, doi: 10.1109/TCYB.2021.3103148.

[63] Q. Wei, J. Lu, T. Zhou, X. Cheng, and F.-Y. Wang, "Event-triggered near-optimal control of discrete-time constrained nonlinear systems with application to a boiler-turbine system," *IEEE Transactions on Industrial Informatics*, vol. 18, no. 6, pp. 3926-3935, 2022.

[64] R. Moghadam and S. Jagannathan, "Optimal adaptive control of uncertain nonlinear continuous-time systems with input and state delays," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 34, no. 6, pp. 3195-3204, 2023.

[65] M. Han, Y. Tian, L. Zhang, J. Wang, and W. Pan, "Reinforcement learning control of constrained dynamic systems with uniformly ultimate boundedness stability guarantee," *Automatica*, vol. 129, p. 109689, 2021, doi: 10.1016/j.automatica.2021.109689.

[66] C.-J. Ong and B. Hou, "Consensus of heterogeneous multi-agent system with input constraints," *Automatica*, vol. 134, p. 109895, 2021.

[67] P. Deptula, Z. I. Bell, F. M. Zegers, R. A. Licitra, and W. E. Dixon, "Approximate optimal influence over an agent through an uncertain interaction dynamic," *Automatica*, vol. 134, p. 109913, 2021.

[68] W. Wang, J. Long, J. Zhou, J. Huang, and C. Wen, "Adaptive backstepping based consensus tracking of uncertain nonlinear systems with event-triggered communication," *Automatica*, vol. 133, 109841, 2021, doi: 10.1016/j.automatica.2021.109841.

[69] S. Li, P. Durdevic, and Z. Yang, "Model-free H∞ tracking control for de-oiling hydrocyclone systems via off-policy reinforcement learning," *Automatica*, vol. 133, p. 109862, 2021.

[70] H. Zhao, X. Meng, and S. Wu, "Distributed edge-based event-triggered coordination control for multi-agent systems," *Automatica*, vol. 132, p. 109797, 2021, doi: 10.1016/j.automatica.2021.109797.

[71] 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, 2021.

[72] V. Rezaei and M. Stefanovic, "Event-triggered cooperative stabilization of multiagent systems with partially unknown interconnected dynamics," *Automatica*, vol. 130, p. 109657, 2021.

[73] P. Li, F. Jabbari, and X.-M. Sun, "Containment control of multi-agent systems with input saturation and unknown leader inputs," *Automatica*, vol. 130, p. 109677, 2021, doi: 10.1016/j.automatica.2021.109677.

[74] B. Wang, H. Ashrafiuon, and S. Nersesov, "Leader–follower formation stabilization and tracking control for heterogeneous planar underactuated vehicle networks," *Systems & Control Letters*, vol. 156, p. 105008, 2021, doi: 10.1016/j.sysconle.2021.105008.

[75] A. Scampicchio, A. Aravkin, and G. Pillonetto, "Stable and robust LQR design via scenario approach," *Automatica*, vol. 129, p. 109571, 2021.

[76] 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, 2021.

[77] C. Li, J. Ding, F. L. Lewis, and T. Chai, "A novel adaptive dynamic programming based on tracking error for nonlinear discrete-time systems," *Automatica*, vol. 129, p. 109687, 2021.

[78] F. Mazenc, M. Malisoff, C. Barbalata, and Z.-P. Jiang, "Event-triggered control using a positive systems approach," *European Journal of Control*, vol. 62, pp. 63-68, 2021.

[79] G. Wang, C. Wang, Y. Ji, and Q. Li, "Practical output consensus control of uncertain nonlinear multi-agent systems without using the higher-order states of neighbours," *IET Control Theory & Applications*, vol. 15, no. 8, pp. 1091-1103, 2021, doi: 10.1049/cth2.12106.

[80] B. Yan, C. Wu, and P. Shi, "Formation consensus for discrete-time heterogeneous multi-agent systems with link failures and actuator/sensor faults," *Journal of the Franklin Institute*, vol. 356, no. 12, pp. 6547-6570, 2019, doi: 10.1016/j.jfranklin.2019.03.028.