SCIENCE CHINA
Information Sciences

• RESEARCH PAPER •

November 2025, Vol. 68, Iss. 11, 210211:1–210211:13
https://doi.org/10.1007/s11432-025-4614-9

Special Topic: Mean-Field Game and Control of Large Population Systems: From Theory to Practice

Zero-sum game control of unmanned aerial vehicle confrontation via reinforcement learning

Zijun LI[1], Yongshuai WANG[2], Guoxing WEN[3] & Chengyi XIA[2*]

[1]*School of Control Science and Engineering, Tiangong University, Tianjin 300387, China*
[2]*School of Artificial Intelligence, Tiangong University, Tianjin 300387, China*
[3]*College of Science, Shandong University of Aeronautics, Binzhou 256600, China*

Received 23 June 2025/Revised 13 August 2025/Accepted 12 September 2025/Published online 30 October 2025

Abstract This paper presents a zero-sum game-based control strategy for the confrontation between the pursuit multi-quadrotor unmanned aerial vehicle (QUAV) and an evaded QUAV via reinforcement learning (RL) and sliding mode control (SMC) techniques. The SMC mechanism drives the attitude states of the multi-QUAV system asymptotically to the pre-defined trajectory. The RL provides a feasible solution to the Hamilton-Jacobi-Isaacs (HJI) equation to obtain the Nash equilibrium in zero-sum games, while conventional analytical methods often struggle with the complexity. Then, under the identifier-double actor-critic (I-DAC) architecture, RL is executed to optimize the consensus control in zero-sum games. The proposed method presents two distinct advantages: (i) adaptive identifier strategies in RL design can compensate for unknown dynamics, and the update rules for actor and critic in RL are significantly simplified; (ii) by integrating RL with the sliding mode mechanism, the Nash equilibrium point can be successfully obtained for both multi-QUAV and single-QUAV zero-sum games when solving the HJI equation. The proposed method will provide an effective game control strategy for unmanned confrontation systems.

Keywords zero-sum game, reinforcement learning, unmanned aerial vehicle, sliding mode control, Nash equilibrium

Citation Li Z J, Wang Y S, Wen G X, et al. Zero-sum game control of unmanned aerial vehicle confrontation via reinforcement learning. Sci China Inf Sci, 2025, 68(11): 210211, https://doi.org/10.1007/s11432-025-4614-9

1 Introduction

At present, the application scenarios for multi-quadrotor unmanned aerial vehicle (multi-QUAV) systems have become increasingly extensive, including military and civilian sectors [1–3]. The control of multi-QUAV systems has emerged as a focal area, particularly in complex adversarial settings [4–6]. Game theory has proven highly effective in strategic choice, where each player is influenced by both its own actions and those of the other participants [7, 8]. A zero-sum game is a strategic interaction in which each player chooses a strategy to maximize its own gain, resulting in losses of an equal total amount for the other players [9, 10]. The control of multi-UAV confrontation systems is fundamentally viewed as a game-theoretic problem, characterized by its zero-sum nature. This property dictates that the system state converges towards an optimal solution. In this solution, the increase in benefit for one party is directly proportional to the decrease in benefit for the other. This characteristic of direct proportionality between the benefits of opposing parties makes zero-sum games ideal for describing the confrontational relationships between pursuers and evaders in multi-QUAV systems.

In the pursuer-evader problem, the pursuer tries to minimize the attitude error between the pursuer and the evader, while the evader tries to maximize the error. The optimal solution to this kind of game theory problem can be obtained by finding its Nash equilibrium or saddle point [11]. The saddle point of a zero-sum game is the optimal intersection of the return functions of both parties in the game, which corresponds to the optimal strategy of each party, making it impossible for either party to obtain a better result by unilaterally changing the strategy [12]. In a nonlinear dynamic environment, saddle points typically require solving the Hamilton-Jacobi-Isaacs (HJI) equation, and it can ensure that both parties achieve the optimal game equilibrium in a complex nonlinear environment [13]. The HJI equation is due to its capability to precisely delineate the dynamic optimization process within a zero-sum game

* Corresponding author (email: cyxia@tiangong.edu.cn)

between two parties, through the introduction of the Hamiltonian function, integration of states, control inputs, and respective cost functions [14].

Although the HJI equation provides a theoretical basis for the optimal control, it is often very difficult to solve analytically in practical applications, especially for highly complex systems and uncertainties [15]. In addition, in some traditional optimal control methods, the reinforcement learning (RL) algorithm is very complex and requires known dynamics about the system, making it difficult to expand and apply [16–19]. To overcome this challenge, this study improves the identifier-actor-critic (IAC) framework [20] in RL so as to generate the novel I-DAC arithmetic through the continuous iterative training of the adaptive neural network. In the context of the game, both participants can derive suitable actions based on the evaluative feedback provided by the system response. So as to continuously improve the system performance. Within this framework, the identifier serves to recognize environmental or state conditions, transmitting this information to the critic for evaluation. The critic, in turn, assesses the situation and provides feedback to the actor, who takes actions or makes decisions based on the critic assessments. This cycle guides the intelligent system in determining the next appropriate action under the current environment [21].

In the multi-QUAV attitude system, the robustness of the control strategy is very important since the system has strong nonlinearity and uncertainty [22]. To achieve stable and robust performance in an uncertain environment, sliding mode control (SMC) presents an effective solution [23]. The SMC guarantees system stability after entering the sliding mode plane by designing a suitable sliding mode surface, and can effectively control multiple state variables and constrain the system state on a predetermined sliding mode hyperplane [24, 25]. In the game control of pursuer and evader, SMC can effectively constrain error dynamics [26]. Combined RL with SMC, a stable and robust control strategy can be realized for the uncertain environment [27].

In this paper, the zero-sum game control of multi-QUAV confrontation via RL and SMC is studied. The primary contributions of this study are summarized as follows.

(i) An optimization approach based on zero-sum game theory is proposed for the attitude game control problem involving multi-QUAV and single-QUAV. In this zero-sum framework, the losses incurred by one party are exactly equal to the gains of the other one. Consequently, the optimal strategies obtained by solving the HJI equations constitute a saddle point equilibrium for the game.

(ii) The I-DAC scheme is presented to solve the HJI equation to obtain the saddle-point solution. Unlike conventional zero-sum game methods, this approach significantly simplifies the optimal game control algorithm by deriving the reinforcement learning weight update laws via a simple positive function that equivalently represents the HJI equation.

(iii) The proposed zero-sum game-based control method eliminates the need for persistent excitation conditions and complete dynamic knowledge, as the adaptive identifier in the I-DAC framework can effectively compensate for unknown dynamic functions, and the RL algorithm can effectively train adaptive parameters to eliminate continuous excitation conditions. Finally, the stability is conducted by using the Lyapunov theory.

2 Preliminaries

2.1 Attitude system description

For an interconnected multi-QUAV system, the attitude dynamic of each QUAV [28] can be expressed using the Newton-Euler formulation, which is

$$\begin{aligned}
\ddot{\phi}_k(t) &= \frac{l\tau_{\phi k}}{I_{xk}} + \dot{\psi}_k(t)\dot{\theta}_k(t)\left(\frac{I_{yk}-I_{zk}}{I_{xk}}\right) - \frac{G_{\phi k}l}{I_{xk}}\dot{\phi}_k(t),\\
\ddot{\psi}_k(t) &= \frac{l\tau_{\psi k}}{I_{yk}} + \dot{\phi}_k(t)\dot{\theta}_k(t)\left(\frac{I_{zk}-I_{xk}}{I_{yk}}\right) - \frac{G_{\psi k}l}{I_{yk}}\dot{\psi}_k(t),\ k=1,2,\ldots,n,\\
\ddot{\theta}_k(t) &= \frac{l\tau_{\theta k}}{I_{zk}} + \dot{\phi}_k(t)\dot{\psi}_k(t)\left(\frac{I_{xk}-I_{yk}}{I_{zk}}\right) - \frac{G_{\theta k}l}{I_{zk}}\dot{\theta}_k(t),
\end{aligned} \tag{1}$$

where $\phi_k(t)$, $\psi_k(t)$ and $\theta_k(t)$ are the roll, pitch and yaw angles, constrained in $\phi_k \in [-\frac{\pi}{2}, \frac{\pi}{2}]$, $\psi_k \in [-\frac{\pi}{2}, \frac{\pi}{2}]$ and $\theta_k \in [-\pi, \pi]$; $\tau_{\phi k}, \tau_{\psi k}$ and $\tau_{\theta k}$ are control torques; I_{xk}, I_{yk} and I_{zk} are rotational inertia on the x, y, z coordinate; $G_{\phi k}, G_{\psi k}$ and $G_{\theta k}$ are drag coefficients, and l is the length from mass center to rotor center.

For designing the optimized game control, the pursuer attitude dynamic model in (1), is reformulated as

$$\dot{x}_{pk}(t) = x_{vk}(t),\ \dot{x}_{vk}(t) = u_{zk} + F_k(x_{pk}, x_{vk}),\ k = 1, 2, \ldots, n, \tag{2}$$

where $x_{pk}(t) = [\phi_k(t), \psi_k(t), \theta_k(t)]^{\mathrm{T}} \in \mathbb{R}^3$ and $x_{vk}(t) = [\dot{\phi}_k(t), \dot{\psi}_k(t), \dot{\theta}_k(t)]^{\mathrm{T}} \in \mathbb{R}^3$, $u_{zk} = [l\tau_{\phi k}/I_{xk}, l\tau_{\psi k}/I_{yk}, l\tau_{\theta k}/I_{zk}]^{\mathrm{T}} \in \mathbb{R}^3$, $F_k(x_{pk}, x_{vk}) = [\dot{\psi}_k(t)\dot{\theta}_k(t)(I_{yk} - I_{zk})/I_{xk} - G_{\phi k}l\dot{\phi}_k(t)/I_{xk}, \dot{\phi}_k(t)\dot{\theta}_k(t)(I_{zk} - I_{xk})/I_{yk} - G_{\psi k}l\dot{\psi}_k(t)/I_{yk}, \dot{\phi}_k(t)\dot{\psi}_k(t)(I_{xk} - I_{yk})/I_{zk} - G_{\theta i}l\dot{\theta}_k(t)/I_{zk}]^{\mathrm{T}} \in \mathbb{R}^3$.

The dynamic mode of the evader attitude is

$$\dot{x}_{ps}(t) = x_{vs}(t),\ \dot{x}_{vs}(t) = u_s + F_s(x_{ps}, x_{vs}), \tag{3}$$

where $x_{ps}(t) = [\phi_s(t), \psi_s(t), \theta_s(t)]^{\mathrm{T}} \in \mathbb{R}^3$, $x_{vs}(t) = [\dot{\phi}_s(t), \dot{\psi}_s(t), \dot{\theta}_s(t)]^{\mathrm{T}} \in \mathbb{R}^3$, and $u_s = [lu_{\phi s}/I_x, lu_{\psi s}/I_y, lu_{\theta s}/I_z]^{\mathrm{T}} \in \mathbb{R}^3$, $F_s(x_{ps}, x_{vs})$ is a continuous nonlinear function.

Definition 1. The multi-QUAV system (1) is said to achieve the second-order pursuer-evader consensus, if $\lim_{t\to\infty} \|x_{pk}(t) - x_{ps}(t)\| = 0$ and $\lim_{t\to\infty} \|x_{vk}(t) - x_{vs}(t)\| = 0$ hold.

Control objective. For the multi-QUAV system (1), the goal is to determine the optimal consensus control, such that (i) the optimal control algorithm based on zero-sum differential game can keep the dynamic equilibrium of the controller at saddle point; (ii) all control signals are guaranteed to be semi-globally uniformly ultimately bounded (SGUUB), ensuring stability and performance within a specified bound; (iii) the consensus described by the pursuer-evader in Definition 1 can be obtained.

2.2 Algebraic graph theory

The communication network within the multi-QUAV system is characterized through an undirected topological graph, denoted as $G = (\Pi, \Psi, A)$, where $A = [a_{ij}] \in \mathbb{R}^{n\times n}$, $\Pi = \{\Pi_1, \Pi_2, \ldots, \Pi_n\}$, $\Psi \subset \Pi \times \Pi$ represent the adjacency matrix, the node set, and edge set, respectively. If there exists a pathway for information communication from node Π_j to node Π_i, then node Π_j is considered to be a neighbor of node Π_i, where the node Π_i is the behalf of the ith agent of multi-QUAV attitude system. Furthermore, the element a_{ij} of the adjacency matrix A is set to 1; otherwise $a_{ij} = 0$ and also $a_{ii} = 0$. The G is said to be an undirected graph if and only if the adjacency matrix A is symmetrical, i.e., $a_{ij} = a_{ii}$.The set of neighbors of Π_i, is denoted by $\Lambda_i = \{j|(\Pi_i, \Pi_j) \in \Psi\}$.

The Laplacian matrix L of the graph G can be constructed as

$$L = \Psi - A, \tag{4}$$

where $\Psi = \mathrm{diag}\{\Psi_1, \Psi_2, \cdots, \Psi_n\}$ and $\Psi_i = \sum_{j=1}^{n} a_{ij}$, $\Psi = \mathrm{diag}\{\sum_{j=1}^{n} a_{1j}, \ldots, \sum_{j=1}^{n} a_{nj}\}$.

Assumption 1. The communication topology of the multi-QUAV system in (1) is represented by an undirected connected graph.

Lemma 1 ([29])**.** If the communication topology graph G is an undirected and connected graph, then the Laplacian matrix L defined in (4) is classified as an irreducible matrix.

Lemma 2 ([29])**.** When the Laplacian matrix L possesses the property of being irreducible, then $\tilde{L} = L + B$, where $B = \mathrm{diag}\{b_1, \ldots, b_m\}$ and $b_1 + b_2 + \cdots + b_m > 0$, is a positive definite matrix.

2.3 Neural network (NN)

In [30], it is proved that a nonlinear and continuous function $K(\varsigma) : \mathbb{R}^n \to \mathbb{R}^m$, which is delineated on a compact domain Ω, the formulation for the NN approximation can be articulated as

$$K(\varsigma) = \omega^{\mathrm{T}}\tau(\varsigma), \tag{5}$$

where $\omega \in \mathbb{R}^{p\times m}$ is the weight matrix associated with the NN, and p denotes the count of neurons comprising the network, and $\tau(\varsigma) = [\tau_1(\varsigma), \ldots, \tau_p(\varsigma)]^{\mathrm{T}}$ represents the basis function vector, on which $\tau_k(\varsigma) = \exp[-(\varsigma - o_k)^{\mathrm{T}}(\varsigma - o_k)/2\rho_k^2]$, $o_k \in \mathbb{R}^n$ and $\rho_k \in \mathbb{R}$ are the centroid of the receptive field and the breadth of the Gaussian function, respectively.

In (5), an optimal weight matrix is denoted as $\omega^* = \arg\min_{\omega\in\mathbb{R}^{p\times m}} \{\sup_{\varsigma\in\Omega} \|K(\varsigma) - \omega^{\mathrm{T}}\tau(\varsigma)\|\}$, then the function $K(\varsigma)$ is redefined as

$$K(\varsigma) = \omega^{*\mathrm{T}}\tau(\varsigma) + \varepsilon(\varsigma), \tag{6}$$

where the approximation error $\varepsilon(\varsigma)$ satisfies $\|\varepsilon(\varsigma)\| \leqslant \upsilon$, and $\upsilon > 0$ is a constant. Furthermore, by selecting an appropriate number of NN neurons, the error $\varepsilon(\varsigma)$ can be sufficiently small.

3 Main results

3.1 Zero-sum game formulation description

Define the tracking errors of pursuer-evader as

$$\zeta_{pk}(t) = x_{pk}(t) - x_{ps}(t),\ \zeta_{vk}(t) = x_{vk}(t) - x_{vs}(t),\ k = 1, 2, \ldots, n. \tag{7}$$

Based on (2) and (3), dynamic of the tracking error can be formulated as

$$\dot{\zeta}_{pk}(t) = \zeta_{vk}(t),\ \dot{\zeta}_{vk}(t) = u_{zk} - u_s + F_k(x_{pk}, x_{vk}) - F_s(x_{ps}, x_{vs}),\ k = 1, 2, \ldots, n. \tag{8}$$

To achieve pursuer-evader consensus for the multi-QUAV system, a sliding mode variable, incorporating attitude errors, is introduced as

$$s_k(\bar{\zeta}_k) = \beta\zeta_{pk}(t) + \zeta_{vk}(t),\ k = 1, 2, \ldots, n, \tag{9}$$

where $\beta > 0$ denotes a constant to be designed later, and $\bar{\zeta}_k(t) = [\zeta_{pk}^{\mathrm{T}}, \zeta_{vk}^{\mathrm{T}}]^{\mathrm{T}} \in \mathbb{R}^{2\times 3}$. It is worth noting that, utilizing the sliding mode mechanism [23], the attitude tracking error is limited to a small zero neighborhood when $s_k(\bar{\zeta}_k) \to 0$.

Drawing upon (8), the dynamics of the sliding mode can be derived as

$$\begin{aligned}\dot{s}_k &= \beta\zeta_{vk}(t) + u_{zk} + F_k(x_{pk}, x_{vk}) - u_s - F_s(x_{ps}, x_{vs})\\ &= F_k^*(x_{pk}, x_{vs}) - F_s^*(x_{ps}, x_{vs}) + u_{zk} - u_s,\ k = 1, 2, \ldots, n,\end{aligned} \tag{10}$$

where $F_k^*(x_{pk}, x_{vs}) = \beta x_{vk}(t) + F_k(x_{pk}, x_{vk})$ and $F_s^*(x_{ps}, x_{vs}) = \beta x_{vs}(t) + F_s(x_{ps}, x_{vs})$.

Define the term that incorporates neighboring states for consensus as

$$E_k^s(t) = \sum_{i\in\Lambda_k} a_{ki}\left(\left(\beta x_{pk} + x_{vk}\right) - \left(\beta x_{pi} + x_{vi}\right)\right) + b_k s_k(t),\ k = 1, 2, \ldots, n, \tag{11}$$

where Λ_k is the neighbor tag set associated with agent k.

By adding and subtracting the term $\beta x_{ps} + x_{vs}$, the consensus term (11) can be reformulated as

$$E_k^s(t) = \sum_{i\in\Lambda_k} a_{ki}\big(s_k(t) - s_i(t)\big) + b_k s_k(t),\ k = 1, 2, \ldots, n. \tag{12}$$

Taking derivative of $E_k^s(t)$ and combining (10) can obtain

$$\dot{E}_k^s(t) = \lambda_k\left(F_k^*(x_{pk}, x_{vs}) - F_s^*(x_{ps}, x_{vs}) + u_{zk} - u_s\right) - \sum_{i\in\Lambda_k} a_{ki}\dot{s}_i(t), \tag{13}$$

where $\lambda_k = \sum_{i\in\Lambda_k} a_{ki} + b_k$.

3.2 Zero-sum differential game-based optimal controller design

Define the performance index as

$$J(0) = \int_0^\infty c\big(s, u_z, u_s\big)\mathrm{d}\nu, \tag{14}$$

where $c(s, u_z, u_s) = s^{\mathrm{T}}(t)\big(\tilde{L}^{\mathrm{T}}\tilde{L} \otimes I_3\big)s(t) \in \mathbb{R}$ is the cost function, $s(t) = [s_1^{\mathrm{T}}, \ldots, s_n^{\mathrm{T}}]^{\mathrm{T}} \in \mathbb{R}^{3n}$, $u_z = [u_{z1}^{\mathrm{T}}, \ldots, u_{zn}^{\mathrm{T}}]^{\mathrm{T}} \in \mathbb{R}^{3n}$, and $\tilde{L} = L + B$.

Definition 2 (Admissible control [19]). A control input pair $\{u_z, u_s\}$ is called as admissible with respect to (14) on the set Ω, if u_z and u_s are continuous to stabilizes (13) on Ω with $u_z(0) = 0$ and $u_s(0) = 0$, and make (14) finite.

By considering that $E_s(t) = \big(\tilde{L} \otimes I_3\big)s(t)$, and $E_s(t) = [E_1^{s\mathrm{T}}(t), \ldots, E_n^{s\mathrm{T}}(t)]^{\mathrm{T}} \in \mathbb{R}^{3n}$, the cost function $c(s, u_z, u_s)$ can be re-described as

$$\begin{aligned} c(s, u_z, u_s) &= s^{\mathrm{T}}(t)\big(\tilde{L}^{\mathrm{T}}\tilde{L} \otimes I_3\big)s(t) + u_z^{\mathrm{T}}(P_1 \otimes I_n)u_z - u_s^{\mathrm{T}}P_2u_s \\ &= E_s^{\mathrm{T}}(t)E_s(t) + u_z^{\mathrm{T}}(P_1 \otimes I_n)u_z - u_s^{\mathrm{T}}P_2u_s, \end{aligned} \tag{15}$$

where $P_1, P_2 \in \mathbb{R}^{3\times 3}$. Then, the distributed performance function is formulated as

$$J(0) = \int_0^\infty c(E_s, u_z, u_s)\mathrm{d}\nu. \tag{16}$$

Utilizing (16), the definition of the distributed performance function is

$$J(E_s) = \int_t^\infty c(E_s, u_z, u_s)\mathrm{d}\nu. \tag{17}$$

Letting $u_z^* \in \mathbb{R}^{3n}$ and $u_s^* \in \mathbb{R}^3$ denote the optimal pursuer and evader controls and considering the zero-sum game theory, the optimal distributed performance index $J^*(E_s) \in \mathbb{R}$ is expressed as

$$J^*(E_s) = \int_t^\infty c(E_s, u_z^*, u_s^*)\mathrm{d}\nu = \min_{u_z\in\psi(\Omega)} \max_{u_s\in\psi(\Omega)} \left\{ \int_t^\infty c(E_s, u_z, u_s)\mathrm{d}\nu \right\}, \tag{18}$$

where $\Omega \in \mathbb{R}^n$ is a given compact set.

Definition 3 ([9]). The control policy denoted by (u_z^*, u_s^*) is the saddle point equilibrium solution in a zero-sum game, if

$$J(E_s, u_z^*, u_s) \leqslant J(E_s, u_z^*, u_s^*) \leqslant J(E_s, u_z, u_s^*). \tag{19}$$

Taking the time derivative of (18) along (13), the distributed HJI equation is

$$\begin{aligned} \tilde{H}_J(E_s, u_z^*, u_s^*, J^*) =& c(E_s, u_z^*, u_s^*) + \frac{\mathrm{d}J^*(E_s)}{\mathrm{d}t} \\ =& \|E_s(t)\|^2 + \sum_{k=1}^n u_{zk}^{\mathrm{T}}P_1u_{zk} - u_s^{\mathrm{T}}P_2u_s + \sum_{k=1}^n \frac{\mathrm{d}J_k^*(E_k^s)}{\mathrm{d}E_k^{s\mathrm{T}}} \Bigg(\lambda_k\big(F_k^*(x_{pk}, x_{vk}) + u_{zk} - u_s \\ & - F_s^*(x_{ps}, x_{vs})\big) - \sum_{i\in\Lambda_k} a_{ki}\big(F_i^*(x_{pi}, x_{vi}) - F_s^*(x_{ps}, x_{vs}) + u_{zi} - u_s\big) \Bigg) = 0, \\ k =& 1, 2, \ldots, n, \end{aligned} \tag{20}$$

where $J_k^*(E_k^s)$ is k-th optimal distributed performance index.

As previously indicated, the optimal game pursuer-evader controls u_z^* and u_s^* must exclusively fulfill the optimal performance function given in (18). Consequently, the optimal solution to the distributed HJI equation presented in (20). Subsequently, the determination of u_s^* can be achieved as

$$\frac{\partial \tilde{H}_J(E_s^s, u_z^*, u_s^*, J^*)}{\partial u_s^*} = 0 \Longrightarrow u_s^* = -\frac{1}{2}P_2^{-1}\eta_s \sum_{k=1}^n \frac{\mathrm{d}J_k^*(E_k^s)}{\mathrm{d}E_k^s}, \tag{21}$$

where $\eta_s = \sum_{k=1}^n b_k$, and $\sum_{k=1}^n \frac{\mathrm{d}J_k^*(E_k^s)}{\mathrm{d}E_k^s} = \frac{\mathrm{d}J^*(E_s^s)}{\mathrm{d}E_s^s}$.

Due to the dynamic coupling game relationship in the game, to avoid ambiguity in the local derivation of the global index, the local optimization problem of a single pursuer is extracted from (20), u_{zm} can be calculated from the partial differential equation as

$$\tilde{H}_m^J(E_m^s, u_{zm}^*, u_s^*, J_m^*) = c_m(E_m^s, u_{zm}^*, u_s^*) + \frac{\mathrm{d}J_m^*(E_m^s)}{\mathrm{d}t}$$

$$=\|E_m^s(t)\|^2+u_{zm}^{\mathrm{T}}P_1u_{zm}-u_s^{\mathrm{T}}P_2u_s+\frac{\mathrm{d}J_m^*(E_m^s)}{\mathrm{d}E_m^{s\mathrm{T}}}\lambda_m(F_m^*(x_{pm},x_{vm})$$
$$-F_s^*(x_{ps},x_{vs})+u_{zm}-u_s)-\frac{\mathrm{d}J_m^*(E_m^s)}{\mathrm{d}E_m^{s\mathrm{T}}}\sum_{i\in\Lambda_m}a_{mi}\dot{s}_i(t),\ m=1,2,\ldots,n, \quad (22)$$

where m is the index of a single pursuer, which is equivalent to the global index k and can traverse all pursuers.

By transforming (22), the optimal strategy u_{zm}^* of any agent among the multiple pursuers is

$$\frac{\partial\tilde{H}_m^J(E_m^s,u_{zm}^*,u_s^*,J_m^*)}{\partial u_{zm}^*}=0\Longrightarrow u_{zm}^*=-\frac{1}{2}P_1^{-1}\lambda_m\frac{\mathrm{d}J_m^*(E_m^s)}{\mathrm{d}E_m^s},\ m=1,2,\ldots,n. \quad (23)$$

Substituting (23) into (22), the HJI equation is derived as

$$\tilde{H}_m^J(E_m^s,u_{zm}^*,u_s^*,J_m^*)=\|E_m^s(t)\|^2-\frac{1}{4}\lambda_m^2\sum_{m=1}^n\frac{\mathrm{d}J_m^*(E_m^s)}{\mathrm{d}E_m^{s\mathrm{T}}}P_1^{-1}\frac{\mathrm{d}J_m^*(E_m^s)}{\mathrm{d}E_m^s}+\frac{1}{4}\eta_s^2\sum_{m=1}^n\frac{\mathrm{d}J_m^*(E_m^s)}{\mathrm{d}E_m^{s\mathrm{T}}}P_2^{-1}\frac{\mathrm{d}J_m^*(E_m^s)}{\mathrm{d}E_m^s}$$
$$+\sum_{m=1}^n\frac{\mathrm{d}J_m^*(E_m^s)}{\mathrm{d}E_m^{s\mathrm{T}}}\Bigg(\lambda_m\big(F_m^*(x_{pm},x_{vm})-F_s^*(x_{ps},x_{vs})\big)-\sum_{i\in\Lambda_m}a_{mi}$$
$$\times\left(F_i^*(x_{pi},x_{vi})-F_s^*(x_{ps},x_{vs})-\frac{1}{2}\lambda_iP_1^{-1}\frac{\mathrm{d}J_i^*(E_i^s)}{\mathrm{d}E_i^s}\right)\Bigg),m=1,2,\ldots,n. \quad (24)$$

3.3 Reinforcement learning (RL) design

To derive the game optimized consensus control for multi-QUAV attitude system (1), an RL is designed and the term $\frac{\mathrm{d}J_m^*(E_m^s)}{\mathrm{d}E_m^{s\mathrm{T}}}$ of (24) is decomposed as

$$\frac{\mathrm{d}J_m^*(E_m^s)}{\mathrm{d}E_m^{s\mathrm{T}}}=\frac{2P_1}{\lambda_m}\mu_mE_m^s(t)+\frac{2P_1}{\lambda_m}F_m^{s*}+\frac{P_1}{\lambda_m}J_m^0(E_m^s,\bar{x}_m),\ m=1,2,\ldots,n, \quad (25)$$

where $J_m^0(E_m^s,\bar{x}_m)=-2\mu_mP_1^{-1}E_m^s(t)-2F_m^{s*}+\lambda_mP_1^{-1}\frac{\mathrm{d}J_m^*(E_m^s)}{\mathrm{d}E_m^{s\mathrm{T}}}$, $F_m^{s*}=F_m^*(x_{pm},x_{vm})-F_s^*(x_{ps},x_{vs})$, $\bar{x}_m=[x_{pm}^{\mathrm{T}},x_{vm}^{\mathrm{T}}]^{\mathrm{T}}$, and $\mu_m>0$ is the design constant.

Substituting (25) into (23) yields

$$u_{zm}^*=-\mu_mE_m^s(t)-F_m^{s*}-\frac{1}{2}J_m^0(E_m^s,\bar{x}_m),\ m=1,2,\ldots,n, \quad (26)$$

which is infeasible due to the uncertainty $J_m^0(E_m^s,\bar{x}_m)$ and F_m^{s*}. Therefore, NNs are harnessed to approximate the terms within the confines of the compact set Ω, which is

$$F_m^{s*}=\Phi_{fm}^{*\mathrm{T}}\xi_{fm}(\bar{x}_m)+\varepsilon_{fm}(\bar{x}_m), \quad (27)$$

$$J_m^0(E_m^s,\bar{x}_m)=\Phi_m^{*\mathrm{T}}\xi_m(E_m^s,\bar{x}_m)+\varepsilon_m(E_m^s,\bar{x}_m),\ m=1,2,\ldots,n, \quad (28)$$

where $\Phi_{fm}^*\in\mathbb{R}^{q_1\times3}$ and $\Phi_m^*\in\mathbb{R}^{q_2\times3}$ are the ideal NN weight vectors, $\xi_{fm}(\bar{x}_m)\in\mathbb{R}^{q_1}$ and $\xi_m(E_m^s,\bar{x}_m)\in\mathbb{R}^{q_2}$ are the basis function vectors. The bounded errors associated with the NN approximations are represented by $\varepsilon_{fm}\in\mathbb{R}^3$ and $\varepsilon_m\in\mathbb{R}^3$, i.e., $\|\varepsilon_{fm}\|\leqslant\varrho_{fm}$ and $\|\varepsilon_m\|\leqslant\varrho_m$.

$$\frac{\mathrm{d}J_m^*(E_m^s)}{\mathrm{d}E_m^{s\mathrm{T}}}=\frac{2P_1}{\lambda_m}\mu_mE_m^s(t)+\frac{2P_1}{\lambda_m}\Big(\Phi_{fm}^{*\mathrm{T}}\xi_{fm}(\bar{x}_m)+\varepsilon_{fm}(\bar{x}_m)\Big)+\frac{P_1}{\lambda_m}\Big(\Phi_m^{*\mathrm{T}}\xi_m(E_m^s,\bar{x}_m)+\varepsilon_m(E_m^s,\bar{x}_m)\Big), \quad (29)$$

$$u_{zm}^*=-\mu_mE_m^s(t)-\Phi_{fm}^{*\mathrm{T}}\xi_{fm}(\bar{x}_m)-\varepsilon_{fm}(\bar{x}_m)-\frac{1}{2}\Phi_m^{*\mathrm{T}}\xi_m(E_m^s,\bar{x}_m)-\frac{1}{2}\varepsilon_m(E_m^s,\bar{x}_m),$$
$$m=1,2,\ldots,n. \quad (30)$$

Since Φ_{fm}^* and Φ_m^* constitute two unknown constant vectors, the feasibility of the optimal control in (29) remains compromised. To find a feasible and optimized control strategy, RL is implemented through the adaptive identifier, actor, and critic NNs.

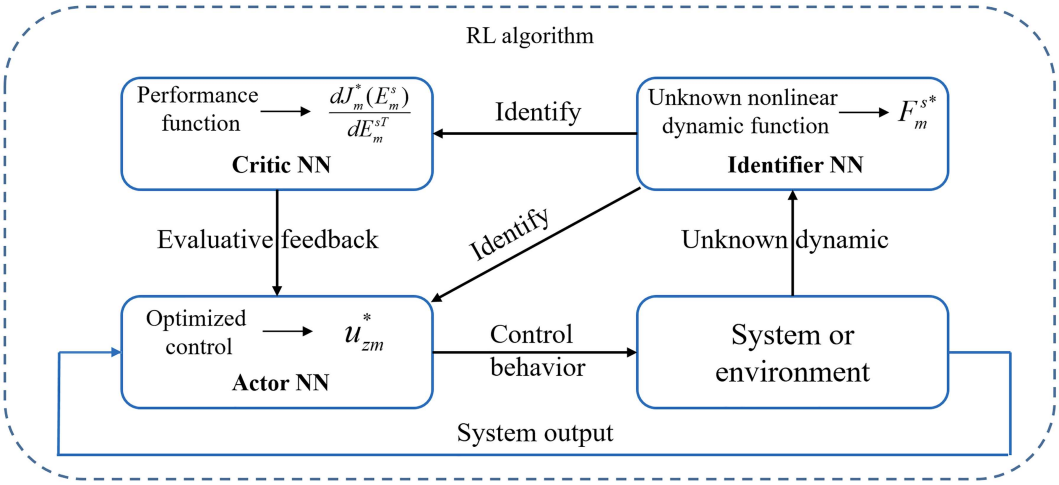

Figure 1 (Color online) Identifier-critic-actor RL framework for evader u_{zm}^*.

For the purpose of approximating the uncertain dynamic function, the identifier NN is constructed as

$$\hat{F}_m^s(\bar{x}_m) = \hat{\Phi}_{fm}^{\mathrm{T}}(t)\xi_{fm}(\bar{x}_m),\ m = 1, 2, \ldots, n, \tag{31}$$

where $\hat{\Phi}_{fm}(t) \in \mathbb{R}^{q_1 \times 3}$ represents the estimation of the ideal weight for the identifier NN, $\hat{F}_m^s(\bar{x}_m)$ denotes the output produced by the adaptive identifier. The weight $\hat{\Phi}_{fm}(t)$ of the adaptive identifier NN is trained according to the following update law as

$$\dot{\hat{\Phi}}_{fm}(t) = \phi_{fm}\Big(\xi_{fm}(\bar{x}_m)E_m^{s\mathrm{T}}(t) - \sigma_{fm}\hat{\Phi}_{fm}(t)\Big),\ m = 1, 2, \ldots, n, \tag{32}$$

where $\phi_{fm}, \sigma_{fm} \in \mathbb{R}$ are the positive constants.

The formulation for the critic NN, tasked with approximating the unknown term $\frac{\mathrm{d}\hat{J}_m^*(E_m^s)}{\mathrm{d}E_m^{s\mathrm{T}}}$ is

$$\frac{\mathrm{d}\hat{J}_m^*(E_m^s)}{\mathrm{d}E_m^{s\mathrm{T}}} = \frac{2P_1}{\lambda_m}\mu_m E_m^s(t) + \frac{2P_1}{\lambda_m}\hat{\Phi}_{fm}^{\mathrm{T}}(t)\xi_{fm}(\bar{x}_m) + \frac{P_1}{\lambda_m}\hat{\Phi}_{cm}^{\mathrm{T}}(t)\xi_m(E_m^s, \bar{x}_m),\ m = 1, 2, \ldots, n, \tag{33}$$

where $\hat{\Phi}_{cm}(t) \in \mathbb{R}^{q_2 \times 3}$ refers to the adaptive weight of the critic NN. The description of the updating rule for $\hat{\Phi}_{cm}(t)$ is outlined as

$$\dot{\hat{\Phi}}_{cm}(t) = -\kappa_{cm}\Big(\xi_m(E_m^s, \bar{x}_m)\xi_m^{\mathrm{T}}(E_m^s, \bar{x}_m) + \sigma_m I_q\Big)\hat{\Phi}_{cm}(t),\ m = 1, 2, \ldots, n, \tag{34}$$

where $\kappa_{cm} > 0$ represents the critic gain constant, I_q denotes the $q_2 \times q_2$ identity matrix and $\sigma_m > 0$ is the design constant.

The actor NN of the optimized consensus control $\hat{u}_{zm}^*$ is

$$\hat{u}_{zm}^* = -\mu_m E_m^s(t) - \hat{\Phi}_{fm}^{\mathrm{T}}(t)\xi_{fm}(\bar{x}_m) - \frac{1}{2}\hat{\Phi}_{am}^{\mathrm{T}}(t)\xi_m(E_m^s, \bar{x}_m),\ m = 1, 2, \ldots, n, \tag{35}$$

where $\hat{\Phi}_{am}(t) \in \mathbb{R}^{q_2 \times 3}$ is the actor adaptive NN weight. The updating rule for $\hat{\Phi}_{am}(t)$ is

$$\dot{\hat{\Phi}}_{am}(t) = -\kappa_{am}\Big(\xi_m(E_m^s, \bar{x}_m)\xi_m^{\mathrm{T}}(E_m^s, \bar{x}_m) + \sigma_m I_q\Big)\Big(\hat{\Phi}_{am}(t) - \hat{\Phi}_{cm}(t)\Big) - \frac{1}{2}\xi_m(E_m^s, \bar{x}_m)\xi_m^{\mathrm{T}}(E_m^s, \bar{x}_m)\hat{\Phi}_{am}(t),\ m = 1, 2, \ldots, n, \tag{36}$$

where $\kappa_{am} > 0$ denotes the actor gain constant. Figure 1 depicts the identifier-critic-actor RL algorithm framework of the pursuer u_{zm}^*.

Design the NN about evader u_s^* as

$$\frac{\mathrm{d}J_s^*(E_s^s)}{\mathrm{d}E_s^{s\mathrm{T}}} = \frac{P_2}{\eta_s}\Phi_s^{*\mathrm{T}}\xi_s(E_s^s, \bar{x}_s) + \frac{P_2}{\eta_s}\varepsilon_s(E_s^s, \bar{x}_s), \tag{37}$$

$$u_s^* = -\frac{1}{2}\Big(\Phi_s^{*\mathrm{T}}\xi_s(E_s^s, \bar{x}_s) + \varepsilon_s(E_s^s, \bar{x}_s)\Big), \tag{38}$$

where $\Phi_s^* \in \mathbb{R}^{q_3\times 3}$ is the ideal NN weight vectors, $\xi_s(E_s^s, \bar{x}_s) \in \mathbb{R}^{q_3}$ is the basis function vectors, $\varepsilon_s(E_s^s, \bar{x}_s) \leqslant \varrho_s$ represents the bounded errors associated with the NN approximations, and $\bar{x}_s = [x_{ps}^{\mathrm{T}}, x_{vs}^{\mathrm{T}}]^{\mathrm{T}}$.

The critic NN of $\mathrm{d}\hat{J}_s^*(E_s^s)/\mathrm{d}E_s^s$ is

$$\frac{\mathrm{d}\hat{J}_s^*(E_s^s)}{\mathrm{d}E_s^{s\mathrm{T}}} = \frac{P_2}{\eta_s}\hat{\Phi}_{cs}^{*\mathrm{T}}\xi_s(E_s^s, \bar{x}_s), \tag{39}$$

where $\hat{\Phi}_{cs}^* \in \mathbb{R}^{q_3\times 3}$ denotes the adaptive weight of the critic NN, the updating rule for $\hat{\Phi}_{cs}(t)$ is

$$\dot{\hat{\Phi}}_{cs}(t) = -\kappa_{cs}\Big(\xi_s(E_s^s, \bar{x}_s)\xi_s^{\mathrm{T}}(E_s^s, \bar{x}_s) + \sigma_s I_n\Big)\hat{\Phi}_{cs}(t), \tag{40}$$

where $\kappa_{cs}, \sigma_s \in \mathbb{R}$ represent two positive design constants that comprise the gain matrix.

The actor NN of u_s^* is

$$\hat{u}_s^* = -\frac{1}{2}\hat{\Phi}_{as}^{\mathrm{T}}(t)\xi_s(E_s^s, \bar{x}_s), \tag{41}$$

where $\hat{u}_s^*$ denotes the escape optimal strategy, the actor adaptive NN weight $\hat{\Phi}_{as}(t) \in \mathbb{R}^{q_3\times 3}$ is

$$\dot{\hat{\Phi}}_{as}(t) = -\kappa_{as}\Big(\xi_s(E_s^s, \bar{x}_s)\xi_s^{\mathrm{T}}(E_s^s, \bar{x}_s) + \sigma_s I_n\Big)\Big(\hat{\Phi}_{as}(t) - \hat{\Phi}_{cs}(t)\Big), \tag{42}$$

where $\kappa_{as} > 0$ is the actor gain constant.

3.4 Main theorem and proof

Lemma 3 ([31])**.** If the positive continuous function $P(t) \in \mathbb{R}$ satisfies $\dot{P}(t) \leqslant -\varpi P(t) + \Xi$, where $\varpi > 0$ and $\Xi > 0$ are constants, then, the following inequality holds true

$$P(t) \leqslant \mathrm{e}^{-\varpi t}P(0) + \frac{\Xi}{\varpi}\big(1 - \mathrm{e}^{-\varpi t}\big). \tag{43}$$

Theorem 1. For the multi-QUAV attitude system described in (1) with the bounded initial values, if the design parameters fulfill the subsequent conditions

$$\mu_m > \frac{3}{4},\ \kappa_{am} > -\frac{1}{2},\ \kappa_{cm} > \kappa_{am},\ \kappa_{cs} > \kappa_{as} > 0,\ \sigma_m > 0,\ \sigma_s > 0, \tag{44}$$

where $\lambda_{min}^{\tilde{L}}$ represents the minimum eigenvalue of $\tilde{L} = L + B$, the designed I-DAC RL algorithm (34)–(39) for optimized consensus control in the zero-sum game can achieve

(i) All control signals are SGUUB;

(ii) The tracking errors $\zeta_{mj}(t)$, $m = 1, 2, \ldots, n$, $j = 1, 2, \ldots, i$, can converge to a small neighborhood of zero.

Proof. The Lyapunov function is constructed as

$$\begin{aligned} V(t) =& \frac{1}{2}s^{\mathrm{T}}(t)(\tilde{L}\otimes I_3)s(t) + \frac{1}{2}\sum_{m=1}^{n}\phi_{fm}^{-1}\mathrm{Tr}\Big\{\tilde{\Phi}_{fm}^{\mathrm{T}}(t)\tilde{\Phi}_{fm}(t)\Big\} + \frac{1}{2}\sum_{m=1}^{n}\mathrm{Tr}\Big\{\tilde{\Phi}_{cm}^{\mathrm{T}}(t)\tilde{\Phi}_{cm}(t)\Big\} \\ &+ \frac{1}{2}\sum_{m=1}^{n}\mathrm{Tr}\Big\{\tilde{\Phi}_{am}^{\mathrm{T}}(t)\tilde{\Phi}_{am}(t)\Big\} + \frac{1}{2}\sum_{m=1}^{n}\mathrm{Tr}\Big\{\tilde{\Phi}_{cs}^{\mathrm{T}}(t)\tilde{\Phi}_{cs}(t)\Big\} + \frac{1}{2}\sum_{m=1}^{n}\Big\{\tilde{\Phi}_{as}^{\mathrm{T}}(t)\tilde{\Phi}_{as}(t)\Big\}, \end{aligned} \tag{45}$$

where $s(t) = [s_1^{\mathrm{T}}, s_2^{\mathrm{T}}, \ldots, s_n^{\mathrm{T}}]^{\mathrm{T}}$, $\tilde{\Phi}_{fm}(t) = \hat{\Phi}_{fm}(t) - \Phi_{fm}^*$, $\tilde{\Phi}_{cm}(t) = \hat{\Phi}_{cm}(t) - \Phi_k^*$, $\tilde{\Phi}_{am}(t) = \hat{\Phi}_{am}(t) - \Phi_k^*$, $\tilde{\Phi}_{cs}(t) = \hat{\Phi}_{cs}(t) - \Phi_k^*$, $\tilde{\Phi}_{as}(t) = \hat{\Phi}_{as}(t) - \Phi_k^*$, I_3 is the 3×3 identity matrix and $\otimes$ is the Kronecker product.

Drawing upon Lemma 2, it can infer that matrix $\tilde{L} = L + B \in R^{n\times n}$ is positive definite. Furthermore, according to $E_s(t) = (\tilde{L}\otimes I_3)s(t)$, the subsequent equation can be derived

$$s^{\mathrm{T}}(t)(\tilde{L}\otimes I_3)s(t) = E_s^{\mathrm{T}}(t)(\tilde{L}\otimes I_3)^{-1}E_s(t) = s^{\mathrm{T}}(t)(\tilde{L}\otimes I_3)(\tilde{L}\otimes I_3)^{-1}(\tilde{L}\otimes I_3)s(t). \tag{46}$$

Utilizing (46) can obtain

$$\frac{1}{\lambda_{max}^{\tilde{L}}}\|E_s(t)\|^2 \leqslant s^{\mathrm{T}}(t)(\tilde{L}\otimes I_3)s(t) \leqslant \frac{1}{\lambda_{min}^{\tilde{L}}}\|E_s(t)\|^2, \tag{47}$$

where $\lambda_{max}^{\tilde{L}}$ and $\lambda_{min}^{\tilde{L}}$ are the maximum and minimum values of the matrix $\tilde{L}$.

Taking the time derivative of $V(t)$ along (10), (32), (34), (36), (40) and (42) and the substitution of (30), (38) and (41) results in

$$\begin{aligned}\dot{V}(t) =& \sum_{m=1}^{n} E_m^{s\mathrm{T}}(t)\Big(\Phi_{fm}^{*\mathrm{T}}\xi_{fm}(\bar{x}_m) + \varepsilon_{fm}(\bar{x}_m) - \mu_m E_m^s(t) - \hat{\Phi}_{fm}^{\mathrm{T}}(t)\xi_{fm}(\bar{x}_m)\\
&- \frac{1}{2}\hat{\Phi}_{am}^{\mathrm{T}}\xi_m(E_m^s,\bar{x}_m)\Big) + \sum_{m=1}^{n}\mathrm{Tr}\Big\{\tilde{\Phi}_{fm}^{\mathrm{T}}(t)\Big(\xi_m(\bar{x}_m)E_m^{s\mathrm{T}}(t) - \sigma_{fm}\hat{\Phi}_{fm}(t)\Big)\Big\}\\
&- \sum_{m=1}^{n}\kappa_{cm}\mathrm{Tr}\Big\{\tilde{\Phi}_{cm}^{\mathrm{T}}(t)\Big(\xi_m(E_m^s,\bar{x}_m)\xi_m^{\mathrm{T}}(E_m^s,\bar{x}_m) + \sigma_m I_q\Big)\hat{\Phi}_{cm}(t)\Big\} - \frac{1}{2}E_s^{\mathrm{T}}\hat{\Phi}_{as}^{\mathrm{T}}(t)\xi_s(E_s^s,\bar{x}_s)\\
&- \sum_{m=1}^{n}\mathrm{Tr}\Big\{\tilde{\Phi}_{am}^{\mathrm{T}}(t)\Big(\kappa_{am}\big(\xi_m(E_m^s,\bar{x}_m)\xi_m^{\mathrm{T}}(E_m^s,\bar{x}_m) + \sigma_m I_q\big)\big(\hat{\Phi}_{am}(t) - \hat{\Phi}_{cm}(t)\big)\Big)\\
&+ \frac{1}{2}\xi_m(E_m^s,\bar{x}_m)\xi_m^{\mathrm{T}}(E_m^s,\bar{x}_m)\hat{\Phi}_{am}(t)\Big\} - \kappa_{cs}\mathrm{Tr}\Big\{\tilde{\Phi}_{cs}^{\mathrm{T}}(t)\Big(\xi_s(E_s^s,\bar{x}_s)\xi_s^{\mathrm{T}}(E_s^s,\bar{x}_s) + \sigma_s I_n\Big)\hat{\Phi}_{cs}(t)\Big\}\\
&- \kappa_{as}\mathrm{Tr}\Big\{\tilde{\Phi}_{as}^{\mathrm{T}}(t)\Big(\xi_s(E_s^s,\bar{x}_s)\xi_s^{\mathrm{T}}(E_s^s,\bar{x}_s) + \sigma_s I_n\Big)\Big(\hat{\Phi}_{as}(t) - \hat{\Phi}_{cs}(t)\Big)\Big\}.\end{aligned} \tag{48}$$

By applying Young's inequality, the subsequent inequalities can be inferred as

$$\begin{aligned}-\frac{1}{2}E_m^{s\mathrm{T}}(t)\hat{\Phi}_{am}^{\mathrm{T}}(t)\xi_m(E_m^s,\bar{x}_m) &\leqslant \frac{1}{4}\|E_m^s(t)\|^2 + \frac{1}{4}\mathrm{Tr}\Big\{\hat{\Phi}_{am}^{\mathrm{T}}(t)\xi_m(E_m^s,\bar{x}_m)\xi_m^{\mathrm{T}}(E_m^s,\bar{x}_m)\hat{\Phi}_{am}(t)\Big\},\\
-\frac{1}{2}E_s^{\mathrm{T}}\hat{\Phi}_{as}^{\mathrm{T}}(t)\xi_s(E_s^s,\bar{x}_s) &\leqslant \frac{1}{4}\|E_s(t)\|^2 + \frac{1}{4}\mathrm{Tr}\Big\{\hat{\Phi}_{as}^{\mathrm{T}}(t)\xi_s(E_s^s,\bar{x}_s)\xi_s^{\mathrm{T}}(E_s^s,\bar{x}_s)\hat{\Phi}_{as}(t)\Big\}.\end{aligned} \tag{49}$$

Inserting inequality (49) into (48) has

$$\begin{aligned}\dot{V}(t) \leqslant& \sum_{m=1}^{m}\left(\frac{3}{4} - \mu_m\right)\|E_m^s(t)\|^2 + \frac{1}{4}\|E_s(t)\|^2 - \sum_{m=1}^{n}\sigma_{fm}\mathrm{Tr}\Big\{\tilde{\Phi}_{fm}^{\mathrm{T}}(t)\hat{\Phi}_{fm}(t)\Big\} + \frac{1}{2}\sum_{m=1}^{n}\|\varepsilon_{fm}(x_m)\|^2\\
&- \sum_{m=1}^{n}\kappa_{cm}\mathrm{Tr}\Big\{\tilde{\Phi}_{cm}^{\mathrm{T}}(t)\Big(\xi_m(E_m^s,\bar{x}_m)\xi_m^{\mathrm{T}}(E_m^s,\bar{x}_m) + \sigma_m I_q\Big)\hat{\Phi}_{cm}(t)\Big\}\\
&- \sum_{m=1}^{n}\mathrm{Tr}\Big\{\tilde{\Phi}_{am}^{\mathrm{T}}(t)\Big(\kappa_{am}\big(\xi_m(E_m^s,\bar{x}_m)\xi_m^{\mathrm{T}}(E_m^s,\bar{x}_m) + \sigma_m I_q\big)\big(\hat{\Phi}_{am}(t) - \hat{\Phi}_{cm}(t)\big)\Big)\Big\}\\
&- \frac{1}{2}\sum_{m=1}^{n}\mathrm{Tr}\Big\{\xi_m(E_m^s,\bar{x}_m)\xi_m^{\mathrm{T}}(E_m^s,\bar{x}_m)\hat{\Phi}_{am}(t)\Big\} - \kappa_{cs}\mathrm{Tr}\Big\{\tilde{\Phi}_{cs}^{\mathrm{T}}(t)\Big(\xi_s(E_s^s,\bar{x}_s)\xi_s^{\mathrm{T}}(E_s^s,\bar{x}_s) + \sigma_s I_n\Big)\\
&\times\hat{\Phi}_{cs}(t)\Big\} - \kappa_{as}\mathrm{Tr}\Big\{\tilde{\Phi}_{as}^{\mathrm{T}}(t)\Big(\xi_s(E_s^s,\bar{x}_s)\xi_s^{\mathrm{T}}(E_s^s,\bar{x}_s) + \sigma_s I_n\Big)\Big(\hat{\Phi}_{as}(t) - \hat{\Phi}_{cs}(t)\Big)\Big\}.\end{aligned} \tag{50}$$

Moreover, the following equations hold

$$\begin{aligned}\kappa_{am}\mathrm{Tr}\Big\{\tilde{\Phi}_{am}^{\mathrm{T}}(t)\Big(\xi_m(E_m^s,\bar{x}_m)\xi_m^{\mathrm{T}}(E_m^s,\bar{x}_m) + \sigma_m I_q\Big)\tilde{\Phi}_{cm}(t)\Big\} &\leqslant \frac{\kappa_{am}}{2}\mathrm{Tr}\Big\{\tilde{\Phi}_{am}^{\mathrm{T}}(t)\Big(\xi_m(E_m^s,\bar{x}_m)\xi_m^{\mathrm{T}}(E_m^s,\bar{x}_m)\\
&+ \sigma_m I_q\Big)\tilde{\Phi}_{am}(t)\Big\} + \frac{\kappa_{am}}{2}\mathrm{Tr}\Big\{\tilde{\Phi}_{cm}^{\mathrm{T}}(t)\Big(\xi_m(E_m^s,\bar{x}_m)\xi_m^{\mathrm{T}}(E_m^s,\bar{x}_m) + \sigma_m I_q\Big)\tilde{\Phi}_{cm}(t)\Big\},\\
\kappa_{as}\mathrm{Tr}\Big\{\tilde{\Phi}_{as}^{\mathrm{T}}(t)\Big(\xi_s(E_s^s,\bar{x}_s)\xi_s^{\mathrm{T}}(E_s^s,\bar{x}_s) + \sigma_s I_n\Big)\tilde{\Phi}_{cs}(t)\Big\} &\leqslant \frac{\kappa_{as}}{2}\mathrm{Tr}\Big\{\tilde{\Phi}_{as}^{\mathrm{T}}(t)\Big(\xi_s(E_s^s,\bar{x}_s)\xi_s^{\mathrm{T}}(E_s^s,\bar{x}_s)\\
&+ \sigma_s I_n\Big)\tilde{\Phi}_{as}(t)\Big\} + \frac{\kappa_{cs}}{2}\mathrm{Tr}\Big\{\tilde{\Phi}_{cs}^{\mathrm{T}}(t)\Big(\xi_s(E_s^s,\bar{x}_s)\xi_s^{\mathrm{T}}(E_s^s,\bar{x}_s) + \sigma_s I_n\Big)\tilde{\Phi}_{cs}(t)\Big\}.\end{aligned} \tag{51}$$

By the weight error relationships $\tilde{\Phi}_{fm,cm,am,as,cs}(t)=\hat{\Phi}_{fm,cm,am,as,cs}(t)-\Phi^*_{fm,m,m,s,s}$, combined with (51), Eq. (50) can be redescribed as

$$\begin{aligned}\dot{V}(t)\leqslant&\sum_{m=1}^{m}\left(\frac{3}{4}-\mu_m\right)\|E_m^s(t)\|^2+\frac{1}{4}\|E_s(t)\|^2-\sum_{m=1}^{n}\frac{\sigma_{fm}}{2}\mathrm{Tr}\Big\{\tilde{\Phi}_{fm}^{\mathrm{T}}(t)\tilde{\Phi}_{fm}(t)\Big\}\\&-\sum_{m=1}^{n}\frac{\kappa_{cm}-\kappa_{am}}{2}\mathrm{Tr}\Big\{\tilde{\Phi}_{cm}^{\mathrm{T}}(t)\Big(\xi_m(E_m^s,\bar{x}_m)\xi_m^{\mathrm{T}}(E_m^s,\bar{x}_m)+\sigma_m I_q\Big)\tilde{\Phi}_{cm}(t)\Big\}\\&-\sum_{m=1}^{n}\frac{2\kappa_{am}+1}{4}\mathrm{Tr}\Big\{\tilde{\Phi}_{am}^{\mathrm{T}}(t)\Big(\xi_m(E_m^s,\bar{x}_m)\xi_m^{\mathrm{T}}(E_m^s,\bar{x}_m)+\sigma_m I_q\Big)\tilde{\Phi}_{am}(t)\Big\}\\&-\frac{\kappa_{cs}-\kappa_{as}}{2}\mathrm{Tr}\Big\{\tilde{\Phi}_{cs}^{\mathrm{T}}(t)\Big(\xi_s(E_s^s,\bar{x}_s)\xi_s^{\mathrm{T}}(E_s^s,\bar{x}_s)+\sigma_s I_n\Big)\tilde{\Phi}_{cs}(t)\Big\}\\&-\frac{\kappa_{as}}{2}\mathrm{Tr}\Big\{\tilde{\Phi}_{as}^{\mathrm{T}}(t)\Big(\xi_s(E_s^s,\bar{x}_s)\xi_s^{\mathrm{T}}(E_s^s,\bar{x}_s)+\sigma_s I_n\Big)\tilde{\Phi}_{as}^{\mathrm{T}}(t)\Big\}+\Xi(t),\end{aligned}\tag{52}$$

where $\Xi(t)=\sum_{m=1}^{n}\mathrm{Tr}\big\{\frac{\sigma_{fm}}{2}\Phi_{fm}^{*\mathrm{T}}\Phi_{fm}^*+\frac{\kappa_{cm}}{2}\Phi_m^{*\mathrm{T}}\big(\xi_m(E_m^s,\bar{x}_m)\xi_m^{\mathrm{T}}(E_m^s,\bar{x}_m)+\sigma_m I_q\big)\Phi_m^*+\frac{1}{4}\Phi_m^{*\mathrm{T}}\xi_m(E_m^s,\bar{x}_m)\times\xi_m^{\mathrm{T}}(E_m^s,\bar{x}_m)\Phi_m^*\big\}+\frac{1}{4}\mathrm{Tr}\big\{\Phi_s^{*\mathrm{T}}\big(\xi_m(E_m^s,\bar{x}_m)\xi_m^{\mathrm{T}}(E_m^s,\bar{x}_m)+\sigma_s I_n\big)\Phi_s^*\big\}+\frac{1}{2}\sum_{m=1}^{n}\|\varepsilon_{fm}(x_m)\|^2$, which is bounded by a constant Θ, i.e., $\|\Xi\|\leqslant\Theta$, because all of its terms are bounded.

According to the condition (46), Eq. (52) is rewritten as

$$\begin{aligned}\dot{V}(t)\leqslant&\sum_{m=1}^{m}\left(\frac{3}{4}-\mu_m\right)\|E_m^s(t)\|^2+\frac{1}{4}\|E_s(t)\|^2-\sum_{m=1}^{n}\frac{\sigma_{fm}}{2}\mathrm{Tr}\Big\{\tilde{\Phi}_{fm}^{\mathrm{T}}(t)\tilde{\Phi}_{fm}(t)\Big\}\\&-\sum_{m=1}^{n}\frac{(\kappa_{cm}-\kappa_{am})\sigma_m}{2}\mathrm{Tr}\Big\{\tilde{\Phi}_{cm}^{\mathrm{T}}(t)\tilde{\Phi}_{cm}(t)\Big\}-\sum_{m=1}^{n}\frac{(2\kappa_{am}+1)\sigma_m}{4}\mathrm{Tr}\Big\{\tilde{\Phi}_{am}^{\mathrm{T}}(t)\tilde{\Phi}_{am}(t)\Big\}\\&-\frac{(\kappa_{cs}-\kappa_{as})\sigma_s}{2}\mathrm{Tr}\Big\{\tilde{\Phi}_{cs}^{\mathrm{T}}(t)\tilde{\Phi}_{cs}(t)\Big\}-\frac{\kappa_{as}\sigma_s}{2}\mathrm{Tr}\Big\{\tilde{\Phi}_{as}^{\mathrm{T}}(t)\tilde{\Phi}_{as}^{\mathrm{T}}(t)\Big\}+\Xi(t).\end{aligned}\tag{53}$$

Letting $\mu=min_{m=1,\ldots,n}\{2(\mu_k-3/4)\}$, $\sigma_f=min_{m=1,\ldots,n}\{\sigma_{fm}\phi_{fm}\}$, $\sigma_{cm}=min_{m=1,\ldots,n}\{(\kappa_{cm}-\kappa_{am})\sigma_m\}$, $\sigma_{am}=min_{m=1,\ldots,n}\{(2\kappa_{am}+1)\sigma_m/2\}$, $\sigma_{cs}=(\kappa_{cs}-\kappa_{as})\sigma_s$, $\sigma_{as}=\kappa_{as}\sigma_s$, then Eq. (53) turns into

$$\begin{aligned}\dot{V}(t)\leqslant&-\frac{\mu}{2}\sum_{m=1}^{n}\|E_m^s(t)\|^2+\frac{1}{4}\|E_s(t)\|^2-\frac{\sigma_f}{2}\sum_{m=1}^{n}\phi_m^{-1}\mathrm{Tr}\big\{\tilde{\Phi}_{fm}^{\mathrm{T}}(t)\tilde{\Phi}_{fm}(t)\big\}-\sum_{m=1}^{n}\frac{\sigma_{cm}}{2}\mathrm{Tr}\Big\{\tilde{\Phi}_{cm}^{\mathrm{T}}(t)\tilde{\Phi}_{cm}(t)\Big\}\\&-\sum_{m=1}^{n}\frac{\sigma_{am}}{2}\mathrm{Tr}\Big\{\tilde{\Phi}_{am}^{\mathrm{T}}(t)\tilde{\Phi}_{am}(t)\Big\}-\frac{\sigma_{cs}}{2}\mathrm{Tr}\Big\{\tilde{\Phi}_{cs}^{\mathrm{T}}(t)\tilde{\Phi}_{cs}(t)\Big\}-\frac{\sigma_{as}}{2}\mathrm{Tr}\Big\{\tilde{\Phi}_{as}^{\mathrm{T}}(t)\tilde{\Phi}_{as}^{\mathrm{T}}(t)\Big\}+\Xi(t).\end{aligned}\tag{54}$$

By applying the inequality (47), Eq. (54) turns into

$$\begin{aligned}\dot{V}(t)\leqslant&-\frac{\mu\lambda_{min}^{\tilde{L}}}{2}s^{\mathrm{T}}(t)(\tilde{L}\otimes I_n)s(t)-\frac{\sigma_f}{2}\sum_{m=1}^{n}\phi_m^{-1}\mathrm{Tr}\big\{\tilde{\Phi}_{fm}^{\mathrm{T}}(t)\tilde{\Phi}_{fm}(t)\big\}-\frac{\sigma_{cm}}{2}\sum_{m=1}^{n}\mathrm{Tr}\Big\{\tilde{\Phi}_{cm}^{\mathrm{T}}(t)\tilde{\Phi}_{cm}(t)\Big\}\\&-\frac{\sigma_{am}}{2}\sum_{m=1}^{n}\mathrm{Tr}\Big\{\tilde{\Phi}_{am}^{\mathrm{T}}(t)\tilde{\Phi}_{am}(t)\Big\}-\frac{\sigma_{cs}}{2}\mathrm{Tr}\Big\{\tilde{\Phi}_{cs}^{\mathrm{T}}(t)\tilde{\Phi}_{cs}(t)\Big\}-\frac{\sigma_{as}}{2}\mathrm{Tr}\Big\{\tilde{\Phi}_{as}^{\mathrm{T}}(t)\tilde{\Phi}_{as}^{\mathrm{T}}(t)\Big\}+\Xi(t).\end{aligned}\tag{55}$$

Letting $\varpi=min\Big\{\mu\lambda_{min}^{\tilde{L}},\sigma_f,\sigma_{cm},\sigma_{am},\sigma_{cs},\sigma_{as}\Big\}$, Eq. (55) can be rewritten as

$$\dot{V}(t)\leqslant-\varpi V(t)+\Xi.\tag{56}$$

By utilizing the Lemma 3 and (56), the following inequality is obtained

$$V(t)\leqslant\mathrm{e}^{-\varpi t}V(0)+\frac{\Xi}{\varpi}\big(1-\mathrm{e}^{-\varpi t}\big).\tag{57}$$

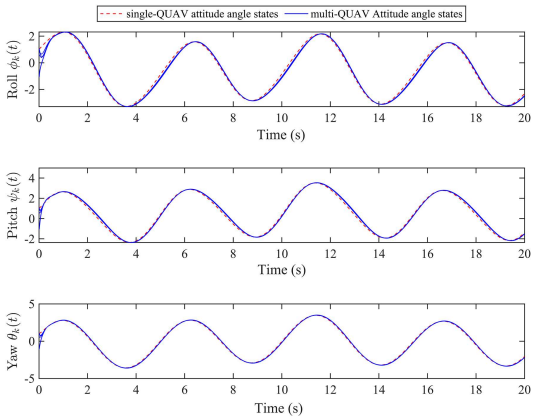

Figure 2 (Color online) Three attitude angle states tracking performance of the pursuer-evader.

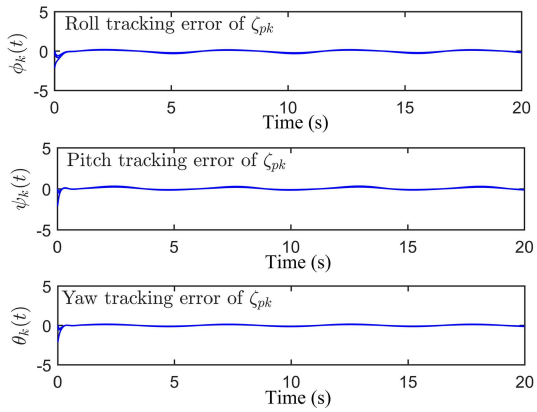

Figure 3 (Color online) The convergence of tracking errors ζ_{pk}.

The aforementioned inequality demonstrates that all error signals, encompassing $s_k(t), \tilde{\Phi}_{fm}(t), \tilde{\Phi}_{cm}(t), \tilde{\Phi}_{am}(t), \tilde{\Phi}_{as}(t), \tilde{\Phi}_{cs}(t)$, are SGUUB, Theorem 1(i) is proved. And then through appropriate tuning of the parameter ϖ, the sliding variable s_k can be driven to converge within a vicinity around zero. According to Lemma 4 in [23], it can be known that condition $s_k(\bar{\zeta}_k) = 0$ can ensure all tracking error states, $\zeta_{mj}(t)$, $m = 1, 2, \ldots, n$, $j = 1, 2 \ldots, i$, to converge to zero as $t \to \infty$, this directly proves Theorem 1(ii).

4 Numerical examples

This part conducts numerical simulations of the multi-QUAV attitude system comprising four QUAVs in the Matlab environment. The parameters within its dynamic model are $I_{zk} = 8.81 \times 10^{-3}$, $I_{yk} = 4.85 \times 10^{-3}$, $I_{xk} = 4.35 \times 10^{-3}$, $l = 0.325$, $G_{\phi k} = G_{\psi k} = G_{\theta k} = 0.6$, $k = 1, \ldots, 4$. The initial states are $x_{pk=1,\ldots,4}(0) = \left[\phi_k(0), \psi_i(0), \theta_k(0)\right]^{\mathrm{T}} = [\pi/3, \pi/3, \pi/3]^{\mathrm{T}}, [\pi/4, \pi/4, \pi/4]^{\mathrm{T}}, [\pi/5, \pi/5, \pi/5]^{\mathrm{T}}, -[\pi/3, \pi/3, \pi/3]^{\mathrm{T}}$ and $x_{vk=1,\ldots,4}(0) = \left[\dot{\phi}_i(0), \dot{\psi}_i(0), \dot{\theta}_k(0)\right]^{\mathrm{T}} = [3, 3, 3], [2, 2, 2], [1, 1, 1], [-1, -1, -1]$. The communication relationship between evader and pursuer can be represented by $B = \mathrm{diag}\{0, 1, 0, 0\}$.

The adjacency matrix described the intercommunication of multi-QUAV attitude system is

$$A = [0, 1, 0, 1; 1, 0, 1, 0; 0, 1, 0, 1; 1, 0, 1, 0].$$

Additionally, the sliding mode variable in (9) is formulated with the parameter $\beta = 8$. Subsequently, the consensus error, which is associated with the adjacency matrix A and the communication weight matrix B, can be derived by consulting (11).

The NN serving as an identifier, corresponding to (31), is configured with nine nodes. The initial values of the NN weights are $\hat{\Phi}_{f1,\ldots,f4}(0) = [0.4, \ldots, 0.4]^{\mathrm{T}} \in \mathbb{R}^9$. The adaptive training law, which corresponds to (32), is formulated using gain values $\phi_{1,\ldots,4} = 0.6$ and $\sigma_{f1,\ldots,f4} = 0.3$. Subsequently, the basis function vector $\xi_{fk}(\bar{x}_k)$, $k = 1, 2, 3, 4$ are constructed by a Gaussian kernel function of width 2.

The actor and critic NNs are configured with 6 nodes that are uniformly distributed within the interval extending from $-[10, 10, 10, 40, 40, 40, 40, 40, 40]^{\mathrm{T}}$ to $[10, 10, 10, 40, 40, 40, 40, 40, 40]^{\mathrm{T}}$. Subsequently, the basis function vector $\xi_k(e_k^s, \bar{x}_k)$, $k = 1, 2, 3, 4$ is obtained using a Gaussian function with a width of 2. The updating laws are formulated with the parameters $\sigma_{1,\ldots,4} = 0.5$, $\kappa_{a1,\ldots,a4} = 0.3$, and $\kappa_{c1,\ldots,c4} = 0.35$. The initial values are $\hat{\Phi}_{a1,\ldots,a4}(0) = [0.4, \ldots, 0.4]^{\mathrm{T}} \in \mathbb{R}^6$ and $\hat{\Phi}_{c1,\ldots,c4}(0) = [0.3, \ldots, 0.3]^{\mathrm{T}} \in \mathbb{R}^6$. Ultimately, by assigning $\mu_{k=1,2,3,4} = 36$, the optimal pursuer controller corresponding to (35) can be derived.

In addition, the NN parameters of u_s are designed. The actor NN in (39) and the critic NN in (41), they both include a total of 6 nodes, respectively. Assign the initial values of the critic and the actor NN weights as $\hat{\Phi}_{as}(0) = 0.4$ and $\hat{\Phi}_{as}(0) = 0.5$, respectively. Set the value of the updating laws parameters associated with (39) and (41) to $\kappa_{as} = 0.32$, $\kappa_{cs} = 0.35$ and $\sigma_s = 0.2$. Additionally, the basis function vector $\xi_s(e_s^s, \bar{x}_s)$ are constructed by a Gaussian kernel function of width 2.

Figures 2–7 exhibit the three attitude angle states of the pursuer-evader have better tracking performance. Figure 2 represents the pursuer composed of four QUAV systems and the evader composed of

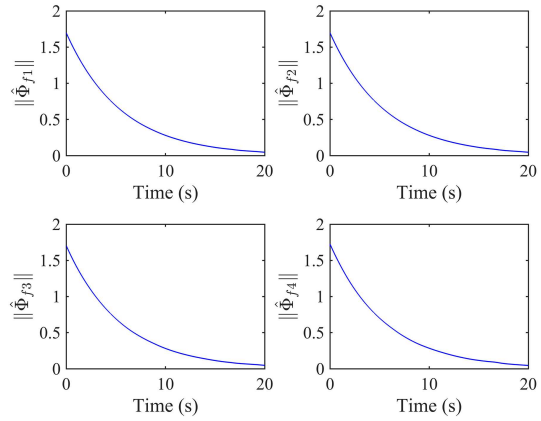

Figure 4 (Color online) The norm $\|\hat{\Phi}_{fm}\|$, $m = 1, 2, 3, 4$ of the NN weights for the identifier.

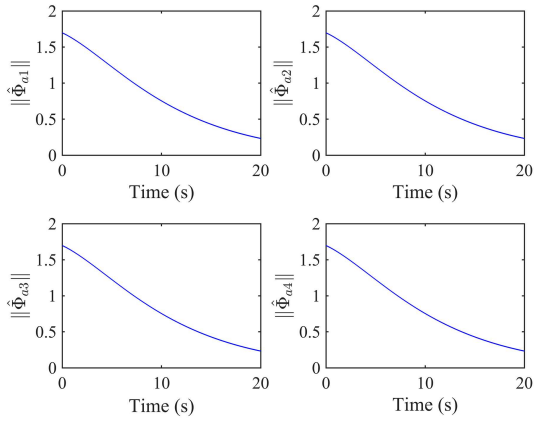

Figure 5 (Color online) The norm $\|\hat{\Phi}_{am}\|$, $m = 1, 2, 3, 4$ of the NN weights for the actor.

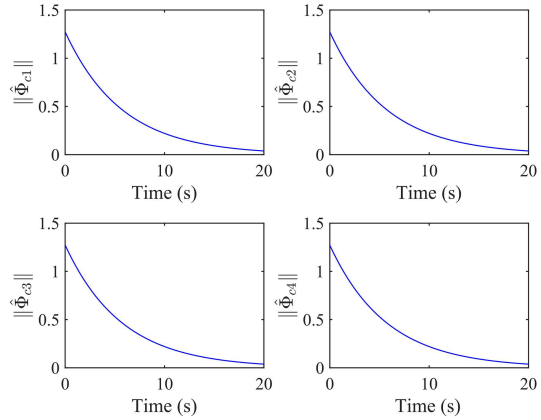

Figure 6 (Color online) The norm $\|\hat{\Phi}_{cm}\|$, $m = 1, 2, 3, 4$ of the NN weights for the critic.

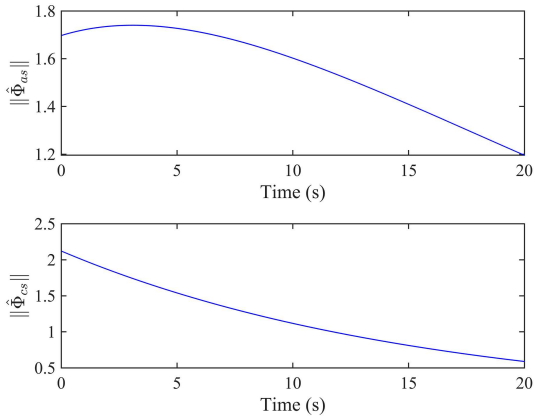

Figure 7 (Color online) The $\|\hat{\Phi}_{cs}\|$ and $\|\hat{\Phi}_{as}\|$ scores represent the critic and actor NN weights norm of u_s^*.

a single QUAV system, and their attitude game following trajectory performance. Figure 3 displays the tracking errors for the three attitude angle states (roll, pitch, and yaw) of the 4 agents, demonstrating their convergence to zero. Figures 4–6 illustrate the boundedness of the parameter matrices for the identifier, critic, and actor NNs with control inputs $u^*_{z1,\ldots,z4}$, respectively. Figure 7 illustrates the boundedness of the parameter matrices of the actor and critic NNs with respect to evader u_s^*.

5 Conclusion

This paper proposes a novel attitude zero-sum game control approach for multi-QUAV systems and single-QUAV system based on the sliding mode control and RL approach. In this game control, the SMC principle is used to design a sliding mode hyperplane to manage multiple state variables with differential relationships. To obtain the saddle point of the zero-sum game, we execute RL within the I-DAC framework, which can derive the optimal control strategy for both parties. As opposed to traditional RL methods, the proposed RL approach features a simpler algorithm and releases the sustained incentives and known dynamics. The designed game control algorithm is applicable to a certain range of multi-QUAV pursuit-evasion problems. The RL-based algorithm we designed provides a new solution for multi-UAV system confrontation. Finally, the effectiveness of the control method is verified through the Lyapunov stability theorem and numerical simulation. In future studies, we will consider the non-zero-sum game control for multi-QUAV attitude systems via sliding mode and fixed-time convergent RL.

Acknowledgements This work was supported by National Natural Science Foundation of China (Grant Nos. 62573316, 62503360),

Natural Science Foundation of Tianjin (Grant No. 25JCQNJC00980), and Natural Science Foundation of Shandong Province (Grant No. ZR2025MS996).

References

1 Li X, Lu X, Chen W, et al. Research on UAVs reconnaissance task allocation method based on communication preservation. IEEE Trans Consumer Electron, 2024, 70: 684–695
2 Shahid S, Zhen Z, Javaid U. Cooperative task assignment of heterogeneous unmanned aerial vehicles for simultaneous multi-directional attack on a moving target. Eng Appl Artif Intell, 2025, 139: 109595
3 Li J, Yang X, Yang Y, et al. Cooperative mapping task assignment of heterogeneous multi-UAV using an improved genetic algorithm. Knowl-Based Syst, 2024, 296: 111830
4 Chen S, Liu G, Zhou Z, et al. Robust multi-agent reinforcement learning method based on adversarial domain randomization for real-world dual-UAV cooperation. IEEE Trans Intell Veh, 2024, 9: 1615–1627
5 Chen Y, Liu G, Zhang Z, et al. Improving physical layer security for multi-UAV systems against hybrid wireless attacks. IEEE Trans Veh Technol, 2024, 73: 7034–7048
6 Zhou C, Kadhim K M R, Zheng X. Multi-UAVs path planning for data harvesting in adversarial scenarios. Comput Commun, 2024, 221: 42–53
7 Gao Q Y, Wu H C, Zhang Y F, et al. Differential game-based analysis of multi-attacker multi-defender interaction. Sci China Inf Sci, 2021, 64: 222302
8 Xia C, Ding S, Wang C, et al. Risk analysis and enhancement of cooperation yielded by the individual reputation in the spatial public goods game. IEEE Syst J, 2016, 11: 1516–1525
9 Long J, Yu D, Wen G, et al. Game-based backstepping design for strict-feedback nonlinear multi-agent systems based on reinforcement learning. IEEE Trans Neural Netw Learn Syst, 2022, 35: 817–830
10 Gong Z, Yang F. Secure tracking control via fixed-time convergent reinforcement learning for a UAV CPS. IEEE CAA J Autom Sin, 2024, 11: 1699–1701
11 Ren H, Jiang B, Ma Y. Zero-sum differential game-based fault-tolerant control for a class of affine nonlinear systems. IEEE Trans Cybern, 2022, 54: 1272–1282
12 Lv S Y, Wu Z, Xiong J. A zero-sum hybrid stochastic differential game with impulse controls. Sci China Inf Sci, 2024, 67: 212209
13 Yang X, Liu D, Ma H, et al. Online approximate solution of HJI equation for unknown constrained-input nonlinear continuous-time systems. Inf Sci, 2016, 328: 435–454
14 Yan R, Duan X, Shi Z, et al. Matching-based capture strategies for 3D heterogeneous multiplayer reach-avoid differential games. Automatica, 2022, 140: 110207
15 Fu H, Liu H H T. Justification of the geometric solution of a target defense game with faster defenders and a convex target area using the HJI equation. Automatica, 2023, 149: 110811
16 Yang F, Gong Z, Wei Q, et al. Secure containment control for multi-UAV systems by fixed-time convergent reinforcement learning. IEEE Trans Cybern, 2025, 55: 1981–1994
17 Cui J, Pan Y, Xue H, et al. Simplified optimized finite-time containment control for a class of multi-agent systems with actuator faults. Nonlinear Dyn, 2022, 109: 2799–2816
18 Zhang Z P, Xia C Y, Qi G Y, et al. Multi-step state-based opacity for unambiguous weighted machines. Sci China Inf Sci, 2024, 67: 212204
19 Wen G, Chen C L P, Ge S S, et al. Optimized adaptive nonlinear tracking control using actor-critic reinforcement learning strategy. IEEE Trans Ind Inf, 2019, 15: 4969–4977
20 Li Z, Song Y, Wen G. Reinforcement learning based optimized sliding-mode consensus control of high-order nonlinear canonical dynamic multiagent system. IEEE Syst J, 2023, 17: 6302–6311
21 Song Y, Li Z, Li B, et al. Optimized leader-follower consensus control using combination of reinforcement learning and sliding mode mechanism for multiple robot manipulator system. Intl J Robust Nonlinear, 2024, 34: 5212–5228
22 Wen G, Yu D, Zhao Y. Optimized fuzzy attitude control of quadrotor unmanned aerial vehicle using adaptive reinforcement learning strategy. IEEE Trans Aerosp Electron Syst, 2024, 60: 6075–6083
23 Wen G, Dou H, Li B. Adaptive fuzzy leader-follower consensus control using sliding mode mechanism for a class of high-order unknown nonlinear dynamic multi-agent systems. Intl J Robust Nonlinear, 2023, 33: 545–558
24 Yu W, Wang H, Cheng F, et al. Second-order consensus in multiagent systems via distributed sliding mode control. IEEE Trans Cybern, 2016, 47: 1872–1881
25 Zhao B R, Peng Y J, Song Y N, et al. Sliding mode control for consensus tracking of second-order nonlinear multi-agent systems driven by brownian motion. Sci China Inf Sci, 2018, 61: 070216
26 Qu Q, Zhang H, Yu R, et al. Neural network-based H^{∞} sliding mode control for nonlinear systems with actuator faults and unmatched disturbances. Neurocomputing, 2018, 275: 2009–2018
27 Zhao H, Zong G, Zhao X, et al. Hierarchical sliding-mode surface-based adaptive critic tracking control for nonlinear multi-player zero-sum games via generalized fuzzy hyperbolic models. IEEE Trans Fuzzy Syst, 2023, 31: 4010–4023
28 Cui G, Yang W, Yu J, et al. Fixed-time prescribed performance adaptive trajectory tracking control for a quav. IEEE Trans Circ Syst II, 2022, 69: 494–498
29 Wen G, Li B. Optimized leader-follower consensus control using reinforcement learning for a class of second-order nonlinear multiagent systems. IEEE Trans Syst Man Cybern Syst, 2021, 52: 5546–5555
30 Park J, Sandberg I W. Universal approximation using radial-basis-function networks. Neural Comput, 1991, 3: 246–257
31 Ge S S, Wang C. Adaptive neural control of uncertain MIMO nonlinear systems. IEEE Trans Neural Netw, 2004, 15: 674–692