Research Article

Intelligent *L2-L∞* Consensus of Multiagent Systems under Switching Topologies via Fuzzy Deep *Q* Learning

Haoyu Cheng ⓘ,[1] **Linpeng Xu,[2,3] Ruijia Song,[4] Yue Zhu,[2,5] and Yangwang Fang[1]**

[1]*Unmanned System Research Institute, Northwestern Polytechnical University, Xi'an, China*
[2]*School of Marine and Technology, Northwestern Polytechnical University, Xi'an, China*
[3]*China Airborne Missile Academy, Luoyang 471009, China*
[4]*School of Astronautics, Northwestern Polytechnical University, Xi'an, China*
[5]*Beijing Institute of Remote Sensing Equipment, Beijing, China*

Correspondence should be addressed to Haoyu Cheng; chenghaoyu@nwpu.edu.cn

The problem of intelligent L_2-L_∞ consensus design for leader-followers multiagent systems (MASs) under switching topologies is investigated based on switched control theory and fuzzy deep Q learning. It is supposed that the communication topologies are time-varying, and the model of MASs under switching topologies is constructed based on switched systems. By employing linear transformation, the problem of consensus of MASs is converted into the issue of L_2-L_∞ control. The consensus protocol is composed of the dynamics-based protocol and learning-based protocol, where the robust control theory and deep Q learning are applied for the two parts to guarantee the prescribed performance and improve the transient performance. The multiple Lyapunov function (MLF) method and mode-dependent average dwell time (MDADT) method are combined to give the scheduling interval, which ensures stability and prescribed attenuation performance. The sufficient existing conditions of consensus protocol are given, and the solutions of the dynamics-based protocol are derived based on linear matrix inequalities (LMIs). Then, the online design of the learning-based protocol is formulated as a Markov decision process, where the fuzzy deep Q learning is utilized to compensate for the uncertainties and achieve optimal performance. The variation of the learning-based protocol is modeled as the external compensation on the dynamics-based protocol. Therefore, the convergence of the proposed protocol can be guaranteed by employing the nonfragile control theory. In the end, a numerical example is given to validate the effectiveness and superiority of the proposed method.

1. Introduction

In recent years, the coordination control of MASs has attracted considerable attention for their broad applications in many fields [1, 2], such as formation control, cooperative attack, and attitude alignment. The MAS consists of a series of agents, which can communicate and interact with each other to realize multiple missions and adapt to the complex environment [3, 4]. In particular, much attention has been paid to the problem of consensus of MASs because of their great potential applications in both economic and military. The purpose of MASs is to construct a relationship between the agents to achieve an agreement for the state/output. In the past decades, fruitful research studies have emerged to contribute to the development in theory and applications. To mention a few, the problem of distributed formation control for MASs is studied in [5], the time-varying formation design for MASs with disturbances is proposed in [6], and the problem of finite-time consensus for switched nonlinear MASs is investigated in [7].

In practical applications, it is well known that the communication topology among the agents may change dramatically over time to adjust to multiple missions and complex environments [8, 9], such as the MASs can realize obstacle avoidance and higher flight efficiency by formation transformation [10, 11]. The design flexibility, security, and performance of convergence will be improved, which motivated the studies on the switching topologies of MASs

[1, 12]. Recently, because of the broad potential applications of switching topologies, considerable significant research studies have been proposed by scholar at home and abroad. The communication topologies among interacting agents will change according to the flight conditions and missions, which can be modeled as switched systems. The switched systems consist of a series of continuous-time (or discrete-time) subsystems and a switching signal, which determines the switching strategy between subsystems. It provides an efficient approach to deal with the problem of fast time-varying conditions. Therefore, it can be inferred that the switching of topologies can be viewed as the switching between subsystems, and it is essential to study the problem of consensus protocol design to make sure the state/output can converge to the given value. In [13], the problem of time-varying formation control of MASs is investigated. The communication topologies switching among given connected topologies and the switching signal depend on the Markovian process. The Lyapunov function method is utilized to analyze the convergence. In the work of [14], the problem of event-triggered leader-following consensus problem for multiagent systems with external disturbances is addressed under switching topologies. A novel distributed event-triggered protocol is proposed to realize disturbance rejection based on extended state observer. The average dwell time (ADT) method is utilized to ensure the stability of the event-triggered protocol. In [15], the time-varying practical formation problem is studied for spacecraft, where switching topologies and time-delays are taken into consideration. Sufficient conditions are provided to ensure that the error system is convergent, which are derived based on the ADT method. It is well known that the research studies mentioned above are proposed to deal with the problem of switching topologies. However, the convergence is guaranteed based on the ADT method. It can be inferred that the common parameters are applied for all subsystems in the ADT method, which will lead to conservativeness. To obtain tighter bounds on dwell time and improve the design flexibility of the algorithm, MDADT is applied during last decades. In [16], the MDADT method and multiple discontinuous Lyapunov function (MDLF) method are combined to analyze the stability of switched systems with unstable modes. The sufficient conditions are established, and the results in existing literature are covered as a special case. The fast switching and slow switching in the framework of MDADT are applied to unstable modes and stable modes. In [17], the global adaptive control algorithm for switched systems is proposed based on the MDADT method. The different properties of subsystems are taken into consideration. Then, the adaptive tracking controller is applied to the nonlinear switched systems with external disturbance and unmodeled dynamics, which illustrates the effectiveness and superiority of the MDADT method. In the work of [18], the event-triggered sliding mode controller is proposed. By employing the MDADT method and event-triggered strategy, less conservative and more practical results are obtained. Sufficient conditions are given to ensure stochastically exponential stability by the aid of the LMI technique. The literature mentioned has provided fruitful

results on consensus protocol design for MASs under switching topologies. However, stability and convergence are ensured by the traditional ADT method. The different properties of subsystems cannot be considered, which will lead to conservativeness. Therefore, how to obtain less restrictive results is still an open and challenging problem, which has been fully investigated, and it has an important value and potential applications in practice.

Moreover, in practical environment, there always exist uncertainties and disturbances, which will lead to performance degradation and even instability [19, 20]. Therefore, it is essential to investigate the robust consensus problem to improve the performance in the uncertain environment [21–23]. In the work of [24], the problem of distributed $H\infty$ containment control for MASs with switching topologies is studied. An observer-based containment control scheme is proposed. The external disturbance and time delay in the environment are taken into consideration, which is more applicable than the traditional method. By employing the Lyapunov function method and LMIs technique, the sufficient existing conditions and solutions of control protocol are given in the form of LMIs. In [25], the problem of time-varying formation of second-order discrete-time MASs under switching topologies and the time delay is investigated. The sufficient conditions are given to ensure MASs accomplish the mission of time-varying formation based on the state transformation method. The time delay and uncertainties are considered. Compared with the existing literature, the proposed can overcome the undesirable response caused by time delay and improve the transient performance. In the work of [26], the problem of formation control for tail-sitters in flight mode transitions is studied. The nonlinear dynamics and uncertainties are considered, and the robust time-varying formation control protocol is proposed. It is proven that the tracking errors can converge to the origin in finite time. The problem of L_2-gain robust protocol for time-varying output formation-containment of MASs is addressed in [27]. The PID-based output-feedback control protocol is provided to ensure that all followers can track a time-varying formation reference, where communication delays and external disturbance are taken into consideration. The asymptotic stability of MASs is proved by the Lyapunov function method. However, as well known, the transient performance and robustness cannot achieve simultaneously. Therefore, we need to make comprise of the transient performance and robustness, which still remains an open and challenging problem.

In addition, with the development of computing ability, the intelligent technique has been an attractable problem during the last decades [28–30]. It is widely applied in the areas of target recognition, machine vision, robotic systems, and controller design [31, 32]. It provides an efficient method to improve the autonomy and design flexibility of the system [33]. The most widely used methods are the deep learning and reinforcement learning. As a combination of deep learning and reinforcement learning, the advantages of deep learning and reinforcement learning are adopted, which include the characteristics of self-fitting and self-learning. In the work of [34], the automatic completion of

multiple peg-in-hole assemble tasks is realized. Because the traditional method requires an accurate contact model and complex analysis, the intelligent control method is formulated by constructing the task as a Markov decision process. The deep deterministic policy gradient (DDPG) algorithm is proposed to accomplish the task to achieve optimal policy and avoid risky actions. In [35], a noninteger PID controller is proposed based on the DDPG algorithm. The measurement noises and external disturbances are taken into consideration. The kinematic controller and dynamic controller are proposed to achieve optimal performance. The DDPG algorithm is given to compensate for the uncertainties and disturbances in the framework of actor-critic. A numerical example is given to illustrate the effectiveness of the proposed method. Cheng et al. [36] proposed the real-time controller for the problem of fuel-optimal moon landing. Because the traditional method cannot meet the demand of high requirements of real-time performance and autonomy, the deep reinforcement learning algorithm is proposed for the real-time optimal control based on actor-indirect method architecture. The deep neural networks are applied for initial guesses, and the efficiency of training data is guaranteed. The literature mentioned above has provided considerable meaningful results in the area of machine learning. However, to the best of the authors' knowledge, the intelligent consensus design for MASs with considerations of stability, robustness, and optimal transient performance has not been fully studied yet. It is essential and important to achieve optimal comprise of robustness and transient performance.

Based on the statement above, it can be inferred that the problem of the improvement of autonomy and design flexibility for the system needs to be studied. The problem of consensus protocol design for MASs under switching topologies has not been fully investigated yet. The design flexibility can be improved by employing tighter bounds on dwell time because less conservative results can be obtained, and it leaves more room to ensure the switching logic stays in the subsystems with better performance for long enough time. Moreover, it is of great importance to combine the advantages of the traditional method and intelligent technique, which can ensure convergence, robustness, and transient performance simultaneously. Therefore, the problem of intelligent L_2-L_∞ consensus design of MASs under switching topologies is investigated. The convergence and robustness are guaranteed by the Lyapunov function method and the MDADT method, which are more applicable. The transient performance is improved by fuzzy deep Q learning, in which the fuzzy reward function is proposed for the complex scheduling process. The main contributions of this study can be summarized as follows:

(1) The L_2-L_∞ consensus protocol of MASs under switching topologies is designed. The problem of L_2-L_∞ consensus of MASs is converted into the problem of stability analysis for switched systems, which is more applicable than the traditional method. The MDADT method and multiple Lyapunov function

method are combined to guarantee the stability and prescribed attenuation performance index, which can obtain tighter bounds on dwell time and less conservative results.

(2) The consensus protocol is composed of the dynamics-based consensus protocol and learning-based consensus protocol. Compared with the traditional method, the proposed strategy can ensure the stability, robustness, and transient performance simultaneously.

(3) The fuzzy reward function is utilized to improve the efficiency of the deep reinforcement learning algorithm. The design of reward function for the traditional method mainly depends on the experience of designer, which will lead to complexity. The fuzzy reward function can improve the data efficiency and ensure optimal performance.

The rest of the study is organized as follows: the preliminaries and problem statement are provided in Section 2; in Section 3, the main results of the study are given; the numerical example is given in Section 4, which is followed by the conclusion in Section 5.

2. Preliminaries and Problem Statement

In this study, it is supposed that MASs are composed of a leader labelled as 0 and n followers labelled as 1, 2, ..., n. The connection topology among n followers can be described as a time-varying model with N_f topologies. We define $\mathscr{G}_{\sigma(k)} = (\mathscr{G}_1, \mathscr{G}_2, \ldots, \mathscr{G}_{N_f})$ as undirected connected graph, respectively. $\mathscr{H} = (1, 2, \ldots, n), n > 1$ represents the set of finite nodes. $s = \sigma(k): [0, \infty) \longrightarrow R = \{1, 2, \ldots, N_f\}$ denotes the switching signal, which is a piecewise continuous function of time and takes value in the finite set $\mathscr{H}$. $\mathscr{A}_{\sigma(k)} = (a_{ij}^{\sigma(k)})_{n \times n}$ and $\mathscr{L}_{\sigma(k)} = (l_{ij}^{\sigma(k)})_{n \times n}$ are the adjacency matrices of the undirected graph $\mathscr{G}_{\sigma(k)}$ and the Laplacian matrix at time instant k, where $a_{ij}^{\sigma(k)}$ stands for the element of adjacency matrix, where $a_{ij}^{\sigma(k)} = 1$ represents that the node i can obtain information from node j, and $l_{ij}^{\sigma(k)}$ is defined in the following equation.

$$\begin{cases} l_{ij}^{\sigma(k)} = -a_{ij}^{\sigma(k)} & i \neq j, \\ l_{ii}^{\sigma(k)} = \sum_{j=1}^{n} a_{ij}^{\sigma(k)} & i = j. \end{cases} \tag{1}$$

Then, for given node $i \in \mathscr{H}$, we can define the neighbors of node i as $\mathscr{N}_{i,\sigma(k)} = \{j \in \mathscr{H}: a_{ij}^{\sigma(k)} = 1\}$.

Another undirected connected graph is defined as $\widetilde{\mathscr{G}}_{\sigma(k)} = (\widetilde{\mathscr{G}}_1, \widetilde{\mathscr{G}}_2, \ldots, \widetilde{\mathscr{G}}_{N_f})$ to indicate the information transformation between the leader and the followers with n nodes. Define a diagonal matrix $\Theta_{\sigma(k)} = \text{diag}\{\theta_1^{\sigma(k)}, \theta_2^{\sigma(k)}, \ldots, \theta_n^{\sigma(k)}\}$, where $\theta_i^{\sigma(k)} = 1$ stands for that the node $i \in \mathscr{H}$ can obtain information from the leader; otherwise, we define $\theta_i^{\sigma(k)} = 0$.

Therefore, MASs with leader-followers can be described as in the following equations:

$$x_0(k+1) = Ax_0(k), \tag{2}$$

$$\begin{cases} x_i(k+1) = Ax_i(k) + Bu_i(k) + D\omega_i(k) \\ z_i(k) = Cv_i(k) \end{cases}, i \in \mathscr{H}, \tag{3}$$

where A, B, C, and D are the system matrices with appropriate dimensions, $x_0(k) = [x_{01}(k), x_{02}(k), \ldots, x_{0p}(k)]^T \in R^p$ represents the state vector of leader, $x_i(k) = [x_{i1}(k), x_{i2}(k), ..., x_{ip}(k)]^T \in R^p$ is the state of the ith follower, $u_i(k) = [u_{i1}(k), u_{i2}(k), \ldots, u_{il}(k)]^T \in R^l$ is the input of the ith follower, $z_i(k) = [z_{i1}(k), z_{i2}(k), \ldots, z_{iq}(k)]^T \in R^q$ stands for the output of the ith follower, and $\omega_i(k) \in R^m$ denotes the external disturbance belonging to $L_2[0, \infty)$. It is supposed that the agent i can obtain information from its neighbors and leader. Therefore, we define $v_i(k)$ as relative state measurements of the ith agent, which can be described as follows:

$$v_i(k) = \sum_{j \in \mathscr{N}_{i,\sigma(k)}} \left(x_i(k) - x_j(k) \right) + \theta_i^{\sigma(k)} \left(x_0(k) - x_i(k) \right). \tag{4}$$

In this study, the control input of the ith agent to ensure the consensus of leader-followers is proposed.

$$u_i(k) = \left(K_{\sigma(k)} + K_{c,\sigma(k)} \right) v_i(k), \tag{5}$$

where $K_{\sigma(k)}$ is the control parameter to be determined by robust control theory, and $K_{c,\sigma(k)}$ is the compensated parameter obtained by deep Q learning. In this study, the gained parameters $K_{c,\sigma(k)}$ are supposed to vary in a finite set with given bounds. The $K_{c,\sigma(k)}$ can be viewed as additional perturbation of $K_{\sigma(k)}$, which can be described as follows:

$$K_{c,\sigma(k)} = M_{\sigma(k)} F_{\sigma(k)} N_{\sigma(k)}, \tag{6}$$

where $M_{\sigma(k)} \in R^{l \times l_\Delta}$ and $N_{\sigma(k)} \in R^{q_\Delta \times q}$ are the known matrices with appropriate dimensions, and $F_{\sigma(k)} \in R^{l_\Delta \times q_\Delta}$ are the unknown matrices with $F_{\sigma(k)}^T F_{\sigma(k)} \leq I$.

For the ith agent, the error of state is defined as $e_i(k) = x_i(k) - x_0(k)$. Then, the closed-loop system can be rewritten as

$$\begin{cases} e(k+1) = \tilde{A}_{\sigma(k)} e(k) + \omega \tilde{B}(k), \\ z(k) = \tilde{C}_{\sigma(k)} e(k), \end{cases} \tag{7}$$

where $e(k) = [e_1^T(k), \ldots, e_n^T(k)]^T$, $z(k) = [z_1^T(k), \ldots, z_n^T(k)]^T$, $\omega(k) = [\omega_1^T(k), \ldots, \omega_n^T(k)]^T$, $\tilde{A}_{\sigma(k)} = I_n \otimes A - (\mathscr{L}_{\sigma(k)} + \Theta_{\sigma(k)}) \otimes B(K_{\sigma(k)} + K_{c,\sigma(k)})$, $\tilde{B} = I_n \otimes D$, and $\tilde{C}_{\sigma(k)} = -(\mathscr{L}_{\sigma(k)} + \Theta_{\sigma(k)}) \otimes C$.

To facilitate the proof, the definitions and lemmas are given as follows.

Definition 1 (see [37]). For given switching signal $\sigma(k)$ and $k_1 > 0$, define $N_{\sigma,s}(0, k_1)$ as the number of switching instants over the time interval $(0, k_1)$. $T_{\sigma s}(0, k_1)$ is set to be the activated time of undirected graph $\mathscr{G}_s$ during $(0, k_1)$. There exist constant scalars $N_0 \geq 0$ and $\tau_{as} > 0$, such that

$$N_{\sigma,s}(0, k_1) \leq N_0 + \frac{T_{\sigma s}(0, k_1)}{\tau_{as}}. \tag{8}$$

Then, τ_{as} is called the mode-dependent average dwell time and $N_{\sigma,s}(0, k_1)$ is the mode-dependent chatter bound, respectively. In this study, we set $N_0 = 0$.

Definition 2 (see [37]). If there exist control protocol in equation (5), all agents asymptotically track the state trajectory of the leader, such that

$$\lim_{k \to \infty} \|x_i(k) - x_0(k)\|^2 = 0, \tag{9}$$

for any initial conditions $x_i(0), i \in \mathscr{H}$

Definition 3 (see [38]). For given constant scalars $0 < \delta < 1$ and $\gamma > 0$, the prescribed $L_2 - L_\infty$ attenuation performance γ is satisfied such that

(1) The MASs in equations (2)-(3) are asymptotically stable when $\omega(k) = 0$.

(2) The following inequation holds for all nonzero $\omega(k) \in l_2(0, \infty]$.

$$\|z(r)\|_{L_\infty}^2 \leq \gamma^2 \|\omega(r)\|_{L_2}^2. \tag{10}$$

Lemma 1 (see [35]). *The matrices $\mathscr{L}_{\sigma(k)} + \Theta_{\sigma(k)}$ are symmetric and positive definite if and only if the graphs $\tilde{\mathscr{G}}_{\sigma(k)}$ are connected for $t \geq 0$. Moreover, there exist a transformation matrix $T_{\sigma(k)}$, such that the following equation holds.*

$$T_{\sigma(k)} \left(\mathscr{L}_{\sigma(k)} + \Theta_{\sigma(k)} \right) T_{\sigma(k)}^{-1} = Z_{\sigma(k)} = \text{diag}\left(\lambda_1^{\sigma(k)}, \ldots, \lambda_n^{\sigma(k)} \right), \tag{11}$$

where $\lambda_i^{\sigma(k)}, i \in \mathscr{H}$ are the nonzero eigenvalues of matrices $\mathscr{L}_{\sigma(k)} + \Theta_{\sigma(k)}$.

Lemma 2 (see [39]). *For given constant $a > 0$ and real matrices Θ, U, V, W, it is concluded that equation (12) is equivalent to equation (13).*

$$\begin{bmatrix} \Theta & U + aV \\ * & -a(W + W^T) \end{bmatrix} < 0, \tag{12}$$

$$\Theta + UW^{-1}V^T + VW^{-T}U^T < 0. \tag{13}$$

Lemma 3 (see [39]). *For given symmetric matrix $\mathscr{T}$ and matrices $\mathscr{M}, \mathscr{N}$, if there exist constant scalar $\varepsilon > 0$, such that*

$$\mathscr{T} + \mathscr{M}\mathscr{M}^T \varepsilon^{-1} + \varepsilon \mathscr{N}^T \mathscr{N} < 0 \tag{14}$$

Then, the following equation holds for any appropriate $\mathscr{F}$ with $\mathscr{F}^T \mathscr{F} \leq I$.

$$\mathscr{T} + \mathscr{M}\mathscr{F}\mathscr{N} + \mathscr{N}^{\mathscr{T}} \mathscr{F}^{\mathscr{T}} \mathscr{M} < 0. \tag{15}$$

3. Main Results

3.1. L_2-L_∞ Consensus Protocol Design.

In this section, the L_2-L_∞ consensus protocol is proposed, and the stability and prescribed performance are guaranteed.

$$\begin{cases} \widetilde{e}(k+1) = \left(I_n \otimes A - Z_{\sigma(k)} \otimes B\left(K_{\sigma(k)} + \Delta K_{\sigma(k)}\right)\right)\widetilde{e}(k) + \left(I_n \otimes D\right)\widetilde{\omega}(k), \\ \widetilde{z}(k) = -\left(Z_{\sigma(k)} \otimes C\right)\widetilde{e}(k), \end{cases} \tag{16}$$

where

$$\begin{aligned} \widetilde{\omega}(k) &= \left(T_{\sigma(k)} \otimes I_l\right)\omega(k), \\ \widetilde{z}(k) &= \left(T_{\sigma(k)} \otimes I_q\right)z(k), \\ \widetilde{e}(k) &= \left(T_{\sigma(k)} \otimes I_p\right)e(k). \end{aligned} \tag{17}$$

Proof. Substituting equation (17) to (7), one can obtain equation (16). It can be inferred that the transformation matrix $T_{\sigma(k)}$ is unique; therefore, we have the following equations.

$$\begin{aligned} \|\widetilde{\omega}(k)\|_{L_2}^2 &= \sum_0^\infty \|\widetilde{\omega}(k)\|_2^2, \\ &= \sum_0^\infty \omega^T(k)\left(T_{\sigma(k)} \otimes I_l\right)^T\left(T_{\sigma(k)} \otimes I_l\right)\omega(k), \\ &= \|\omega(k)\|_{L_2}^2, \end{aligned} \tag{18}$$

$$\begin{aligned} \|\widetilde{z}(r)\|_{L_\infty}^2 &= \sup\left\{\widetilde{z}^T(k)\widetilde{z}(k)\right\}, \\ &= \sup\left\{z^T(k)\left(T_{\sigma(k)} \otimes I_l\right)^T\left(T_{\sigma(k)} \otimes I_l\right)z(k)\right\}, \\ &= \|z(r)\|_{L_\infty}^2. \end{aligned} \tag{19}$$

It is obvious that the problem of robust consensus protocol design can be converted to the controller design of (16). □

Remark 1. The system in equation (16) consists of the independent system in equation (20). Therefore, the stability of equation (7) is equivalent to the stability of n subsystems in equation (20); the attempt to ensure the prescribed attenuation performance of (7) can be converted to guarantee the attenuation performance of (16).

$$\begin{cases} \widetilde{e}_i(k+1) = \overline{A}_i\widetilde{e}_i(k) + D\widetilde{\omega}_i(k) \\ \widetilde{z}_i(k) = \overline{C}_i\widetilde{e}_i(k) \end{cases}, \quad i \in \mathcal{H}, \tag{20}$$

where $\overline{A}_i = (A - \lambda_i^{\sigma(k)}B(K_{\sigma(k)} + K_{c,\sigma(k)}))$, and $\overline{C}_i = -\lambda_i^{\sigma(k)}C$.

Lemma 4. *For given constant scalars $0 < \delta < 1$, $\gamma > 0$. The system in (7) with control input in (5) is asymptotic stable with L_2-L_∞ attenuation performance γ if and only if the following equation holds.*

In Theorem 1, the sufficient conditions to guarantee the stability and prescribed attenuation performance index are presented.

Theorem 1. *For given constant scalars $\mu_s > 1$, $0 < \delta_s < 1$, $\gamma > 0$, if there exist Lyapunov functions $V_{i,s}\left(\widetilde{e}_i(k)\right)$, $i \in \mathcal{H}$, and class functions κ_1, κ_2, the switched systems in equation (20) with MDADT satisfying equation (25) are globally uniformly asymptotically stable with prescribed L_2-L_∞ attenuation performance γ, such that*

$$\kappa_1\left(\left\|\widetilde{e}_i(k)\right\|\right) \le V_{i,s}\left(\widetilde{e}_i(k)\right) \le \kappa_2\left(\left\|\widetilde{e}_i(k)\right\|\right), \tag{21}$$

$$V_{i,s}\left(\widetilde{e}_i(k)\right) \le \mu_s V_{i,r}\left(\widetilde{e}_i(k)\right), \quad s \ne r, \tag{22}$$

$$\Delta V_{i,s}\left(\widetilde{e}_i(k)\right) \le -\delta_s V_{i,s}\left(\widetilde{e}_i(k)\right) + \|\widetilde{\omega}(k)\|_2^2, \tag{23}$$

$$\left\|\widetilde{z}_i(k)\right\|_2^2 \le \gamma^2 V_{i,s}\left(\widetilde{e}_i(k)\right), \tag{24}$$

$$\tau_{as} \ge \tau_{as}^* = -\frac{\ln \mu_s}{\ln\left(1 - \delta_s\right)}. \tag{25}$$

Proof. The entire proof can be divided into two steps.

(1) The stability of equation (20).

The switching instants in the time interval $(0, k)$ are set to be $k_1, k_2, \ldots, k_t$ with $k_{t+1} = k$. Then, (26) holds when $\|\omega_i(k)\| \equiv 0$.

$$V_{i,\sigma(k_t)}\left(\widetilde{e}_i(k)\right) \le \left(1 - \delta_{\sigma(k_t)}\right)^{k - k_t}V_{i,\sigma(k_t)}\left(\widetilde{e}_i(k_t)\right). \tag{26}$$

Together with (22), we can conclude that

$$V_{i,\sigma(k_t)}\left(\widetilde{e}_i(k)\right) \le \mu_{\sigma(k_t)}\left(1 - \delta_{\sigma(k_t)}\right)^{k - k_t}V_{i,\sigma(k_t)}\left(\widetilde{e}_i(k_t)\right). \tag{27}$$

Based on equations (26)-(27), the following equation can be obtained by iteration.

$$
\begin{aligned}
V_{i,\sigma(k_t)}\big(\widetilde{e}_i(k)\big) &\leq \mu_{\sigma(k_t)}\Big(1-\delta_{\sigma(k_t)}\Big)^{k-k_t}V_{i,\sigma(k_t)}\big(\widetilde{e}_i(k_t)\big), \\
&\leq \mu_{\sigma(k_t)}\mu_{\sigma(k_{t-1})}\Big(1-\delta_{\sigma(k_t)}\Big)^{k-k_t}\Big(1-\delta_{\sigma(k_{t-1})}\Big)^{k_t-k_{t-1}}V_{i,\sigma(k_{t-1})}\big(\widetilde{e}_i(k_{t-1})\big), \\
&\cdots \\
&\leq \prod_{r=0}^{t}\left\{\mu_{\sigma(k_r)}\Big(1-\delta_{\sigma(k_r)}\Big)^{k_{r+1}-k_r}\right\}V_{i,\sigma(0)}\big(\widetilde{e}_i(0)\big), \\
&\leq \prod_{s=1}^{N_f}\left\{\mu_{\sigma(k_s)}^{N_{\sigma,s}(0,k)}\Big(1-\delta_{\sigma(k_s)}\Big)^{T_{\sigma s}(0,k)}\right\}V_{i,\sigma(0)}\big(\widetilde{e}_i(0)\big).
\end{aligned}
\tag{28}
$$

Combining with Definition 1, we have

$$
\begin{aligned}
V_{i,\sigma(k_t)}\big(\widetilde{e}_i(k)\big) &\leq \exp\left\{\sum_{s=1}^{N_f}\Big(N_{\sigma,s}(0,k)\ln\mu_{\sigma(k_s)}+T_{\sigma s}(0,k)\ln\Big(1-\delta_{\sigma(k_s)}\Big)\Big)\right\}V_{i,\sigma(0)}\big(\widetilde{e}_i(0)\big), \\
&\leq \exp\left\{\sum_{s=1}^{N_f}\Big(\frac{T_{\sigma s}(0,k)}{\tau_{as}}\ln\mu_{\sigma(k_s)}+T_{\sigma s}(0,k)\ln\Big(1-\delta_{\sigma(k_s)}\Big)\Big)\right\}V_{i,\sigma(0)}\big(\widetilde{e}_i(0)\big), \\
&\leq \exp\left\{\sum_{s=1}^{N_f}\Big(T_{\sigma s}(0,k)\Big(\frac{\ln\mu_{\sigma(k_s)}}{\tau_{as}}+\ln\Big(1-\delta_{\sigma(k_s)}\Big)\Big)\Big)\right\}V_{i,\sigma(0)}\big(\widetilde{e}_i(0)\big).
\end{aligned}
\tag{29}
$$

Then, we can obtain (29) based on (21).

$$
\left\|\widetilde{e}_i(k)\right\|^2 \leq \kappa_1^{-1}\kappa_2\exp\left\{\sum_{s=1}^{N_f}\Big(T_{\sigma s}(0,k)\Big(\frac{\ln\mu_{\sigma(k_s)}}{\tau_{as}}+\ln\Big(1-\delta_{\sigma(k_s)}\Big)\Big)\Big)\right\}\left\|\widetilde{e}_i(0)\right\|^2.
\tag{30}
$$

Therefore, the system in (20) with MDADT satisfying (25) is globally uniformly asymptotically stable.

(2) The system in equation (20) has prescribed L_2-L_∞ attenuation performance γ.

Together with equations (22)-(23), one has

$$
\begin{aligned}
V_{i,\sigma(k_t)}\big(\widetilde{e}_i(k)\big) &\leq \Big(1-\delta_{\sigma(k_t)}\Big)^{k-k_t}V_{i,\sigma(k_t)}\big(\widetilde{e}_i(k_t)\big)+\sum_{r=k_t}^{k-1}\Big(1-\delta_{\sigma(k_t)}\Big)^{k-r-1}\|\widetilde{\omega}(r)\|_2^2, \\
&\leq \mu_{\sigma(k_t)}\Big(1-\delta_{\sigma(k_t)}\Big)^{k-k_t}V_{i,\sigma(k_{t-1})}\big(\widetilde{e}_i(k_t)\big)+\sum_{r=k_t}^{k-1}\Big(1-\delta_{\sigma(k_t)}\Big)^{k-r-1}\|\widetilde{\omega}(r)\|_2^2.
\end{aligned}
\tag{31}
$$

Then, one can obtain the equation as follows by iteration.

$$V_{i,\sigma(k_t)}\big(\widetilde{e}_i(k)\big) \le \mu_{\sigma(k_t)}\Big(1-\delta_{\sigma(k_t)}\Big)^{k-k_t} V_{i,\sigma(k_{t-1})}\big(\widetilde{e}_i(k_t)\big) + \sum_{r=k_t}^{k-1}\Big(1-\delta_{\sigma(k_t)}\Big)^{k-r-1}\|\widetilde{\omega}(r)\|_2^2,$$

$$\le \mu_{\sigma(k_t)}\Big(1-\delta_{\sigma(k_t)}\Big)^{k-k_t}\left[\mu_{\sigma(k_{t-1})}\Big(1-\delta_{\sigma(k_{t-1})}\Big)^{k_t-k_{t-1}}V_{i,\sigma(k_{t-2})}\big(\widetilde{e}_i(k_{t-1})\big), + \sum_{r=k_{t-1}}^{k_t-1}\Big(1-\delta_{\sigma(k_{t-1})}\Big)^{k-r-1}\|\widetilde{\omega}(r)\|_2^2\right] + \sum_{r=k_t}^{k-1}\Big(1-\delta_{\sigma(k_t)}\Big)^{k-r-1}\|\widetilde{\omega}(r)\|_2^2,$$

$$\cdots$$

$$\le \mu_{\sigma(k_1)}\cdots\mu_{\sigma(k_t)}\Big(1-\delta_{\sigma(k_0)}\Big)^{k_1-k_0}\cdots\Big(1-\delta_{\sigma(k_t)}\Big)^{k-k_t}V_{i,\sigma(0)}\big(\widetilde{e}_i(0)\big),$$

$$+ \mu_{\sigma(k_1)}\cdots\mu_{\sigma(k_t)}\Big(1-\delta_{\sigma(k_0)}\Big)^{k_1-k_0}\cdots\Big(1-\delta_{\sigma(k_t)}\Big)^{k-k_t}\sum_{r=k_0}^{k_1-1}\Big(1-\delta_{\sigma(k_0)}\Big)^{k_1-r-1}\|\widetilde{\omega}(r)\|_2^2,$$

$$+ \mu_{\sigma(k_2)}\cdots\mu_{\sigma(k_t)}\Big(1-\delta_{\sigma(k_1)}\Big)^{k_2-k_1}\cdots\Big(1-\delta_{\sigma(k_t)}\Big)^{k-k_t}\sum_{r=k_1}^{k_2-1}\Big(1-\delta_{\sigma(k_1)}\Big)^{k_2-r-1}\|\widetilde{\omega}(r)\|_2^2,$$

$$+\cdots+\sum_{r=k_t}^{k-1}\Big(1-\delta_{\sigma(k_t)}\Big)^{k-r-1}\|\widetilde{\omega}(r)\|_2^2,$$

$$\le \Big(1-\delta_{\sigma(k_t)}\Big)^{k-k_t}\prod_{r=0}^{t-1}\left\{\mu_{\sigma(k_{r+1})}\Big(1-\delta_{\sigma(k_r)}\Big)^{k_{r+1}-k_r}\right\}V_{i,\sigma(0)}\big(\widetilde{e}_i(0)\big),$$

$$+ \mu_{\sigma(k_t)}\Big(1-\delta_{\sigma(k_t)}\Big)^{k-k_t}\sum_{q=1}^{t-1}\left(\prod_{p=q}^{t-1}\mu_{\sigma(k_p)}\Big(1-\delta_{\sigma(k_p)}\Big)^{k_{p+1}-k_p}\sum_{r=k_{q-1}}^{k_q-1}\Big(1-\delta_{\sigma(k_{q-1})}\Big)^{k_q-r-1}\widetilde{\omega}(r)\|_2^2\right),$$

$$+ \mu_{\sigma(k_t)}\Big(1-\delta_{\sigma(k_t)}\Big)^{k-k_t}\sum_{r=k_{t-1}}^{k_t-1}\Big(1-\delta_{\sigma(k_{t-1})}\Big)^{k_t-r-1}\|\widetilde{\omega}(r)\|_2^2,$$

$$+ \sum_{r=k_t}^{k-1}\Big(1-\delta_{\sigma(k_t)}\Big)^{k-r-1}\|\widetilde{\omega}(r)\|_2^2,$$

$$\le \prod_{s=1}^{N_f}\left\{\mu_{\sigma(k_s)}^{N_{\sigma s}(0,k)}\Big(1-\delta_{\sigma(k_s)}\Big)^{T_{\sigma s}(0,k)}\right\}V_{i,\sigma(0)}\big(\widetilde{e}_i(0)\big),$$

$$+ \sum_{q=1}^{t}\left(\prod_{s=1}^{N_f}\left\{\mu_{\sigma(k_s)}^{N_{\sigma s}(k_q,k)}\Big(1-\delta_{\sigma(k_s)}\Big)^{T_{\sigma s}(k_q,k)}\right\}\sum_{r=k_{q-1}}^{k_q-1}\Big(1-\delta_{\sigma(k_{q-1})}\Big)^{k_q-r-1}\|\widetilde{\omega}(r)\|_2^2\right), + \sum_{r=k_t}^{k-1}\Big(1-\delta_{\sigma(k_t)}\Big)^{k-r-1}\|\widetilde{\omega}(r)\|_2^2.$$

$$(32)$$

Substituting the equation above into (8), one can obtain that

$$V_{i,\sigma(k_t)}\big(\widetilde{e}_i(k)\big) \le \exp\left\{\sum_{s=1}^{N_f}T_{\sigma s}(0,k)\left(\frac{\ln\mu_{\sigma(k_s)}}{\tau_{as}}+\ln\Big(1-\delta_{\sigma(k_s)}\Big)\right)\right\}V_{i,\sigma(0)}\big(\widetilde{e}_i(0)\big),$$

$$+ \sum_{q=1}^{t}\left\{\exp\left(\sum_{s=1}^{N_f}T_{\sigma s}(k_q,k)\left(\frac{\ln\mu_{\sigma(k_s)}}{\tau_{as}}+\ln\Big(1-\delta_{\sigma(k_s)}\Big)\right)\right),\times\sum_{r=k_{q-1}}^{k_q-1}\Big(1-\delta_{\sigma(k_{q-1})}\Big)^{k_q-r-1}\|\widetilde{\omega}(r)\|_2^2\right\},$$

$$+ \sum_{r=k_t}^{k-1}\Big(1-\delta_{\sigma(k_t)}\Big)^{k-r-1}\|\widetilde{\omega}(r)\|_2^2.$$

$$(33)$$

According to the conditions $\mu_s > 1$, $0 < \delta_s < 1$, and (25), we have

$$0 < 1 - \delta_{\sigma(k_s)} < 1, \tag{34}$$

$$0 < \exp\left(\sum_{s=1}^{N_f} T_{\sigma s}(k_q, k)\left(\frac{\ln \mu_{\sigma(k_s)}}{\tau_{as}} + \ln\left(1 - \delta_{\sigma(k_s)}\right)\right)\right) < 1. \tag{35}$$

Combining equations (32)–(34), one can obtain equation (35).

$$V_{i,\sigma(k_t)}\left(\widetilde{e}_i(k)\right) \le \sum_{q=1}^{t}\sum_{r=k_{q-1}}^{k_q-1} \|\widetilde{\omega}(r)\|_2^2 + \sum_{r=k_t}^{k-1} \|\widetilde{\omega}(r)\|_2^2,$$

$$= \sum_{r=0}^{k_t-1} \|\widetilde{\omega}(r)\|_2^2 + \sum_{r=k_t}^{k-1} \|\widetilde{\omega}(r)\|_2^2, \tag{36}$$

$$= \sum_{r=0}^{k-1} \|\widetilde{\omega}(r)\|_2^2.$$

Together with (24), it is obvious that

$$\|\widetilde{z}_i(k)\|_2^2 \le \gamma^2 V_{i,s}\left(\widetilde{e}_i(k)\right),$$

$$\le \gamma^2 \sum_{r=0}^{k-1} \|\widetilde{\omega}(r)\|_2^2, \tag{37}$$

which implies that $\sup\left\{\|\widetilde{z}(r)\|_2^2\right\} \le \gamma^2 \|\widetilde{\omega}(r)\|_{L_2}^2$, and the proof is complete. $\square$

Corollary 1. *For given constant scalars $\mu_s > 1$, $0 < \delta_s < 1$, $\gamma > 0$, if there exist positive-definite matrices $P_{i,s} \in R^{p \times p}$ satisfying equations (37)–(39), the switched systems in equation (20) with MDADT satisfying equation (25) are globally uniformly asymptotically stable with prescribed L_2-L_∞ attenuation performance γ.*

$$P_{i,s} \le \mu_s P_{i,r}, \quad s \ne r, \tag{38}$$

$$\begin{bmatrix} \overline{A}_i^T P_{i,s} \overline{A}_i - (1 - \delta_s)P_{i,s} & \overline{A}_i^T P_{i,s} D \\ * & D^T P_{i,s} D - I \end{bmatrix} \le 0, \tag{39}$$

$$(\lambda_i^s)^2 C^T C < \gamma^2 P_{i,s}. \tag{40}$$

Proof. The Lyapunov function $V_{i,s}\left(\widetilde{e}_i(k)\right)$, $i \in \mathscr{H}$, is defined as follows:

$$V_{i,s}\left(\widetilde{e}_i(k)\right) = \widetilde{e}_i^T(k) P_{i,s} \widetilde{e}_i(k). \tag{41}$$

According to (20) and (39), we can conclude that (38) is equivalent to (24). Along the trajectory of $V_{i,s}\left(\widetilde{e}_i(k)\right)$, one has

$$\Delta V_{i,s}\left(\widetilde{e}_i(k)\right) = \widetilde{e}_i^T(k+1) P_{i,s} \widetilde{e}_i(k+1) - \widetilde{e}_i^T(k) P_{i,s} \widetilde{e}_i(k),$$

$$= \left(\overline{A}_i \widetilde{e}_i(k) + D\widetilde{\omega}_i(k)\right)^T \times P_{i,s}\left(\overline{A}_i \widetilde{e}_i(k) + D\widetilde{\omega}_i(k)\right),$$

$$- \widetilde{e}_i^T(k) P_{i,s} \widetilde{e}_i(k),$$

$$= \widetilde{e}_i^T(k)\left[\overline{A}_i^T P_{i,s} \overline{A}_i - P_{i,s}\right]\widetilde{e}_i(k) + \widetilde{\omega}_i^T(k)\left[D^T P_{i,s} D\right]\widetilde{\omega}_i(k),$$

$$+ \widetilde{e}_i^T(k)\overline{A}_i^T P_{i,s} D\widetilde{\omega}_i(k) + \widetilde{\omega}_i^T(k)D^T P_{i,s}\overline{A}_i \widetilde{e}_i(k). \tag{42}$$

Together with equations (40)-(41), we have

$$\Delta V_{i,s}\left(\widetilde{e}_i(k)\right) + \delta_s V_{i,s}\left(\widetilde{e}_i(k)\right) - \widetilde{\omega}_i^T(k)\widetilde{\omega}_i(k),$$

$$= \widetilde{e}_i^T(k)\left[\overline{A}_i^T P_{i,s}\overline{A}_i - (1 - \delta_s)P_{i,s}\right]\widetilde{e}_i(k) + \widetilde{e}_i^T(k)\overline{A}_i^T P_{i,s} D\widetilde{\omega}_i(k),$$

$$+ \widetilde{\omega}_i^T(k)D^T P_{i,s}\overline{A}_i\widetilde{e}_i(k) + \widetilde{\omega}_i^T(k)\left[D^T P_{i,s} D_i - I\right]\widetilde{\omega}_i(k),$$

$$= \begin{bmatrix} \widetilde{e}_i(k) \\ \widetilde{\omega}_i(k) \end{bmatrix}^T \begin{bmatrix} \overline{A}_i^T P_{i,s}\overline{A}_i - (1 - \delta_s)P_{i,s} & \overline{A}_i^T P_{i,s} D \\ * & D^T P_{i,s} D - I \end{bmatrix}\begin{bmatrix} \widetilde{e}_i(k) \\ \widetilde{\omega}_i(k) \end{bmatrix},$$

$$\le 0. \tag{43}$$

According to Theorem 1, we can conclude that the system in (20) with MDADT satisfying (25) is globally uniformly asymptotically stable with prescribed L_2-L_∞ attenuation performance γ.

Based on Theorem 1 and Corollary 1, the solutions of consensus protocol are given in Theorem 2. $\square$

Theorem 2. *For given constant scalars $\mu_s > 1$, $0 < \delta_s < 1$, $\gamma > 0$, $a_s > 0$, and $\varepsilon_s > 0$, if there exist positive-definite matrices $P_{i,s} \in R^{p \times p}$, matrices $X_s \in R^{l \times l}$, $Y_s \in R^{l \times q}$, the MASs in (2)-(3) with control input in equation (5) are asymptotically stable with prescribed L_2-L_∞ attenuation performance γ such that equation (43) holds.*

$$\begin{bmatrix} \Xi_{11} & 0 & \Xi_{13} & a_s\lambda_i^s Y_s^T & 0 \\ * & -I & D^T P_{i,s} & 0 & 0 \\ * & * & -(I_{n-1}\otimes P_s) & BX_s - P_{i,s}B & P_{i,s}BM_s \\ * & * & * & -a_s(X_s + X_s^T) & 0 \\ * & * & * & * & -\varepsilon_s I \end{bmatrix} < 0. \quad (44)$$

The parameters of control protocol can be derived in (43).

$$K_s = X_s^{-1} Y_s, \quad (45)$$

where $\Gamma_{11} = -(1-\delta_s)P_{i,s} + \varepsilon_s(\lambda_i^s)^2 N_s^T N_s,$ $\Gamma_{13} = A^T P_{i,s} - \lambda_i^s Y_s^T B^T.$

Proof. According to Schur complement, it is obvious that equation (43) is equivalent to equation (45).

$$\begin{bmatrix} \Xi_{11} & 0 & \Xi_{13} & a_s\lambda_i^s Y_s^T \\ * & -I & D^T P_{i,s} & 0 \\ * & * & -P_{i,s} & BX_s - P_{i,s}B \\ * & * & * & -a_s(X_s + X_s^T) \end{bmatrix} + \varepsilon_s^{-1}\begin{bmatrix} 0 \\ 0 \\ P_{i,s}BM_s \\ 0 \end{bmatrix}\begin{bmatrix} 0 \\ 0 \\ P_{i,s}BM_s \\ 0 \end{bmatrix}^T,$$

$$= \begin{bmatrix} \Xi_{11} & 0 & \Xi_{13} & a_s\lambda_i^s Y_s^T \\ * & -I & D^T P_{i,s} & 0 \\ * & * & -P_{i,s} + \varepsilon_s^{-1}P_{i,s}BM_sM_s^T B^T P_{i,s} & BX_s - P_{i,s}B \\ * & * & * & -a_s(X_s + X_s^T) \end{bmatrix},$$

$$< 0. \quad (46)$$

Define $\Theta_s = \mathscr{T}_s + \varepsilon_s^{-1}\mathscr{M}_s^T + \varepsilon_s\mathscr{N}_s^T\mathscr{N}_s, U_s = \begin{bmatrix} 0 & 0 & X_s^T B^T - B^T P_{i,s} \end{bmatrix}^T,$ and $V_s = \begin{bmatrix} \lambda_i^s Y_s & 0 & 0 \end{bmatrix}^T,$ where $\mathscr{T}_s = \begin{bmatrix} -(1-\delta_s)P_{i,s} & 0 & A^T P_{i,s} - \lambda_i^s Y_s^t B^T \\ * & -I & D^T P_{i,s} \\ * & * & -P_{i,s} \end{bmatrix}, \mathscr{M}_s = \begin{bmatrix} 0 \\ 0 \\ P_{i,s}BM_s \end{bmatrix},$ and $\mathscr{N}_s = \begin{bmatrix} -\lambda_i^s N_s & 0 & 0 \end{bmatrix}.$

Together with Lemma 2, we have

$$\Theta_s + U_s X_s^{-1} V_s^T + V_s X_s^{-T} U_s^T,$$
$$= \mathscr{T}_s + \varepsilon_s^{-1}\mathscr{M}_s\mathscr{M}_s^T + \varepsilon_s\mathscr{N}_s^T\mathscr{N}_s + U_s X_s^{-1} V_s^T + V_s X_s^{-T} U_s^T < 0. \quad (47)$$

Moreover, based on Lemma 3, one has

$$\mathscr{T} + \mathscr{M}\mathscr{F}\mathscr{N} + \mathscr{N}^T\mathscr{F}^T\mathscr{M}^T + U_s X_s^{-1} V_s^T + V_s X_s^{-T} U_s^T$$
$$= \begin{bmatrix} -(1-\delta_s)P_{i,s} & 0 & \overline{A}_i^T P_{i,s} \\ * & -I & D^T P_{i,s} \\ * & * & -P_{i,s} \end{bmatrix} < 0. \quad (48)$$

According to Schur complement, it is obvious that (46) is equivalent to (37), which completes the proof. □

4. Compensated Consensus Protocol Design Based on FDQL

In this section, the learning-based consensus protocol is proposed based on deep reinforcement learning, where fuzzy deep Q learning is utilized. The stability and prescribed attenuation performance are guaranteed by the robust control theory, and the learning-based control protocol is introduced to improve the transient performance and realize optimal control policy. The output of the learning-based control protocol can be viewed as an additional variation of robust consensus protocol. The online scheduling of control protocol is established as a Markovian process. Therefore, the advantages of robust control theory and deep reinforcement learning are combined.

It is well known that reinforcement learning is composed of state, action, agent, and environment. The state of kth step is defined as s_k, and the chosen action is supposed to be a_k; then, the reward function r_k and the state s_{k+1} are generated based on the interaction with the environment. Therefore, the optimal control policy can be obtained by maximum the reward function.

To improve the convergence of consensus protocol, the state is defined as $s_k = [\widetilde{e}_i(k) \ \widetilde{z}_i(k)]$ and the action is defined as $a_k = [K_{c,\sigma(k)}]$.

In Q learning, the deep neural network is utilized to approximate the action-state value function $Q^*(s_k, a_k)$, which can be described as

$$f(s_k, a_k, \omega) = Q^*(s_k, a_k), \quad (49)$$

where $f(s_k, a_k, \omega)$ denotes the function of deep neural networks.

The action is chosen based on the maximum Q value:

$$a^* = \arg\max_a Q(s_k, a_k). \quad (50)$$

There exist two neural networks in the deep Q learning algorithm, whose structures are the same and can be called as the critic neural network and target neural network. The parameters of the critic neural network are updated based on temporal-difference learning. The output of the critic neural network is defined as $Q(s_k, a_k, \omega)$ and the output of the target neural network is defined as $Q(s_k, a_k, \omega^-)$. Therefore, the parameters of the critic neural network are updated based on the equation as follows:

$$Q(s_k, a_k, \omega^-) = L_r[R + \gamma_s\max_{a'}(Q(s', a', \omega^-))] + (1 - L_r)Q(s_k, a_k, \omega), \quad (51)$$

where L_r is the learning rate, γ_s denotes the discount factor, R represents the reward of state transition from s_k to s' through action a_k, and $\max_{a'}(Q(s', a', \omega^-))$ stands for the maximum Q value of the target neural network.

It can be inferred that the reward function has an important influence on the final performance. The design of traditional deep Q learning mainly depends on the experience of designers, which can not achieve optimal performance and will improve the computational complexity. In this study, the reward function is applied to design the

reward function. The input value of fuzzy reward function can be divided into five categories, which can be described as VB, B, N, G, and VG. The five categories represent very bad, bad, normal, good, and very good. In this study, it is supposed that there are four followers. Therefore, the inputs of the fuzzy reward system are set to be $|e_1|$, $|e_2|$, $|e_3|$, and $|e_4|$. It can be inferred that each fuzzy set includes 25 rules, and the total number of the fuzzy rules is 75. The output of the fuzzy reward function is limited in the interval $[-1,0)$, and the defuzzifier of the fuzzy reward function is defined as

$$f(z) = \frac{\sum_{p=1}^{25} C^i \prod_{q=1}^{m} s_q^p (z_q)}{\sum_{p=1}^{25} \prod_{q=1}^{m} s_q^p (z_q)}. \tag{52}$$

Based on the statement above, the learning-based consensus protocol design algorithm can be summarized as follows:

Remark 2. The FDQN algorithm proposed in this study can improve the transient convergence performance of MASs. The output of the deep Q network is supposed to be variation of parameters of consensus protocol. As well known, the design of reward function in the traditional method depends of the experience of the designers. To overcome the problem, the fuzzy reward function is developed to improve the learning efficiency in this study.

5. Numerical Example

In this section, an example is provided to illustrate the effectiveness of the method. The model of MASs is constructed as follows:

$$A = \begin{bmatrix} -1 \cdot 3 & -1 \cdot 5 \\ 1 \cdot 3 & 1 \cdot 2 \end{bmatrix},$$

$$B = \begin{bmatrix} 1 \cdot 2 \\ 1 \cdot 5 \end{bmatrix}, \tag{53}$$

$$C = \begin{bmatrix} 1 \cdot 1 & 0 \cdot 9 \end{bmatrix},$$

$$D = \begin{bmatrix} 0 \cdot 13 & 0 \cdot 12 \end{bmatrix}.$$

The external disturbance is

$$d(k) = 0 \cdot 1 e^{-0.2k} \cos(0 \cdot 4k). \tag{54}$$

The switching topologies are shown in Figure 1. Then, we can obtain the Laplace matrices as follows:

$$L_1 = \begin{bmatrix} 1 & 0 & 0 & -1 \\ 0 & 1 & -1 & 0 \\ 0 & -1 & 1 & 0 \\ -1 & 0 & 0 & 1 \end{bmatrix},$$

$$\tag{55}$$

$$L_2 = \begin{bmatrix} 1 & 0 & 0 & -1 \\ 0 & 1 & 0 & -1 \\ 0 & 0 & 1 & -1 \\ -1 & -1 & -1 & 3 \end{bmatrix}.$$

The parameters of switching topologies are given as follows:

$$\begin{cases} a_1 = 1 \cdot 01 \\ a_2 = 1 \cdot 11, \\ a_3 = 1 \cdot 08 \end{cases} \begin{cases} \mu_1 = 1 \cdot 25 \\ \mu_2 = 1 \cdot 31, \\ \mu_3 = 1 \cdot 38 \end{cases} \begin{cases} \delta_1 = 0 \cdot 51 \\ \delta_2 = 0 \cdot 52. \\ \delta_3 = 0 \cdot 53 \end{cases} \tag{56}$$

Therefore, we can obtain MDADT according to (25).

$$\tau_1 = 0 \cdot 3128, \ \tau_2 = 0 \cdot 3679, \ \tau_3 = 0 \cdot 4266. \tag{57}$$

It is well known that the ADT method can be viewed as a special case of the MDADT method. Therefore, it can be inferred that $\tau_a = \max\{\tau_{ai}\} = 0 \cdot 4266$. It is obvious that tighter bounds on dwell time and less conservative results can be obtained. Then, we define the attenuation performance index $\gamma = 0 \cdot 9$, and we can obtain the parameters of consensus protocol based on Theorem 2.

The switching logic is shown in Figure 2. In order to illustrate the effectiveness and superiority of the proposed method, the traditional ADT method and MDADT method are given as comparisons. From the statement above, we have realized that MDADT can obtain tighter bounds and less conservative results. Moreover, the comparisons of state response of the ADT method and MDADT method are shown in Figures 3–6. The state responses of MASs with ADT switching topologies are shown in Figures 3-4. The state responses of MASs with MDADT switching topologies are shown in Figures 5-6. We can see that the transient performance of the ADT method is better than that of the MDADT method because the different characteristics of subsystems are taken into consideration, which will no doubt improve the design flexibility and make it more applicable for practical conditions.

Validate the superiority of the proposed method. The state response of the proposed method is shown in Figures 7–11. The state responses of the proposed method are shown in Figures 7-8. We can conclude that the transient performance can be improved by the aid of fuzzy deep Q learning. The advantages of the traditional method and intelligent method are combined. Compared with the traditional method, the transient performance can be improved, and compared with the intelligent method, stability and training efficiency can be guaranteed. The attenuation performance index is shown in Figure 9, from which we can see that the robustness of the proposed is ensured. The episodes reward response is shown in Figure 10, and we can see that the reward function of the fuzzy deep Q learning algorithm can converge to the neighbor of the origin, which demonstrates the effectiveness of the algorithm in this study. In addition, the response of the action is shown in Figure 11, from which we can see that the learning-based consensus protocol is provided to compensate the additional input caused by the uncertainties.

Based on the statement above, we can conclude that the convergence, robustness, and prescribed attenuation performance index are guaranteed. The less conservative results and tighter bounds on dwell time can be obtained by the MDADT method. The transient performance of the system can be improved based on the fuzzy deep Q

(1) Design the dynamics-based consensus protocol according to Theorem 2
(2) Define the bounds of learning-based consensus protocol parameters
(3) Initialize the weights of the Q value network
(4) Initialize the weights of the target Q value network
(5) Initialize the replay buffer R, episode = 0
(6) **for** episode = 1 to M do
(7)   Initialize a random state s_1 and receive the initial observation
(8)   For t = 1 to K do
(9)   Select an action based on the state and reward function
(10)   Execute the action a_k, and then, one can obtain the reward function r_k and the state s_{k+1}
(11)   Store the pair (s_k, a_k, r_k, s_{k+1}) in the replay buffer
(12)   Sample a random minibatch of transitions (s_m, a_m, r_m, s_{m+1}) from the replay buffer
(13)   Update the target Q value function
(14)   Update the weights of the target Q value network
(15)   **end for**
(16) **end for**

ALGORITHM 1: Learning-based control protocol design based on FDQN.

(a)   (b)
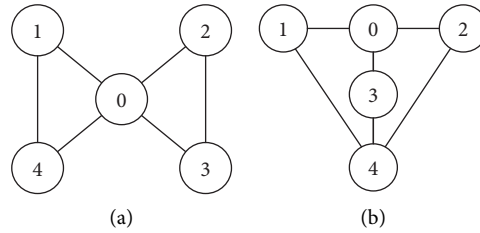
FIGURE 1: Switching topologies of MASs. (a) Interacting topology $\mathcal{G}_1$. (b) Interacting topology $\mathcal{G}_2$.

—— $\sigma(k)$
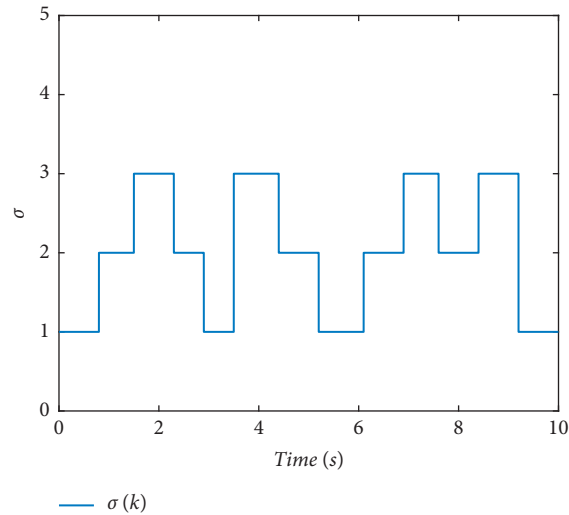
FIGURE 2: The switching logic.

learning algorithm. It is worth mentioning that the traditional robust cannot make comprised of robustness and transient performance, and the intelligent method always cannot guarantee convergence. By employing the proposed method, convergence, robustness, and transient performance are guaranteed simultaneously.
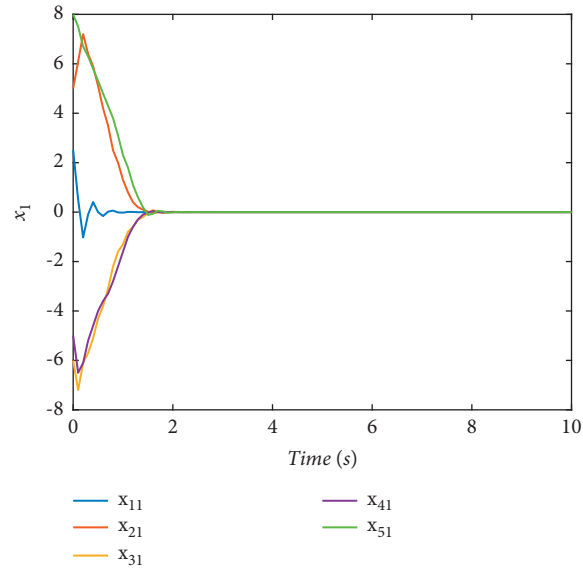
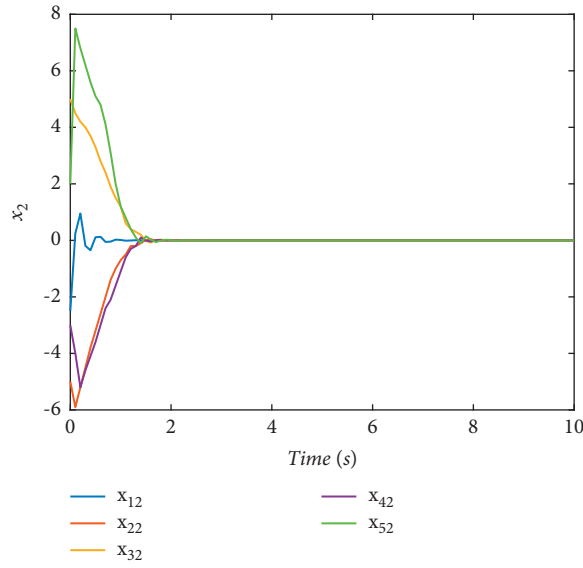FIGURE 3: The state response of $(x)_1$ under the ADT method.

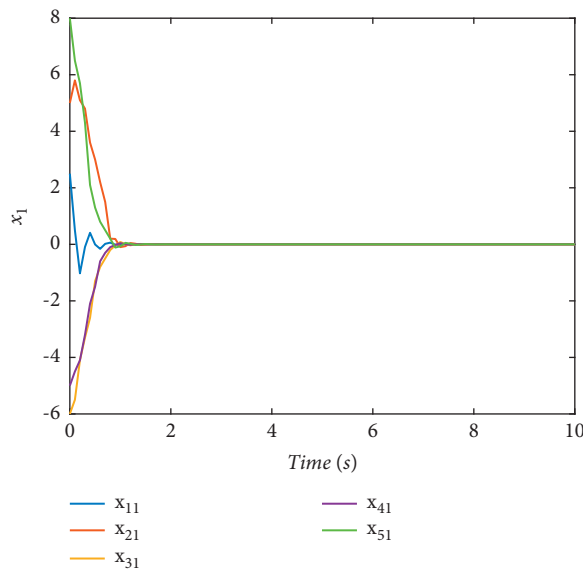FIGURE 4: The state response of $(x)_2$ under the ADT method.

FIGURE 5: The state response of $(x)_1$ under the MDADT method.
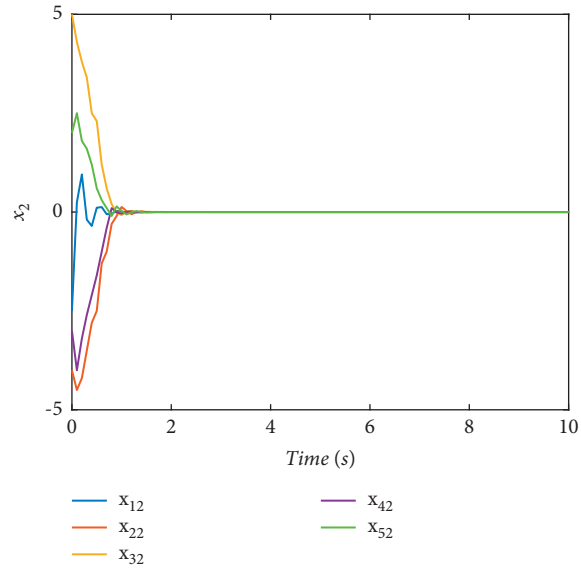
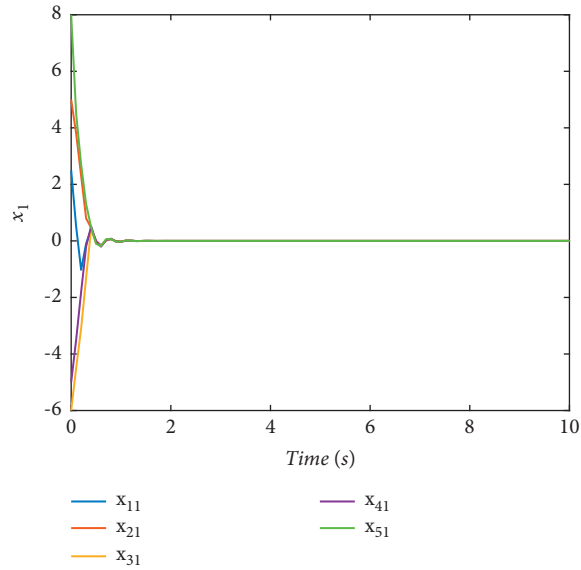FIGURE 6: The state response of $(x)_2$ under the MDADT method.

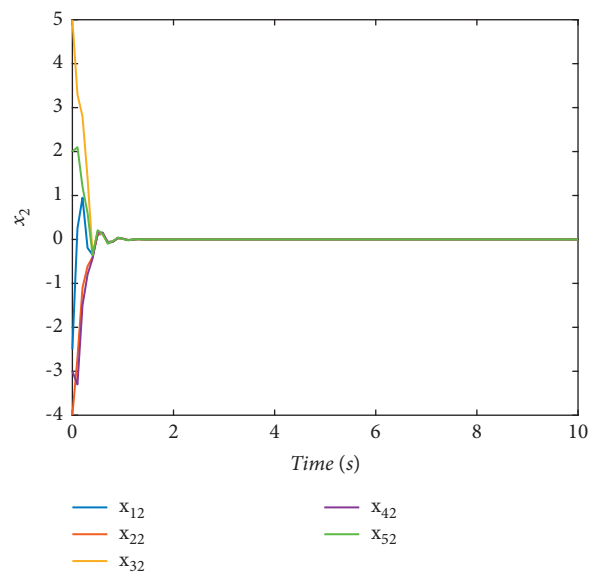FIGURE 7: The state response of $(x)_1$ with the proposed method.

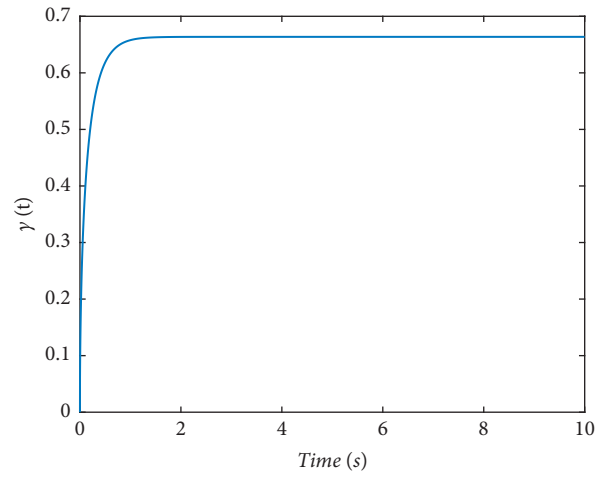FIGURE 8: The state response of $(x)_2$ with the proposed method.
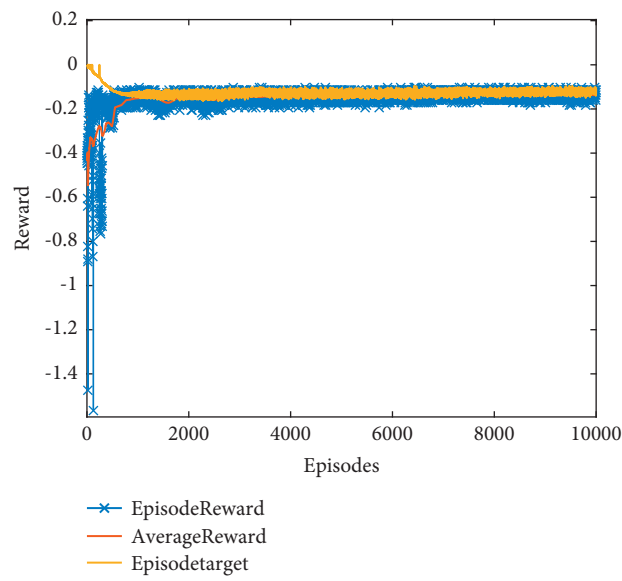
Figure 9: The response of attenuation performance index.
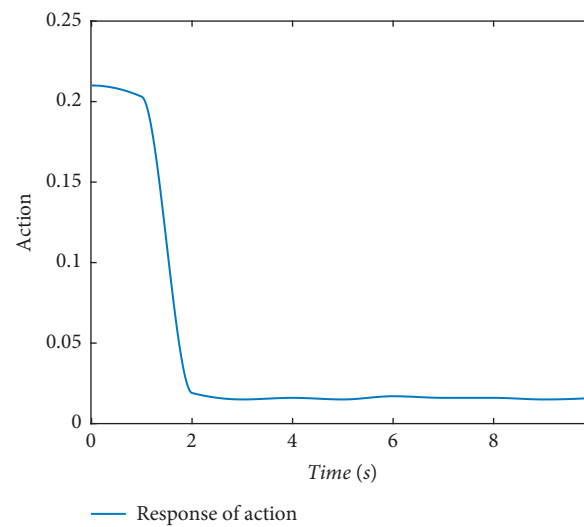

Figure 10: The response of episodes reward.

Figure 11: The response of the action.

6. Conclusions

The problem of intelligent L_2-L_∞ consensus design for MASs under switching topologies is investigated in this study. The switching topologies of MASs are modeled as switched system theory by employing linear transformation. Then, the problem of consensus protocol design can be converted to the problem of L_2-L_∞ control. To ensure the convergence, robustness, and transient performance simultaneously, the proposed consensus protocol is composed of dynamics-based consensus protocol and learning-based consensus protocol, which provides baseline and compensation of uncertainties. The baseline of consensus protocol is obtained by dynamics-based consensus protocol, which is provided based on the MDADT method and MLF method. The scheduling interval of learning-based protocol is given by nonfragile control theory. Then, the learning-based consensus protocol is proposed based on the fuzzy deep Q learning algorithm to improve the transient performance and achieve optimal policy, where the fuzzy reward function is introduced to improve the learning efficiency.

Data Availability

The data used to support the findings of this study are included within the article.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

References

[1] Y. Zhao, Q. Duan, G. Wen, D. Zhang, and B. Wang, "Time-varying formation for general linear multiagent systems over directed topologies: a fully distributed adaptive technique," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 51, no. 1, pp. 532–541, 2021.

[2] S. Zhou, X. Dong, Y. Hua, J. Yu, and Z. Ren, "Predefined formation-containment control of high-order multi-agent systems under communication delays and switching topologies," *IET Control Theory & Applications*, vol. 15, no. 12, pp. 1661–1672, 2021.

[3] D. Zhang, Y. Tang, Z. Ding, and F. Qian, "Event-based resilient formation control of multiagent systems," *IEEE Transactions on Cybernetics*, vol. 51, no. 5, pp. 2490–2503, 2021.

[4] K. Zhang, B. Zhou, H. Jiang, and G. Duan, "Finite-time output regulation by bounded linear time-varying controls with applications to the satellite formation flying," *International Journal of Robust and Nonlinear Control*, vol. 32, pp. 1–21, 2021.

[5] H. Liu, Y. Wang, and J. Xi, "Completely distributed formation control for networked quadrotors under switching communication topologies," *Systems & Control Letters*, vol. 147, Article ID 104841, 2021.

[6] L. Wang, J. Xi, M. He, and G. Liu, "Robust time-varying formation design for multiagent systems with disturbances: extended-state-observer method," *International Journal of Robust and Nonlinear Control*, vol. 30, no. 7, pp. 2796–2808, 2020.

[7] W. Zou, P. Shi, Z. Xiang, and Y. Shi, "Finite-time consensus of second-order switched nonlinear multi-agent systems," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 31, no. 5, pp. 1757–1762, 2020.

[8] Y. Yang, X. Si, D. Yue, and Y.-C. Tian, "Time-varying formation tracking with prescribed performance for uncertain nonaffine nonlinear multiagent systems," *IEEE Transactions on Automation Science and Engineering*, vol. 18, no. 4, pp. 1778–1789, 2021.

[9] J. Yu, X. Dong, Q. Li, J. Lv, and Z. Ren, "Distributed adaptive cooperative time-varying formation tracking guidance for multiple aerial vehicles system," *Aerospace Science and Technology*, vol. 117, Article ID 106925, 2021.

[10] W. Zou, C. Zhou, J. Guo, and Z. Xiang, "Global adaptive leader-following consensus for second-order nonlinear multiagent systems with switching topologies," *IEEE Transactions on Circuits and Systems II: Express Briefs*, vol. 68, no. 2, pp. 702–706, 2021.

[11] P. Zhang, H. Xue, S. Gao, and J. Zhang, "Distributed adaptive consensus tracking control for multi-agent system with communication constraints," *IEEE Transactions on Parallel and Distributed Systems*, vol. 32, no. 6, pp. 1293–1306, 2021.

[12] T.-M. Nguyen, Z. Qiu, T. H. Nguyen, M. Cao, and L. Xie, "Persistently excited adaptive relative localization and time-varying formation of robot swarms," *IEEE Transactions on Robotics*, vol. 36, no. 2, pp. 553–560, 2020.

[13] B. Li, G. Wen, Z. Peng, S. Wen, and T. Huang, "Time-varying formation control of general linear multi-agent systems under Markovian switching topologies and communication noises," *IEEE Transactions on Circuits and Systems II: Express Briefs*, vol. 68, no. 4, pp. 1303–1307, 2021.

[14] R. Mu, A. Wei, H. Li, and Z. M. Wang, "Event-triggered leader-following consensus for multi-agent systems with external disturbances under fixed and switching topologies," *IET Control Theory & Applications*, vol. 14, no. 11, pp. 1486–1496, 2020.

[15] Y. Yuan, Y. Wang, and L. Guo, "Sliding-mode-observer-based time-varying formation tracking for multispacecrafts subjected to switching topologies and time-delays," *IEEE Transactions on Automatic Control*, vol. 66, no. 8, pp. 3848–3855, 2021.

[16] X. Zhao, P. Shi, Y. Yin, and S. K. Nguang, "New results on stability of slowly switched systems: a multiple discontinuous Lyapunov function approach," *IEEE Transactions on Automatic Control*, vol. 62, no. 7, pp. 3502–3509, 2017.

[17] B. Niu, P. Zhao, J. Liu, H. Ma, and Y. Liu, "Global adaptive control of switched uncertain nonlinear systems: an improved MDADT method," *Automatica*, vol. 115, Article ID 108872, 2020.

[18] H. Yan, H. Zhang, X. Zhan, Y. Wang, S. Chen, and F. Yang, "Event-triggered sliding mode control of switched neural networks with mode-dependent average dwell time," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 51, no. 2, pp. 1233–1243, 2021.

[19] D. Yu and C. L. P. Chen, "Smooth transition in communication for swarm control with formation change," *IEEE Transactions on Industrial Informatics*, vol. 16, no. 11, pp. 6962–6971, 2020.

[20] H. Xiao and C. L. Philip Chen, "Time-varying nonholonomic robot consensus formation using model predictive based protocol with switching topology," *Information Sciences*, vol. 567, pp. 201–215, 2021.

[21] Y. Ren and Z. Hou, "Robust model-free adaptive iterative learning formation for unknown heterogeneous non-linear multi-agent systems," *IET Control Theory & Applications*, vol. 14, no. 4, pp. 654–663, 2020.

[22] Y. Xu, S. Zhao, D. Luo, and Y. You, "Affine formation maneuver control of high-order multi-agent systems over directed networks," *Automatica*, vol. 118, Article ID 109004, 2020.

[23] H. Liu, S. Li, G. Li, and H. Wang, "Robust adaptive control for fractional-order financial chaotic systems with system uncertainties and external disturbances," *Information Technology and Control*, vol. 46, no. 2, pp. 246–259, 2017.

[24] D. Wang, Y. Huang, S. Guo, and W. Wang, "Distributed H_∞ containment control of multi-agent systems over switching topologies with communication time delay," *International Journal of Robust and Nonlinear Control*, vol. 30, no. 1, pp. 1–12, 2020.

[25] J. Wang, L. Han, X. Li, X. Dong, Q. Li, and Z. Ren, "Time-varying formation of second-order discrete-time multi-agent systems under non-uniform communication delays and switching topology with application to UAV formation flying," *IET Control Theory & Applications*, vol. 14, no. 14, pp. 1947–1956, 2020.

[26] D. Liu, H. Liu, F. L. Lewis, and K. P. Valavanis, "Robust time-varying formation control for tail-sitters in flight mode transitions," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 51, no. 7, pp. 4102–4111, 2021.

[27] G. De Tommasi, D. G. Lui, A. Petrillo, and S. Santini, "A L2-gain robust PID-like protocol for time-varying output formation-containment of multi-agent systems with external disturbance and communication delays," *IET Control Theory & Applications*, vol. 15, no. 9, pp. 1169–1184, 2021.

[28] X. Wang, Y. Gu, Y. Cheng, A. Liu, and C. L. P. Chen, "Approximate policy-based accelerated deep reinforcement learning," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 31, no. 6, pp. 1820–1830, 2020.

[29] Z. Peng, D. Wang, T. Li, and M. Han, "Output-feedback cooperative formation maneuvering of autonomous surface vehicles with connectivity preservation and collision avoidance," *IEEE Transactions on Cybernetics*, vol. 50, no. 6, pp. 2527–2535, 2020.

[30] H. Liu, Y. Pan, S. Li, and Y. Chen, "Synchronization for fractional-order neural networks with full/under-actuation using fractional-order sliding mode control," *International Journal of Machine Learning and Cybernetics*, vol. 9, no. 7, pp. 1219–1232, 2018.

[31] H. Li, S. Chen, D. Izzo, and H. Baoyin, "Deep networks as approximators of optimal low-thrust and multi-impulse cost in multitarget missions," *Acta Astronautica*, vol. 166, pp. 469–481, 2020.

[32] T. Tan, F. Bao, Y. Deng, A. Jin, Q. Dai, and J. Wang, "Cooperative deep reinforcement learning for large-scale traffic grid signal control," *IEEE Transactions on Cybernetics*, vol. 50, no. 6, pp. 2687–2700, 2020.

[33] V. Shalumov, "Cooperative online guide-launch-guide policy in a target-missile-defender engage-ment using deep reinforcement learning," *Aerospace Science and Technology*, vol. 104, Article ID 105996, 2020.

[34] J. Xu, Z. Hou, W. Wang, B. Xu, K. Zhang, and K. Chen, "Feedback deep deterministic policy gradient with fuzzy reward for robotic multiple peg-in-hole assembly tasks," *IEEE Transactions on Industrial Informatics*, vol. 15, no. 3, pp. 1658–1667, 2019.

[35] M. Gheisarnejad and M. H. Khooban, "An intelligent non-integer PID controller-based deep reinforcement learning: implementation and experimental results," *IEEE Transactions on Industrial Electronics*, vol. 68, no. 4, pp. 3609–3618, 2021.

[36] L. Cheng, Z. Wang, and F. Jiang, "Real-time control for fuel-optimal Moon landing based on an interactive deep reinforcement learning algorithm," *Astrodynamics*, vol. 3, no. 4, pp. 375–386, 2019.

[37] T. Shi, P. Shi, and L. Zhang, "Distributed L_2-L_∞ consensus of multi-agent systems under asynchronous switching topologies," *International Journal of Control*, vol. 95, no. 2, pp. 544–553, 2022.

[38] X. Li, C. Zhou, J. Zhou, Z. Wang, and J. Xia, "Couple-group L2-l∞ consensus of nonlinear multi-agent systems with markovian switching topologies," *International Journal of Control, Automation and Systems*, vol. 17, no. 3, pp. 575–585, 2019.

[39] J. Zhou, Y. Wang, X. Zheng, Z. Wang, and H. Shen, "Weighted $${\mathcal {H}}_{\infty }$$ H ∞ consensus design for stochastic multi-agent systems subject to external disturbances and ADT switching topologies," *Nonlinear Dynamics*, vol. 96, no. 2, pp. 853–868, 2019.