Research Article | Open Access

Intelligence & Robotics

Liu *et al. Intell. Robot.* **2026**, *6*(4), 868-90 **DOI:**10.20517/ir.2026.38

Optimal consensus control of Markov jump multi-agent systems under switching topology

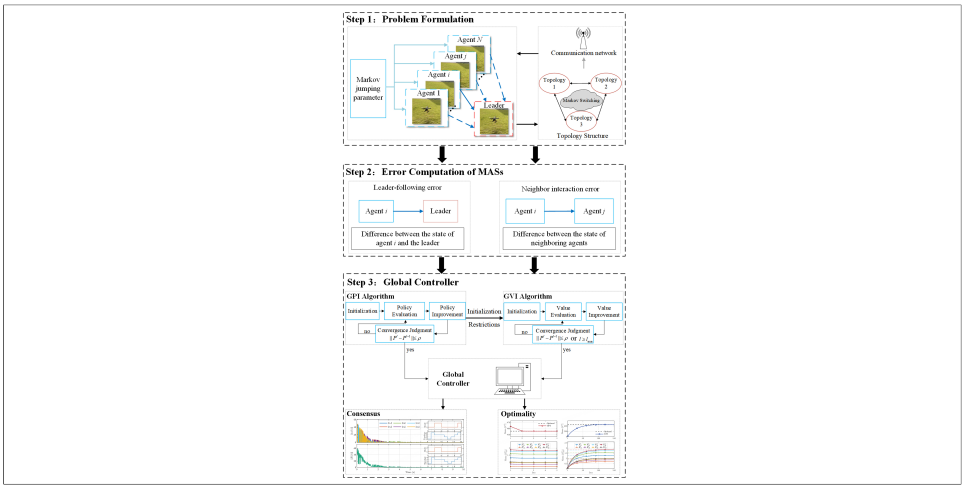
Bowen Liu, Wei Xing, Haoyue Yang, Junfeng Zhang

Keywords:
Optimal control, Markov jump multi-agent systems, Markov-switching topology, adaptive dynamic programming

Citation: Liu, B.; Xing, W.; Yang, H.; Zhang, J. Optimal consensus control of Markov jump multi-agent systems under switching topology. *Intell. Robot.* **2026**, *6*(4), 868-90. https://dx.doi.org/10.20517/ir.2026.38

Received: 31 Jul 2026
First Decision: 8 Sep 2026
Revised: 8 Oct 2026
Accepted: 26 Oct 2026
Published: 28 Oct 2026

Academic Editor:
Hongtian Chen
Copy Editor:
Pei-Yun Wang
Production Editor:
Pei-Yun Wang



Abstract

This paper investigates the optimal consensus control problem for multi-agent systems with Markov jump characteristics in system parameters and subject to Markov switching in communication topology. Firstly, the cost function is constructed to characterize the impact of topology switching and jumps in system parameters on system performance. Then, to tackle the impact of neighbor coupling terms on the performance of the cost function, the global algebraic Riccati equation is derived to determine the optimal control policy. Within the adaptive dynamic programming framework, global form policy iteration and value iteration algorithm are proposed to solve the global control policy. Subsequently, the convergence of the proposed iterative algorithm is theoretically proved, and the stability of the system is established through the Lyapunov stability theorem. Finally, the effectiveness of the presented optimal control scheme is verified through simulation examples.

1. INTRODUCTION

In recent years, distributed collaborative control for multi-agent systems (MASs) has attracted extensive attention in control theory and industrial engineering[1,2]. MASs consist of multiple nodes that coordinate to accomplish complex tasks[3,4]. Consensus

Check for updates

College of Information and Communication Engineering, Hainan University, Haikou 570228, Hainan, China.

Correspondence to: Dr. Haoyue Yang, College of Information and Communication Engineering, Hainan University, Haikou 570228, Hainan, China. E-mail: haoyueyang@hainanu.edu.cn

control is a fundamental problem in this field. Its main objective is to design distributed control protocols based on local neighborhood information[5-7]. The literature[7] investigated efficient communication consensus in nonlinear MASs and proposed an adaptive temporal partitioning method to reduce unnecessary information exchange while maintaining consensus performance. Distributed consensus control has been extensively implemented in various practical scenarios, such as unmanned aerial vehicle (UAV) swarm formation, smart grid synchronization, and autonomous industrial manufacturing[8,9]. Since resource allocation must be considered during task execution[10,11], the research focus on MASs has increasingly shifted toward optimal consensus control[12,13]. Specifically, a cost function is constructed and minimized to achieve optimal control. This approach ensures task execution while reducing energy consumption[13].

Physical events such as actuator degradation, sudden payload variations, and unexpected failures frequently cause random transitions within the internal dynamic parameters of the agents. Markov jump systems provide a mathematical framework to describe these structural variations. In this framework, a Markov chain dictates the transitions among different operational modes[14,15]. Recently, Markov jump systems have attracted widespread attention, and numerous results have been reported for both single systems[16,17] and MASs[18,19]. Specifically, the literature[17] focused on fault detection in fuzzy Markov switching systems and proposed an adaptive memory-triggering mechanism to overcome the limitation on communication resources. For MASs, interference attenuation and bump suppression were achieved in[19] by constructing a cost function and an error-state cost, and a balance between stability and performance of bumpless transfer control was achieved using a weighted optimization criterion. In practice, MASs are subject not only to parameter jumps but also to frequently experience changes in their communication topologies due to external environmental factors[20,21]. Consequently, MASs with switching topology have been extensively studied[22,23]. For example, the consensus problem for MASs with Markov switching topology and time-varying delays was investigated in[22], where necessary and sufficient consensus conditions were established by considering stochastic noises and two types of network-induced delays. In[23], the scholars transformed a time-delay system into a delay-free system by introducing state variables and designed a value iteration (VI) algorithm according to the inputs and outputs of the system to solve the consensus problem for MASs. Notably, most existing studies focus solely on jumping parameters in single systems or switching topology in MASs. However, current research on consensus issues in MASs under sudden changes in system parameters and communication topology switching is insufficient, which motivated the work presented in this paper.

On the other hand, practical systems must not only ensure stability during task execution but also optimize their operational costs. Consequently, optimal control problems have attracted significant attention for both single systems[24,25] and MASs[26,27]. Specifically, a novel quasi-model network was developed in[24] to solve the problem of system model pretraining, and an algorithm based on the adaptive dynamic programming (ADP) method was developed to address the optimal control problem for unknown continuous-time affine systems. Meanwhile, the literature[26] designed an optimal consensus control scheme for MASs based on the Hamilton-Jacobi-Bellman (HJB) equation and proposed a novel policy iteration (PI) algorithm for online learning of the solution of the HJB equation. At present, the core of most methods for solving optimal consensus problems is to derive a coupled HJB equation. However, obtaining analytical solutions is difficult due to factors such as the nonlinearity of the cost function kernel matrix. To overcome this challenge, various ADP-based PI[28,29] and VI[30,31] algorithms have been proposed to approximate the solutions of the HJB equation. For example, an iterative policy based on an adaptive evaluator was used in[29] to find the state-feedback, infinite-horizon optimal control solution for linear continuous-time systems, and this algorithm converges to the optimal control solution without requiring an explicit, pre-determined model of the system's internal dynamics. The literature[30] solved the linear quadratic regulator problem using the VI algorithm and proposed a discounted cost function to achieve optimal control. However, these studies overlook the impact of neighbor coupling terms (NCTs) on system performance. Therefore, an improved

ADP method is required to address the issue of NCTs, which raises two questions. (1) How to design an improved ADP method to avoid interference from NCTs? (2) How to design suitable PI and VI algorithms to derive the optimal control policy? This motivated the work presented in this paper.

Based on the previous discussion, the optimal consensus control problem of MASs satisfying both Markov jump dynamics and Markov-switching topology is solved through the ADP approach. The main novelties and contributions are summarized as follows.

• The existing studies[16,23] investigated the optimal control problem for single or multiple systems subject to jumping system parameters or communication topologies. In this paper, the problem of optimal consensus for MASs involving jumping parameters and switching topologies is discussed. Then, a cost function is formulated to evaluate the impact of these two types of switching on system performance.

• Unlike[32], the algebraic Riccati equation (ARE) cannot be successfully derived due to the influence of NCTs. This paper derives a global ARE to solve the optimal control policy under the influence of NCTs. Based on the ADP framework, we develop global policy iteration (GPI) and global value iteration (GVI) algorithms to approximate the controller gain.

The remainder of this paper is structured as follows. Section 2 shows the problem formulation and preliminaries. GPI and GVI algorithms are proposed to find optimal control policy, with the convergence of the algorithms and the stability of the system analyzed in Section 3. The validity of the theoretical results obtained through numerical simulation was verified in Section 4. Finally, Section 5 summarizes the work presented in this paper.

2. PROBLEM FORMULATION AND PRELIMINARIES

2.1. Notation

The symbol $\otimes$ represents the Kronecker product. $\mathbb{R}^n$ denotes the set of n-dimensional real vectors. $R^{n\times n}$ is the set of $n \times n$ real matrices. I_N is the N-dimensional identity matrix. $\lambda_{min}\{\cdot\}$ and $\lambda_{max}\{\cdot\}$ represent the smallest and largest eigenvalues of $\cdot$, respectively.

2.2. Graph theory

Define a directed graph $\vartheta = (\mathcal{V}, \mathcal{E}, \mathcal{A})$ to represent the interactive relationships between a leader and N followers. $\mathcal{V} = \{v_j : j = 1, \ldots, N\}$ and $\mathcal{E} = \{(v_i, v_j)|i, j \in \mathcal{V}\}$ indicate the set of nodes and the set of edges, respectively. Within this graph, the associated adjacency matrix is $\mathcal{A} = [a_{ij}] \in \mathbb{R}^{N\times N}$, where the element $a_{ij} > 0$ if the edge $(v_i, v_j) \in \mathcal{E}$ exists, with agent j acts as the neighbor of agent i under this condition, whereas $a_{ij} = 0$. The neighborhood set of v_i is defined as $N_i = \{j|(i, j) \in \mathcal{E}\}$. The degree matrix is denoted as $\mathcal{D} = \text{diag}\{d_1, d_2, \ldots, d_N\}$, where $d_i = \Sigma_{j\in N_i} a_{ij}$. Then, the Laplacian matrix is expressed as $\mathcal{L} = \mathcal{D} - \mathcal{A}$. The pinning matrix $\mathcal{B} = \text{diag}\{b_1, b_2, \ldots, b_N\}$ characterizes the connection relationship between followers and the leader, where $b_i = 1$ if agent i receives information from the leader directly; otherwise $b_i = 0$. The graph contains a directed spanning tree if a path exists from the leader to all followers.

For a time-variant topology, the dynamic communication structure at time step k is $\vartheta(k) = (\mathcal{V}, \mathcal{E}(k), \mathcal{A}(k))$, and all possible topological forms in the system constitute a finite set $\{\vartheta_1, \vartheta_2, \ldots, \vartheta_f\}$. A Markov switching signal $\varsigma(k) : [0, \infty] \to \ell \in F \triangleq \{1, 2, \ldots, f\}$ is utilized to describe the switching of the topologies. Accordingly, a_{ij}^{ℓ}, b_i^{ℓ}, and $l_{ii}^{\ell} = \Sigma_{j=1}^{N} a_{ij}$ are elements of the adjacency matrix $\mathcal{A}$, the pinning matrix $\mathcal{B}$, and the Laplacian matrix $\mathcal{L}$ in the topology ℓ, respectively. It is said that $\vartheta(k) = \vartheta_\ell$ if and only if $\varsigma(k) = \ell$. ϑ_{un} is indicated as $\vartheta_{un} = \cup_{\ell\in F}\vartheta_\ell$, which is employed to represent the union of all possible topologies. $\Omega = [\Omega_{ij}]_{f\times f}$ indicates the transition probability matrix of a Markov process, where $\Omega_{ij} = Pr\{\varsigma(k+1) = j|\varsigma(k) = i\}$, $\forall i, j \in F$. χ is the stationary distribution for $\{\varsigma(k)\}$, then $\Omega^T\chi = \chi$.

Assumption 1 Under the framework of Markov switching topology, the associated union graph is assumed to contain a directed spanning tree, with the switching signal $\varsigma(k)$ governed by an ergodic Markov chain.

2.3. Problem formulation

Consider MASs consisting of one leader and N followers. The dynamic evolution equations for each follower and the leader can be expressed as follows:

$$x_i(k+1) = A_g x_i(k) + B_g u_i(k), \tag{1a}$$

$$x_0(k+1) = A_g x_0(k), \tag{1b}$$

where $x_i(k) \in \mathbb{R}^n$ is the state of the i-th agent, $x_0(k) \in \mathbb{R}^n$ is the state of the leader and $u_i(k) \in \mathbb{R}^m$ denotes the system input. g represents a simplified symbol for a Markov chain $\{g(k)\}$, which takes values in a finite set $\mathcal{S}=\{1, 2, 3, \ldots, m\}$.

The transition probability of the Markov chain$\{g(k)\}$ is defined as

$$\pi_{gp} = \Pr\{\zeta(k+1) = p \mid \zeta(k) = g\}, \tag{2}$$

where π_{gp} is the transition probability from mode g to mode p at the next time step with $\pi_{gp} \geq 0$ for all $g, p \in S$ and $\Sigma_{p=1}{}^{m}\pi_{gp} = 1$ for each mode g.

Denote the tracking error of the local neighborhood for the i-th agent by

$$e_i(k) = \sum_{j=1}^{N} a_{ij}^{\ell}(x_i(k) - x_j(k)) + b_i^{\ell}(x_i(k) - x_0(k)). \tag{3}$$

Then, the dynamic evolution equation for the error system of agent i is expressed as follows:

$$\begin{aligned} e_i(k+1) &= \sum_{j=1}^{N} a_{ij}^{\ell}(x_i(k+1) - x_j(k+1)) + b_i^{\ell}(x_i(k+1) - x_0(k+1)) \\ &= \sum_{j=1}^{N} a_{ij}^{\ell}(A_g x_i(k) + B_g u_i(k) - A_g x_j(k) - B_g u_j(k)) + b_i^{\ell}(A_g x_i(k) + B_g u_i(k) - A_g x_0(k)) \\ &= A_g e_i(k) + (l_{ii}^{\ell} + b_i^{\ell}) B_g u_i(k) - \sum_{j=1}^{N} B_g a_{ij}^{\ell} u_j(k). \end{aligned} \tag{4}$$

Remark 1 It is worth noting that both A_g and B_g depend on the Markov jump process $g(k)$, while $a_{ij}{}^{\ell}$, $b_i{}^{\ell}$, and $l_{ii}{}^{\ell}$ rely on the switching signal $\varsigma(k)$ of the topology. Since the probability distribution of the signal $\varsigma(k)$ for topology switching follows a stationary distribution, the topology graph among agents is guaranteed not to remain in a fixed mode indefinitely. This property further ensures the joint connectivity of the Markov switching topology graph considered in this paper. In addition, the Markov jumping signal $g(k)$ and the Markov switching signal $\varsigma(k)$ are independent. Note that if they are not independent, the calculation of transition probabilities during modal transitions may be affected, which could in turn influence the reconstruction of the cost function and the analysis of system stability. This issue is left for future research.

To characterize the cost of the agent, the cost function of the i-th agent in mode g is defined as follows:

$$V_{i,g}(e_i(k)) = \sum_{z=k}^{\infty} (e_i^T(z) Q_{ig} e_i(z) + u_i^T(z) R_{ig} u_i(z)) = \sum_{z=k}^{\infty} C_{ig}(e_i(z), u_i(z)), \tag{5}$$

where Q_{ig} is the semidefinite matrix and R_{ig} is the positive definite matrix.

Remark 2 In the cost function (5), the utility function C_{ig} of each agent consists of two components: the tracking error (3) and the energy consumed by the control input $u_i(k)$. This paper aims to make all followers reach consensus with the leader and minimize the cost function (5) of agent i at the same time. To achieve the optimization goal of the system, it is crucial to select appropriate weight parameters in the utility function C_{ig}. According to the classical linear quadratic regulator theory[33], the weighting matrix Q_{ig} directly affects the convergence rate of the system states, while R_{ig} is closely related to the control input energy. Therefore, practical implementation requires a reasonable balance between system energy consumption and convergence performance.

Based on the above analysis, the main objective of this paper is to design an optimal controller $u_i(k)$ that enables the MASs to achieve optimal consensus.

2.4. Problem transformation

In order to solve the optimal control problem in Markov jump MASs, a cost function is defined for each agent under different modes in Equation (5). Since mode jumps can affect system performance and complicate the stability analysis, the cost function is reformulated. Meanwhile, the optimal control policy is difficult to express in a linear state-feedback form due to the NCTs. For this reason, the method of restructuring the cost function is adopted to settle the problem caused by the jumping parameter. Subsequently, a global approach is proposed to address the interference caused by NCTs.

Inspired by[16], the transition probability (2) is utilized to reconstruct the cost function for each agent, as shown below:

$$\begin{cases} V_I(e_i(k)) = \pi_{11}V_{i,1}(e_i(k)) + \pi_{12}V_{i,2}(e_i(k)) + \cdots + \pi_{1m}V_{i,m}(e_i(k)) \\ V_{II}(e_i(k)) = \pi_{21}V_{i,1}(e_i(k)) + \pi_{22}V_{i,2}(e_i(k)) + \cdots + \pi_{2m}V_{i,m}(e_i(k)) \\ \vdots \\ V_m(e_i(k)) = \pi_{m1}V_{i,1}(e_i(k)) + \pi_{m2}V_{i,2}(e_i(k)) + \cdots + \pi_{mm}V_{i,m}(e_i(k)). \end{cases} \tag{6}$$

Hence, the control problem for MASs is formulated as a multiobjective optimization problem. The above multi-objective problem is transformed into a single-objective problem by employing the weighted sum method. The reconstructed cost function is written as

$$V_i(e_i(k)) = \lambda_1 V_I(e_i(k)) + \lambda_2 V_{II}(e_i(k)) + \cdots + \lambda_m V_m(e_i(k)), \tag{7}$$

where $\lambda_g > 0$ is a weight coefficient and $\Sigma_{g=1}{}^{m}\lambda_g = 1$. λ_g represents the weights associated with the system parameter modes. Its values can be generated by using the transition probabilities π_{gp} between system modes. Moreover, in practice, users can influence system performance by adjusting λ_g.

With the Equations (6) and (7), the cost function of each agent is derived as follows:

$$\begin{aligned} V_i(e_i(k)) &= \lambda_1 V_I(e_i(k)) + \lambda_2 V_{II}(e_i(k)) + \cdots + \lambda_m V_m(e_i(k)) \\ &= (\lambda_1\pi_{11} + \lambda_2\pi_{21} + \cdots + \lambda_m\pi_{m1})\, V_{i,1}(e_i(k)) \\ &\quad + (\lambda_1\pi_{12} + \lambda_2\pi_{22} + \cdots + \lambda_m\pi_{m2})\, V_{i,2}(e_i(k)) \\ &\quad + \cdots + (\lambda_1\pi_{1m} + \cdots + \lambda_m\pi_{mm})\, V_{i,m}(e_i(k)) \\ &= D_1 V_{i,1}(e_i(k)) + D_2 V_{i,2}(e_i(k)) + \cdots + D_m V_{i,m}(e_i(k)) \\ &= \sum_{g=1}^{m} D_g \sum_{z=k}^{\infty} \left(e_i(z)^T \left(Q_{ig}\right) e_i(z) + u_i(z)^T \left(R_{ig}\right) u_i(z) \right) \end{aligned} \tag{8}$$

$$
\begin{aligned}
&= e_i(k)^T Q_i e_i(k) + u_i(k)^T \mathcal{R}_i u_i(k) + \sum_{p=1}^{m} D_g V_{i,g}(e_i(k+1)) \\
&= e_i(k)^T Q_i e_i(k) + u_i(k)^T \mathcal{R}_i u_i(k) + V_i(e_i(k+1)),
\end{aligned}
$$

where $D_g = \Sigma_{p=1}^{m} \lambda_p \pi_{pg} > 0$. $c_{ig}(e_i(k), u_i(k)) = e_i(k)^T Q_i e_i(k) + u_i(k)^T \mathcal{R}_i u_i(k)$ is the utility function, where $Q_i = \Sigma_{g=1}^{m} D_g Q_{ig}$ and $\mathcal{R}_i = \Sigma_{g=1}^{m} D_g R_{ig}$.

Due to the interference from the NCTs, it is difficult to directly establish a linear relationship between the controller $u_i(k)$ and the tracking error $e_i(k)$. To overcome this limitation, the global forms of the controller and error are proposed to establish a linear relationship between the controller and the error system.

The global error vector $E(k)$ and the global control vector $U(k)$ are defined as

$$
E(k) = \begin{bmatrix} e_1(k) \\ e_2(k) \\ \vdots \\ e_N(k) \end{bmatrix}, \quad U(k) = \begin{bmatrix} u_1(k) \\ u_2(k) \\ \vdots \\ u_N(k) \end{bmatrix}. \tag{9}
$$

From the Equation (4), the dynamic evolution equation of the global error system is established as follows:

$$
\begin{aligned}
E(k+1) =& \begin{bmatrix} A_g & 0 & \cdots & 0 \\ 0 & A_g & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & A_g \end{bmatrix} \begin{bmatrix} e_1(k) \\ e_2(k) \\ \vdots \\ e_N(k) \end{bmatrix} + \begin{bmatrix} (l_{11}^{\ell} + b_1^{\ell}) B_g & 0 & \cdots & 0 \\ 0 & (l_{22}^{\ell} + b_2^{\ell}) B_g & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & (l_{NN}^{\ell} + b_N^{\ell}) B_g \end{bmatrix} \begin{bmatrix} u_1(k) \\ u_2(k) \\ \vdots \\ u_N(k) \end{bmatrix} \\
& - \begin{bmatrix} 0 & a_{12}^{\ell} B_g & \cdots & a_{1N}^{\ell} B_g \\ a_{21}^{\ell} B_g & 0 & \cdots & a_{2N}^{\ell} B_g \\ \vdots & \vdots & \ddots & \vdots \\ a_{N1}^{\ell} B_g & a_{N2}^{\ell} B_g & \cdots & 0 \end{bmatrix} \begin{bmatrix} u_1(k) \\ u_2(k) \\ \vdots \\ u_N(k) \end{bmatrix} \\
=& \begin{bmatrix} A_g & 0 & \cdots & 0 \\ 0 & A_g & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & A_g \end{bmatrix} \begin{bmatrix} e_1(k) \\ e_2(k) \\ \vdots \\ e_N(k) \end{bmatrix} + \begin{bmatrix} \sum_{j=1}^{N} a_{1j}^{\ell} B_g & -a_{12}^{\ell} B_g & \cdots & -a_{1N}^{\ell} B_g \\ -a_{21}^{\ell} B_g & \sum_{j=1}^{N} a_{2j}^{\ell} B_g & \cdots & -a_{2N}^{\ell} B_g \\ \vdots & \vdots & \ddots & \vdots \\ -a_{N1}^{\ell} B_g & -a_{N2}^{\ell} B_g & \cdots & \sum_{j=1}^{N} a_{Nj}^{\ell} B_g \end{bmatrix} \begin{bmatrix} u_1(k) \\ u_2(k) \\ \vdots \\ u_N(k) \end{bmatrix} \\
& + \begin{bmatrix} b_1^{\ell} B_g & 0 & \cdots & 0 \\ 0 & b_2^{\ell} B_g & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & b_N^{\ell} B_g \end{bmatrix} \begin{bmatrix} u_1(k) \\ u_2(k) \\ \vdots \\ u_N(k) \end{bmatrix} \\
=& (I_N \otimes A_g) E(k) + ((\mathcal{L}^{\ell} + \mathcal{B}^{\ell}) \otimes B_g) U(k) \\
=& \bar{A}_g E(k) + \bar{B}_{\ell g} U(k).
\end{aligned} \tag{10}
$$

where $\bar{A}_g = I_N \otimes A_g$ and $\bar{B}_{\ell g} = (\mathcal{L}^{\ell} + \mathcal{B}^{\ell}) \otimes B_g$.

3. OPTIMAL CONSENSUS CONTROL FOR MASS UNDER SWITCHING TOPOLOGY

In this section, a global cost function is derived to solve the problem of interference from NCTs in the derivation of the optimal control policy. Then, the optimization problem is converted into a minimization problem in mathematics using Bellman's optimality principle. Based on this, GPI and GVI algorithms are developed to find the optimal control policy. Finally, the monotonicity of the cost function kernel matrix is derived to prove the convergence of the algorithm, and the Lyapunov stability theorem is applied to verify the system's stability.

3.1. Optimal consensus control

From the Equations (8) and (9), the global cost function is derived as below:

$$\begin{aligned} V(E(k)) &= E^T(k)QE(k) + U^T(k)RU(k) + V(E(k+1)) \\ &= C(E(k), U(k)) + V(E(k+1)), \end{aligned} \tag{11}$$

where the matrix Q, the matrix R, and the global kernel matrix P are expressed as follows:

$$Q = \begin{bmatrix} Q_1 & 0 & \cdots & 0 \\ 0 & Q_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & Q_N \end{bmatrix}, \quad R = \begin{bmatrix} \mathcal{R}_1 & 0 & \cdots & 0 \\ 0 & \mathcal{R}_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \mathcal{R}_N \end{bmatrix}, \quad P = \begin{bmatrix} P_{11} & P_{12} & \cdots & P_{1N} \\ P_{21} & P_{22} & \cdots & P_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ P_{N1} & P_{N2} & \cdots & P_{NN} \end{bmatrix}, \tag{12}$$

where the matrix P_{ii} is the kernel matrix of the cost function $V_i(e_i(k))$ and is the positive definite matrix. Since the cost function can be expressed in a quadratic form $V(E(k)) = E^T(k)PE(k)$, Equation (11) is rewritten as

$$E^T(k)PE(k) = E^T(k)QE(k) + U^T(k)RU(k) + E^T(k+1)PE(k+1). \tag{13}$$

By Bellman's optimality principle, the optimal cost function $V^*(E(k))$ is defined as

$$V^*(E(k)) = \min_{U(k)} \left[C(E(k), U(k)) + V^*(E(k+1)) \right]. \tag{14}$$

Then, the Hamilton function is constructed as follows:

$$\begin{aligned} H[E(k), U(k)] &= E^T(k)QE(k) + U^T(k)RU(k) + E^T(k+1)PE(k+1) - E^T(k)PE(k) \\ &= E^T(k)(Q + \bar{A}_g^T P \bar{A}_g - P)E(k) + 2E^T(k)\bar{A}_g^T P \bar{B}_{\ell g} U(k) + U^T(k)(R + \bar{B}_{\ell g}^T P \bar{B}_{\ell g})U(k). \end{aligned} \tag{15}$$

Moreover, the first-order necessary condition requires that $\frac{\partial H[E(k),U(k)]}{\partial U(k)} = 0$. Furthermore, the equation $(R + \bar{B}_{\ell g}{}^T P \bar{B}_{\ell g})U(k) + \bar{B}_{\ell g}{}^T P \bar{A}_g E(k) = 0$ holds and the global optimal controller is described as

$$U^*(k) = -(R + \bar{B}_{\ell g}^T P^* \bar{B}_{\ell g})^{-1} \bar{B}_{\ell g}^T P^* \bar{A}_g E(k) = -\bar{K}_{\ell g}^* E(k), \tag{16}$$

where $\bar{K}^*_{\ell g}$ is the global optimal controller gain and P^* is the unique solution to the ARE is expressed as

$$P = Q + \bar{A}_g^T P \bar{A}_g - \bar{A}_g^T P \bar{B}_{\ell g}(R + \bar{B}_{\ell g}^T P \bar{B}_{\ell g})^{-1} \bar{B}_{\ell g}^T P \bar{A}_g. \tag{17}$$

Remark 3 The researchers in[23,28,34] derived the corresponding Hamilton function by constructing a cost function for each agent and determined the optimal control policy. A local neighborhood tracking error system is established. It is worth noting that, due to the influence of the NCTs, the methods of[23,28,34] cannot derive the ARE based on the local cost function, which makes it difficult to derive the optimal control policy. To this end, a global cost function is developed to derive the global ARE in this paper. Subsequently, GPI and GVI algorithms are proposed to obtain approximate solutions for the global ARE and thereby determine the optimal controller. Compared with the aforementioned literature, the GPI and GVI algorithms developed in this paper are intended to address the problem that the ARE cannot be established due to NCTs. In addition, since matrix P in Equation (17) exhibits nonlinear characteristics, its exact solution is difficult to obtain.

Remark 4 The problem of optimal consensus control of MASs is considered in[28], where the optimal control policy is derived as $u_i^*(k) = -\frac{\alpha}{2}(d_i + b_i)R_{ii}^{-1}B_i^{\mathrm{T}}\frac{\partial V_i^*(e_i(k+1))}{\partial e_i(k+1)}$. By expanding the gradient section, one can obtain

$$u_i^*(k) = -\alpha(d_i + b_i)[R_{ii} + \alpha(d_i + b_i)^2 R_{ii}^{-1} B_i^T P_i B_i]^{-1} R_{ii}^{-1} B_i^T P_i [Ae_i(k) - \sum_{j \in \mathcal{N}_i} a_{ij} B_j u_j(k)].$$

Notably, the above equation contains the term $\Sigma_{j\in N_i} a_{ij}B_j u_j(k)$. This causes $e_j(k)$ (assume $u_j(k) = K_j e_j(k)$) to appear when converting the Bellman equation through a quadratic value function, thereby preventing the successful derivation of the ARE. To tackle this issue, this paper defines the global cost function and successfully derives the global ARE suitable for the system.

3.2. GPI algorithm and GVI algorithm

To calculate the solution of the global ARE, GPI and GVI algorithms are proposed below. Proofs of the convergence of the two algorithms will be provided in Section 3.3.

Combining the Equations (13) and (16), the Lyapunov equation is given by

$$P = Q + \bar{K}_{\ell g}^T R \bar{K}_{\ell g} + (\bar{A}_g - \bar{B}_{\ell g}\bar{K}_{\ell g})^T P (\bar{A}_g - \bar{B}_{\ell g}\bar{K}_{\ell g}). \tag{18}$$

Definition 1 (Admissible Control) If there exists a control policy $U(k)$ that ensures the cost function (11) is bounded and makes the system (10) stable, then this policy is an admissible control.

The specific implementation details of the proposed GPI and GVI algorithms are detailed in Algorithm 1 and Algorithm 2, respectively.

Algorithm 1 GPI Algorithm

1) Initialization: Define l as the iteration index and set $l = 0$. Start from an admissible controller gain $\bar{K}_{\ell g}^0$. Select a sufficiently small calculation accuracy ρ as the convergence criterion.
2) Policy Evaluation: The iterative kernel matrix P^l is updated by

$$P^l = Q + (\bar{K}_{\ell g}^l)^T R \bar{K}_{\ell g}^l + (\bar{A}_g - \bar{B}_{\ell g}\bar{K}_{\ell g}^l)^T P^l (\bar{A}_g - \bar{B}_{\ell g}\bar{K}_{\ell g}^l). \tag{19}$$

3) Policy Improvement: The iterative controller gain $\bar{K}_{\ell g}^{l+1}$ is updated as the following scheme:

$$\bar{K}_{\ell g}^{l+1} = (R + \bar{B}_{\ell g}^T P^l \bar{B}_{\ell g})^{-1} \bar{B}_{\ell g}^T P^l \bar{A}_g. \tag{20}$$

4) If $||P^l - P^{l-1}|| \le \rho$, terminate the iteration. Otherwise, set $l = l + 1$ and return to 2) .
5) Return the optimal controller gain $\bar{K}_{\ell g}^* = \bar{K}_{\ell g}^l$ and the kernel matrix $P^* = P^l$.

Algorithm 2 GVI Algorithm

1) Initialization: Define l as the iteration index and set $l = 0$. Start from an arbitrary matrix P^0 and calculate the corresponding controller gain $\bar{K}_{\ell g}^0$ via the equation (16). Select a sufficiently small calculation accuracy ρ and a maximum iteration times l_{max}.
2) Value Evaluation: The iterative kernel matrix P^{l+1} is updated based on the controller gain $\bar{K}_{\ell g}^l$ as follows:

$$P^{l+1} = Q + (\bar{K}_{\ell g}^l)^T R \bar{K}_{\ell g}^l + (\bar{A}_g - \bar{B}_{\ell g}\bar{K}_{\ell g}^l)^T P^l (\bar{A}_g - \bar{B}_{\ell g}\bar{K}_{\ell g}^l). \tag{21}$$

3) Value Improvement: The iterative controller gain $\bar{K}_{\ell g}^{l+1}$ is updated as the following scheme:

$$\bar{K}_{\ell g}^{l+1} = (R + \bar{B}_{\ell g}^T P^{l+1} \bar{B}_{\ell g})^{-1} \bar{B}_{\ell g}^T P^{l+1} \bar{A}_g. \tag{22}$$

4) If $||P^l - P^{l-1}|| \le \rho$ or $l \ge l_{max}$, terminate the iteration. Otherwise, set $l = l + 1$ and return to 2).
5) Return the optimal controller gain $\bar{K}_{\ell g}^* = \bar{K}_{\ell g}^{l+1}$ and the kernel matrix $P^* = P^{l+1}$.

Since the GPI algorithm requires the prior determination of an admissible policy, it exhibits a certain degree of conservatism. In comparison, the VI algorithm allows the utilization of any initial policy. Hence, the following algorithm is designed to solve for an approximate solution of ARE.

Remark 5 In contrast to traditional local controller $ui^*(k) = -(R_i + B_i^T P_i B_i)^{-1} B_i^T P_i^* A_i e_i(k)$, the global optimal controller is provided in the Equation (16). The reason is that it is difficult for traditional local controllers to utilize the ARE approach for solving the optimal controller, whereas the global controller in (16) mitigates the influence of NCTs. In fact, the controller $u_i(k)$ of each agent can also be derived in this paper: $u_i(k) = -K_{ii} e_i(k) - \Sigma_{j=1}^{N} K_{ij} e_j(k)$, where $K_{ii} = (l_{ii}^{\ell} + b_i^{\ell})[\mathcal{R}_i + (l_{ii}^{\ell} + b_i^{\ell})^2 B_g^T P_{ii} B_g]^{-1} B_g^T P_{ii} A_g$ and $K_{ij} = (l_{ii}^{\ell} + b_i^{\ell}) a_{ij}^{\ell} [\mathcal{R}_i + (l_{ii}^{\ell} + b_i^{\ell})^2 B_g^T P_{ii} B_g]^{-1} B_g^T P_{ii} B_g K_j$; K_j is the optimal controller gain for the j-th agent and its specific form is $u_j(k) = -K_j e_j(k)$. Furthermore, the global controller gain $\bar{K}_{\ell g}$ in (16) is expressed in the form of K_{ii} and K_{ij} as follows:

$$\bar{K}_{\ell g} = \begin{bmatrix} K_{11} & K_{12} & \cdots & K_{1N} \\ K_{21} & K_{22} & \cdots & K_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ K_{N1} & K_{N2} & \cdots & K_{NN} \end{bmatrix}.$$

3.3. Properties of iterative algorithms and stability of system

This subsection analyzes the algorithms' monotonicity and determines their convergence properties. Then, the proof of the stability for the closed-loop system is given.

Lemma 1 The iterative cost function $V^l(E(k))$ of GPI is a monotonically nonincreasing sequence, that is, the following relationship holds: $V^{l+1}(E(k)) \leq V^l(E(k))$.

Proof The matrix ψ^l is designed to be

$$\psi^l = I_N \otimes A_g - [(\mathcal{L}^{\ell} + \mathcal{B}^{\ell}) \otimes B_g] \bar{K}_{\ell g}^l = \bar{A}_g - \bar{B}_{\ell g} \bar{K}_{\ell g}^l. \tag{23}$$

Hence, the Equation (19) can be written as

$$P^l = Q + (\bar{K}_{\ell g}^l)^T R \bar{K}_{\ell g}^l + \{\psi^l\}^T P^l \{\psi^l\}. \tag{24}$$

A function of a matrix is defined as follows:

$$\Phi(P^l, K) = Q + K^T R K + (\bar{A}_g - \bar{B}_{\ell g} K)^T P^l (\bar{A}_g - \bar{B}_{\ell g} K). \tag{25}$$

According to Equation (16), the following conclusion can be drawn: $(R + \bar{B}_{\ell g}^T P^l \bar{B}_{\ell g}) \bar{K}_{\ell g}^{l+1} - \bar{B}_{\ell g} P^l \bar{A}_g = 0$.

A matrix Π is expressed as $\Pi = R + \bar{B}_{\ell g}^T P^l \bar{B}_{\ell g}$ and the Equation (25) can be rewritten as follows:

$$\Phi(P^l, K) = Q + \bar{A}_g^T P^l \bar{A}_g - (\bar{K}_{\ell g}^{l+1})^T \Pi \bar{K}_{\ell g}^{l+1} + (K - \bar{K}_{\ell g}^{l+1})^T \Pi (K - \bar{K}_{\ell g}^{l+1}). \tag{26}$$

- Assuming that $K = \bar{K}_{\ell g}^{l+1}$, the Equation (25) can be expressed as $\Phi(P^l, \bar{K}_{\ell g}^{l+1}) = Q + \bar{A}_g^T P^l \bar{A}_g - (\bar{K}_{\ell g}^{l+1})^T \Pi (\bar{K}_{\ell g}^{l+1})$;
- When K is the gain of any controller, the following condition is true:

$$\Phi(P^l, K) - \Phi(P^l, \bar{K}_{\ell g}^{l+1}) = (K - \bar{K}_{\ell g}^{l+1})^T \Pi (K - \bar{K}_{\ell g}^{l+1}). \tag{27}$$

Due to the relationship that $\Pi \succeq 0$, it concludes that $\Phi(P^l, \bar{K}_{\ell g}^{l}) \succeq \Phi(P^l, \bar{K}_{\ell g}^{l+1})$. The difference form of the kernel matrix for the cost function is expressed by

$$\Delta P^l = (\bar{K}_{\ell g}^l)^T R \bar{K}_{\ell g}^l + \{\psi^l\}^T P^l \{\psi^l\} - (\bar{K}_{\ell g}^{l+1})^T R \bar{K}_{\ell g}^{l+1} - \{\psi^{l+1}\}^T P^l \{\psi^{l+1}\} \succeq 0. \tag{28}$$

Therefore, it can be concluded that: $P^l \succeq P^{l+1}$. Using the relationship between the cost function and the kernel matrix in Equation (13), the monotonicity of the cost function $V^l(E(k))$ follows.

Lemma 2 Define the iterative cost function V^l of GVI as

$$V^{l+1}(E(k)) = C(E(k), U^l(k)) + V^l(E(k+1)). \tag{29}$$

If the kernel matrix P^{l+1} and controller gain $\bar{K}_{\ell g}{}^l$ are updated based on (21) and (22), the following statements hold:

• If $P^1 - P^0 \succ 0$, then $V^{l+1}(E(k)) \geq V^l(E(k))$.
• If $P^1 - P^0 \prec 0$, then $V^{l+1}(E(k)) \leq V^l(E(k))$.

Proof It is first demonstrated that the cost function sequence is nondecreasing under the assumption $P^1 - P^0 \succ 0$. Define a new sequence γ^l with $\gamma^0(E(k)) = V^0(E(k)) = E^T(k)P^0E(k) = 0$ as follows:

$$\gamma^{l+1}(E(k)) = C(E(k), U^{l+1}(k)) + \gamma^l(E(k+1)). \tag{30}$$

Due to $P^1 - P^0 \succ 0$, it can derive $E^T(k)(P^1 - P^0)E(k) \geq 0$. From $\gamma^0(E(k)) = V^0(E(k))$, we have

$$V^1(E(k)) - \gamma^0(E(k)) = V^1(E(k)) - V^0(E(k)) = E(k)^T(P^1 - P^0)E(k) \geq 0, \tag{31}$$

which means

$$V^1(E(k)) \geq \gamma^0(E(k)). \tag{32}$$

Assuming that $V^l(E(k)) \geq \gamma^{l-1}(E(k))$ is true, it follows that $V^l(E(k+1)) \geq \gamma^{l-1}(E(k+1))$ holds. Considering

$$\begin{aligned} V^{l+1}(E(k)) - \gamma^l(E(k)) &= C(E(k), U^l(k)) + V^l(E(k+1)) - C(E(k), U^l(k)) - \gamma^{l-1}(E(k+1)) \\ &= V^l(E(k+1)) - \gamma^{l-1}(E(k+1)) \geq 0, \end{aligned} \tag{33}$$

which indicates

$$V^{l+1}(E(k)) \geq \gamma^l(E(k)). \tag{34}$$

Based on the Equation (21), a function is given by

$$F(X, K) = Q + K^T RK + (\bar{A}_g - \bar{B}_{\ell g}K)^T X(\bar{A}_g - \bar{B}_{\ell g}K). \tag{35}$$

Select the matrix M as the kernel matrix of the Equation (30), the following equation holds: $\gamma^l(E(k)) = E^T(k)M^lE(k)$. By the conditions defined above, one can conclude that: $M^0 = P^0$. Subsequently, assume that the relationship between M^l and P^l satisfies $M^l \succeq P^l$ when $l \geq 1$. When condition $K = \bar{K}_{\ell g}{}^l$ is selected, the following conclusion can be obtained:

$$F(M^l, \bar{K}_{\ell g}^l) \succeq F(P^l, \bar{K}_{\ell g}^l). \tag{36}$$

From the Equation (21) and the Equation (36), the following conclusion is obtained: $M^{l+1} \succeq P^{l+1}$, which means $\gamma^{l+1}(E(k)) \geq V^{l+1}(E(k))$. Therefore, the inequality $\gamma^l(E(k)) \geq V^l(E(k))$ holds.

Then, the following conclusion can be derived:

$$V^{l+1}(E(k)) \geq \gamma^{l}(E(k)) \geq V^{l}(E(k)), \tag{37}$$

which implies

$$V^{l+1}(E(k)) \geq V^{l}(E(k)). \tag{38}$$

Hence, the first part of the proof is completed by mathematical induction.

Next, prove that the sequence of the cost function is nonincreasing by assuming $P^{1} - P^{0} \prec 0$. Define a new sequence Γ^{l} with $\Gamma^{1}(E(k)) = V^{1}(E(k)) = E^{T}(k)P^{1}E(k)$ as follows:

$$\Gamma^{l+1}(E(k)) = C(E(k), U^{l-1}(k)) + \Gamma^{l}(E(k+1)). \tag{39}$$

Due to $P^{1} - P^{0} \prec 0$, it can derive $E^{T}(k)(P^{1} - P^{0})E(k) \leq 0$. From $\Gamma^{1}(E(k)) = V^{1}(E(k))$, we have

$$\gamma^{1}E(k) - V^{0}(E(k)) = V^{1}(E(k)) - V^{0}(E(k)) = E(k)^{T}(P^{1} - P^{0})E(k) \leq 0, \tag{40}$$

which yields

$$\gamma^{1}E(k) \leq V^{0}(E(k)). \tag{41}$$

Assuming that $\Gamma^{l}(E(k)) \leq V^{l-1}(E(k))$ is true, it follows that $\Gamma^{l}(E(k+1)) \leq V^{l-1}(E(k+1))$ holds.

Considering the following equation:

$$\begin{aligned} \Gamma^{l+1}(E(k)) - V^{l}(E(k)) &= C(E(k), U^{l-1}(k)) + \Gamma^{l}(E(k+1)) - C(E(k), U^{l-1}(k)) - V^{l-1}(E(k+1)) \\ &= \Gamma^{l}(E(k+1)) - V^{l}(E(k+1)) \leq 0, \end{aligned} \tag{42}$$

which means

$$V^{l}(E(k)) \geq \Gamma^{l+1}(E(k)). \tag{43}$$

The proof of $\Gamma^{l+1}(E(k)) \geq V^{l+1}(E(k))$ is similar to the proof given above and will not be repeated here. Then, the inequality can be represented as

$$V^{l}(E(k)) \geq \Gamma^{l+1}(E(k)) \geq V^{l+1}(E(k)). \tag{44}$$

Therefore, the second part is proven to be complete using mathematical induction.

Remark 6 In[30], the monotonicity of the VI algorithm is proved by constructing two specific functions and comparing them with the cost function, but a complete proof of $\gamma^{l}(E(k)) \geq V^{l}(E(k))$ is not provided. A rigorous proof of this inequality is given in Lemma 2 of this paper. The proof shows that the monotone property of the cost function relies on the comparison between P^{1} and P^{0}. Once this relationship is established, the monotonic behavior is preserved throughout the iterations.

Theorem 1 Suppose the sequences $\bar{K}_{\ell g}^{\ l}$ and P^l are computed by Algorithm 1 and Algorithm 2. Then, the iterative controller gain $\bar{K}_{\ell g}^{\ l}$ and the iterative kernel matrix P^l converge to the optimal controller gain $\bar{K}_{\ell g}^{\ *}$ and the optimal kernel matrix P^*.

Proof Since the iterative cost function in Algorithm 1 and the iterative cost function in Algorithm 2 when $P^1 - P^0 \prec 0$ are both nonincreasing, the convergence of these two cases is analyzed first. With Lemma 1, Lemma 2, and Equation (13), the nonincreasing property of the iterative kernel matrix P^l is obtained as $P^{l+1} \preceq P^l$. Subsequently, it can be derived from Equation (18) that $P^l \succeq 0$.

As the matrix sequence $\{P^l\}$ is monotonically bounded, there exists a limit $\lim_{l\to\infty} P^l = \underline{P}$. According to Algorithms 1 and 2, $\bar{K}_{\ell g}^{\ l} = \underline{K}$ when $l \to \infty$. Then, the Equations (17) and (19) are transformed as

$$\underline{P} = Q + \underline{K}^T R \underline{K} + (A_g - \bar{B}_{\ell g}\underline{K})^T \underline{P}(\bar{A}_g - \bar{B}_{\ell g}\underline{K}). \tag{45}$$

As a result, the limit matrix $\underline{P}$ satisfies the ARE. Since the proof of case $P^1 - P^0 \succ 0$ in Lemma 2 is similar to the above process, it will not be repeated in the subsequent derivation. Under the above analysis, it is concluded that the iterative controller gain $\bar{K}_{\ell g}^{\ l}$ and iterative kernel matrix P^l converge to the optimal solutions $\bar{K}_{\ell g}^{\ *}$ and P^*, respectively.

Next, the convergence of the iterative kernel matrix P^l in Algorithm 2 when $P^1 - P^0 \succ 0$ is analyzed. From Equation (30), the following equation can be derived as

$$\begin{aligned}
\gamma^{l+1}(E(k)) =& C(E(k), U^{l+1}(k)) + \gamma^l(E(k+1)) \\
=& C(E(k), U^{l+1}(k)) + C(E(k+1), U^l(k+1)) + \gamma^{l-1}(E(k+2)) \\
& \vdots \\
=& C(E(k), U^{l+1}(k)) + C(E(k+1), U^l(k+1)) + \cdots \\
& + C(E(k+l), U^1(k+l)) + \gamma^0(E(k+l+1)) \\
=& \sum_{\kappa=0}^{l} C(E(k+\kappa), U^{l+1-\kappa}(k+\kappa)),
\end{aligned} \tag{46}$$

where $\gamma^0(E(k+l+1)) = 0$. Thus, the inequality is shown from (45) as follows:

$$\gamma^{l+1}(E(k)) \le \lim_{l\to\infty} \sum_{\kappa=0}^{l} C(E(k+\kappa), U^{l+1-\kappa}(k)) = \bar{V}. \tag{47}$$

By the Equation (37), it follows that $V^{l+1}(E(k)) \le \gamma^{l+1}(E(k)) \le \bar{V}$. Furthermore, the relationship is derived as $P^{l+1} \preceq \bar{P}$, where $\bar{P}$ satisfies the equation $\bar{V} = E^T(k)PE(k)$. Hence, it is obtained that $\bar{P}$ is an upper bound of P^l. Then, the convergence of the iterative kernel matrix P^l is proven by Equation (13) and Lemma 2. Therefore, it is concluded that the iterative controller gain $\bar{K}_{\ell g}^{\ l}$ and iterative kernel matrix P^l converge to the optimal solutions $\bar{K}_{\ell g}^{\ *}$ and P^*, respectively.

Theorem 2 Assuming that the $U^*(k)$ is provided by the Equation (16) and $V^*(E(k))$ complies with the Equation (14), then the system (10) converges to an asymptotically stable state, and the consensus of the agents is reached.

Proof Select optimal cost function $V^*(E(k))$ as a Lyapunov function, whose specific form as

$$V^*(E(k)) = E^T(k)QE(k) + (U^*(k))^T RU^*(k) + V^*(E(k+1)), \tag{48}$$

where $Q \succeq 0$ and $R \succ 0$. Subsequently, by analyzing the Equation (48), the kernel matrix P^* of $V^*(E(k))$ satisfies $P^* \succ 0$. Define the minimum eigenvalue $\underline{p} = \lambda_{min}\{P^*\} > 0$ and the maximum eigenvalue $\overline{p} = \lambda_{max}\{P^*\} > 0$ of matrix P^*, respectively. Thus, the following equation is derived:

$$\underline{p}\|E(k)\|^2 \leq V^*(E(k)) \leq \overline{p}\|E(k)\|^2. \tag{49}$$

Furthermore, from the inequality $\underline{p}||E(k)||^2 \leq V^*(E(k))$, it follows that $V^*(E(k)) \to \infty$ when $||E(k)|| \to \infty$. Then, $V^*(E(k))$ is radially unbounded.

From Equations (10), (16), and (23), the difference of the Lyapunov function is defined as

$$\begin{aligned}\mathbb{E}[\Delta V^*(E(k))] &\triangleq \mathbb{E}\left[V^*(E(k+1))\right] - \mathbb{E}\left[V^*(E(k))\right] \\ &= E^T(k)\left[(\psi^*)^T P^* \psi^* - P^*\right] E(k) \\ &= -E^T(k)\left[Q + (\bar{K}^*_{\ell g})^T R \bar{K}^*_{\ell g}\right] E(k) \\ &= -E^T(k)\bar{W}^*_{\ell g} E(k).\end{aligned} \tag{50}$$

In particular, the above conditions imply that $\bar{W}_{\ell g}^{\ *}$ is a positive definite matrix and the minimum eigenvalue of $\bar{W}_{\ell g}^{\ *}$ is defined as $\underline{w} = \lambda_{min}\{\bar{W}\}_{\ell g}^{\ *} > 0$. Then, the Equation (50) is expressed as the following form:

$$\Delta V^*(E(k)) \leq -\underline{w}\|E(k)\|^2 \leq 0. \tag{51}$$

Combining Equation (51) and the Lyapunov stability theorem, the state of the error system (10) is asymptotically stable.

4. SIMULATION EXAMPLES

This section verifies the developed algorithms and the designed controller through two simulation examples. In particular, Example 1 demonstrates the numerical simulation and compares it with methods in the existing literature. Then, Example 2 validates the effectiveness of the proposed approach in practical systems.

Example 1: Consider a MAS consisting of one leader and six followers. Specifically, the system matrices are shown by

$$A_1 = \begin{bmatrix} 0.9898 & 0.0198 \\ -0.0198 & 0.9898 \end{bmatrix}, \quad B_1 = \begin{bmatrix} 0.1 & 0.15 \\ 0.1 & 0 \end{bmatrix}, \quad A_2 = \begin{bmatrix} 1.0119 & 0.0202 \\ -0.0202 & 1.0119 \end{bmatrix}, \quad B_2 = 0.6B_1.$$

Thereafter, the transition probability (2) is chosen in matrix form as $[\pi_{gp}]_{2\times 2} = \begin{bmatrix} 0.5 & 0.5 \\ 0.6 & 0.4 \end{bmatrix}$. Assuming that the system is more likely to be in Mode 1, the weight coefficient is specified as $\lambda_1 = 0.7$ and $\lambda_2 = 0.3$. Set the initial status value for leader and six followers as $x_0(0) = [8, -6]^T$, $x_1(0) = [20, -15]^T$, $x_2(0) = [-15, 18]^T$, $x_3(0) = [18, -12]^T$, $x_4(0) = [-12, 20]^T$, $x_5(0) = [22, -18]^T$, and $x_6(0) = [-18, 22]^T$. In addition, the weight matrices in the cost functions of each agent are selected as follows:

$$Q_{i1} = \begin{bmatrix} 5 & 0 \\ 0 & 5 \end{bmatrix}, \quad R_{i1} = \begin{bmatrix} 4 & 0 \\ 0 & 4 \end{bmatrix}, \quad Q_{i2} = \begin{bmatrix} 20 & 0 \\ 0 & 20 \end{bmatrix}, \quad R_{i2} = \begin{bmatrix} 25 & 0 \\ 0 & 25 \end{bmatrix}.$$

Select an appropriate stationary distribution $\chi > 0$ to ensure the ergodicity of the Markov parameter jump process. Correspondingly, the transition probability matrix for the topology switching process $\varsigma(k)$ is selected as $\Omega = \begin{bmatrix} 0.4 & 0.3 & 0.2 & 0.1 \\ 0.1 & 0.5 & 0.2 & 0.2 \\ 0.2 & 0.2 & 0.5 & 0.1 \\ 0.3 & 0.1 & 0.3 & 0.3 \end{bmatrix}$. Thus, the switching topology in Figure 1 is jointly connected and includes a directed spanning tree. The sampling time is $T_s = 0.01$ s. According to Algorithm 1 and Algorithm 2, the global

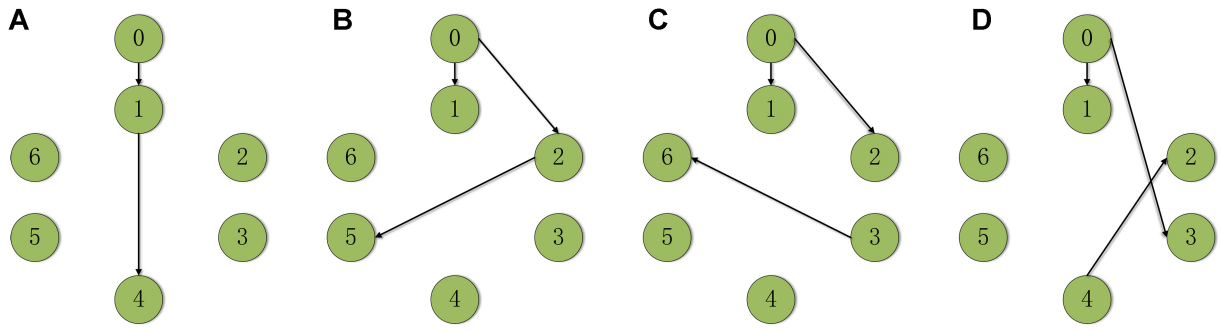


Figure 1. Switch topology diagram.

optimal controller gain $\bar{K}_{\ell g}^{*}$ and optimal global kernel matrix P^{*} are given by

$$\bar{K}_{\ell g}^{*}=\begin{pmatrix} 0.35 & 0.55 & 0.04 & 0.04 & 0 & 0 & 0.16 & 0.18 & 0.04 & 0.04 & 0 & 0 \\ 0.87 & -0.32 & 0.09 & -0.03 & 0 & 0 & 0.36 & -0.11 & 0.09 & -0.02 & 0 & 0 \\ 0.04 & 0.04 & 0.13 & 0.27 & 0 & 0 & 0.06 & 0.07 & 0.07 & 0.10 & 0 & 0 \\ 0.09 & -0.03 & 0.35 & -0.14 & 0 & 0 & 0.13 & -0.04 & 0.18 & -0.07 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0.31 & 0.52 & 0 & 0 & 0 & 0 & 0.12 & 0.14 \\ 0 & 0 & 0 & 0 & 0.82 & -0.31 & 0 & 0 & 0 & 0 & 0.28 & -0.09 \\ -0.04 & -0.04 & -0.13 & -0.27 & 0 & 0 & -0.06 & -0.07 & -0.07 & -0.10 & 0 & 0 \\ -0.09 & 0.03 & -0.35 & 0.14 & 0 & 0 & -0.13 & 0.04 & -0.18 & 0.07 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}, \ell = 4, g = 1.$$

$$P^{*}=\begin{pmatrix} 69.03 & -28.40 & 0 & 0 & 0 & 0 & 13.95 & -5.12 & 0 & 0 & 0 & 0 \\ -28.40 & 101.72 & 0 & 0 & 0 & 0 & -5.37 & 15.35 & 0 & 0 & 0 & 0 \\ 0 & 0 & 66.19 & -28.83 & 0 & 0 & 0 & 0 & 14.93 & -5.85 & 0 & 0 \\ 0 & 0 & -28.83 & 101.75 & 0 & 0 & 0 & 0 & -6.14 & 17.60 & 0 & 0 \\ 0 & 0 & 0 & 0 & 150.56 & -54.20 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & -54.20 & 194.24 & 0 & 0 & 0 & 0 & 0 & 0 \\ 13.95 & -5.37 & 0 & 0 & 0 & 0 & 127.94 & -41.55 & 0 & 0 & 0 & 0 \\ -5.12 & 15.35 & 0 & 0 & 0 & 0 & -41.55 & 150.89 & 0 & 0 & 0 & 0 \\ 0 & 0 & 14.93 & -6.14 & 0 & 0 & 0 & 0 & 112.50 & -38.21 & 0 & 0 \\ 0 & 0 & -5.85 & 17.60 & 0 & 0 & 0 & 0 & -38.21 & 138.26 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 102.87 & -35.72 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -35.72 & 129.63 \end{pmatrix}.$$

The 3D diagram in Figure 2 presents the trajectories of all agents, while Figure 3 displays the state curves for each agent. Figure 4 shows the curves of the norms of the local tracking error $e_i(k)$ and the global tracking error $E(k)$. From Figures 2-4, it can be observed that under the action of the global controller, the states of all follower agents successfully track the dynamic trajectory of the leader, and the local and global tracking errors converge to 0 in approximately 8 s. This indicates that leader-follower consensus has been achieved and demonstrates the validity of the proposed control policy.

Figures 5 and 6 plot the curve of the norm of the kernel matrix P^{l} and the norm of the control gain matrix $\bar{K}_{\ell g}^{\ l}$ for GPI and GVI algorithms during the iteration process. As illustrated in the figures, after corresponding iterations, both the GPI and GVI algorithms converge stably to the optimal values for the kernel matrix and the control gain matrix. As the number of iterations l increases, the norm of the matrix errors (P^{l+1} - P^{l}) and ($\bar{K}_{\ell g}^{\ l+1}$ - $\bar{K}_{\ell g}^{\ l}$) converge to 0. This demonstrates that the proposed Algorithm 1 and Algorithm 2 effectively find an optimal control policy for MASs. The proposed control policy and algorithm have been repeatedly validated as effective.

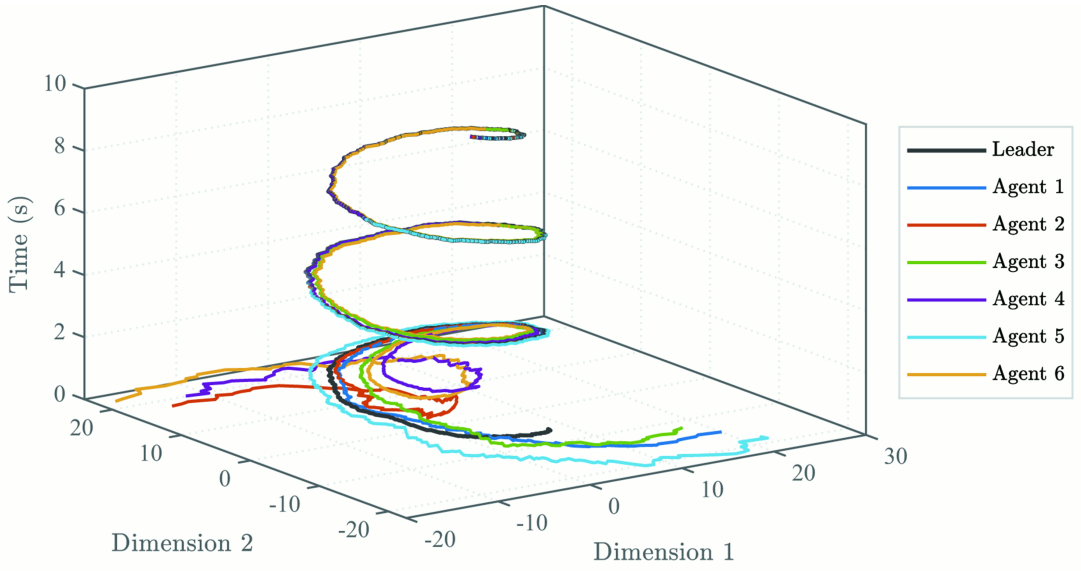


Figure 2. 3D diagram of system states under Markov jump of the parameter and Markov switching topology.

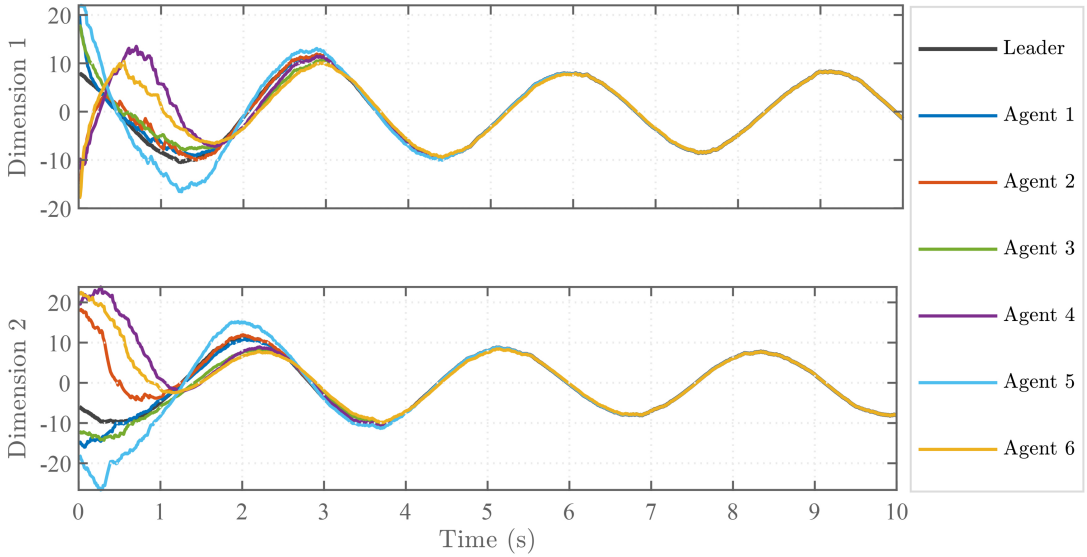


Figure 3. The state curve of each agent under the optimal controller.

Table 1 compares the characteristics of the presented approach in this paper with those of existing studies[16,23,28]. It should be noted that × indicates that the literature considers NCTs in the system, but the influence of NCTs cannot be eliminated through the ARE framework, making it impossible to derive the optimal controller. By contrast, √ denotes the case where the literature considers NCTs in the system, but the influence of NCTs can be eliminated through the ARE framework. Prior research has proposed several algorithms for finding approximate solutions to the optimal controller. Nevertheless, the optimal consensus problem for Markov jumping parameter under the Markov switching topology has not been thoroughly investigated. Specifically, the literature[16] studied the stability of Markov jumping parameter systems, while the literature[23] investigated the optimal consensus of MASs under a Markov switching topology. In this paper, GPI and GVI algorithms are proposed to successfully calculate an approximate solution for the optimal controller in both of these cases. Notably, the methods in[23,28] are affected by NCTs when solving for the ARE. Thus, they are unable to successfully derive the ARE, and the optimal control policy cannot be

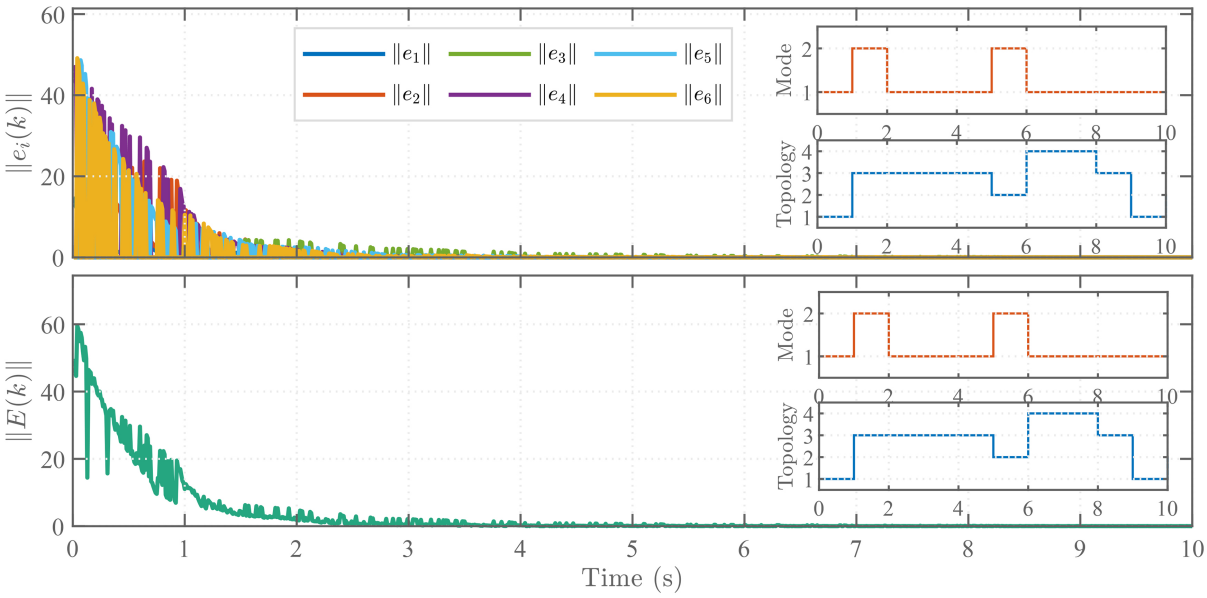


Figure 4. The norms of $e_i(k)$ and $E(k)$ under jumping signals $g(k)$ and $\varsigma(k)$.

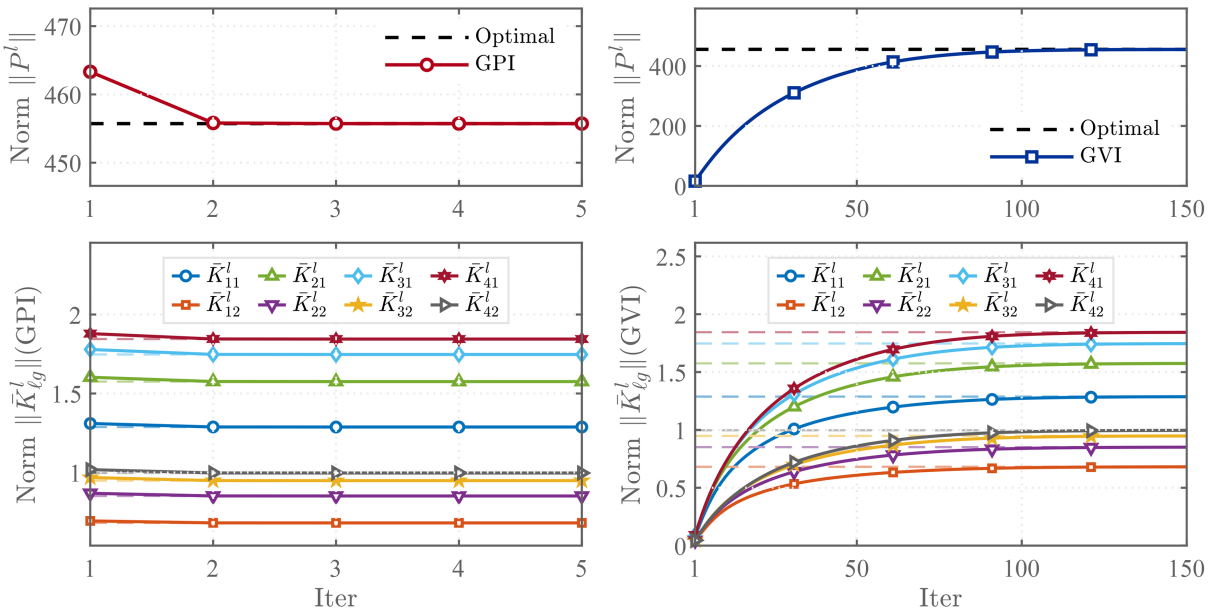


Figure 5. Convergence of GPI and GVI algorithms. GPI: Global policy iteration; GVI: global value iteration.

computed. The global ARE using the global method is successfully derived in this paper. Based on this, two global algorithms are proposed for approximating the optimal control policy.

Then, Table 2 describes the consensus comparison between this paper and the literature[23,35,36] under Markov jumping parameter and Markov switching topology, as well as the optimality comparison under the ARE framework. In contrast, although the method employed in the literature[23] achieves consensus under Markov jumping parameter and Markov switching topology, it fails to achieve optimality within the ARE framework. Meanwhile, the methods used in the literature[35,36] cannot achieve consensus under Markov jumping parameter and Markov switching topology. By comparison, the method proposed in this paper achieves

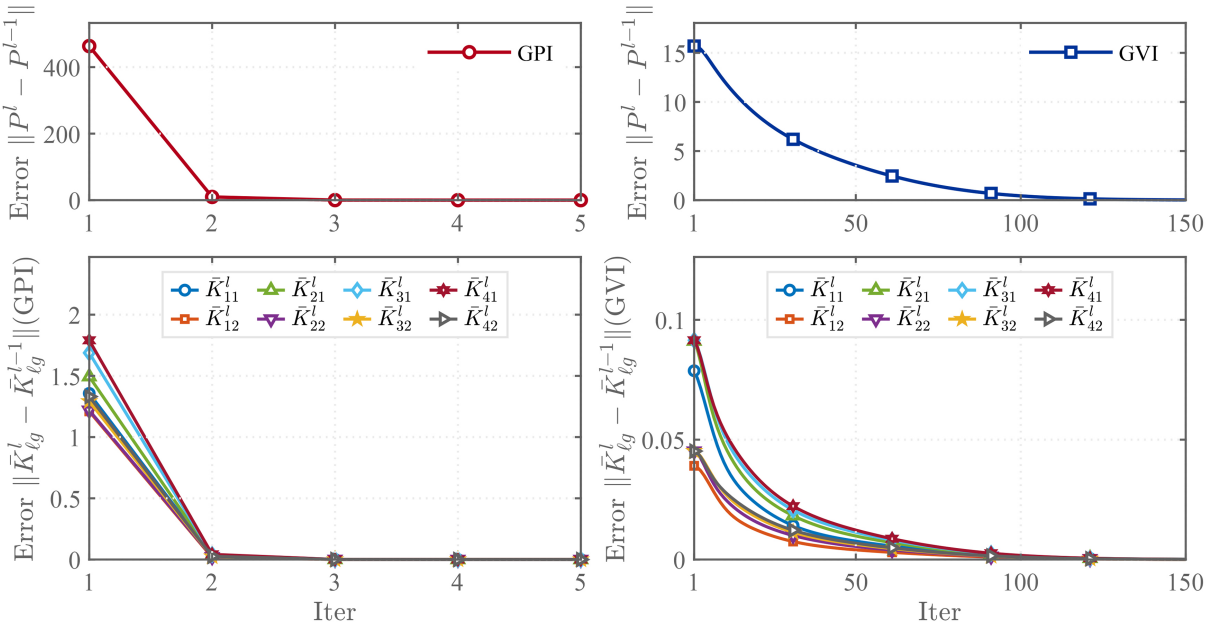


Figure 6. Iterative error in GPI and GVI algorithms. GPI: Global policy iteration; GVI: global value iteration.

Table 1. Comparison with literature

Method	Markov jumping parameter	Markov switching topology	PI	VI	NCTs
[16]	Yes	No	No	Yes	-
[23]	No	Yes	No	Yes	×
[28]	No	No	Yes	No	×
Our	Yes	Yes	Yes	Yes	√

PI: Policy iteration; VI: value iteration; NCTs: neighbor coupling terms.

Table 2. Performance comparison under Markov jumping parameter and Markov switching topology

Method	Consensus	Optimality under ARE
[23]	Yes	No
[35]	No	-
[36]	No	No
Our	Yes	Yes

ARE: Algebraic Riccati equation.

optimal consensus within the ARE framework. This further verifies that the proposed approach effectively resolves problems that have not been sufficiently investigated in earlier literature.

Subsequently, as shown in Figures 7 and 8, a comparison is conducted between the proposed algorithm and the VI algorithm based on gradient and neural network approximation presented in the literature[23], subject to both Markov jumping parameter and Markov switching topology. Correspondingly, the convergence times of the norm of the kernel matrix and the norm of the tracking error for each agent under the two algorithms are compared using identical parameter settings and illustrated in Figures 7 and 8. As shown in Figures 7 and 8, the algorithm proposed in this paper can mitigate the effects of NCTs while ensuring comparable performance to the VI algorithm in[23].

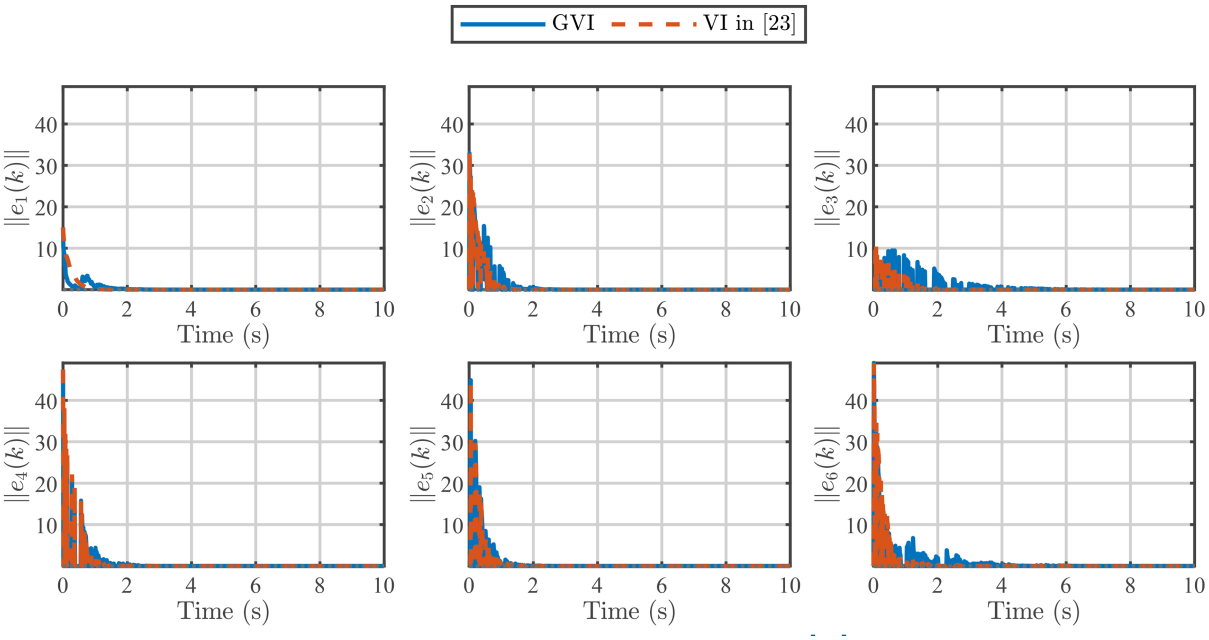


Figure 7. The state responses of each agent for VI in[23] and GVI in this paper. VI: Value iteration; GVI: global value iteration.

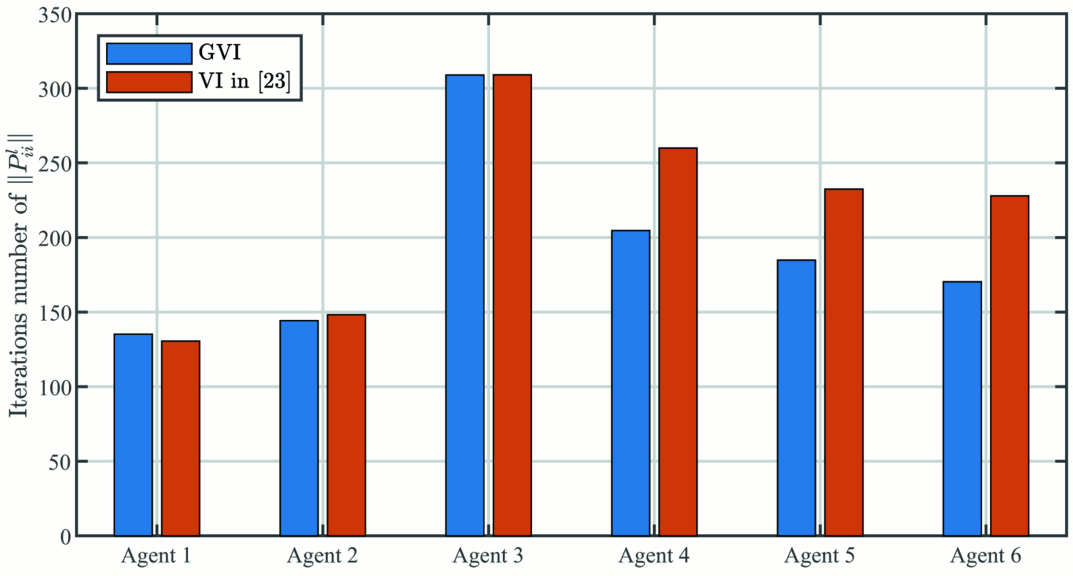


Figure 8. Iterations number of $||P_i^l||$.

Example 2: To further demonstrate the effectiveness of the proposed optimal consensus control method in practical applications, this example applies a practical example related to UAVs[37,38]. According to the literature[37] and[38], the dynamics of the agent represented by a UAV can be expressed as

$$\begin{bmatrix} \dot{s}_i \\ \ddot{s}_i \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ -\varkappa_g & -c_g \end{bmatrix} \begin{bmatrix} s_i - r_i \\ \dot{s}_i \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix} u_i, \tag{52}$$

where $\varkappa_g$ and c_g represent the damping and spring coefficients, respectively. The cooperative formation problem of UAVs is considered. s_i and $\dot{s}_i$ denote the position and velocity of the ith UAV, while r_i is introduced to characterize the desired relative displacement among the UAVs. The state of the agent is represented as $x_i = [s_i - r_i, s_i]^T$, $i = 0, 1, \ldots, N$. Next, discretization converts the continuous-time dynamics into

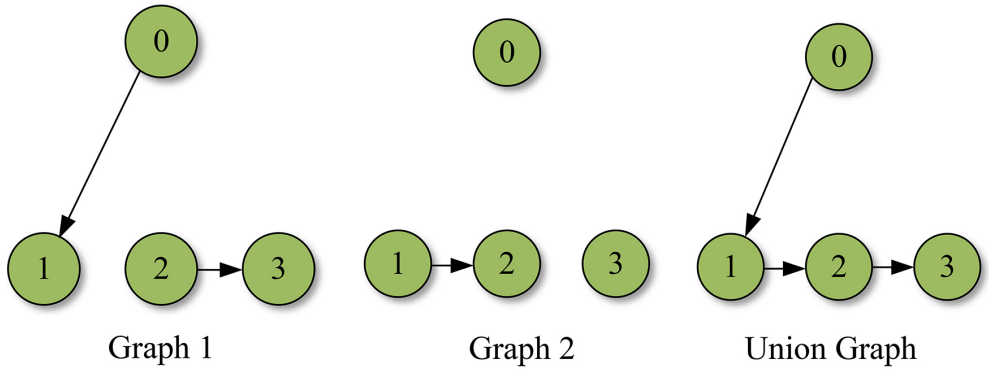


Figure 9. The switching topology graph.

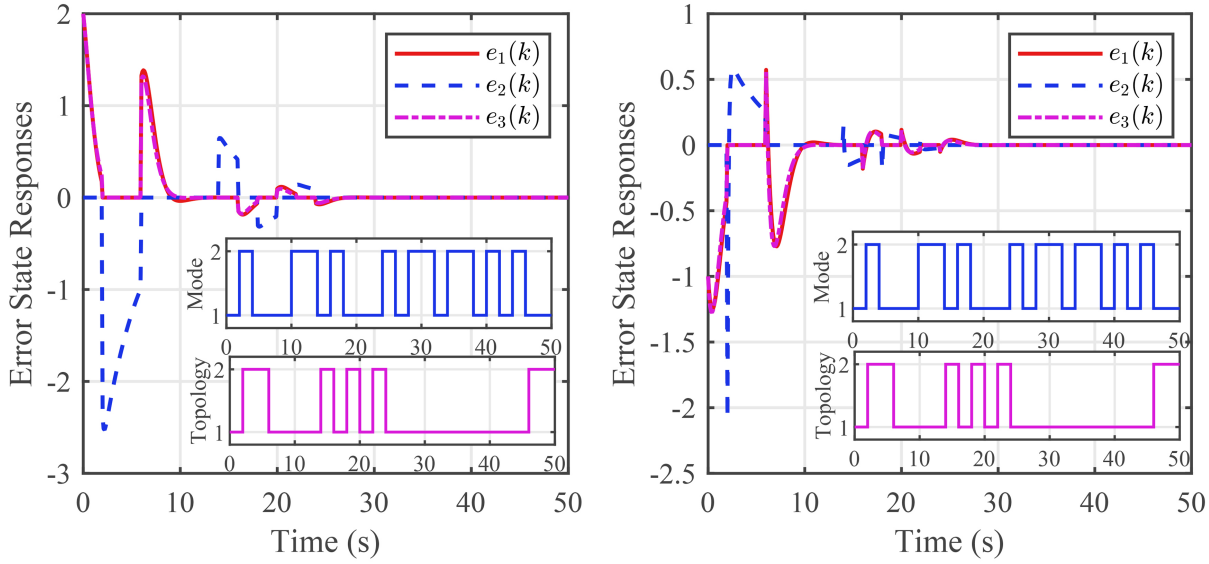


Figure 10. The local tracking error $e_i(k)$ under jumping signals $g(k)$ and $\varsigma(k)$.

system (1a). Two operating modes are adopted to represent variations in the UAV dynamics. The corresponding system matrices are chosen as

$$A_1 = \begin{bmatrix} 0.9960 & 0.0984 \\ -0.0787 & 0.9665 \end{bmatrix}, \quad B_1 = \begin{bmatrix} 0.0049 & 0 \\ 0.0984 & 0 \end{bmatrix}, \quad A_2 = \begin{bmatrix} 0.9965 & 0.0991 \\ -0.0694 & 0.9816 \end{bmatrix}, \quad B_2 = \begin{bmatrix} 0.0050 & 0 \\ 0.0991 & 0 \end{bmatrix}.$$

The jump between the two UAV modes is governed by the Markov chain $g(k)$, whose transition probability matrix is chosen as $[\pi_{pq}]_{2\times 2} = \begin{bmatrix} 0.5 & 0.5 \\ 0.6 & 0.4 \end{bmatrix}$.

This paper employs the switching topology to describe the exchange of information among agents. Figure 9 shows the switching topology diagram and the joint connect diagram. The transition probability matrix for the topology switching signal $\varsigma(k)$ is selected as $\Omega = \begin{bmatrix} 0.7 & 0.3 \\ 0.4 & 0.6 \end{bmatrix}$.

The initial states of the leader agent and the follower agent are as follows: $x_0(0) = [-2, 4]^T$, $x_1(0) = [0, 3]^T$, $x_2(0) = [2, 2]^T$, $x_3(0) = [4, 1]^T$ by choosing suitable s_i and r_i, $i = 0, 1, \ldots, N$. Using the optimal controller (16), the tracking error of local neighborhood and the global error system are depicted in Figures 10 and 11, respectively. Subsequently, Figure 12 illustrates the convergence of the GPI and GVI algorithms. From the simulation results, it can be found that both algorithms converge as the number of iterations increases.

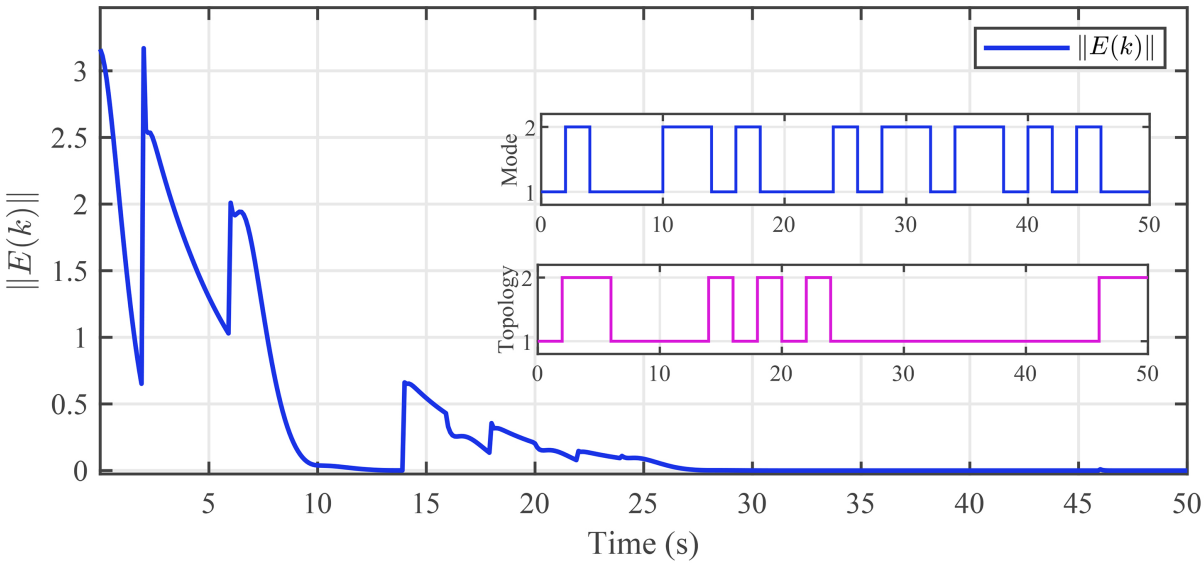


Figure 11. The Norm of global tracking error $E(k)$ under jumping signals $g(k)$ and $\varsigma(k)$.

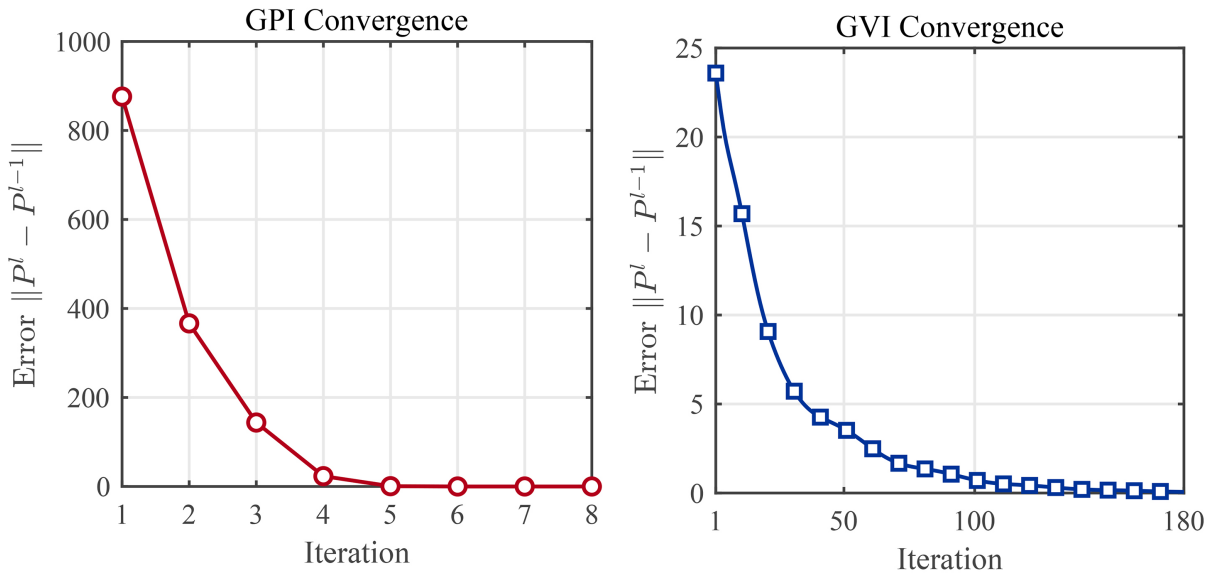


Figure 12. Iterative errors of GPI and GVI algorithms. GPI: Global policy iteration; GVI: global value iteration.

Through the simulation example, the state curves of the leader and each follower are obtained and are shown in Figure 13. As can be seen from the figure, under the designed consensus control policy, the followers always reach consensus with the leader agent.

5. CONCLUSION

The problem of optimal control for MASs with Markov jumping parameter under the Markov switching topology has been investigated. Since optimizing the analysis becomes complicated when defining a cost function for each agent in each mode, the cost function restructuring approach has been employed. Then, a global cost function is constructed to alleviate the influence of NCTs on the system performance and the global ARE is derived. Based on this, GPI and GVI algorithms inspired by ADP framework are proposed to

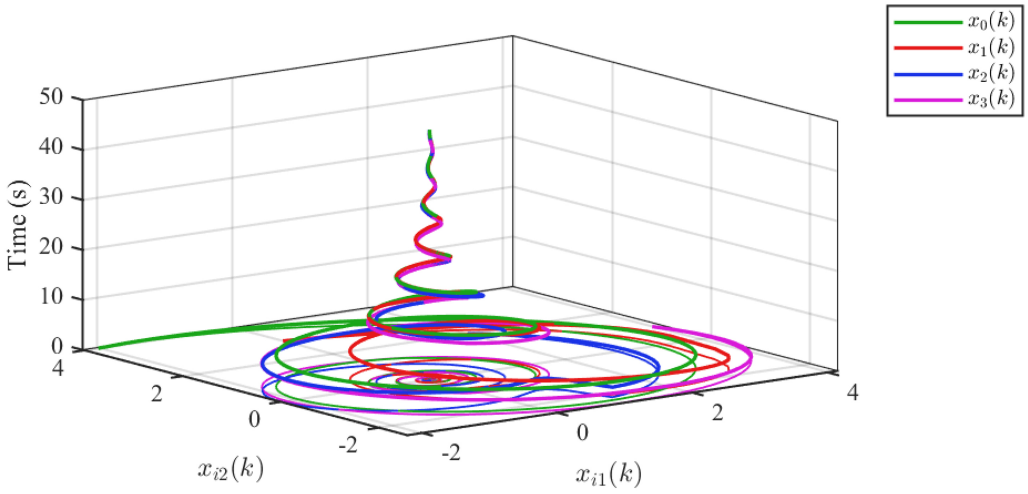


Figure 13. The state curves of all agents in the 3D plane.

approximate the optimal controller gain. Subsequently, the convergence of the algorithms and the stability of the system have been proven. Finally, simulations demonstrate that the proposed control policy and algorithms achieve comparable performance to reported in the existing literature and mitigate the impact of NCTs, while ensuring optimal consensus under Markov jumping parameter and Markov switching topology.

In future research, the topic of optimal consensus in heterogeneous MASs is of great significance. The issue of interdependence between two Markov jump processes also warrants further study. In addition, the practical system may suffer from cyber attacks (e.g., denial-of-service attacks and false data injection attacks). This means that it is important to study the impact of cyber attacks on the performance of MASs.

DECLARATIONS

Authors' contributions

Made substantial contributions to mathematical derivations, algorithm design, and simulation, wrote and edited the original manuscript: Liu, B.

Provided key feedback for this study and assisted in the completion of the simulation: Xing, W.

Conceived and proposed this study: Yang, H.

Checked the results and assisted in the improvement of the manuscript: Zhang, J.

Availability of data and materials

All data used to support the findings of this study are presented in the simulation section.

AI and AI-assisted tools statement

Not applicable.

Financial support and sponsorship

This work was supported by National Natural Science Foundation of China (62463007, 62463005, and 62563006), Natural Science Foundation of Hainan (625RC710 and 625MS047), Postdoctoral Research Funding of Hainan Province (2025-79), and Postgraduate Innovative Research Funding of Hainan Province (Hyb2025-22 and Hyb2025-31).

Conflicts of interest

All authors declared that there are no conflicts of interest.

Ethical approval and consent to participate.

Not applicable.

Consent for publication.

Not applicable.

Copyright

© The Author(s) 2026.

REFERENCES

1. Olfati-Saber, R.; Fax, J. A.; Murray, R. M. Consensus and cooperation in networked multi-agent systems. *Proc. IEEE.* **2007**, *95*, 215-33. DOI
2. Liang, Z.; Cao, J.; Jiang, S.; Xu, H. Hierarchical reinforcement learning with partner modeling for distributed multi-agent cooperation. *IEEE. Trans. Parallel. Distrib. Syst.* **2024**, 1-13. DOI
3. Peng, C.; Zhang, A.; Li, J. Neuro-adaptive cooperative control for high-order nonlinear multi-agent systems with uncertainties. *Int. J. Appl. Math. Comput. Sci.* **2021**, *31*, 635-45. DOI
4. An, R.; Wang, Y.; Zhao, Y.; Zhang, J. Consensus of high-order multiagent systems with binary-valued communications and switching topologies. *IEEE. Trans. Control. Netw. Syst.* **2025**, *12*, 1369-80. DOI
5. Li, B.; Zhao, L.; Wen, S. Periodic event-triggered consensus of stochastic multi-agent systems under switching topology. *AI. Sci. Eng.* **2025**, *1*, 147-56. DOI
6. He, W.; Chen, X.; Zhang, M.; Sun, Y.; Sekiguchi, A.; She, J. Data-driven optimal consensus control for switching multiagent systems via joint communication graph. *IEEE. Trans. Ind. Inf.* **2024**, *20*, 5959-68. DOI
7. Zhang, L.; Li, Y.; Ahn, C. K.; Liang, H. Adaptive temporal partitioning for communication-efficient consensus in nonlinear multiagent systems. *IEEE. Trans. Ind. Informat.* **2026**, 1-12. DOI
8. Jin, N.; Xu, J.; Zhang, H. Distributed optimal consensus control of multiagent systems involving state and control dependent multiplicative noise. *IEEE. Trans. Autom. Control.* **2023**, *68*, 7787-94. DOI
9. Wan, X.; Guo, Y.; Wu, X. Differentially private consensus for multi-agent systems under switching topology. *IEEE. Trans. Circuits. Syst. II.* **2023**, *70*, 3499-503. DOI
10. Xing, W.; Zhao, X.; Basar, T.; Xia, W. Security investment in cyber-physical systems: stochastic games with asymmetric information and resource-constrained players. *IEEE. Trans. Autom. Control.* **2022**, *67*, 5384-91. DOI
11. Yang, C.; Zhang, X.; Zhang, L.; et al. Coordinated energy-efficient walking assistance for paraplegic patients by using the exoskeleton-walker system. *Intell. Robot.* **2024**, *4*, 107-24. DOI
12. Yang, A.; Liang, X.; Zhang, J.; Hou, Y.; Wang, N. Global optimal consensus for UAV swarms with time-varying objective functions and bounded input constraints. *IEEE. Trans. Aerosp. Electron. Syst.* **2024**, *60*, 3822-36. DOI
13. Jia, C.; Chen, F. Feasibility of optimal formation under multiplicative noise with energy and time constraints. *IEEE. Trans. Autom. Sci. Eng.* **2025**, *22*, 18522-30. DOI
14. Shen, H.; Wang, Y.; Wang, J.; Park, J. H. A fuzzy-model-based approach to optimal control for nonlinear Markov jump singularly perturbed systems: a novel integral reinforcement learning scheme. *IEEE. Trans. Fuzzy. Syst.* **2023**, *31*, 3734-40. DOI
15. Tao, Y. Y.; Che, W. W.; Wu, Z. G.; Xu, Y.; Dong, S. Asynchronous control of 2-D Markov jump roesser systems with nonideal transition probabilities. *IEEE. Trans. Cybern.* **2024**, *54*, 6643-52. DOI
16. Zhong, X.; He, H.; Zhang, H.; Wang, Z. Optimal control for unknown discrete-time nonlinear Markov jump systems using adaptive dynamic programming. *IEEE. Trans. Neural. Netw. Learn. Syst.* **2014**, *25*, 2141-55. DOI
17. Feng, L.; Yang, G. Adaptive memory-event-triggered fault detection for interval type-2 fuzzy Markov jump systems with sensor saturation. *IEEE. Trans. Fuzzy. Syst.* **2024**, *32*, 2089-101. DOI
18. Wang, X.; Park, J. H.; Yang, H. An improved protocol to consensus of delayed MASs with UNMS and aperiodic DoS cyber-attacks. *IEEE. Trans. Netw. Sci. Eng.* **2021**, *8*, 2506-16. DOI
19. Wang, J.; Dong, J. Robust predictive bipartite bumpless control for Markov jump MASs under switching topology: application to rotary-wing air vehicles. *IEEE. Trans. Veh. Technol.* **2025**, *74*, 7020-32. DOI
20. Elhamamsy, A.; Aghili, F.; Aghdam, A. G. Connectivity preservation and collision avoidance in multi-agent systems using model predictive control. *IEEE. Trans. Netw. Sci. Eng.* **2023**, *10*, 1779-91. DOI
21. Yao, J.; Li, X.; Zhang, Y.; Ji, J.; Ren, G.; Ren, Z. Obstacle avoidance control for unmanned tracked vehicle formation with communication constraints. *IEEE. Trans. Autom. Sci. Eng.* **2025**, *22*, 19735-49. DOI
22. Shang, Y. Consensus of noisy multiagent systems with Markovian switching topologies and time-varying delays. *Math. Probl. Eng.* **2015**, *2015*, 1-13. DOI
23. Ji, L.; Wang, C.; Zhang, C.; Wang, H.; Li, H. Optimal consensus model-free control for multi-agent systems subject to input delays and switching topologies. *Inf. Sci.* **2022**, *589*, 497-515. DOI

24. Han, X.; Zhao, X.; Karimi, H. R.; Wang, D.; Zong, G. Adaptive optimal control for unknown constrained nonlinear systems with a novel quasi-model network. *IEEE. Trans. Neural. Netw. Learn. Syst.* **2022**, *33*, 2867-78. DOI
25. Chen, C.; Lewis, F. L.; Li, B. Homotopic policy iteration-based learning design for unknown linear continuous-time systems. *Automatica* **2022**, *138*, 110153. DOI
26. Cai, Y.; Zhang, H.; Zhang, K.; Liu, C. Fuzzy adaptive dynamic programming-based optimal leader-following consensus for heterogeneous nonlinear multi-agent systems. *Neural. Comput. Appl.* **2020**, *32*, 8763-81. DOI
27. Yang, H.; Zhang, H.; Wang, Z.; Huang, C.; Yan, H. Fault-tolerant optimal consensus for multiagent systems: a fuzzy-based game approach. *IEEE. Trans. Aerosp. Electron. Syst.* **2024**, *60*, 6839-52. DOI
28. Peng, Z.; Zhao, Y.; Hu, J.; Ghosh, B. K. Data-driven optimal tracking control of discrete-time multi-agent systems with two-stage policy iteration algorithm. *Inf. Sci.* **2019**, *481*, 189-202. DOI
29. Vrabie, D.; Pastravanu, O.; Abu-Khalaf, M.; Lewis, F. Adaptive optimal control for continuous-time linear systems based on policy iteration. *Automatica* **2009**, *45*, 477-84. DOI
30. Wang, D.; Ren, J.; Ha, M.; Qiao, J. System stability of learning-based linear optimal control with general discounted value iteration. *IEEE. Trans. Neural. Netw. Learn. Syst.* **2023**, *34*, 6504-14. DOI PubMed
31. Ren, Y.; Wang, Q.; Duan, Z. Optimal leader-following consensus control of multi-agent systems: a neural network based graphical game approach. *IEEE. Trans. Netw. Sci. Eng.* **2022**, *9*, 3590-601. DOI
32. Zhang, H.; Zhang, J.; Yang, G.; Luo, Y. Leader-based optimal coordination control for the consensus problem of multiagent differential games via fuzzy adaptive dynamic programming. *IEEE. Trans. Fuzzy. Syst.* **2015**, *23*, 152-63. DOI
33. Lewis, F. L.; Vrabie, D. L.; Syrmos, V. L. *Optimal control*. Wiley; 2012. DOI
34. Zhang, H.; Jiang, H.; Luo, Y.; Xiao, G. Data-driven optimal consensus control for discrete-time multi-agent systems with unknown dynamics using reinforcement learning method. *IEEE. Trans. Ind. Electron.* **2017**, *64*, 4091-100. DOI
35. Ding, J.; Sheng, J.; Li, F.; Xu, J.; Ling, Q. Consensus of multi-agent Markovian jump systems under event-triggered control. In *2023 42nd Chinese Control Conference (CCC)*, Tianjin, China. Jul 24-26, 2023. IEEE; 2023. pp. 5682-6. DOI
36. Yao, Z.; Zhu, Q.; Qin, P.; Luo, M. Homotopic reinforcement learning for distributed consensus control of stochastic Markov jump multi-agent systems. *J. Supercomput.* **2025**, *81*, 7885. DOI
37. Michael, N.; Schwager, M.; Kumar, V.; Rus, D. An experimental study of time scales and stability in networked multi-robot systems. In *Experimental robotics*; Khatib, O.; Kumar, V.; Sukhatme, G.; Eds.; Berlin: Springer Berlin Heidelberg; 2014. pp. 631-43. DOI
38. Wang, X.; Yang, H.; Zhong, S. Reliable asynchronous control for delayed multi-agent systems with stochastic faults and switching topologies. *Int. J. Syst. Sci.* **2021**, *52*, 3214-25. DOI

Disclaimer/Publisher's Note: All statements, opinions, and data contained in this publication are solely those of the individual author(s) and contributor(s) and do not necessarily reflect those of OAE and/or the editor(s). OAE and/or the editor(s) disclaim any responsibility for harm to persons or property resulting from the use of any ideas, methods, instructions, or products mentioned in the content.



© The Author(s) 2026. Open Access This article is licensed under a Creative Commons Attribution 4.0 International License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, sharing, adaptation, distribution and reproduction in any medium or format, for any purpose, even commercially, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.