Distributed Formation Control for Second-Order Nonlinear Multiagent Systems Using Predictor-Based Accelerated Fuzzy Learning

Jinzhao Miao, Jinliang Liu, *Senior Member, IEEE*, Lijuan Zha, Engang Tian, *Senior Member, IEEE*, and Chen Peng, *Senior Member, IEEE*

***Abstract*—This article investigates the distributed formation control problem of second-order multiagent systems (MASs) subject to unknown nonlinear dynamics. A predictor-based fuzzy learning framework is developed, in which fuzzy logic systems approximate the unknown nonlinear dynamics and a two-layer learning mechanism accelerates weight adaptation by utilizing the velocity prediction error. A distributed observer is designed for jointly connected switching directed topologies to handle limited leader accessibility, allowing agents to estimate the global reference trajectory using solely local information. Furthermore, an actor–critic reinforcement learning architecture is integrated to realize distributed optimal formation control, in which the value function is reformulated with a compensatory term that exploits available system structural information, thereby enhancing policy optimization efficiency and accelerating convergence. Simulation results demonstrate that the proposed approach exhibits improved robustness and accuracy in both dynamics approximation and position tracking when compared with conventional learning-based methods.**

***Index Terms*—Formation control, fuzzy logic systems (FLSs), multiagent systems (MASs), reinforcement learning.**

I. Introduction

SECOND-ORDER multiagent systems (MASs) have emerged as a fundamental research paradigm in cooperative control, aiming to achieve coordinated behaviors among multiple interacting dynamical agents governed by nonlinear dynamics [1]. Owing to their inherent position–velocity coupling, second-order MASs provide an effective and realistic framework for modeling a wide range of physical systems with inertial characteristics. Typical coordination objectives include consensus [2], synchronization, formation maintenance [3], flocking, and cooperative tracking, which play a pivotal role in practical applications such as uncrewed aerial vehicles, uncrewed surface vessels (USVs), autonomous vehicle platoons, satellite clusters, robotic networks, and distributed energy systems. Compared with linear or first-order models, second-order nonlinear MASs offer a more faithful representation of real-world systems [4]. However, this increased modeling fidelity inevitably brings substantial challenges, including strong nonlinearities, unmodeled dynamics, parameter uncertainties, and external disturbances [5]. These issues significantly complicate controller synthesis, particularly in distributed settings where only local information exchange is available, thereby strongly motivating the development of distributed, robust, and adaptive formation control strategies for second-order nonlinear MASs.

Among various cooperative control problems, formation control has attracted particular attention [6], [7], which enables groups of agents to achieve and maintain prescribed geometric configurations while performing complex tasks [8], [9]. Formation control not only enhances mission efficiency and coverage but also improves safety and robustness by enabling distributed task execution [10]. In the past decade, a considerable amount of research has been devoted to the formation control of MASs, and a variety of methodologies have been proposed. In [11], an adaptive output-feedback formation control method is proposed for stochastic nonlinear MASs, ensuring unknown dynamics identification and interagent collision avoidance. In [12], an event-triggered predefined-time formation control method is proposed for second-order MASs with reduced communication. A stochastic feedback-based formation control approach for second-order MASs is proposed in [13], enabling smooth trajectories and reduced actuator load. An active disturbance rejection formation control method for MASs subject to input constraints is proposed in [14], combining an intermediate estimator and an improved model-predictive controller. More recently, learning-based methods [15], [16] have been introduced, leveraging data-driven modeling and approximation capabilities to enhance adaptability in dynamic environments. Despite these advances, many existing results still rely on restrictive assumptions, such as Lipschitz continuity of nonlinear dynamics and availability of precise system models, which may not hold in practical MASs [17]. These theoretical constraints significantly limit the capacity of the system to cope with inherent

Received 31 December 2025; revised 28 February 2026; accepted 7 April 2026. Date of publication 13 April 2026; date of current version 4 June 2026. This work was supported by the National Natural Science Foundation of China under Grant 62373252 and Grant 62273174. Recommended by Associate Editor F. Deng. *(Corresponding author: Jinliang Liu.)*

Jinzhao Miao and Jinliang Liu are with the School of Computer Science, Nanjing University of Information Science and Technology, Nanjing 210044, China (e-mail: miaojinzhao2022@163.com; liujinliang@vip.163.com).

Lijuan Zha is with the School of Science, Nanjing Forestry University, Nanjing 210037, China (e-mail: zhalijuan@vip.163.com).

Engang Tian is with the School of Optical-Electrical and Computer Engineering, University of Shanghai for Science and Technology, Shanghai 200093, China (e-mail: tianengang@163.com).

Chen Peng is with the School of Mechatronic Engineering and Automation, Department of Automation, Shanghai University, Shanghai 200444, China (e-mail: c.peng@shu.edu.cn).

Digital Object Identifier 10.1109/TFUZZ.2026.3682687

nonlinearities, unmodeled uncertainties, and time-varying disturbances. Consequently, there is a compelling need to develop formation control frameworks integrated with adaptive learning capabilities without depending on accurate model information.

To cope with the challenges posed by unknown nonlinear dynamics in second-order MASs, fuzzy logic systems (FLSs), which are widely recognized as universal approximators [18], [19], are commonly adopted to estimate and compensate for uncertain nonlinearities [20]. In previous studies, FLS-based adaptive schemes have been widely applied to approximate unknown nonlinearities in MASs [21], [22]. However, their performance is often constrained by approximation errors and limited adaptability. To address these issues, Shen et al. [23] proposed a fuzzy weight-based reinforcement learning scheme for nonlinear servo systems, which integrates a switching event-triggered strategy to ensure predefined-time tracking performance. Moreover, Wang et al. [24] developed a novel hybrid reinforcement Q-learning method for nonlinear Markov jump systems, utilizing a parallel learning algorithm to solve the adaptive fuzzy H_∞ control problem. In addition, other works have explored adaptive fuzzy weight updating strategies [25], [26]. However, regardless of such progress, a significant number of existing methods still rely on fixed fuzzy rules or purely local feedback. These structural rigidities inevitably compromise the approximation accuracy of unknown nonlinearities in complex second-order MASs. Motivated by these deficiencies, contemporary research efforts have focused on constructing more flexible and adaptive fuzzy learning strategies, aiming to significantly improve the capability of MASs to handle nonlinear dynamics and external disturbances.

Reinforcement learning has emerged as an effective approach for solving the optimal control problem [27], enabling control policies to be learned directly from system interactions and reward signals [28]. Among them, the actor–critic framework has received considerable attention, as it simultaneously approximates the control policy and the value function, facilitating efficient policy evaluation and iterative improvement. In recent years, actor–critic methods have been successfully applied to consensus, formation control, and cooperative coordination of MASs, demonstrating enhanced adaptability and convergence in dynamic and uncertain environments [29]. In [30], an actor–critic neural network framework was constructed to realize model-free optimal consensus control for discrete-time nonlinear MASs with unknown dynamics. In [31], a reinforcement learning control algorithm was developed for USVs with modeling uncertainties, where actor–critic networks were combined with backstepping design to achieve optimized tracking performance. Actor–critic reinforcement learning has undoubtedly yielded substantial progress in addressing optimal control problems. However, achieving stable, fast-converging, and fully distributed optimal control for second-order nonlinear MASs with unknown dynamics continues to present significant theoretical hurdles.

Motivated by these limitations, this work develops a distributed control framework that leverages actor–critic reinforcement learning to achieve control policies, employing FLS to approximate the unknown nonlinear dynamics. The main contributions of this article are summarized as follows.

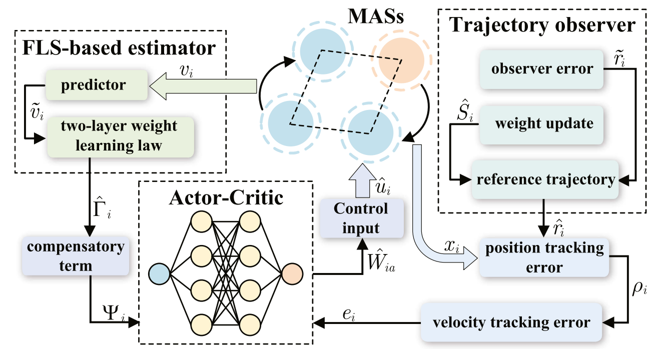

Fig. 1. Structure of the distributed formation control scheme.

1) A novel predictor-based fuzzy learning framework is developed, in which FLSs serve as universal approximators for unknown nonlinear dynamics. To enhance approximation accuracy and accelerate convergence, a two-layer learning mechanism is proposed that exploits the velocity prediction error.
2) An actor–critic reinforcement learning scheme is established to achieve optimal formation control. By reformulating the value function with a compensatory term that leverages available system structural information, the proposed approach accelerates convergence in both policy optimization and value function learning.
3) A comprehensive formation control framework is constructed by integrating the distributed reference trajectory observer with the predictor-based fuzzy learning scheme and the actor–critic reinforcement learning architecture, enabling each agent to achieve distributed optimal formation control despite limited access to the global reference trajectory.

Notation: $\mathbb{R}^n$ denotes the n-dimensional Euclidean space, and $\mathbb{R}^+$ represents the set of all positive real numbers. $\mathbb{R}^{n\times m}$ denotes the set of all real matrices with dimension $n \times m$. M^{T} denotes the transpose of matrix M. $\|\cdot\|$ represents the Euclidean norm. $\|\cdot\|_F$ denotes the Frobenius norm. diag$\{\cdot\}$ represents a diagonal matrix with the specified elements on its main diagonal. $\lambda_{\min}(\cdot)$ denotes the minimum eigenvalue of a given matrix. $\sigma_{\min}(\cdot)$ denotes the minimum singular value of a given matrix. $\mathrm{rand}(n)$ represents an n-dimensional vector containing random values drawn from a uniform distribution in the interval $(0, 1)$. $\mathrm{atan2}(z_1, z_2)$ denotes the four-quadrant inverse tangent of z_1 and z_2.

II. Preliminaries and Problem Formulation

A. Graph Theory and Switching Topology

Let $\mathcal{G}(\eta) = (\mathcal{V}, \mathcal{E}(\eta))$ denote a switching directed graph, where the piecewise-constant switching signal $\eta(t) : [0, \infty) \to \mathcal{P} = \{1, 2, \ldots, p\}$ specifies the active communication topology at time t, and p represents the total number of admissible communication subgraphs. The node set $\mathcal{V} = \{1, \ldots, N\}$ corresponds to the N agents, and the time-varying edge set $\mathcal{E}(\eta)$ and adjacency matrix $\mathcal{A}(\eta) = [a_{ij}(\eta)]_{N\times N}$ are determined by η. Specifically, $a_{ij}(\eta) > 0$ if agent i can receive information from agent j at time t, and $a_{ij}(\eta)=0$ otherwise. The Laplacian matrix associated with $\mathcal{G}(\eta)$ is defined as $\mathcal{L}(\eta)=\mathcal{D}(\eta)-\mathcal{A}(\eta)$, where $\mathcal{D}(\eta)=\mathrm{diag}\{d_1(\eta),\ldots,d_N(\eta)\}$ with $d_i(\eta)=\sum_{j=1}^{N}a_{ij}(\eta)$. To characterize intermittent communication, jointly connected switching topologies are considered, where the union graph over each interval $[t_k,t_{k+1})$ is defined as $\mathcal{G}([t_k,t_{k+1}))=(\mathcal{V},\cup_{t\in[t_k,t_{k+1})}\mathcal{E}(\eta))$.

Assumption 1: The switching communication topology $\mathcal{G}(\eta)$ is jointly connected; that is, for every interval $[t_k,t_{k+1})$, the union graph $\mathcal{G}([t_k,t_{k+1}))$ contains a spanning tree with the leader as its root.

B. System Model

A class of second-order nonlinear MASs is considered, where each agent is described by position dynamics and nonlinear velocity dynamics subject to external disturbances. The dynamics of agent i are modeled as

$$\begin{cases}\dot{x}_i(t)=A_i(x_i)v_i(t)\\ \dot{v}_i(t)=f_i(v_i(t))+B_i(x_i)u_i(t)+\omega_i(t)\end{cases}\tag{1}$$

where $x_i(t)\in\mathbb{R}^n$ denotes the position of agent i, $v_i(t)\in\mathbb{R}^n$ represents the velocity vector, and $u_i(t)\in\mathbb{R}^n$ is the control input. $A_i(x_i)\in\mathbb{R}^{n\times n}$ and $B_i(x_i)\in\mathbb{R}^{n\times n}$ are the system matrices that facilitate coordinate transformations between different reference frames. $f_i(v_i(t))\in\mathbb{R}^n$ denotes the intrinsic nonlinear velocity dynamics and unmodeled uncertainties, and $\omega_i(t)\in\mathbb{R}^n$ is the external environmental disturbances.

Remark 1: The system matrices $A_i(x_i)$ and $B_i(x_i)$ in (1) are employed to map velocities and control inputs between different reference frames, such as body-fixed and inertial frames. In particular, if $A_i(x_i)=I_n$ and $B_i(x_i)=I_n$, the dynamics (1) reduces to the conventional second-order MASs models studied in the papers [3], [8], [32]. By suitably selecting $A_i(x_i)$ and $B_i(x_i)$, this dynamics (1) can readily accommodate a wide range of practical second-order MASs, including USVs and other vehicles with orientation dependent dynamics. Compared with standard second-order agent models, this formulation offers enhanced modeling flexibility.

C. Control Objective

The reference trajectory $r(t)\in\mathbb{R}^n$ is generated by $\dot{r}(t)=f_0(r(t))$. Each agent i is required to asymptotically track the reference trajectory $x_0(t)$ and maintain a prescribed formation pattern, given by

$$\lim_{t\to\infty}\|x_i(t)-r(t)-d_i\|\le\varpi_i,\quad i=1,\ldots,N\tag{2}$$

where $d_i\in\mathbb{R}^n$ denotes the desired formation displacement of agent i relative to the reference trajectory, and $\varpi_i\in\mathbb{R}^+$ represents a bounded error tolerance.

Definition 1: Consider the nonlinear system $\dot{z}(t)=g(z(t))$ with $z(t)\in\mathbb{R}^n$. The system state $z(t)$ is said to be uniformly ultimately bounded (UUB) if there exist constants $\kappa>0$ and $T>0$, independent of the initial condition $z(0)$, such that $\|z(t)\|\le\kappa$ for all $t\ge T$.

III. Main Results

To provide a clear overview of the proposed approach, the structure of the distributed formation control scheme is illustrated in Fig. 1. The proposed framework combines a predictor-based fuzzy learning approach for approximating unknown dynamics, a distributed reference trajectory observer via local neighbor communication, and an actor–critic reinforcement learning scheme for distributed optimal formation control. The details of each component are presented in the following sections.

A. Predictor-Based Fuzzy Learning Approach

To address the unknown nonlinear dynamics $f_i(v_i)$, the FLS is adopted as a universal approximator [33]. The membership function of the jth fuzzy rule with respect to the lth element of v_i is defined as

$$\mu_{ij}(v_{il})=\exp\left(-\frac{(v_{il}-h_{ij})^2}{2w_{ij}^2}\right)\tag{3}$$

where $h_{ij}\in\mathbb{R}$ and $w_{ij}\in\mathbb{R}^+$ denote the center and width of the Gaussian membership function, respectively.

Based on the fuzzy inference mechanism, the nonlinear function $f_i(v_i)$ can be approximated as

$$f_i(v_i(t))=\frac{\sum_{j=1}^{m}\Gamma_{ij}\prod_{l=1}^{n}\mu_{ij}(v_{il})}{\sum_{j=1}^{m}\prod_{l=1}^{n}\mu_{ij}(v_{il})}+\epsilon_i(t)\tag{4}$$

where Γ_{ij} is the adjustable parameter vector associated with the jth rule, m is the number of fuzzy rules, and $\epsilon_i(t)$ denotes the approximation error.

By introducing the fuzzy basis function

$$\psi_{i,\ell}(v_i)=\frac{\prod_{l=1}^{n}\mu_{i,\ell}(v_{il})}{\sum_{j=1}^{m}\prod_{l=1}^{n}\mu_{ij}(v_{il})},\quad \ell=1,2,\ldots,m\tag{5}$$

the dynamics (4) can be expressed in a compact form as

$$f_i(v_i)=\Gamma_i^{\mathrm{T}}\psi_i(v_i)+\epsilon_i(t)\tag{6}$$

where $\Gamma_i=[\Gamma_{i1}^{\mathrm{T}},\Gamma_{i2}^{\mathrm{T}},\ldots,\Gamma_{im}^{\mathrm{T}}]^{\mathrm{T}}\in\mathbb{R}^{m\times n}$ is the ideal FLS weight and $\psi_i(v_i)=[\psi_{i1}(v_i),\psi_{i2}(v_i),\ldots,\psi_{im}(v_i)]^{\mathrm{T}}\in\mathbb{R}^m$ is the fuzzy basis function vector.

Accordingly, the FLS-based estimate of the unknown nonlinear dynamics is given by

$$\hat{f}_i(v_i)=\hat{\Gamma}_i^{\mathrm{T}}(t)\psi_i(v_i)\tag{7}$$

where $\hat{\Gamma}_i(t)$ denotes the estimate of the ideal FLS weight Γ_i.

To improve the convergence speed and approximation accuracy of FLS-based estimate (7), a predictor-based accelerated learning scheme is proposed for the adaptive online tuning of the FLS weight parameters $\hat{\Gamma}_i(t)$. The predictor of the velocity vector v_i is described by

$$\dot{\hat{v}}_i(t)=\hat{\Gamma}_i^{\mathrm{T}}(t)\psi_i(v_i)+B_i(x_i)u_i(t)-K_{i1}\tilde{v}_i(t)\tag{8}$$

where $\tilde{v}_i(t)=\hat{v}_i(t)-v_i(t)$ represents the velocity prediction error and $K_{i1}\in\mathbb{R}^{n\times n}$ is a positive definite gain matrix.

The dynamics of the prediction error can be expressed as

$$\dot{\tilde{v}}_i(t)=\tilde{\Gamma}_i^{\mathrm{T}}(t)\psi_i(v_i)-K_{i1}\tilde{v}_i(t)-\epsilon_i(t)\tag{9}$$

where $\tilde{\Gamma}_i(t) = \hat{\Gamma}_i(t) - \Gamma_i(t)$ is the weight estimation error.

To adaptively update the FLS weights, a two-layer predictor-based learning law is given by

$$\dot{\hat{\Gamma}}_i(t) = -c_{i1}\left(\hat{\Gamma}_i(t) - \Theta_i(t)\right) \tag{10}$$

$$\dot{\Theta}_i(t) = -c_{i2}\frac{\psi_i(v_i)}{1 + \psi_i^{\mathrm{T}}(v_i)\psi_i(v_i)}\theta_i^{\mathrm{T}}(t) \tag{11}$$

where $\theta_i(t) = \dot{\tilde{v}}_i(t) + K_{i1}\tilde{v}_i(t)$ and $c_{i1}, c_{i2} \in \mathbb{R}^+$ are the design learning rates.

Assumption 2: The environmental disturbances and the approximation errors are bounded. There exist positive constants $\bar{\omega}_i$ and $\bar{\epsilon}_i$ such that the inequalities $\|\omega_i(t)\| \le \bar{\omega}_i$ and $\|\epsilon_i(t)\| \le \bar{\epsilon}_i$ hold for all $t \ge 0$ and $i = 1, \ldots, N$.

B. Reference Trajectory Observer Design

Given the switching communication topology $\mathcal{G}(\eta)$, agents are subject to intermittent access to the leader or their neighbors. To address this challenge, a distributed observer is designed to simultaneously estimate the reference trajectory and learn its unknown dynamics, which are modeled as

$$f_0(r(t)) = S^T\psi_0(r) + \vartheta(t) \tag{12}$$

where $S \in \mathbb{R}^{m\times n}$ is the ideal FLS weight, $\psi_0(r) \in \mathbb{R}^m$ is the fuzzy basis function vector, and $\vartheta(t)$ denotes the bounded approximation error satisfying $\|\vartheta(t)\| \le \bar{\vartheta}$.

Since the true reference trajectory $r(t)$ is unavailable to the followers, the observer constructs the approximate dynamics using the estimated weight matrix $\hat{S}_i(t)$

$$\hat{f}_0(\hat{r}_i(t)) = \hat{S}_i^T(t)\psi_0(\hat{r}_i). \tag{13}$$

To achieve consensus on the reference trajectory under the jointly connected switching topology $\mathcal{G}(\eta)$, the composite observer error $\tilde{r}_i(t)$ is defined as

$$\tilde{r}_i(t) = \sum_{j=1}^{N} a_{ij}(\eta)(\hat{r}_i(t) - \hat{r}_j(t)) + \beta_i(\eta)(\hat{r}_i(t) - r(t)) \tag{14}$$

where $\beta_i(\eta) > 0$ indicates that agent i has direct access to the leader at time t. Then, the distributed reference trajectory observer is designed as

$$\dot{\hat{r}}_i(t) = \hat{S}_i^T(t)\psi_0(\hat{r}_i) - K_{i2}(t)\tilde{r}_i(t) \tag{15}$$

where $K_{i2}(t) = \varrho_{i1} + \varrho_{i2}e^{-\varrho_{i3}t}$ is the observer gain with $\varrho_{i1}, \varrho_{i2}, \varrho_{i3} \in \mathbb{R}^+$.

To capture the true leader dynamics, the weight matrix $\hat{S}_i(t)$ is updated via a normalized gradient adaptation law

$$\dot{\hat{S}}_i(t) = -c_{i3}\hat{S}_i(t) - c_{i4}\tilde{\psi}_0(\hat{r}_i)\tilde{r}_i^{\mathrm{T}}(t) \tag{16}$$

where $\tilde{\psi}_0(\hat{r}_i) = \psi_0(\hat{r}_i)/(1 + \psi_0^{\mathrm{T}}(\hat{r}_i)\psi_0(\hat{r}_i))$ and $c_{i3}, c_{i4} \in \mathbb{R}^+$ are the design parameters.

Let $\delta_i(t) = \hat{r}_i(t) - r(t)$ be the local observer error of agent i relative to the true reference trajectory. The global error vector is denoted as $\tilde{r} = [\tilde{r}_1^{\mathrm{T}}, \ldots, \tilde{r}_N^{\mathrm{T}}]^{\mathrm{T}}$ and $\delta = [\delta_1^{\mathrm{T}}, \ldots, \delta_N^{\mathrm{T}}]^{\mathrm{T}}$. The observer error (14) can be written as

$$\tilde{r}(t) = (\mathcal{H}(\eta) \otimes I_n)\,\delta(t) \tag{17}$$

where $\mathcal{H}(\eta) = \mathcal{L}(\eta) + \mathcal{B}(\eta)$ is positive definite according to Assumption 1. The time derivative of $\delta_i(t)$ is given by

$$\dot{\delta}_i(t) = \hat{S}_i^{\mathrm{T}}(t)\psi_0(\hat{r}_i) - K_{i2}(t)\tilde{r}_i(t) - S^{\mathrm{T}}\psi_0(r) - \vartheta(t). \tag{18}$$

Adding and subtracting $S^{\mathrm{T}}\psi_0(\hat{r}_i)$ to (18) yields

$$\dot{\delta}_i(t) = \tilde{S}_i^{\mathrm{T}}(t)\psi_0(\hat{r}_i) - K_{i2}(t)\tilde{r}_i(t) + \Omega_i(t) \tag{19}$$

where $\Omega_i(t) = S^{\mathrm{T}}(\psi_0(\hat{r}_i) - \psi_0(r)) - \vartheta(t)$.

Assumption 3: The residual error of the reference trajectory observer is bounded. There exists a positive constant $\bar{\Omega}_i$ such that the inequality $\|\Omega_i(t)\| \le \bar{\Omega}_i$ holds for all $t \ge 0$.

C. Distributed Formation Controller Design

Based on the estimated reference trajectory $\hat{r}_i(t)$, the position tracking error of agent i is defined as

$$\rho_i(t) = x_i(t) - \hat{r}_i(t) - d_i. \tag{20}$$

To achieve formation tracking, a desired velocity command $v_i^d(t)$ is designed based on (20) as

$$v_i^d(t) = A_i^{-1}(x_i)\left(-K_{i3}\rho_i(t) + \dot{\hat{r}}_i(t)\right) \tag{21}$$

where $K_{i3} \in \mathbb{R}^{n\times n}$ is a positive definite gain matrix.

In order to design an optimal controller based on the desired velocity command $v_i^d(t)$ in (21), the velocity tracking error of agent i is defined as

$$e_i(t) = v_i(t) - v_i^d(t). \tag{22}$$

To simultaneously ensure error convergence and reasonable control cost, the following performance index is introduced:

$$J_i(e_i, u_i) = \int_0^{\infty} U_i(e_i, u_i)\mathrm{d}s \tag{23}$$

where $U_i(e_i, u_i) = e_i^{\mathrm{T}}(t)P_ie_i(t) + u_i^{\mathrm{T}}(t)Q_iu_i(t)$ is the instantaneous cost function, with $P_i \in \mathbb{R}^{n\times n}$ and $Q_i \in \mathbb{R}^{n\times n}$ being symmetric positive definite weighting matrices.

The associated value function is given by

$$V_i(e_i) = \int_t^{\infty} U_i(e_i, u_i)\mathrm{d}s \tag{24}$$

and its optimal form is expressed as

$$V_i^*(e_i) = \min_{u_i\in\Pi_{u_i}} \int_t^{\infty} U_i(e_i, u_i)\mathrm{d}s \tag{25}$$

where Π_{u_i} denotes the admissible control set.

By optimal control theory [27], the optimal value function $V_i^*(e_i)$ satisfies the Hamilton–Jacobi–Bellman equation

$$\begin{aligned} H_i(e_i, u_i) = U_i(e_i, u_i) + (V_{e_i}^*(e_i))^{\mathrm{T}}\,(f_i(v_i) \\ + B_i(x_i)u_i(t) - \dot{v}_i^d(t) + \omega_i(t)) \end{aligned} \tag{26}$$

where $V_{e_i}^*(e_i) = \partial V_i^*(e_i)/\partial e_i$. Setting $\partial H_i(e_i, u_i)/\partial u_i = 0$, the corresponding optimal control law can be derived as

$$u_i^*(e_i) = -\frac{1}{2}Q_i^{-1}B_i^{\mathrm{T}}(x_i)V_{e_i}^*(e_i). \tag{27}$$

To accelerate the learning of optimal control policies and enhance the robustness, the value function is reformulated by incorporating available system structural information as

$$V_i^*(e_i) = \tilde{V}_i(e_i) + \Psi_i(t) \tag{28}$$

where $\tilde{V}_i(e_i) = V_i^*(e_i) - \Psi_i(t)$ with $\Psi_i(t) \in \mathbb{R}$ being a compensatory term to be designed.

Since the optimal value function is generally difficult to obtain analytically, it is common to approximate it using radial basis function neural networks [34]. Accordingly, the optimal value function (28) can be represented by

$$V_i^*(e_i) = W_i^{\mathrm{T}}\phi_i(e_i) + \Psi_i(t) + \varsigma_i(t) \tag{29}$$

where $W_i \in \mathbb{R}^q$ is the ideal neural networks weight, $\phi_i(e_i) \in \mathbb{R}^q$ denotes the radial basis functions, and $\varsigma_i(t) \in \mathbb{R}$ is the approximation error.

An online-updated critic network is employed to learn the ideal neural network weights W_i, formulated as

$$\hat{V}_i(e_i) = \hat{W}_{ic}^{\mathrm{T}}(t)\phi_i(e_i) + \Psi_i(t) \tag{30}$$

where $\hat{W}_{ic}(t) \in \mathbb{R}^q$ denotes the critic network weight. The gradient of $\hat{V}_i(e_i)$ with respect to e_i is given by

$$\hat{V}_{e_i}(e_i) = \phi_{e_i}^{\mathrm{T}}(e_i)\hat{W}_{ic}(t) + \Psi_{e_i}(e_i) \tag{31}$$

with $\phi_{e_i}(e_i) = \partial\phi_i(e_i)/\partial e_i(t)$ and $\Psi_{e_i}(e_i) = \partial\Psi_i(t)/\partial e_i(t)$.

Based on (27) and (31), the corresponding optimal control policy is expressed as

$$\hat{u}_i(e_i) = -\frac{1}{2}Q_i^{-1}B_i^{\mathrm{T}}(x_i)\left(\phi_{e_i}^{\mathrm{T}}(e_i)\hat{W}_{ic}(t) + \Psi_{e_i}(e_i)\right). \tag{32}$$

To facilitate real-time approximation of the optimal value function, the critic network weight $\hat{W}_{ic}(t)$ is updated by

$$\dot{\hat{W}}_{ic}(t) = -b_{i1}\phi_i(e_i)\phi_i^{\mathrm{T}}(e_i)\hat{W}_{ic}(t) \tag{33}$$

where $b_{i1} \in \mathbb{R}^+$ denotes the learning rate of the critic network.

Motivated by the command velocity tracking error $e_i(t)$ in (22), the position tracking error $\rho_i(t)$ in (20), and the FLS-based estimate of nonlinear dynamics $\hat{f}_i(v_i)$ in (7), the compensatory term $\Psi_i(t)$ is constructed as

$$\begin{aligned}\Psi_i(t) = &\left(K_{i3}e_i(t) + 2K_{i4}\rho_i(t) + 2\hat{\Gamma}_i^{\mathrm{T}}(t)\psi_i(v_i)\right)^{\mathrm{T}} \\ &\times (B_i^{-1}(x_i))^{\mathrm{T}}Q_iB_i^{-1}(x_i)e_i(t)\end{aligned} \tag{34}$$

where $K_{i3} = K_{i3}^{\mathrm{T}} \in \mathbb{R}^{n\times n}$ and $K_{i4} \in \mathbb{R}^{n\times n}$ are positive definite gain matrices.

The derivative of the compensatory term $\Psi_i(t)$ with respect to the tracking error e_i, $\Psi_{e_i}(e_i)$ is further expanded as

$$\begin{aligned}\Psi_{e_i}(e_i) = 2(B_i^{-1}(x_i))^{\mathrm{T}}Q_iB_i^{-1}(x_i)&\Big(K_{i3}e_i(t) \\ &+ K_{i4}\rho_i(t) + \hat{\Gamma}_i^{\mathrm{T}}(t)\psi_i(v_i)\Big).\end{aligned} \tag{35}$$

By substituting (35) into (32), the final actor network-based control law is derived as

$$\begin{aligned}\hat{u}_i(e_i) = &-\frac{1}{2}Q_i^{-1}B_i^{\mathrm{T}}(x_i)\phi_{e_i}^{\mathrm{T}}(e_i)\hat{W}_{ia}(t) - B_i^{-1}\big(K_{i3}e_i(t) \\ &+K_{i4}\rho_i(t) + \hat{\Gamma}_i^{\mathrm{T}}(t)\psi_i(v_i)\big)\end{aligned} \tag{36}$$

where $\hat{W}_{ia}(t) \in \mathbb{R}^q$ denotes the actor network weight.

To achieve online adaptation, the actor network weight $\hat{W}_{ia}(t)$ in (36) is updated by

$$\begin{aligned}\dot{\hat{W}}_{ia}(t) = &-\phi_i(e_i)\phi_i^{\mathrm{T}}(e_i)\Big(b_{i2}\hat{W}_{ic}(t) \\ &+ b_{i3}\left(\hat{W}_{ia}(t) - \hat{W}_{ic}(t)\right)\Big)\end{aligned} \tag{37}$$

where $b_{i2}, b_{i3} \in \mathbb{R}^+$ denote the learning rates.

D. Stability Analysis

Theorem 1: Consider the FLS-based estimate (7) implemented with the velocity predictor (8). The prediction error $\tilde{v}_i$ and weight estimation error $\tilde{\Gamma}_i$ are guaranteed to be uniformly bounded, provided that the gain matrix K_{i1} and the learning rates c_{i1}, c_{i2} are selected to satisfy $c_{i1} \geq 2c_{i2}$ and $2\lambda_{\min}(K_{i1})\bar{\psi}_i(v_i) \geq 1$ where $\bar{\psi}_i(v_i) = 1/(1+\psi_i^{\mathrm{T}}(v_i)\psi_i(v_i))$.

Proof: Construct the Lyapunov candidate function as

$$L_{iv} = \frac{1}{2}\tilde{v}_i^{\mathrm{T}}\tilde{v}_i + \frac{1}{c_{i2}}\|\Theta_i - \Gamma_i\|_F^2 + \frac{1}{c_{i2}}\|\hat{\Gamma}_i - \Theta_i\|_F^2. \tag{38}$$

Taking the time derivative of L_{iv} yields

$$\begin{aligned}\dot{L}_{iv} = &\tilde{v}_i^{\mathrm{T}}\left(\tilde{\Gamma}_i^{\mathrm{T}}\psi_i(v_i) - K_{i1}\tilde{v}_i - \epsilon_i\right) - 2\,\mathrm{tr}\big((\Theta_i - \Gamma_i)^{\mathrm{T}} \\ &\times\tilde{\psi}_i(v_i)\theta_i^{\mathrm{T}}\big) - \frac{2}{c_{i2}}\mathrm{tr}\Big((\hat{\Gamma}_i - \Theta_i)^{\mathrm{T}} \\ &\times\left(c_{i1}(\hat{\Gamma}_i - \Theta_i) - c_{i2}\tilde{\psi}_i(v_i)\theta_i^{\mathrm{T}}\right)\Big)\end{aligned} \tag{39}$$

where $\tilde{\psi}_i(v_i) = \psi_i(v_i)/(1+\psi_i^{\mathrm{T}}(v_i)\psi_i(v_i))$.

Substituting the relation $\Gamma_i = \hat{\Gamma}_i - \tilde{\Gamma}_i$ into (39) yields

$$\begin{aligned}\dot{L}_{iv} = &\tilde{v}_i^{\mathrm{T}}\left(\tilde{\Gamma}_i^{\mathrm{T}}\psi_i(v_i) - K_{i1}\tilde{v}_i - \epsilon_i\right) - 2\,\mathrm{tr}\left(\tilde{\Gamma}_i^{\mathrm{T}}\tilde{\psi}_i(v_i)\theta_i^{\mathrm{T}}\right) \\ &-\frac{2c_{i1}}{c_{i2}}\mathrm{tr}\left(\left(\hat{\Gamma}_i - \Theta_i\right)^{\mathrm{T}}\left(\hat{\Gamma}_i - \Theta_i\right)\right) \\ &+ 4\,\mathrm{tr}\left((\hat{\Gamma}_i - \Theta_i)^{\mathrm{T}}\tilde{\psi}_i(v_i)\theta_i^{\mathrm{T}}\right).\end{aligned} \tag{40}$$

Substituting the relationship $\tilde{\Gamma}_i^{\mathrm{T}}\psi_i(v_i) = \theta_i + \epsilon_i$ derived from (9) into (40) yields

$$\begin{aligned}\dot{L}_{iv} = &\tilde{v}_i^{\mathrm{T}}(\theta_i - K_{i1}\tilde{v}_i) - 2\bar{\psi}_i(v_i)\,\mathrm{tr}\left(\theta_i\theta_i^{\mathrm{T}} + \epsilon_i\theta_i^{\mathrm{T}}\right) \\ &-\frac{2c_{i1}}{c_{i2}}\mathrm{tr}\left((\hat{\Gamma}_i - \Theta_i)^{\mathrm{T}}(\hat{\Gamma}_i - \Theta_i)\right) \\ &+ 4\bar{\psi}_i(v_i)\,\mathrm{tr}\left((\hat{\Gamma}_i - \Theta_i)^{\mathrm{T}}\psi_i(v_i)\theta_i^{\mathrm{T}}\right).\end{aligned} \tag{41}$$

Furthermore, applying the square completion technique under the condition $c_{i1} \geq 2c_{i2}$ yields

$$\begin{aligned}\dot{L}_{iv} \leq &\|\tilde{v}_i\|\|\theta_i\| - \lambda_{\min}(K_{i1})\|\tilde{v}_i\|^2 - \bar{\psi}_i(v_i)\|\theta_i\|^2 \\ &-\frac{2c_{i1}}{c_{i2}}\bar{\psi}_i(v_i)\|\hat{\Gamma}_i - \Theta_i\|_F^2 \\ &-\bar{\psi}_i(v_i)\left(2\|\hat{\Gamma}_i - \Theta_i\|_F\psi_i(v_i) - \|\theta_i\|\right)^2.\end{aligned} \tag{42}$$

Based on Young's inequality, we have

$$\|\tilde{v}_i\|\|\theta_i\| \leq \frac{\lambda_{\min}(K_{i1})}{2}\|\tilde{v}_i\|^2 + \frac{1}{2\lambda_{\min}(K_{i1})}\|\theta_i\|^2. \tag{43}$$

Substituting (43) into (42) yields

$$\dot{L}_{iv} \leq -\frac{\lambda_{\min}(K_{i1})}{2}\|\tilde{v}_i\|^2 + \left(\frac{1}{2\lambda_{\min}(K_{i1})} - \bar{\psi}_i(v_i)\right)\|\theta_i\|^2 - \frac{2c_{i1}}{c_{i2}}\bar{\psi}_i(v_i)\|\hat{\Gamma}_i - \Theta_i\|_F^2. \tag{44}$$

Provided that the condition $1/(2\lambda_{\min}(K_{i1})) - \bar{\psi}_i(v_i) \leq 0$ holds, it follows from (44) that $\dot{L}_{iv} \leq 0$. This implies that $L_{iv}(t) \leq L_{iv}(0)$, which guarantees $\|\tilde{v}_i\| \leq \sqrt{2L_{iv}(0)}$ and $\|\Theta_i - \Gamma_i\|_F, \|\hat{\Gamma}_i - \Theta_i\|_F \leq \sqrt{c_{i2}L_{iv}(0)}$. By applying the triangle inequality to the decomposition $\tilde{\Gamma}_i = (\Theta_i - \Gamma_i) + (\hat{\Gamma}_i - \Theta_i)$, we obtain $\|\tilde{\Gamma}_i\|_F \leq 2\sqrt{c_{i2}L_{iv}(0)}$. According to Definition 1, these results confirm that the errors $\tilde{v}_i$ and $\tilde{\Gamma}_i$ are uniformly bounded, which completes the proof.

Theorem 2: Consider the distributed reference trajectory observer described by (15) and the adaptive weight update law (16). Under Assumption 1, the distributed observer error $\tilde{r}_i$ and the weight estimation error $\tilde{S}_i = \hat{S}_i - S$ are UUB, provided that $\varrho_{i1} > 1$ and $c_{i3} > c_{i4}\|\psi_0(\hat{r}_i)\|^2$.

Proof: Consider the following Lyapunov candidate function:

$$L_r = \frac{1}{2}\delta^{\mathrm{T}}\left(\mathcal{H}(\eta) \otimes I_n\right)\delta + \frac{1}{2c_{i4}}\sum_{i=1}^{N}\mathrm{tr}\left(\tilde{S}_i^{\mathrm{T}}\tilde{S}_i\right). \tag{45}$$

Taking the time derivative of L_r and substituting the adaptive update law (16) yields

$$\dot{L}_r = \sum_{i=1}^{N}\tilde{r}_i^{\mathrm{T}}\left(\tilde{S}_i^{\mathrm{T}}\psi_0(\hat{r}_i) - K_{i2}\tilde{r}_i + \Omega_i\right) - \frac{1}{c_{i4}}\sum_{i=1}^{N}\mathrm{tr}\left(c_{i3}\tilde{S}_i^{\mathrm{T}}\hat{S}_i + c_{i4}\tilde{S}_i^{\mathrm{T}}\tilde{\psi}_0(\hat{r}_i)\tilde{r}_i^{\mathrm{T}}\right). \tag{46}$$

Using the property $\mathrm{tr}(\tilde{S}_i^{\mathrm{T}}\psi_0(\hat{r}_i)\tilde{r}_i^{\mathrm{T}}) = \tilde{r}_i^{\mathrm{T}}\tilde{S}_i^{\mathrm{T}}\psi_0(\hat{r}_i)$, and (46) can be rewritten as

$$\dot{L}_r = \sum_{i=1}^{N}\tilde{r}_i^{\mathrm{T}}\left(\Omega_i - K_{i2}\tilde{r}_i\right) - \frac{c_{i3}}{c_{i4}}\sum_{i=1}^{N}\mathrm{tr}\left(\tilde{S}_i^{\mathrm{T}}\hat{S}_i\right) + \sum_{i=1}^{N}\tilde{r}_i^{\mathrm{T}}\tilde{S}_i^{\mathrm{T}}\psi_0(\hat{r}_i)\psi_0^{\mathrm{T}}(\hat{r}_i)\tilde{\psi}_0(\hat{r}_i). \tag{47}$$

By applying Young's Inequality, we obtain

$$\tilde{r}_i^{\mathrm{T}}\Omega_i \leq \frac{1}{2}\|\tilde{r}_i\|^2 + \frac{1}{2}\bar{\Omega}_i^2 \tag{48}$$

$$\tilde{r}_i^{\mathrm{T}}\tilde{S}_i^{\mathrm{T}}\tilde{\psi}_0(\hat{r}_i) \leq \frac{1}{2}\|\tilde{r}_i\|^2 + \frac{1}{2}\|\psi_0(\hat{r}_i)\|^2\|\tilde{S}_i\|_F^2 \tag{49}$$

$$-\frac{c_{i3}}{c_{i4}}\mathrm{tr}\left(\tilde{S}_i^{\mathrm{T}}\hat{S}_i\right) \leq -\frac{c_{i3}}{2c_{i4}}\|\tilde{S}_i\|_F^2 + \frac{c_{i3}}{2c_{i4}}\|S\|_F^2. \tag{50}$$

Note that $0 \leq \psi_0^{\mathrm{T}}(\hat{r}_i)\tilde{\psi}_0(\hat{r}_i) < 1$. Substituting (48)–(50) into (47) yields

$$\dot{L}_r \leq -\sum_{i=1}^{N}\left(K_{i2} - 1\right)\|\tilde{r}_i\|^2 - \sum_{i=1}^{N}\left(\frac{c_{i3}}{2c_{i4}} - \frac{1}{2}\|\psi_0(\hat{r}_i)\|^2\right)\|\tilde{S}_i\|_F^2 + \Omega_r \tag{51}$$

where $\Omega_r = \sum_{i=1}^{N}\left(\frac{1}{2}\bar{\Omega}_i^2 + \frac{c_{i3}}{2c_{i4}}\|S\|_F^2\right)$.

Let $G_1 = \min_{i=1,2,\ldots,N}(K_{i2}-1)$ and $G_2 = \min_{i=1,2,\ldots,N}(\frac{c_{i3}}{2c_{i4}} - \frac{1}{2}\|\psi_0(\hat{r}_i)\|^2)$. Provided that the parameters are selected such that $\varrho_{i1} > 1$ and $c_{i3} > c_{i4}\|\psi_0(\hat{r}_i)\|^2$, we have $G_1 > 0$ and $G_2 > 0$. Consequently, $\dot{L}_r < 0$ whenever either $\|\tilde{r}\| > \sqrt{\Omega_r/G_1}$ or $\|\tilde{S}\|_F > \sqrt{\Omega_r/G_2}$. According to Definition 1, this implies that the errors $\tilde{r}$ and $\tilde{S}$ are UUB.

Under Assumption 1, for every piecewise constant switching signal $\eta \in \mathcal{P}$, the graph contains a spanning tree rooted at the leader, ensuring that $\mathcal{H}(\eta)$ is positive definite. Since the set of admissible topologies $\mathcal{P}$ is finite, there exists $\min_{k\in\mathcal{P}}\lambda_{\min}(\mathcal{H}(k)) > 0$. Recalling the error relationship $\tilde{r}(t) = (\mathcal{H}(\eta) \otimes I_n)\delta(t)$ yields

$$\|\tilde{r}(t)\| \geq \min_{k\in\mathcal{P}}\sigma_{\min}\left(\mathcal{H}(k)\right)\|\delta(t)\|. \tag{52}$$

Consequently, the established boundedness of $\|\tilde{r}(t)\|$ directly guarantees the boundedness of the distributed trajectory estimation error $\|\delta(t)\|$.

Theorem 3: Consider the second-order nonlinear MASs (1), along with the FLS-based estimate (7), the reference trajectory observer (15), the adaptive critic and actor network weight update laws (33) and (37), and the distributed formation controller (36). Under Assumptions 1, the closed-loop error dynamics (20) and (22) are guaranteed to be UUB, provided that there exist gain matrices K_{i3}, K_{i4}, positive scalars b_{i1}, b_{i2}, b_{i3}, and a symmetric positive definite matrix Q_i such that the following inequalities are satisfied:

$$\begin{cases}\lambda_{\min}(K_{i3}) - \frac{1}{2}\|K_{i4}\|^2 - \frac{1}{2} \geq 0\\ \lambda_{\min}(K_{i3}) - \frac{1}{2}\|A_i(x_i)\|^2 - 3 \geq 0\\ b_{i1} + b_{i2} - b_{i3} \geq 0\\ b_{i3} - \|\tilde{Q}_i\|^2 \geq 0\end{cases} \tag{53}$$

where $\tilde{Q}_i = B_i(x_i)Q_i^{-1}B_i^{\mathrm{T}}(x_i)\phi_{e_i}^{\mathrm{T}}(e_i)$.

Proof: Choose the Lyapunov candidate function as

$$L_{ie} = \frac{1}{2}\left(\rho_i^{\mathrm{T}}\rho_i + e_i^{\mathrm{T}}e_i + \mathrm{tr}\left(\tilde{W}_{ic}^{\mathrm{T}}\tilde{W}_{ic}\right) + \mathrm{tr}\left(\tilde{W}_{ia}^{\mathrm{T}}\tilde{W}_{ia}\right)\right) \tag{54}$$

where $\tilde{W}_{ic} = \hat{W}_{ic} - W_i$ and $\tilde{W}_{ia} = \hat{W}_{ia} - W_i$. Taking the time derivative of L_{ie} yields

$$\dot{L}_{ie} = \rho_i^{\mathrm{T}}\dot{\rho}_i + e_i^{\mathrm{T}}\dot{e}_i + \mathrm{tr}\left(\tilde{W}_{ic}^{\mathrm{T}}\dot{\tilde{W}}_{ic}\right) + \mathrm{tr}\left(\tilde{W}_{ia}^{\mathrm{T}}\dot{\tilde{W}}_{ia}\right). \tag{55}$$

It follows from (20) and (21) that

$$\rho_i^{\mathrm{T}}\dot{\rho}_i = \rho_i^{\mathrm{T}}A_i(x_i)e_i - \rho_i^{\mathrm{T}}K_{i3}\rho_i$$

$$\leq \left(\frac{1}{2}-\lambda_{\min}(K_{i3})\right)\|\rho_i\|^2+\frac{1}{2}\|A_i(x_i)\|^2\|e_i\|^2. \quad (56)$$

Similarly, for the velocity tracking error e_i, we have

$$\begin{aligned}
e_i^{\mathrm{T}}\dot{e}_i &= -\frac{1}{2}e_i^{\mathrm{T}}\tilde{Q}_i\hat{W}_{ia}-e_i^{\mathrm{T}}K_{i3}e_i-e_i^{\mathrm{T}}K_{i4}\rho_i\\
&\quad -e_i^{\mathrm{T}}\tilde{\Gamma}_i^{\mathrm{T}}(t)\psi_i(v_i)-e_i^{\mathrm{T}}\dot{v}_i^d+e_i^{\mathrm{T}}\omega_i+e_i^{\mathrm{T}}\epsilon_i\\
&\leq -(\lambda_{\min}(K_{i3})-3)\|e_i\|^2+\frac{1}{2}\|K_{i4}\|^2\|\rho_i\|^2\\
&\quad +\frac{1}{2}\|\tilde{Q}_i\|^2\|\hat{W}_{ia}\|^2+\frac{1}{2}\|\tilde{\Gamma}_i^{\mathrm{T}}(t)\psi_i(v_i)\|^2\\
&\quad +\frac{1}{2}\|\dot{v}_i^d\|^2+\frac{1}{2}\|\omega_i\|^2+\frac{1}{2}\|\epsilon_i\|^2. \quad (57)
\end{aligned}$$

For the weight estimation errors $\tilde{W}_{ic}$ and $\tilde{W}_{ia}$, according to the adaptive laws (33) and (37), we have

$$\begin{aligned}
&\mathrm{tr}\left(\tilde{W}_{ic}^{\mathrm{T}}\dot{\tilde{W}}_{ic}\right)+\mathrm{tr}\left(\tilde{W}_{ia}^{\mathrm{T}}\dot{\tilde{W}}_{ia}\right)\\
&= -\mathrm{tr}\left(b_{i1}\tilde{W}_{ic}^{\mathrm{T}}\phi_i(e_i)\phi_i^{\mathrm{T}}(e_i)\hat{W}_{ic}\right)\\
&\quad -\mathrm{tr}\Big((b_{i2}-b_{i3})\tilde{W}_{ia}^{\mathrm{T}}\phi_i(e_i)\phi_i^{\mathrm{T}}(e_i)\hat{W}_{ic}\\
&\quad +b_{i3}\tilde{W}_{ia}^{\mathrm{T}}\phi_i(e_i)\phi_i^{\mathrm{T}}(e_i)\hat{W}_{ia}\Big). \quad (58)
\end{aligned}$$

Applying Young's inequality yields

$$\tilde{W}_{ic}^{\mathrm{T}}\phi_i\phi_i^{\mathrm{T}}\hat{W}_{ic}\leq\frac{1}{2}\tilde{W}_{ic}^{\mathrm{T}}\phi_i\phi_i^{\mathrm{T}}\tilde{W}_{ic}+\frac{1}{2}\hat{W}_{ic}^{\mathrm{T}}\phi_i\phi_i^{\mathrm{T}}\hat{W}_{ic} \quad (59)$$

$$\tilde{W}_{ia}^{\mathrm{T}}\phi_i\phi_i^{\mathrm{T}}\hat{W}_{ic}\leq\frac{1}{2}\tilde{W}_{ia}^{\mathrm{T}}\phi_i\phi_i^{\mathrm{T}}\tilde{W}_{ia}+\frac{1}{2}\hat{W}_{ic}^{\mathrm{T}}\phi_i\phi_i^{\mathrm{T}}\hat{W}_{ic}. \quad (60)$$

Using $\tilde{W}_{ia}=\hat{W}_{ia}-W_i$, we can get

$$\begin{aligned}
\tilde{W}_{ia}^{\mathrm{T}}\phi_i\phi_i^{\mathrm{T}}\hat{W}_{ia} &= \frac{1}{2}\tilde{W}_{ia}^{\mathrm{T}}\phi_i\phi_i^{\mathrm{T}}\tilde{W}_{ia}+\frac{1}{2}\hat{W}_{ia}^{\mathrm{T}}\phi_i\phi_i^{\mathrm{T}}\hat{W}_{ia}\\
&\quad -\frac{1}{2}W_i^{\mathrm{T}}\phi_i\phi_i^{\mathrm{T}}W_i. \quad (61)
\end{aligned}$$

Substituting (56)–(61) into (55) leads to

$$\begin{aligned}
\dot{L}_{ie}\leq &-\left(\lambda_{\min}(K_{i3})-\frac{1}{2}\|K_{i4}\|^2-\frac{1}{2}\right)\|\rho_i\|^2\\
&-\left(\lambda_{\min}(K_{i3})-\frac{1}{2}\|A_i(x_i)\|^2-3\right)\|e_i\|^2\\
&-\frac{b_{i1}}{2}\|\phi_i(e_i)\|^2\left\|\tilde{W}_{ic}\right\|^2-\frac{b_{i2}}{2}\|\phi_i(e_i)\|^2\left\|\tilde{W}_{ia}\right\|^2\\
&-\frac{1}{2}(b_{i1}+b_{i2}-b_{i3})\|\phi_i(e_i)\|^2\left\|\hat{W}_{ic}\right\|^2\\
&-\frac{1}{2}\left(b_{i3}-\left\|\tilde{Q}_i\right\|^2\right)\|\phi_i(e_i)\|^2\left\|\hat{W}_{ia}\right\|^2\\
&+\frac{1}{2}\left\|\tilde{\Gamma}_i^{\mathrm{T}}(t)\psi_i(v_i)\right\|^2+\frac{1}{2}\left\|\dot{v}_i^d\right\|^2+\frac{1}{2}\bar{\omega}_i^2\\
&+\frac{1}{2}\bar{\epsilon}_i^2+\frac{b_{i3}}{2}\|\phi_i(e_i)\|^2\|W_i\|^2. \quad (62)
\end{aligned}$$

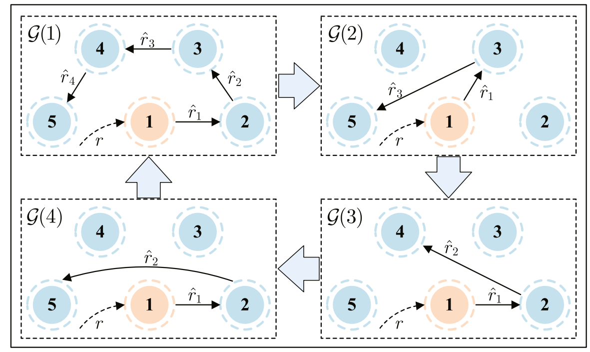

Fig. 2. Switching communication topology of USVs formation.

Under the conditions in (53), (62) can be rewritten as

$$\dot{L}_{ie}\leq-\Phi_iL_{ie}+\Lambda_i \quad (63)$$

where $\Phi_i=\min\{2\lambda_{\min}(K_{i3})-\|K_{i4}\|^2-1,\ 2\lambda_{\min}(K_{i3})-\|A_i(x_i)\|^2-6,\ b_{i1}\|\phi_i(e_i)\|^2, b_{i2}\|\phi_i(e_i)\|^2\}$, $\Lambda_i=(\|\dot{v}_i^d\|^2+\bar{\omega}_i^2+\bar{\epsilon}_i^2+\|\tilde{\Gamma}_i^{\mathrm{T}}(t)\psi_i(v_i)\|^2+b_{i3}\|\phi_i(e_i)\|^2\|W_i\|^2)/2$.

Furthermore, it follows from (63) that:

$$L_{ie}(t)\leq L_{ie}(0)e^{-\Phi_it}+\frac{\Lambda_i}{\Phi_i}\left(1-e^{-\Phi_it}\right) \quad (64)$$

which implies that $\|\rho_i\|, \|e_i\|, \|\tilde{W}_{ic}\|, \|\tilde{W}_{ia}\|\leq\sqrt{2\Lambda_i/\Phi_i}$.

According to Definition 1, these bounds indicate that the closed-loop error dynamics (20) and (22) are UUB, which completes the proof.

IV. Simulation Example

In this section, the effectiveness and robustness of the proposed control strategy are assessed through Monte Carlo simulations on a formation of $N=5$ USVs, where $M=100$ independent trials are performed over a simulation time of $T=100\,\mathrm{s}$ to evaluate system performance under statistical uncertainties. For each USV, the state vectors are defined as $x_i=[x_{i1},x_{i2},x_{i3}]^{\mathrm{T}}$ and $v_i=[v_{i1},v_{i2},v_{i3}]^{\mathrm{T}}$, where the detailed dynamic model parameters are adopted from [6]. To assess robustness against model mismatches, random parametric uncertainties are introduced by perturbing the inertial and damping matrices as $\tilde{M}_i=(0.9+0.2\,\mathrm{rand}(1))M_i$ and $\tilde{D}_i=(0.9+0.2\,\mathrm{rand}(1))D_i$, where M_i and D_i denote the nominal inertial and damping matrices. The reference trajectory $r(t)$ is generated by $r(t)=[0.14\sqrt{3}t-1.5\sin(0.07t),\ 0.14t+1.5\sqrt{3}\sin(0.07t),\ \mathrm{atan2}(0.14+0.105\sqrt{3}\cos(0.07t)\,,0.14\sqrt{3}-0.105\cos(0.07t))]$. A switching communication topology $\mathcal{G}(\eta)$, where $\eta\in\{1,2,3,4\}$ denotes the active communication graph shown in Fig. 2, is employed and periodically alternates among four subgraphs with a switching period of 1.0 s.

The external environmental disturbance $\omega_i(t)$ is described by $\omega_i(t)=\gamma_i([0.4\sin(0.5t),\ 0.2\cos(0.4t),\ 0.05\sin(0.1t)]^{\mathrm{T}}+0.05\,\mathrm{rand}(3))$, where $\gamma_i=0.1$ is the disturbance intensity. The FLS is constructed with $m=6$ fuzzy rules, where the centers are evenly spaced in the interval $[-1,1]$. The number of neurons in the actor and critic networks is set to $q=8$. The

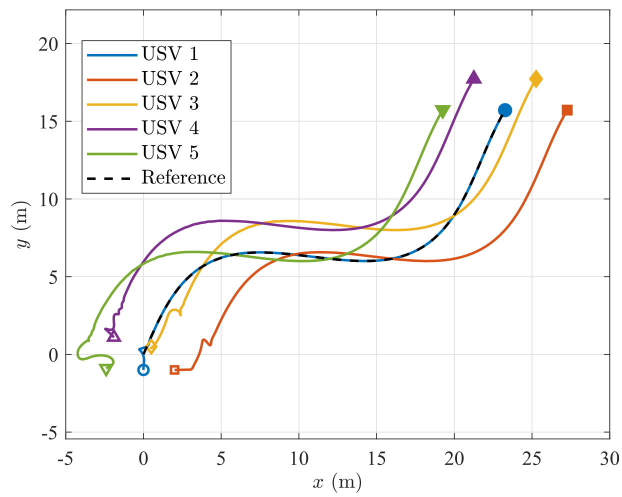

Fig. 3. Actual trajectories of the five USVs.

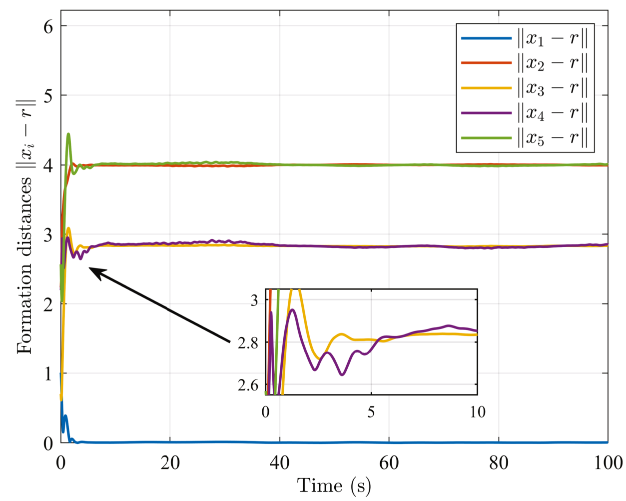

Fig. 4. Time evolution of the formation distances $\|x_i - r\|$ for USVs.

TABLE I
PERFORMANCE METRICS UNDER DIFFERENT DISTURBANCE INTENSITIES

Disturbance γ_i	**1.0**	**2.0**	**3.0**	**4.0**	**5.0**
Observer MSE_1	0.0071	0.0076	0.0075	0.0072	0.0069
Approximation MSE_2	1.8930	1.8793	1.9455	1.9375	1.8460
Tracking MSE_3	0.0211	0.0220	0.0248	0.0281	0.0325

TABLE II
PERFORMANCE METRICS UNDER DIFFERENT COMMUNICATION DELAYS

Delay τ_i	**0.1**	**0.2**	**0.3**	**0.4**	**0.5**
Observer MSE_1	0.0160	0.0329	0.0546	0.0811	0.1162
Approximation MSE_2	4.2202	6.1261	7.6183	8.4901	9.0602
Tracking MSE_3	0.0289	0.0471	0.0692	0.0963	0.1336

initial weights for the FLS, actor, and critic networks are initialized to 0 or small values. The gain matrices are selected as $K_{i1} = 1.1I_3$, $K_{i3} = 3.6I_3$, and $K_{i4} = 2.4I_3$. The parameters determining K_{i2} are selected as $\varrho_{i1} = 2.1$, $\varrho_{i2} = 23.9$, and $\varrho_{i3} = 0.8$. The weighting matrix chosen as $Q_i = 1.9I_3$. The learning rates are set as $b_{i1} = 2.0$, $b_{i2} = 2.0$, $b_{i3} = 2.0$, $c_{i1} = 44.0$, and $c_{i2} = 21.0$. The desired formation displacements for the five USVs are defined as $d_1 = [0, 0, 0]^{\mathrm{T}}$, $d_2 = [4, 0, 0]^{\mathrm{T}}$, $d_3 = [2, 2, 0]^{\mathrm{T}}$, $d_4 = [-2, 2, 0]^{\mathrm{T}}$, $d_5 = [-4, 0, 0]^{\mathrm{T}}$. The initial positions of the USVs are set as $x_1(0) = [0, -1, 0]^{\mathrm{T}}$, $x_2(0) = [2, -1, 0]^{\mathrm{T}}$, $x_3(0) = [0.5, 0.5, 0]^{\mathrm{T}}$, $x_4(0) = [-1.9, 1.1, 0]^{\mathrm{T}}$, $x_5(0) = [-2.4, -0.9, 0]^{\mathrm{T}}$. The initial velocities are set to $v_i(0) = [0, 0, 0]^{\mathrm{T}}$ for all USVs.

The simulation results are presented in Figs. 3–10 and Tables I and II. Fig. 3 visualizes the actual trajectories of the five USVs together with the reference trajectory. Despite starting from widely dispersed initial positions, all agents rapidly align their motion and converge toward the reference trajectory under the proposed distributed control strategy. The USVs not only track the dynamic reference smoothly but also form and preserve the prescribed geometric formation pattern dictated by the displacement vectors d_i. This result qualitatively demonstrates the effectiveness of the proposed formation controller in handling second-order nonlinear dynamics and achieving coordinated motion.

Fig. 4 illustrates the temporal evolution of the relative distances between each USV and the reference trajectory. It can be observed that the formation distances for all USVs converge rapidly to their respective steady-state values within approximately 5 s, demonstrating the fast transient response achieved by the proposed control scheme.

To evaluate the overall performance of the proposed method, three MSE-based metrics are introduced to assess observer accuracy, dynamics approximation quality, and formation tracking error. The observer MSE_1 is defined as

$$\mathrm{MSE}_1 = \frac{1}{MNT} \sum_{k=1}^{M} \sum_{i=1}^{N} \int_0^T \|\hat{r}_i^{(k)}(t) - r(t)\|^2 \mathrm{d}t \tag{65}$$

where $\hat{r}_i^{(k)}(t)$ denotes the estimated reference trajectory in the kth run. The approximation MSE_2 is defined as

$$\mathrm{MSE}_2 = \frac{1}{MNT} \sum_{k=1}^{M} \sum_{i=1}^{N} \int_0^T \|\hat{f}_i^{(k)}(v_i) - f_i^{(k)}(v_i)\|^2 \mathrm{d}t \tag{66}$$

where $\hat{f}_i^{(k)}(v_i)$ is the approximated dynamics and $f_i^{(k)}(v_i)$ represents the actual dynamics in the kth run. The formation tracking MSE_3 is defined as

$$\mathrm{MSE}_3 = \frac{1}{MNT} \sum_{k=1}^{M} \sum_{i=1}^{N} \int_0^T \|x_i^{(k)}(t) - r(t) - d_i\|^2 \mathrm{d}t \tag{67}$$

where $x_i^{(k)}(t)$ represents the actual position in the kth run.

Table I summarizes the performance metrics of the system under varying disturbance intensities. As the disturbance level γ_i increases, the observer error MSE_1 and approximation error MSE_2 remain virtually invariant, and the tracking error MSE_3 shows only a marginal increase. These results confirm the exceptional robustness of the proposed control scheme, demonstrating the capacity to effectively suppress external disturbances and maintain high-precision formation tracking.

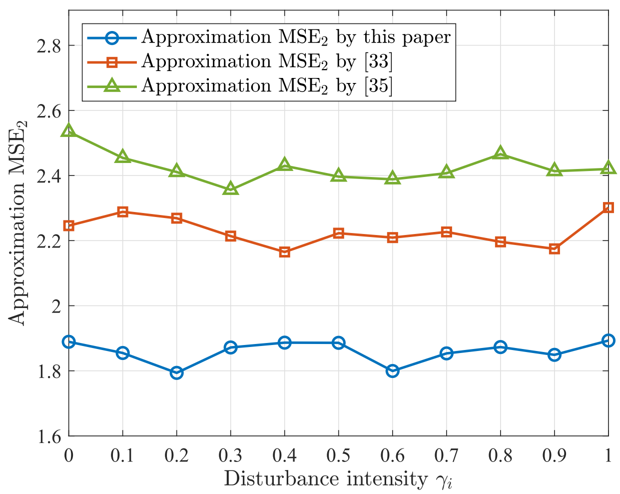

Fig. 5. Approximation MSE_2 under different disturbance intensities γ_i.

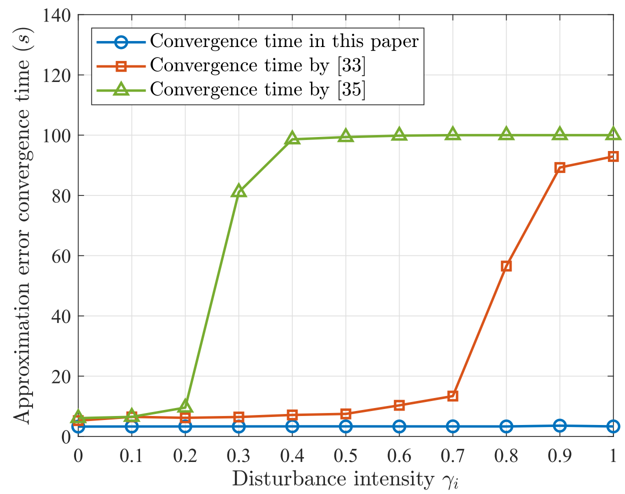

Fig. 6. Comparison of approximation error convergence time under different disturbance intensities γ_i.

Table II evaluates the performance metrics of the proposed control scheme under varying communication delays. For USVs operating under typical communication conditions like standard VHF radio networks [5] and maritime satellite links [28], delays between $\tau_i = 0.1$ s and 0.5 s reflect communication constraints ranging from mild to severe. Although the tracking error MSE_3 increases to 0.1336 as the delay τ_i grows, such a position deviation is sufficiently small compared to the prescribed formation displacements d_i, ensuring that the desired geometric configuration is maintained. These results substantiate the practical robustness of the proposed architecture, confirming that it can achieve effective formation tracking even under maritime communication constraints.

Fig. 5 presents a comparative analysis of the approximation MSE_2 against varying disturbance intensity coefficients $\gamma_i \in [0, 1.0]$. The proposed predictor-based fuzzy learning approach consistently achieves the lowest approximation error, whereas the methods reported in [33] and [35] exhibit substantially higher error levels, with the average error reduced by more than 16%. The observed performance superiority demonstrates that the proposed two-layer learning mechanism substantially improves the ability of the FLS to identify unknown nonlinear dynamics and suppress external perturbations.

Fig. 6 illustrates the convergence time of the approximation error under different disturbance intensities γ_i, where convergence is defined as the time required for the error to decrease below 0.1. The results show that the proposed predictor-based method exhibits strong robustness, consistently achieving a rapid convergence of approximately 3.6 s across all tested disturbance levels. In contrast, the performance of the comparison methods deteriorates markedly as the disturbance increases: the method in [35] is highly sensitive to perturbations and fails to meet the convergence criterion once $\gamma_i \geq 0.4$, and the method in [33] also experiences a substantial increase in convergence time when $\gamma_i \geq 0.6$. This result provides strong evidence that the proposed two-layer learning mechanism effectively enables rapid and high precision approximation of unknown nonlinear dynamics in complex environments.

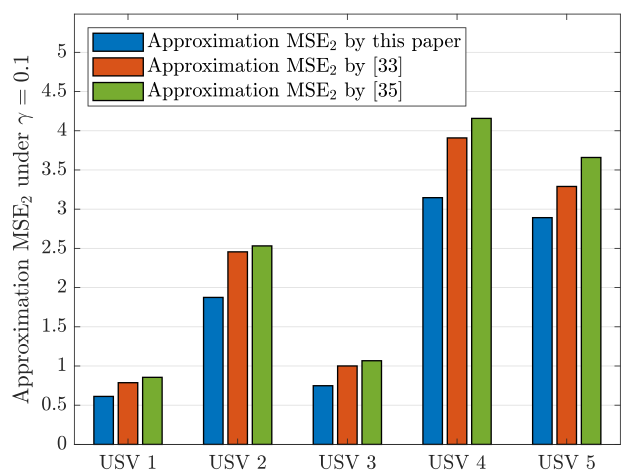

Fig. 7. Approximation MSE_2 for USVs under disturbance intensity $\gamma_i = 0.1$.

Fig. 7 presents the approximation MSE_2 for each individual USV under disturbance intensity $\gamma_i = 0.1$. The results show that the proposed predictor-based fuzzy learning strategy consistently achieves the lowest approximation error across all five USVs, reducing the MSE_2 by at least 12% relative to the methods in [33] and [35]. This uniform improvement across all agents highlights the effectiveness of the two-layer learning mechanism in enhancing the approximation accuracy of unknown nonlinear dynamics for every USV.

Fig. 8 illustrates the variation of the tracking MSE_3 with respect to the disturbance intensity $\gamma_i \in [0, 1.0]$. Across all disturbance levels, the proposed predictor-based fuzzy learning strategy maintains an approximately constant error of 0.02, whereas the methods in [36] and [37] produce markedly higher errors, resulting in an average tracking error reduction of more than 54%. Fig. 9 presents a comparative analysis of the tracking error convergence time under disturbance intensities varying from $\gamma_i = 0$ to 1.0, where convergence is defined as the time required for the tracking error to settle within 0.1. The proposed

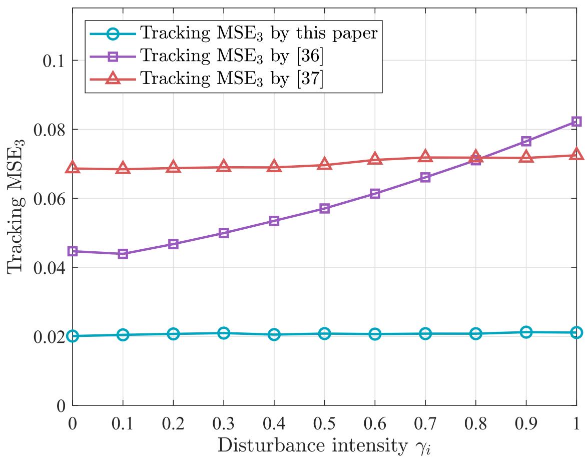

Fig. 8. Tracking MSE$_3$ under different disturbance intensities γ_i.

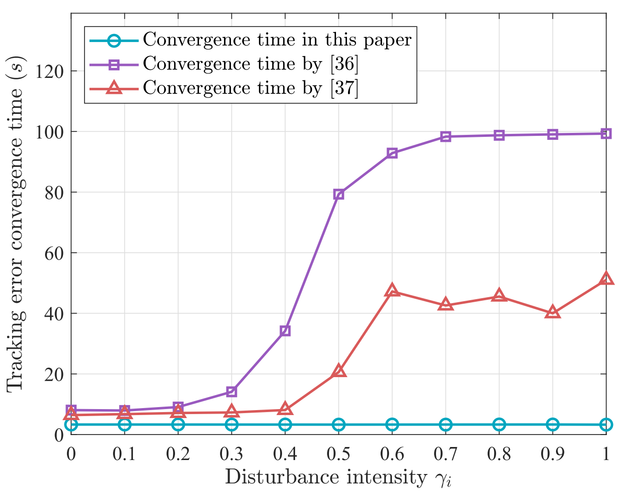

Fig. 9. Comparison of tracking error convergence time under different disturbance intensities γ_i.

method demonstrates strong robustness and efficiency, sustaining a consistently low convergence time of approximately 3.4 s across the full disturbance range. In contrast, the method in [36] fails to achieve convergence within the simulation horizon when $\gamma_i \geq 0.7$, and the approach in [37] becomes increasingly unstable, with convergence times fluctuating between 40 s and 50 s for $\gamma_i > 0.4$. From Figs. 8 and 9, the pronounced performance advantage indicates that the actor–critic learning control framework, augmented with the compensatory term, effectively alleviates the approximation burden of the critic network and ensures reliable tracking accuracy even under increasingly severe perturbations.

Fig. 10 illustrates the tracking MSE$_3$ for each individual USV under a disturbance intensity of $\gamma_i = 0.1$. The results indicate that the proposed formation control strategy consistently yields the lowest tracking errors across all five USVs, reducing the MSE$_3$ by at least 37% relative to the methods in [36] and [37]. This consistent performance advantage among all USVs confirms that incorporating the compensatory term into the actor–critic framework effectively optimizes the control policy for each

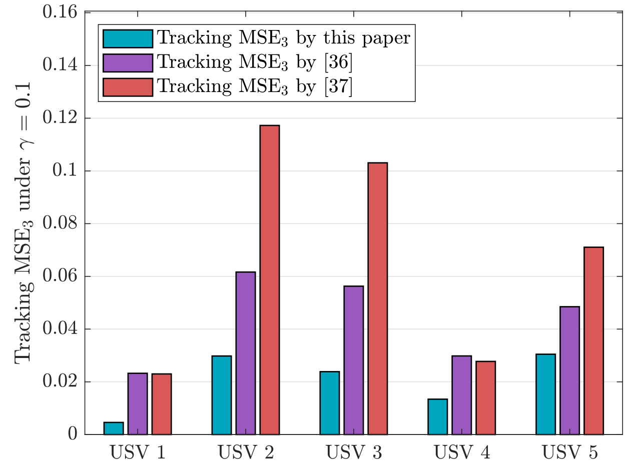

Fig. 10. Tracking MSE$_3$ for USVs under disturbance intensity $\gamma_i = 0.1$.

member of the formation, thereby ensuring precise formation maintenance.

V. Conclusion

This article has presented a novel distributed formation control architecture for second-order nonlinear MASs subject to unknown dynamics and limited communication. The core contribution lies in the seamless integration of a predictor-based fuzzy learning strategy with an actor–critic reinforcement learning framework. By leveraging velocity prediction errors, the proposed two-layer learning mechanism significantly accelerates the approximation of unknown nonlinearities. Furthermore, the introduction of a structural compensatory term reformulates the value function approximation into a residual learning task, thereby expediting policy convergence. Extensive simulations on USV formations confirm that the proposed method achieves superior tracking accuracy and rapid convergence even under high intensity disturbances.

References

[1] D. Jin and Z. Xiang, "Predefined-time consensus for second-order nonlinear multiagent systems via sliding mode technique," *IEEE Trans. Fuzzy Syst.*, vol. 32, no. 8, pp. 4534–4541, Aug. 2024.

[2] R. Zhang, S. Li, C. K. Ahn, and M. Chadli, "Fully distributed adaptive fuzzy consensus for heterogeneous switched nonlinear multiagent systems under state-dependent switchings," *IEEE Trans. Fuzzy Syst.*, vol. 32, no. 2, pp. 607–620, Feb. 2024.

[3] J. Zhang, Y. Fu, and J. Fu, "Optimal formation control of second-order heterogeneous multiagent systems using adaptive predefined-time strategy," *IEEE Trans. Fuzzy Syst.*, vol. 32, no. 4, pp. 2390–2402, Apr. 2024.

[4] P. Gong and C. Ki Ahn, "Robust adaptive fuzzy fixed-time distributed time-varying optimization in unbalanced nonlinear second-order multiagent networks," *IEEE Trans. Fuzzy Syst.*, vol. 33, no. 8, pp. 2730–2742, Aug. 2025.

[5] D. Palma, "Enabling the maritime Internet of Things: Coap and 6lowpan performance over vhf links," *IEEE Internet Things J.*, vol. 5, no. 6, pp. 5205–5212, Dec. 2018.

[6] L. Zha, J. Miao, J. Liu, E. Tian, and C. Peng, "Neural predictor-based formation control for unmanned surface vehicles under aperiodic dos attacks: A noncooperative game approach," *IEEE Trans. Veh. Technol.*, vol. 74, no. 10, pp. 15013–15025, Oct. 2025.

[7] J. Liu, S. Mao, L. Zha, E. Tian, and C. Peng, "Privacy-preserving and collision-free formation control for heterogeneous multiagent systems," *IEEE Trans. Consum. Electron.*, to be published, doi: 10.1109/TCE.2026.3667772.

[8] X. Jiang, L. You, N. Zhang, M. Chi, and H. Yan, "Impulsive formation tracking of nonlinear fuzzy multiagent systems with input saturation constraints," *IEEE Trans. Fuzzy Syst.*, vol. 32, no. 10, pp. 5728–5736, Oct. 2024.

[9] J. Liu, Z. Liu, Y. Li, E. Tian, and C. Peng, "Privacy-protected formation control for unmanned surface vehicles: A neural predictor-based noncooperative game framework," *IEEE Trans. Veh. Technol.*, to be published, doi: 10.1109/TVT.2026.3651837.

[10] S. Sui, D. Shen, S. Tong, and C. L. P. Chen, "State-observer-based adaptive fuzzy event-triggered formation control for nonlinear multiagent system," *IEEE Trans. Emerg. Topics Comput. Intell.*, vol. 8, no. 5, pp. 3327–3338, Oct. 2024.

[11] W. Wu and S. Tong, "Collision-free adaptive fuzzy formation control for stochastic nonlinear multiagent systems," *IEEE Trans. Syst., Man, Cybern. Syst.*, vol. 53, no. 9, pp. 5454–5465, Sep. 2023.

[12] M. Xu, X. Chen, and F. Hao, "Scalable formation control for second-order multiagent systems: An event-triggered predefined-time strategy," *IEEE Trans. Syst., Man, Cybern. Syst.*, vol. 55, no. 2, pp. 1466–1477, Feb. 2025.

[13] K. Luo, B. Zhang, and W. Meng, "Formation control of second-order multiagent system via stochastic control input," *IEEE Trans. Ind. Informat.*, vol. 20, no. 4, pp. 5550–5561, Apr. 2024.

[14] C. Gu, X. Wu, W.-A. Zhang, H. Ni, and S. X. Ding, "Active disturbance rejection formation control for multiagent systems with input constraints," *IEEE Trans. Control Netw. Syst.*, vol. 12, no. 1, pp. 38–50, Mar. 2025.

[15] J. Zhang, Y. Fu, and J. Fu, "Adaptive finite-time optimal formation control for second-order nonlinear multiagent systems," *IEEE Trans. Syst., Man, Cybern. Syst.*, vol. 53, no. 10, pp. 6132–6144, Oct. 2023.

[16] J. Lan, Y.-J. Liu, D. Yu, G. Wen, S. Tong, and L. Liu, "Time-varying optimal formation control for second-order multiagent systems based on neural network observer and reinforcement learning," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 35, no. 3, pp. 3144–3155, Mar. 2024.

[17] Y. Huang, Z. Meng, and J. Sun, "Distributed constrained optimization for second-order multiagent systems via event-based communication," *IEEE Trans. Syst., Man, Cybern. Syst.*, vol. 54, no. 9, pp. 5317–5326, Sep. 2024.

[18] W. Bi, C. Zhang, S. Sui, S. Tong, and C. L. P. Chen, "Fixed-time fuzzy adaptive bipartite output consensus tracking for nonlinear coopetition mass," *IEEE Trans. Fuzzy Syst.*, vol. 32, no. 5, pp. 2775–2785, May 2024.

[19] Y. Yan, H. Zhang, J. Sun, and Y. Wang, "Sliding mode control based on reinforcement learning for T-S fuzzy fractional-order multiagent system with time-varying delays," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 35, no. 8, pp. 10368–10379, Aug. 2024.

[20] L. Zha, T. Huang, J. Liu, E. Tian, X. Xie, and C. Peng, "Secure state estimation for interval type-2 fuzzy systems with FDI attacks and event-triggered WTOD protocol," *IEEE Trans. Syst., Man, Cybern. Syst.*, vol. 55, no. 8, pp. 5320–5331, Aug. 2025.

[21] J. Wang, J. Liu, Y. Li, C. L. P. Chen, Z. Liu, and F. Li, "Prescribed time fuzzy adaptive consensus control for multiagent systems with dead-zone input and sensor faults," *IEEE Trans. Autom. Sci. Eng.*, vol. 21, no. 3, pp. 4016–4027, Jul. 2024.

[22] Y. Xu, H. Liang, T. Li, Y. Long, Y. Cheng, and D. Wang, "Adaptive fuzzy resilient control of nonlinear multiagent systems under DoS attacks: A dynamic event-triggered method," *IEEE Trans. Fuzzy Syst.*, vol. 32, no. 6, pp. 3568–3580, Jun. 2024.

[23] H. Shen, W. Zhao, J. Cao, J. H. Park, and J. Wang, "Predefined-time event-triggered tracking control for nonlinear servo systems: A fuzzy weight-based reinforcement learning scheme," *IEEE Trans. Fuzzy Syst.*, vol. 32, no. 8, pp. 4557–4569, Aug. 2024.

[24] J. Wang, J. Wu, H. Shen, J. Cao, and L. Rutkowski, "Fuzzy H_∞ control of discrete-time nonlinear Markov jump systems via a novel hybrid reinforcement Q-learning method," *IEEE Trans. Cybern.*, vol. 53, no. 11, pp. 7380–7391, Nov. 2023.

[25] H.-Y. Zhu, Y.-X. Li, and S. Tong, "Fuzzy adaptive tracking of constrained nonlinear systems with event-sampling reinforcement learning," *IEEE Trans. Fuzzy Syst.*, vol. 32, no. 2, pp. 536–546, Feb. 2024.

[26] Q. Fu, Z. Pu, Y. Pan, T. Qiu, and J. Yi, "Fuzzy feedback multiagent reinforcement learning for adversarial dynamic multiteam competitions," *IEEE Trans. Fuzzy Syst.*, vol. 32, no. 5, pp. 2811–2824, May 2024.

[27] J. Liu, N. Zhang, Y. Li, X. Xie, E. Tian, and J. Cao, "Learning-based event-triggered tracking control for nonlinear networked control systems with unmatched disturbance," *IEEE Trans. Syst., Man, Cybern. Syst.*, vol. 53, no. 5, pp. 3230–3240, May 2023.

[28] D. Agnew, A. Rice-Bladykas, and J. Mcnair, "Detection of zero-day attacks in a software-defined leo constellation network using enhanced network metric predictions," *IEEE Open J. Commun. Soc.*, vol. 5, pp. 6611–6634, 2024, doi: 10.1109/OJCOMS.2024.3481965.

[29] Y. Meng, C. Liu, Q. Wang, and L. Tan, "Cooperative advantage actor–critic reinforcement learning for multiagent pursuit-evasion games on communication graphs,"," *IEEE Trans. Artif. Intell.*, vol. 5, no. 12, pp. 6509–6523, Dec. 2024.

[30] Q. Liu, H. Yan, M. Wang, Z. Li, and S. Liu, "Data-driven optimal bipartite consensus control for second-order multiagent systems via policy gradient reinforcement learning," *IEEE Trans. Cybern.*, vol. 54, no. 6, pp. 3468–3478, Jun. 2024.

[31] L. Chen, S.-L. Dai, and C. Dong, "Adaptive optimal tracking control of an underactuated surface vessel using actor–critic reinforcement learning," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 35, no. 6, pp. 7520–7533, Jun. 2024.

[32] J. Lan, Y.-J. Liu, T. Xu, S. Tong, and L. Liu, "Adaptive fuzzy fast finite-time formation control for second-order MASs based on capability boundaries of agents," *IEEE Trans. Fuzzy Syst.*, vol. 30, no. 9, pp. 3905–3917, Sep. 2022.

[33] K. Li, K. Feng, and Y. Li, "Fuzzy adaptive fault-tolerant formation control for USVs with intermittent actuator faults," *IEEE Trans. Intell. Veh.*, vol. 9, no. 3, pp. 4445–4455, Mar. 2024.

[34] G. Wen, W. Hao, W. Feng, and K. Gao, "Optimized backstepping tracking control using reinforcement learning for quadrotor unmanned aerial vehicle system," *IEEE Trans. Syst., Man, Cybern. Syst.*, vol. 52, no. 8, pp. 5004–5015, Aug. 2022.

[35] R. Lu et al., "Fuzzy actor–critic reinforcement learning for unmanned surface vessels flexible tracking formation control," *IEEE Trans. Fuzzy Syst.*, vol. 33, no. 10, pp. 3666–3680, Oct. 2025.

[36] N. Wang, Y. Gao, and X. Zhang, "Data-driven performance-prescribed reinforcement learning control of an unmanned surface vehicle," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 32, no. 12, pp. 5456–5467, Dec. 2021.

[37] N. Wang, Y. Gao, H. Zhao, and C. K. Ahn, "Reinforcement learning-based optimal tracking control of an unknown unmanned surface vehicle," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 32, no. 7, pp. 3034–3045, Jul. 2021.

Jinzhao Miao received the B.S. and M.S. degrees in computer science and technology from the Nanjing University of Finance and Economics, Nanjing, China, in 2020 and 2025, respectively. He is currently working toward the Ph.D. degree in computer science and technology with the School of Computer, Nanjing University of Information Science and Technology, Nanjing, China.

His research interests include networked control systems, reinforcement learning, and neural networks.

Jinliang Liu (Senior Member, IEEE) received the Ph.D. degree in control theory and control engineering from the School of Information Science and Technology, Donghua University, Shanghai, China, in 2011.

He was a Postdoctoral Research Associate with the School of Automation, Southeast University, Nanjing, China, from 2013 to 2016. He was a Visiting Researcher/Scholar with the Department of Mechanical Engineering, University of Hong Kong, Hong Kong, from 2016 to 2017. He was a Visiting Scholar with the Department of Electrical Engineering, Yeungnam University, Gyeongsan, South Korea, from 2017 to 2018. From 2011 to 2023, he was an Associate Professor and then a Professor with the Nanjing University of Finance and Economics, Nanjing, China. In 2023, he joined the Nanjing University of Information Science and Technology, Nanjing, China, where he is currently a Professor with the School of Computer Science. His research interests include cyber-physical systems, uncrewed systems, privacy preservation, and game theory.

Lijuan Zha received the Ph.D. degree in control science and engineering from Donghua University, Shanghai, China, in 2018.

From 2017 to 2024, she was a Lecturer and then an Associate Professor with the College of Information Engineering, Nanjing University of Finance and Economics, Nanjing, China. In 2024, she joined Nanjing Forestry University, Nanjing, China, where she is currently a Professor with the College of Science. From 2018 to 2023, she was a Postdoctoral Research Associate with the School of Mathematics, Southeast University, Nanjing. Her current research interests include networked control systems, neural networks, complex dynamical systems, and multiagent systems.

Engang Tian (Senior Member, IEEE) received the B.S. degree in mathematics from Shandong Normal University, Jinan, China, in 2002, the M.Sc. degree in operations research and cybernetics from Nanjing Normal University, Nanjing, China, in 2005, and the Ph.D. degree in control theory and control engineering from Donghua University, Shanghai, China, in 2008.

From 2011 to 2012, he was a Postdoctoral Research Fellow with the Hong Kong Polytechnic University, Hong Kong. From 2015 to 2016, he was a Visiting Scholar with the Department of Information Systems and Computing, Brunel University London, Uxbridge, U.K. From 2008 to 2018, he was an Associate Professor and then a Professor with the School of Electrical and Automation Engineering, Nanjing Normal University. In 2018, he was an Eastern Scholar by the Municipal Commission of Education, Shanghai, and joined the University of Shanghai for Science and Technology, Shanghai, China, where he is currently a Professor with the School of Optical-Electrical and Computer Engineering. He has authored or coauthored more than 100 papers in refereed international journals. His research interests include networked control systems, cyber attack, as well as nonlinear stochastic control and filtering.

Chen Peng (Senior Member, IEEE) received the B.S. and M.S. degrees in coal preparation and the Ph.D. degree in control theory and control engineering from the Chinese University of Mining Technology, Xuzhou, China, in 1996, 1999, and 2002, respectively.

From 2002 to 2004, he was a Postdoctoral Research Fellow in Applied Math with Nanjing Normal University, Nanjing, China. From 2004 to 2005, he was a Research Associate with The University of Hong Kong, Hong Kong. From 2006 to 2007, he was a Visiting Scholar with the Queensland University of Technology, Brisbane, QLD, Australia. From 2010 to 2012, he was a Postdoctoral Research Fellow with Central Queensland University, Rockhampton, QLD, Australia. He is currently a Professor with the School of Mechatronic Engineering and Automation, Shanghai University, Shanghai, China. His current research interests include networked control systems, multiagent systems, power systems, and interconnected systems.

Dr. Peng is currently an Associate Editor for a number of international journals, including the IEEE TRANSACTIONS ON INDUSTRIAL INFORMATICS, *Information Sciences*, and *Transactions of the Institute of Measurement and Control*. He was named a Highly Cited Researcher in 2020, 2021, and 2022 by Clarivate Analytics.