


Article

Integral Reinforcement-Learning-Based Optimal Containment Control for Partially Unknown Nonlinear Multiagent Systems

Qiuye Wu, Yongheng Wu and Yonghua Wang * 

School of Automation, Guangdong University of Technology, Guangzhou 510006, China

* Correspondence: wangyonghua@gdut.edu.cn

Abstract: This paper focuses on the optimal containment control problem for the nonlinear multiagent systems with partially unknown dynamics via an integral reinforcement learning algorithm. By employing integral reinforcement learning, the requirement of the drift dynamics is relaxed. The integral reinforcement learning method is proved to be equivalent to the model-based policy iteration, which guarantees the convergence of the proposed control algorithm. For each follower, the Hamilton–Jacobi–Bellman equation is solved by a single critic neural network with a modified updating law which guarantees the weight error dynamic to be asymptotically stable. Through using input–output data, the approximate optimal containment control protocol of each follower is obtained by applying the critic neural network. The closed-loop containment error system is guaranteed to be stable under the proposed optimal containment control scheme. Simulation results demonstrate the effectiveness of the presented control scheme.

Keywords: adaptive dynamic programming; integral reinforcement learning; containment control; multiagent systems; neural networks



Citation: Wu, Q.; Wu, Y.; Wang, Y. Integral Reinforcement-Learning-Based Optimal Containment Control for Partially Unknown Nonlinear Multiagent Systems. *Entropy* **2023**, *25*, 221. <https://doi.org/10.3390/e25020221>

Academic Editors: Qiang Zhang and Yifeng Zeng

Received: 20 December 2022

Revised: 21 January 2023

Accepted: 22 January 2023

Published: 23 January 2023



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Distributed coordination control of multiagent systems (MASs) has drawn expansive interest due to its potential application on agricultural irrigation [1], disaster rescue [2], microgrid scheduling [3], marine survey [4] and wireless communication [5]. The distributed coordination control aims to guarantee that all agents which exchange local information by communicating with their neighbors reach an agreement on some variables of interest [6]. Over the last decade, containment control has received increasing attention because of its remarkable performance in addressing the secure control issues, such as hazardous material treatment [7] and fire rescue [8]. The goal of containment control is to drive the followers to enter and keep within the convex hull spanned by multiple leaders. Numerous interesting and significant results of containment control have been presented. Reference [9] developed a fuzzy-observer-based backstepping control to achieve the containment of MASs. An adaptive funnel containment control was proposed in [10], where the containment errors converged to an adjustable funnel boundary. In practical applications, containment control has been developed for autonomous surface vehicles [4], unmanned aerial vehicles [11] and spacecrafts [12]. Notice that most of the aforementioned works have ignored the control performance with a minimum of energy consumption.

It is well-known that the Riccati equation or the Hamilton–Jacobi–Bellman equation (HJBE) are solved to acquire the optimal control for linear or nonlinear systems [13], respectively. In other words, the Riccati equation is a particular case of the HJBE. As a classical optimization algorithm, dynamic programming (DP) [14] is regarded as an effective way to obtain the optimal solution of the HJBE. However, as the dimension of state variables increases, the computation of the DP approach expands as a geometric series, which arouses the dilemma of the “curse of dimensionality”. With the success of AlphaGo, reinforcement learning (RL) has stimulated increasing enthusiasm from scholars

to tackle the “curse of dimensionality” problem [15]. As is synonymous with RL, adaptive DP (ADP) [16] forward-in-time-solves the optimal control problem with the aid of neural network (NN)-based approximators. Moreover, ADP has been increasingly exploited for the optimal coordination control of MASs. Reference [17] established a cooperative policy iteration (PI) algorithm to solve the differential graphical games of linear MASs. In the nonlinear case, Reference [18] investigated the consensus problem via model-based PI with a generalized fuzzy hyperbolic critic structure. An event-triggered ADP-based optimal coordination control was proposed for the communication load and the commutation consumption was reduced [19]. To tackle the optimal containment control (OCC) problem, a finite-time fault-tolerant control was proposed via model-based PI [20]. In the presence of state constraints, Reference [21] presented a proper barrier function to transform the state constraint problem into an unconstrained case, thereafter the event-triggered OCC protocols were obtained. In Reference [22], distributed RL was applied to handle an OCC problem with collision avoidance of nonholonomic mobile robots. When the accurate model of the plant is not obtained, system identification is always employed. It should be pointed out that system identification is intractable for responding to dynamic changes of systems in time, which brings inevitable identification errors.

Recently, the integral RL (IRL) method was adopted to relax the accurate model requirement of the plant by constructing the integral Bellman equation [23,24]. An actor–critic architecture was adopted to execute the IRL algorithm, in which an actor NN learned the optimal control strategy and a critic NN was devoted to approximating the optimal value function. In the presence of heterogeneous linear MASs (HLMASs), the IRL method was developed to handle the robust OCC problem [25]. An adaptive output-feedback method was developed for the containment control for HLMASs via the IRL algorithm [26]. In Reference [27], the off-policy IRL-based OCC scheme was presented for unknown HLMASs with active leaders. However, the OCC problem of the nonlinear MASs with partially unknown dynamics has rarely been investigated via the IRL method. Moreover, the actor–critic architecture requires constructing the actor NN, which makes the control structure more complex. It is crucial to develop an IRL-based OCC scheme by implementing a simplified control structure. In addition, most of the aforementioned OCC approaches ensure the weight estimation error of the critic NN is uniformly ultimately bounded (UUB) only, which may degrade the control performance. All the above concerns motivated our research.

Inspired by the aforementioned works, we developed an IRL-based OCC scheme with asymptotically stable critic structure for partially unknown nonlinear MASs. The main contributions are reflected as follows.

- (1) Different from existing control schemes [9,20], an IRL method is introduced to construct the integral Bellman equation without the system identification. Furthermore, IRL proves to be equivalent to model-based PI, which guarantees the convergence of the developed control algorithm.
- (2) The IRL-based OCC scheme is implemented by a critic-only architecture for nonlinear MASs with unknown drift dynamics, rather than by an actor–critic architecture for linear MASs [25–27]. Thus, the proposed scheme simplifies the control structure.
- (3) In contrast to the existing OCC schemes [20–22] which guarantee the weight errors to be UUB, a modified weight-updating law is presented to tune the critic NN weights, whose weight error dynamic is asymptotically stable.

This paper is organized as follows. In Section 2, graph theory and its application to the containment of MASs are outlined. In Section 3, the IRL-based OCC scheme and its convergence proof are presented for nonlinear MASs. Then, the stability of the closed-loop containment error systems is analyzed in detail. In Section 4, two simulation examples demonstrate the effectiveness of the proposed scheme. In Section 5, concluding remarks are drawn.

2. Preliminaries and Problem Description

2.1. Graph Theory

For a network with N agents, the information interactions among agents are reflected by a weighted graph $\mathcal{G} = (V, \varepsilon, \mathbb{A})$ with the nonempty finite set of nodes $V = \{v_1, \dots, v_N\}$, the edge set $\varepsilon \subseteq V \times V$ and the nonnegative weighted adjacency matrix $\mathbb{A} = [a_{ip}]$. If node v_i links to node v_p , the edge $(v_i, v_p) \in \varepsilon$ is available with $a_{ip} > 0$; otherwise, $a_{ip} = 0$. For a node v_i , the node v_p is named as a neighbor of v_i when $(v_i, v_p) \in \varepsilon$. In this way, $N_i = \{v_p \in V: (v_p, v_i) \in \varepsilon\}$ represents the set of all neighbors of v_i . Denote the Laplacian matrix as $L = D - \mathbb{A} = [l_{ip}]$, where $D = \text{diag}\{d_{11}, d_{22}, \dots, d_{NN}\}$, $d_{ii} = \sum_{p=1}^{N_i} a_{ip}$ and l_{ip} satisfies

$$l_{ip} = \begin{cases} \sum_{q=1, q \neq i}^{N_i} a_{iq}, & i = p, \\ -a_{ip}, & i \neq p. \end{cases}$$

It implies that each row sum of L equals to zero. A sequence of edges described by $(v_1, v_2), (v_3, v_4), \dots$ with $v_i \in V$ is defined as a directed path. For arbitrary $(v_i, v_p) \in V$, a directed graph is strongly connected, if there is a directed path from v_i to v_p , while the directed graph is said to contain a spanning tree if there exists a directed path from a root node to every other nodes with respect to \mathcal{G} . This paper focuses on a strongly connected digraph with a spanning tree.

2.2. Problem Description

Consider the leader–follower nonlinear MASs in the form of the graph \mathcal{G} with M leaders and N followers, where the node dynamic of the i th follower is modeled by

$$\dot{x}_i = f(x_i(t)) + g_i(x_i(t))\mu_i(t), \quad (1)$$

where $x_i \in \mathbb{R}^n$ is the state vector for the i th follower, $\mu_i \in \mathbb{R}^m$ is the control input vector, $i = 1, 2, \dots, N$, and the nonlinear functions $f(x_i) \in \mathbb{R}^n$ and $g_i(x_i) \in \mathbb{R}^{n \times m}$ represent the unknown drift dynamic and the control input matrix, respectively. Denote the global state vector as $x = [x_1^T, x_2^T, \dots, x_N^T]^T \in \mathbb{R}^{N \times n}$.

Assumption 1. $f(x_i)$ and $g_i(x_i)$ are Lipschitz continuous on the compact set Ω_i with $f(0) = 0$ and the system (1) is controllable.

Define the node dynamic of the j th leader as

$$\dot{r}_j = h_j(r_j(t)), \quad (2)$$

where $r_j \in \mathbb{R}^n$ stands for the state vector of the j th leader, $j = 1, 2, \dots, M$ and $h_j(r_j) \in \mathbb{R}^n$ satisfies Lipschitz continuity.

Definition 1 (Convex hull [8]). A set $\mathcal{C} \subseteq \mathbb{R}^{M \times n}$ is convex if for any $y_1, y_2 \in \mathcal{C}$ and $\forall \rho \in (0, 1)$, $((1 - \rho)y_1 + \rho y_2) \in \mathcal{C}$. A convex hull of a finite set $Y = \{y_1, y_2, \dots, y_M\}$ is the minimal convex set, i.e., $\text{Co}(Y) = \{\sum_{j=1}^M \rho_j y_j \mid y_j \in Y, \rho_j \in \mathbb{R}, \rho_j \geq 0, \sum_{j=1}^M \rho_j = 1\}$.

The containment control aims to find a set of distributed control protocols $\mu = \{\mu_1, \mu_2, \dots, \mu_N\}$ such that all followers stay in the convex hull formed by the leaders, i.e., $x_i(t) \rightarrow \text{Co}(Y)$ with $Y = \{r_1, r_2, \dots, r_M\}$. For the i th follower, the local neighborhood containment error e_i is formulated as

$$\begin{aligned}
e_i &= \sum_{p \in N_i} a_{ip}(x_i - x_p) + \sum_{j=1}^M b_{ij}(x_i - r_j) \\
&= d_{ii}x_i - \sum_{p \in N_i} a_{ip}x_p + \sum_{j=1}^M b_{ij}(x_i - r_j),
\end{aligned} \tag{3}$$

where $e_i \in \mathbb{R}^n$, $b_{ij} \geq 0$ represents the pinning gain. Define $B_j = \text{diag}[b_{1j}, \dots, b_{ij}, \dots, b_{Nj}] \in \mathbb{R}^{N \times N}$. In fact, the connection between the i th follower and the j th leader is available if and only if $b_{ij} > 0$. Denote the communication graph as $\mathcal{G}_x = (\mathcal{G}, x)$. The global containment error vector of \mathcal{G}_x is

$$e = (G \otimes I_n)x + ((B(I_M \otimes 1_N)) \otimes I_n)\bar{r},$$

where $e = [e_1^T, e_2^T, \dots, e_N^T]^T \in \mathbb{R}^{N \times n}$, $\bar{r} = [r_1^T, r_2^T, \dots, r_M^T]^T \in \mathbb{R}^{M \times n}$, $G = L + B(1_M \otimes I_N)$, I_n represents the n -dimension identity matrix, 1_M stands for the M -dimensional column vector whose every element equals to 1 and $B = [B_1, B_2, \dots, B_M] \in \mathbb{R}^{N \times NM}$. Considering (1), (2) and (3), for the i th follower, the local neighborhood containment error dynamic is formulated as

$$\dot{e}_i = F_i + c_i g_i(x_i)\mu_i + \sum_{p \in N_i} a_{ip} g_p(x_p)\mu_p, \tag{4}$$

where $c_i = (d_{ii} + \sum_{j=1}^M b_{ij})$ and $F_i = c_i f(x_i) - \sum_{p \in N_i} a_{ip} f(x_p) - \sum_{j=1}^M b_{ij} h_j(r_j)$. For the i th follower, the local neighborhood containment error is dominated not only by local states and local control inputs, but also by the information from its neighbors and the leaders. In order to implement the synchronization of the partially unknown nonlinear MASs (i.e., $e_i \rightarrow 0$), an IRL-based OCC scheme is designed in the next subsection.

3. IRL-Based OCC Scheme

3.1. Optimal Containment Control

For the local neighborhood containment error dynamic (4), define the cost function as

$$J_i(e_i(0)) = \int_0^\infty P_i(e_i(\xi), \mu_i(\xi), \mu_{-i}(\xi)) d\xi, \tag{5}$$

where $P_i(e_i, \mu_i, \mu_{-i}) = e_i^T Q_i e_i + \sum_{p \in \{N_i, i\}} \mu_p^T R_{ip} \mu_p$ is a utility function, $\mu_{-i} = \{\mu_p | p \in N_i\}$ represents a set of the local control protocols from the neighbors of node v_i , and $Q_i \in \mathbb{R}^{n \times n}$ and $R_{ip} \in \mathbb{R}^{m \times m}$ are the positive definite matrices.

Definition 2 (Admissible control policies [17]). The feedback control policies $\mu_i(e_i)$ ($i \in \mathcal{I}$) are defined to be admissible with respect to (5) on a compact set Ω_i , denoted by $\mu_i(e_i) \in \mathcal{A}(\Omega_i)$, if $\mu_i(e_i)$ is continuous on Ω_i with $\mu_i(0) = 0$, $\mu_i(e_i)$ stabilizes (4) on Ω_i and $J_i(e_i(0))$ is finite $\forall e_i(0) \in \Omega_i$.

Definition 3 (Nash equilibrium [17]). An N -tuple admissible control policy $\mu^*(e) = \{\mu_1^*(e_1), \mu_2^*(e_2), \dots, \mu_N^*(e_N)\}$ is said to constitute a Nash equilibrium solution in graph \mathcal{G}_x , if the following N inequalities are satisfied

$$J_i(e_i, \mu_i^*, \mu_{-i}^*) \leq J_i(e_i, \mu_i, \mu_{-i}^*), \quad i = 1, 2, \dots, N,$$

where $\mu_{-i}^* = \{\mu_1^*, \dots, \mu_{i-1}^*, \mu_{i+1}^*, \dots, \mu_N^*\}$.

This paper aims to find an N -tuple optimal admissible control policy $\mu^*(e)$ to minimize the cost function (5) for each follower such that the Nash equilibrium solution in \mathcal{G}_x (i.e., the OCC protocols) is obtained.

For arbitrary $\mu_i(e_i) \in \mathcal{A}(\Omega_i)$ of the i th follower, define the value function

$$C_i(e_i(t)) = \int_t^\infty P_i(e_i(\xi), \mu_i(\xi), \mu_{-i}(\xi)) d\xi. \quad (6)$$

When (6) is finite, then the Bellman equation is

$$0 = e_i^T Q_i e_i + \sum_{p \in \{N_i, i\}} \mu_p^T R_{ip} \mu_p + \nabla C_i^T(e_i) \left(F_i + c_i g_i(x_i) \mu_i + \sum_{p \in N_i} a_{ip} g_p(x_p) \mu_p \right), \quad (7)$$

where $V_i(0) = 0$ and $\nabla C_i(e_i) = \partial C_i(e_i) / \partial e_i$. For the i th follower, the local Hamiltonian is

$$H_i(e_i, \mu_i, \mu_{-i}, C_i(e_i)) = e_i^T Q_i e_i + \sum_{p \in \{N_i, i\}} \mu_p^T R_{ip} \mu_p + \nabla C_i^T(e_i) \left(F_i + c_i g_i(x_i) \mu_i + \sum_{p \in N_i} a_{ip} g_p(x_p) \mu_p \right).$$

Define the optimal value function as

$$C_i^*(e_i) = \min_{\mu_i \in \mathcal{A}(\Omega_i)} C_i(e_i). \quad (8)$$

According to [13], the optimal value function $C_i^*(e_i)$ satisfies the HJBE as follows

$$0 = \min_{\mu_i \in \mathcal{A}(\Omega_i)} H_i(e_i, \mu_i, \mu_{-i}, C_i^*(e_i)). \quad (9)$$

The local OCC protocol is

$$\begin{aligned} \mu_i^*(e_i) &= \arg \min_{\mu_i \in \mathcal{A}(\Omega_i)} H_i(e_i, \mu_i, \mu_{-i}, C_i^*(e_i)) \\ &= -\frac{1}{2} c_i R_{ii}^{-1} g_i^T(x_i) \nabla C_i^*(e_i). \end{aligned} \quad (10)$$

It should be mentioned that the analytical solution of the HJBE is intractable to obtain since $C_i^*(e_i)$ is unknown. According to [15], the solution of the HJBE is successively approximated through a sequence of iterations with policy evaluation

$$\begin{aligned} 0 &= e_i^T Q_i e_i + \sum_{p \in \{N_i, i\}} \mu_p^{(k-1)T} R_{ip} \mu_p^{(k-1)} \\ &\quad + \nabla C_i^{(k)T}(e_i) \left(F_i + c_i g_i(x_i) \mu_i^{(k-1)} + \sum_{p \in N_i} a_{ip} g_p(x_p) \mu_p^{(k-1)} \right), \end{aligned} \quad (11)$$

and policy improvement

$$\mu_i^{(k)} = -\frac{1}{2} c_i R_{ii}^{-1} g_i^T(x_i) \nabla C_i^{(k)}(e_i), \quad (12)$$

where (k) represents the k th iteration index with $k \in \mathbb{N}^+$.

From (11), we can see that the policy evaluation requires the accurate mathematical model of (1). However, the accurate mathematical model is always difficult to obtain in practice. To break this bottleneck, the IRL method is developed to relax the requirement of the accurate model in the policy evaluation.

3.2. Integral Reinforcement Learning

For $t_\tau > 0$, (6) can be rewritten as

$$C_i(e_i(t)) = \int_t^{t+t_\tau} \left(e_i^\top(\xi) Q_i e_i(\xi) + \sum_{p \in \{N_{i,i}\}} \mu_p^\top(\xi) R_{ip} \mu_p(\xi) \right) d\xi + C_i(e_i(t+t_\tau)). \quad (13)$$

Based on the integral Bellman Equation (13), $V_i^*(e_i)$ and μ_i^* satisfy

$$0 = \int_t^{t+t_\tau} \left(e_i^\top(\xi) Q_i e_i(\xi) + \sum_{p \in \{N_{i,i}\}} \mu_p^{*\top}(\xi) R_{ip} \mu_p^*(\xi) \right) d\xi + C_i^*(e_i(t+t_\tau)) - C_i^*(e_i(t)). \quad (14)$$

Compared to (7), the policy evaluation (14) is not required for the accurate system dynamics in (1).

Theorem 1. Let $C_i^{(k)}(e_i) \geq 0$, $C_i^{(k)}(0) = 0$ and $\mu_i^{(k)} \in \mathcal{A}(\Omega_i)$. $C_i^{(k)}(e_i)$ is the solution of the integral Bellman equation

$$0 = \int_t^{t+t_\tau} e_i^\top(\xi) Q_i e_i(\xi) d\xi + \int_t^{t+t_\tau} \sum_{p \in \{N_{i,i}\}} \mu_p^{(k-1)\top}(\xi) R_{ip} \mu_p^{(k-1)}(\xi) d\xi + C_i^{(k)}(e_i(t+t_\tau)) - C_i^{(k)}(e_i(t)), \quad (15)$$

if and only if $C_i^{(k)}(e_i)$ is the only solution of (11).

Proof of Theorem 1. Considering (11), the time derivative of $C_i^{(k)}(e_i)$ corresponding to (4) is transformed as

$$\begin{aligned} \frac{dC_i^{(k)}(e_i)}{dt} &= \nabla C_i^{(k)}(e_i) \left(F_i + c_i g_i(x_i) \mu_i^{(k-1)} + \sum_{p \in N_i} a_{ip} g_p(x_p) \mu_p^{(k-1)} \right) \\ &= -e_i^\top Q_i e_i - \sum_{p \in \{N_{i,i}\}} \mu_p^{(k-1)\top} R_{ip} \mu_p^{(k-1)}. \end{aligned} \quad (16)$$

Integrate on both sides of (16) within $[t, t+t_\tau]$, that is

$$\begin{aligned} C_i^{(k)}(e_i(t+t_\tau)) - C_i^{(k)}(e_i(t)) &= - \int_t^{t+t_\tau} e_i^\top(\xi) Q_i e_i(\xi) d\xi \\ &\quad - \int_t^{t+t_\tau} \sum_{p \in \{N_{i,i}\}} \mu_p^{(k-1)\top}(\xi) R_{ip} \mu_p^{(k-1)}(\xi) d\xi. \end{aligned} \quad (17)$$

According to the derivation of (16) and (17), if $C_i^{(k)}(e_i)$ is the solution of (11), $C_i^{(k)}(e_i)$ satisfies the integral Bellman Equation (15). Next, we verify the uniqueness of the solution $C_i^{(k)}(e_i)$.

Supposing that $Y_i^{(k)}(e_i)$ is another solution of (11) with $Y_i^{(k)}(0) = 0$. Similar to the mathematical operation of (16), we have

$$\frac{dY_i^{(k)}(e_i)}{dt} = -e_i^\top Q_i e_i - \sum_{p \in \{N_{i,i}\}} \mu_p^{(k-1)\top} R_{ip} \mu_p^{(k-1)}. \quad (18)$$

Subtracting (16) into (18) yields

$$\frac{d}{dt} \left(Y_i^{(k)}(e_i) - C_i^{(k)}(e_i) \right) = 0. \quad (19)$$

Solving (19), we have $Y_i^{(k)}(e_i) - C_i^{(k)}(e_i) = \varsigma_i$ with $\varsigma_i \in \mathbb{R}$ a real constant. For $e_i = 0$, we have $\varsigma_i = Y_i^{(k)}(0) - C_i^{(k)}(0) = 0$. That is to say, $Y_i^{(k)}(e_i) = C_i^{(k)}(e_i)$. One can derive that $C_i^{(k)}(e_i)$ is the unique solution. In summary, $C_i^{(k)}(e_i)$ is the unique solution of (15) if and only if $C_i^{(k)}(e_i)$ is the only solution of (11). \square

Theorem 1 reveals that the IRL algorithm with (15) and (12) theoretically equals to the model-based PI algorithm, whose relevant convergence analysis was provided in [15]. Hence, the IRL algorithm can be guaranteed to be convergent.

Theorem 2. *Considering the nonlinear MAS with partially unknown dynamic as (1), the local neighborhood containment error dynamic as (4) and the optimal value function $C_i^*(e_i)$ as (8), the closed-loop containment error system is guaranteed to be asymptotically stable under the local OCC protocol (10). Furthermore, the containment control is achieved with a set of the OCC protocols $\{\mu_1^*, \mu_2^*, \dots, \mu_N^*\}$ if there is a spanning tree in the directed graph.*

Proof of Theorem 2. Selecting the Lyapunov function candidate as $C_i^*(e_i)$. Combining (7), (8) and (10), then

$$\begin{aligned} \nabla C_i^{*\top}(e_i)F_i &= -\nabla C_i^{*\top}(e_i)\left(c_i g_i(x_i)\mu_i^* + \sum_{p \in N_i} a_{ip} g_p(x_p)\mu_p^*\right) \\ &\quad - e_i^\top Q_i e_i - \sum_{p \in \{N_i, i\}} \mu_p^{*\top} R_{ip} \mu_p^*. \end{aligned} \quad (20)$$

Substituting (20) into the time derivative of $V_i^*(e_i)$, then

$$\begin{aligned} \dot{C}_i^*(e_i) &= \nabla C_i^{*\top}(e_i)\left(F_i + c_i g_i(x_i)\mu_i^* + \sum_{p \in N_i} a_{ip} g_p(x_p)\mu_p^*\right) \\ &= -e_i^\top Q_i e_i - \sum_{p \in \{N_i, i\}} \mu_p^{*\top} R_{ip} \mu_p^*. \end{aligned}$$

Therefore, $\dot{C}_i^*(e_i) \leq 0$. One can conclude that the closed-loop containment error system (4) is asymptotically stable with the local OCC protocol (10). Since a spanning tree exists in the directed graph, the containment control of the nonlinear MAS with partially unknown dynamic can be achieved. \square

3.3. Critic NN Implementation

Based on the Stone–Weierstrass approximation theorem, on the compact set Ω_i , the optimal function $C_i^*(e_i)$ and its partial gradient can be established by a critic NN as

$$C_i^*(e_i) = \phi_i^{*\top} \sigma_i(e_i) + \omega_i(e_i), \quad (21)$$

$$\nabla C_i^*(e_i) = \nabla \sigma_i^\top(e_i) \phi_i^* + \nabla \omega_i(e_i), \quad (22)$$

where $\phi_i^* \in \mathbb{R}^{l_i}$ represents the ideal weight, $\sigma_i(\cdot) \in \mathbb{R}^{l_i}$ represents the activation function, l_i represents the number of hidden neurons and $\omega_i(e_i)$ stands for the reconstruction error.

Since the ideal weight vector is unknown, the approximation of $C_i^*(e_i)$ and $\nabla C_i^*(e_i)$ are expressed as

$$\begin{aligned} \hat{C}_i(e_i) &= \hat{\phi}_i^\top \sigma_i(e_i), \\ \nabla \hat{C}_i(e_i) &= \nabla \sigma_i^\top(e_i) \hat{\phi}_i, \end{aligned} \quad (23)$$

where $\nabla\sigma_i(e_i) = \partial\sigma_i(e_i)/\partial e_i$ and $\hat{\phi}_i \in \mathbb{R}^{l_i}$ represents the estimation of ϕ_i^* . Then, the local OCC protocol (10) can be approximated by

$$\hat{\mu}_i(e_i) = -\frac{1}{2}c_i R_{ii}^{-1} g_i^T(x_i) \nabla\sigma_i^T(e_i) \hat{\phi}_i. \quad (24)$$

The approximate local Hamiltonian is

$$\begin{aligned} e_{ci} = & \int_t^{t+t_\tau} \left(e_i^T(\xi) Q_i e_i(\xi) + \sum_{p \in \{N_{i,i}\}} \hat{\mu}_p^T(\xi) R_{ip} \hat{\mu}_p(\xi) \right) d\xi \\ & + \hat{\phi}_i^T \underbrace{(\sigma_i(e_i(t+t_\tau)) - \sigma_i(e_i(t)))}_{\theta_i}. \end{aligned} \quad (25)$$

Combining (14) and (21) with (25) yields

$$\begin{aligned} e_{ci} = & \int_t^{t+t_\tau} \left(e_i^T(\xi) Q_i e_i(\xi) + \sum_{p \in \{N_{i,i}\}} \hat{\mu}_p^T(\xi) R_{ip} \hat{\mu}_p(\xi) \right) d\xi \\ & - \int_t^{t+t_\tau} \left(e_i^T(\xi) Q_i e_i(\xi) + \sum_{p \in \{N_{i,i}\}} \mu_p^{*T}(\xi) R_{ip} \mu_p^*(\xi) \right) d\xi \\ & + \hat{\phi}_i^T \theta_i - \phi_i^{*T} \theta_i - \omega_i(e_i(t+t_\tau)) + \omega_i(e_i(t)) \\ = & \int_t^{t+t_\tau} \sum_{p \in \{N_{i,i}\}} (\hat{\mu}_p(\xi) + \mu_p^*(\xi))^T R_{ip} (\hat{\mu}_p(\xi) - \mu_p^*(\xi)) d\xi \\ & - \tilde{\phi}_i^T \theta_i - \omega_i(e_i(t+t_\tau)) + \omega_i(e_i(t)) \\ = & -\tilde{\phi}_i^T \theta_i + \Phi_i, \end{aligned} \quad (26)$$

where $\tilde{\phi}_i = \phi_i^* - \hat{\phi}_i$ represents the weight estimation error and $\Phi_i = \int_t^{t+t_\tau} \sum_{p \in \{N_{i,i}\}} (\hat{\mu}_p(\xi) + \mu_p^*(\xi))^T R_{ip} (\hat{\mu}_p(\xi) - \mu_p^*(\xi)) d\xi - \omega_i(e_i(t+t_\tau)) + \omega_i(e_i(t))$.

Assumption 2. Φ_i is bounded by η_i , i.e., $\|\Phi_i\| \leq \eta_i$ with $\eta_i > 0$.

In order to tune $\hat{\phi}_i$, the steepest descent algorithm is employed to minimize $E_{ci} = \frac{1}{2}e_{ci}^2$. A modified updating law of $\hat{\phi}_i$ is

$$\dot{\hat{\phi}}_i = -l_{ci} \frac{\theta_i}{(1 + \theta_i^T \theta_i)^2} (e_{ci} - \hat{\eta}_i) \quad (27)$$

where $l_{ci} > 0$ and $\hat{\eta}_i$, the estimation of η_i , can be updated by

$$\dot{\hat{\eta}}_i = l_{si} \frac{\tilde{\phi}_i^T \theta_i}{(1 + \theta_i^T \theta_i)^2}, \quad (28)$$

where $l_{si} > 0$ is a design constant. Considering (26) and (27), the weight estimation error is updated by

$$\dot{\tilde{\phi}}_i = -l_{ci} \frac{\theta_i}{(1 + \theta_i^T \theta_i)^2} (\tilde{\phi}_i^T \theta_i - \Phi_i + \hat{\eta}_i). \quad (29)$$

Theorem 3. Considering the nonlinear MAS with partially unknown dynamic as (1), the local neighborhood containment error dynamic as (4) and the critic NN with the modified updating laws (27) and (28), then $\tilde{\phi}_i$ is guaranteed to be asymptotically stable.

Proof of Theorem 3. Define $\tilde{\eta}_i = \eta_i - \hat{\eta}_i$. Choose the Lyapunov function candidate as

$$\Xi_{ci} = \frac{1}{2l_{ci}} \tilde{\phi}_i^T \tilde{\phi}_i + \frac{1}{2l_{si}} \tilde{\eta}_i^2. \quad (30)$$

According to (28), $\tilde{\eta}_i$ is updated by

$$\dot{\tilde{\eta}}_i = -l_{si} \frac{\tilde{\phi}_i^T \theta_i}{(1 + \theta_i^T \theta_i)^2}. \quad (31)$$

Considering (29) and (31), the time derivative of (30) is

$$\begin{aligned} \dot{\Xi}_{ci} &= \frac{1}{l_{ci}} \tilde{\phi}_i^T \dot{\tilde{\phi}}_i + \frac{1}{l_{si}} \tilde{\eta}_i \dot{\tilde{\eta}}_i \\ &= -\frac{\tilde{\phi}_i^T \theta_i}{(1 + \theta_i^T \theta_i)^2} (\tilde{\phi}_i^T \theta_i - \Phi_i + \hat{\eta}_i) - \frac{\tilde{\phi}_i^T \theta_i}{(1 + \theta_i^T \theta_i)^2} \tilde{\eta}_i \\ &= -\tilde{\phi}_i^T \Psi_i \tilde{\phi}_i + \frac{\tilde{\phi}_i^T \theta_i}{(1 + \theta_i^T \theta_i)^2} (\Phi_i - \hat{\eta}_i - \tilde{\eta}_i), \end{aligned} \quad (32)$$

where $\Psi_i = \theta_i \theta_i^T / (1 + \theta_i^T \theta_i)^2$. According to Assumption 2, (32) is derived as

$$\begin{aligned} \dot{\Xi}_{ci} &\leq -\lambda_{\min}(\Psi_i) \|\tilde{\phi}_i\|^2 + \frac{\tilde{\phi}_i^T \theta_i}{(1 + \theta_i^T \theta_i)^2} (\|\Phi_i\| - \eta_i) \\ &\leq -\lambda_{\min}(\Psi_i) \|\tilde{\phi}_i\|^2. \end{aligned}$$

It indicates $\dot{\Xi}_{ci} \leq 0$. Therefore, one can conclude that $\tilde{\phi}_i$ is ensured to be asymptotically stable. \square

Under the framework of the critic-only architecture, the IRL-based OCC scheme is presented. For each follower, the local neighborhood containment error (3) is established by communicating with its neighbors and the leaders. The value function of each follower is approximated by the critic NN (23), whose weights are tuned by a modified weight updating law (27). Based on (1), (3) and (23), the local OCC protocol (24) is obtained. The structural diagram of the developed IRL-based OCC scheme is shown in Figure 1.

Remark 1. In the actor–critic architecture, the optimal value function and the optimal control policy are approximated by a critic NN and an actor NN, respectively. While for the critic-only architecture, the optimal value function is approximated by a critic NN and the optimal control policy is directly obtained by combining (10) and (22). Hence, the critic-only architecture keeps the same performance as the actor–critic one. In contrast, the critic-only architecture utilizes a single critic NN only, which implies that the control structure is simplified and the computation burden is reduced.

3.4. Stability Analysis

Assumption 3. ϕ_i^* , $\tilde{\phi}_i$, $\nabla \sigma_i(\cdot)$ and $\nabla \omega_i(\cdot)$ are norm-bounded, i.e.,

$$\|\phi_i^*\| \leq \phi_{iM}, \quad \|\tilde{\phi}_i\| \leq \tilde{\phi}_{iM}, \quad \|\nabla \sigma_i(\cdot)\| \leq \bar{\sigma}_{iM}, \quad \|\nabla \omega_i(\cdot)\| \leq \bar{\omega}_{iM}, \quad \|g_i(\cdot)\| \leq \bar{g}_{iM},$$

where ϕ_{iM} , $\tilde{\phi}_{iM}$, $\bar{\sigma}_{iM}$, $\bar{\omega}_{iM}$ and \bar{g}_{iM} are positive constants.

Theorem 4. Considering the nonlinear MAS with partially unknown dynamics as (1), the local neighborhood containment error dynamic as (4), the optimal value function as (8) and the critic NN which is updated by (27) and (28), the local containment control protocol (24) can guarantee the closed-loop containment error system (4) to be UUB.

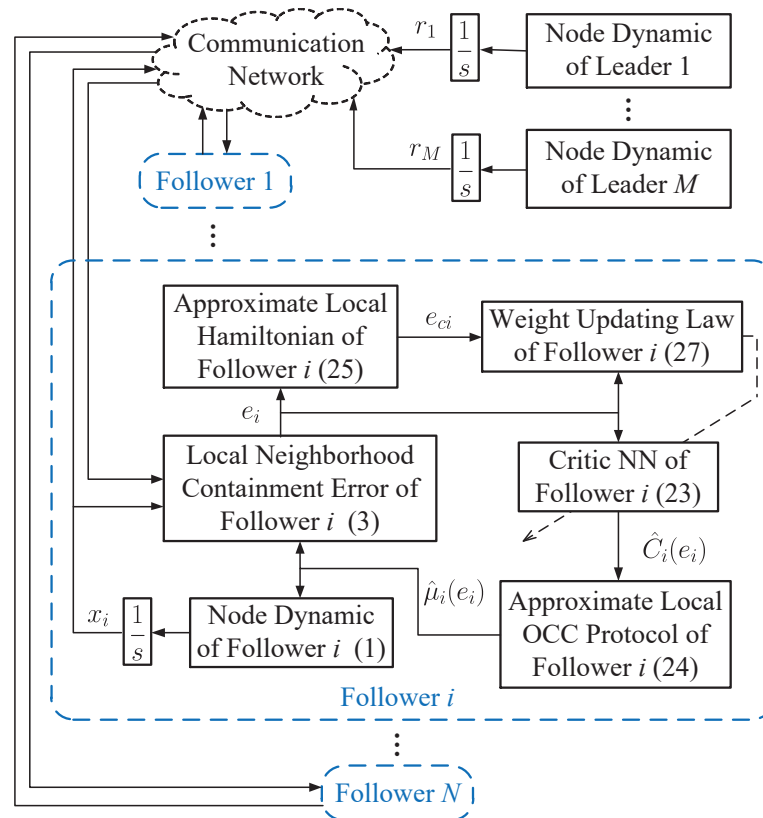


Figure 1. Structural diagram of the developed IRL-based OCC scheme.

Proof of Theorem 4. The Lyapunov function candidate is chosen as

$$\Xi_i = C_i^*(e_i). \quad (33)$$

Considering (20), (21) and Assumption 3, the time derivative of (33) corresponding to (4) is

$$\begin{aligned} \dot{\Xi}_i &= \dot{C}_i^*(e_i) \\ &= \nabla C_i^{*T}(e_i) \left(F_i + c_i g_i(x_i) \hat{\mu}_i + \sum_{p \in N_i} a_{ip} g_p(x_p) \hat{\mu}_p \right) \\ &= \nabla C_i^{*T}(e_i) \left(c_i g_i(x_i) (\hat{\mu}_i - \mu_i^*) + \sum_{p \in N_i} a_{ip} g_p(x_p) (\hat{\mu}_p - \mu_p^*) \right) - e_i^T Q_i e_i - \sum_{p \in \{N_i, i\}} \mu_p^{*T} R_{ip} \mu_p^* \\ &\leq \|\nabla C_i^{*T}(e_i)\| \left(\|c_i g_i(x_i) (\hat{\mu}_i - \mu_i^*)\| + \sum_{p \in N_i} \|a_{ip} g_p(x_p) (\hat{\mu}_p - \mu_p^*)\| \right) - \lambda_{\min}(Q_i) \|e_i\|^2 \\ &\leq (\bar{\sigma}_{iM} \phi_{iM} + \bar{\omega}_{iM}) \left(c_i \bar{g}_{iM} \|\hat{\mu}_i - \mu_i^*\| + \sum_{p \in N_i} a_{ip} \bar{g}_{pM} \|\hat{\mu}_p - \mu_p^*\| \right) - \lambda_{\min}(Q_i) \|e_i\|^2. \end{aligned} \quad (34)$$

Notice that

$$\begin{aligned} \|\hat{\mu}_i - \mu_i^*\| &= \left\| -\frac{1}{2} \mathcal{R}_{ii}^{-1} c_i g_i^T(x_i) \nabla \sigma_i^T(e_i) \hat{\phi}_i + \frac{1}{2} \mathcal{R}_{ii}^{-1} c_i g_i^T(x_i) (\nabla \sigma_i^T(e_i) \phi_i^* + \nabla \omega_i(e_i)) \right\| \\ &= \left\| \frac{1}{2} \mathcal{R}_{ii}^{-1} c_i g_i^T(x_i) (\nabla \sigma_i^T(e_i) \tilde{\phi}_i + \nabla \omega_i(e_i)) \right\| \\ &\leq \frac{c_i \bar{g}_{iM}}{2 \|\mathcal{R}_{ii}\|} (\bar{\sigma}_{iM} \bar{\phi}_{iM} + \bar{\omega}_{iM}). \end{aligned}$$

Then, (34) becomes

$$\begin{aligned} \dot{\Xi}_i \leq & (\bar{\sigma}_{iM}\phi_{iM} + \bar{\omega}_{iM}) \left(\frac{c_i^2 \bar{\sigma}_{iM}^2}{2\|\mathcal{R}_{ii}\|} (\bar{\sigma}_{iM}\bar{\phi}_{iM} + \bar{\omega}_{iM}) + \sum_{p \in N_i} \frac{c_p a_{ip} \bar{\sigma}_{pM}^2}{2\|\mathcal{R}_{pp}\|} (\bar{\sigma}_{pM}\bar{\phi}_{pM} + \bar{\omega}_{pM}) \right) \\ & - \lambda_{\min}(Q_i) \|e_i\|^2. \end{aligned} \quad (35)$$

Let $\Pi_{i1} = \frac{c_i^2 \bar{\sigma}_{iM}^2}{2\|\mathcal{R}_{ii}\|} (\bar{\sigma}_{iM}\bar{\phi}_{iM} + \bar{\omega}_{iM}) + \sum_{p \in N_i} \frac{c_p a_{ip} \bar{\sigma}_{pM}^2}{2\|\mathcal{R}_{pp}\|} (\bar{\sigma}_{pM}\bar{\phi}_{pM} + \bar{\omega}_{pM})$. Thus, (35) turns to

$$\begin{aligned} \dot{\Xi}_i & \leq \underbrace{(\bar{\sigma}_{iM}\phi_{iM} + \bar{\omega}_{iM})\Pi_{i1}}_{\Pi_{i2}} - \lambda_{\min}(Q_i) \|e_i\|^2 \\ & = \Pi_{i2} - \lambda_{\min}(Q_i) \|e_i\|^2. \end{aligned}$$

It shows $\dot{\mathcal{L}}_{i2} < 0$ if e_i lies outside the compact set

$$\Omega_{e_i} = \left\{ e_i : \|e_i\| \leq \sqrt{\frac{\Pi_{i2}}{\lambda_{\min}(Q_i)}} \right\}.$$

Therefore, the closed-loop containment error system (4) is UUB under the local containment control protocol (24). \square

Remark 2. In Assumption 1, we know that the nonlinear functions $f(x)$ and $g_i(x)$ are Lipschitz continuous on a compact set Ω_i containing the origin, $f(0) = 0$. It indicates that the developed control scheme is effective in a compact set Ω_i . If the system states are outside this compact set, this scheme might be invalid. In Theorem 4, we analyzed the system stability within such a compact set via the Lyapunov direct method, which means the closed-loop system is stable in the compact set under the developed IRL-based OCC scheme.

4. Simulation Study

This section provides two simulation examples to support the developed IRL-based OCC scheme.

4.1. Example 1

Consider a six-node graph network connected by three leader nodes. The directed topology of the graph is displayed in Figure 2.

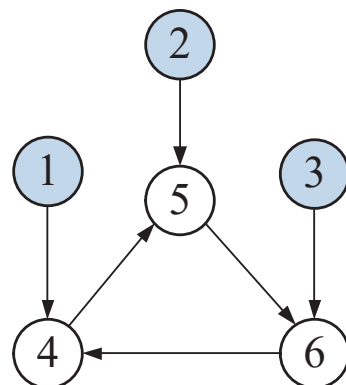


Figure 2. The directed topology of example 1.

As displayed in Figure 2, nodes 1–3 stand for the leaders 1–3 and nodes 4–6 represent the followers 1–3. In (3), the edge weights and pinning gains were set to 0.5. The node dynamic of the j th leader is described as $\dot{r}_j = \bar{A}r_j$, where $r_j = [r_{j1}, r_{j2}]^T \in \mathbb{R}^2$ represents the state vector, $j = 1, 2, 3$ and

$$\bar{A} = \begin{bmatrix} 0.1 & -1 \\ 1 & -0.1 \end{bmatrix}.$$

For the i th follower, the node dynamic is formulated as $\dot{x}_i = \bar{A}x_i + \bar{B}_i\mu_i$, where $x_i = [x_{i1}, x_{i2}]^T \in \mathbb{R}^2$ and $\mu_i \in \mathbb{R}$ with $i = 1, 2, 3$, $\bar{B}_1 = [-1.5, 1]^T$, $\bar{B}_2 = [-1, 1]^T$ and $\bar{B}_3 = [-1, -0.5]^T$. The local neighborhood containment error vector $e_i = [e_{i1}, e_{i2}]^T \in \mathbb{R}^2$ is calculated by (3).

In the simulation, $C_i(e_i)$ was reconstructed by a critic NN with a 2–5–1 structure. The activation function was described as $\sigma_i(e_i) = [e_{i1}^2, e_{i1}e_{i2}, e_{i2}^2, e_{i1}^2e_{i2}, e_{i2}^2e_{i1}]^T$. The initialization of the node dynamics were characterized as $x_1(0) = [0.50, -1.00]^T$, $x_2(0) = [1.00, -0.50]^T$, $x_3(0) = [0.80, -0.30]^T$, $r_1(0) = [0.62, 0.83]^T$, $r_2(0) = [0.45, 0.40]^T$ and $r_3(0) = [0.30, 0.22]^T$. The related parameters were chosen as $Q_i = 5I_2$, $R_{ip} = R_{ii} = 1$, $l_{ci} = 0.1$ and $l_{si} = 0.1$.

The simulation results are shown in Figures 3–5 using the developed IRL-based OCC protocols. The evolution procedure of the local neighborhood containment errors for triple followers is shown in Figure 3, which indicates that the local neighborhood containment errors were regulated to zero under the developed control protocols. Thus, the containment control of MAS could be reached. Figures 4 and 5 depict the state curves of the leaders and the followers, where all followers moved and stayed within the region formed by the envelope curves. It implies that the satisfactory performance of the containment control was acquired. The state curves of the followers and the leaders are displayed as 2-D phase plane plot in Figure 6 and the region enveloped by the three leaders v_1, v_2 and v_3 is shown at three different instants ($t = 16.0$ s, 20.3 s and 25.0 s). We can observe from Figure 6 that the followers converged to the convex hull.

4.2. Example 2

Consider the nonlinear MAS consisting of three single-link robot arms and triple leader nodes. A rigid link is attached to each robot arm via a gear train to a direct current motor [28]. In Figure 2, the directed topology among these robot arms is shown. We chose the values of all edge weights and pinning gain as 1.

The state trajectories of the leaders is given by $r_1 = [0.6 \sin(t), 0.6 \cos(t)]^T$, $r_2 = [0.4 \sin(t + \frac{\pi}{6}), 0.4 \cos(t + \frac{\pi}{6})]^T$ and $r_3 = [0.2 \sin(t - \frac{\pi}{6}), 0.2 \cos(t - \frac{\pi}{6})]^T$. The single-link robot arm for each follower can be described as

$$\mathcal{J}\ddot{z}_i + \bar{B}\dot{z}_i + \bar{M}gl \sin(z_i) = u_i, \quad (36)$$

where $\mathcal{J} = 9 \text{ kg}\cdot\text{m}^2$, $\bar{B} = 30.5$, $\bar{M} = 1 \text{ kg}$, $l = 1 \text{ m}$, $g = 9.8 \text{ m/s}^2$ and $i = 1, 2, 3$. The notations of the model (36) are defined in Table 1.

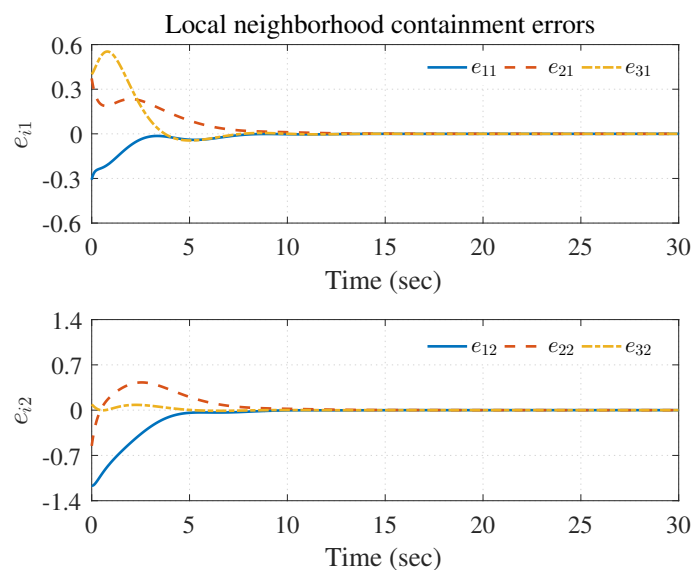


Figure 3. Local neighborhood containment errors e_i .

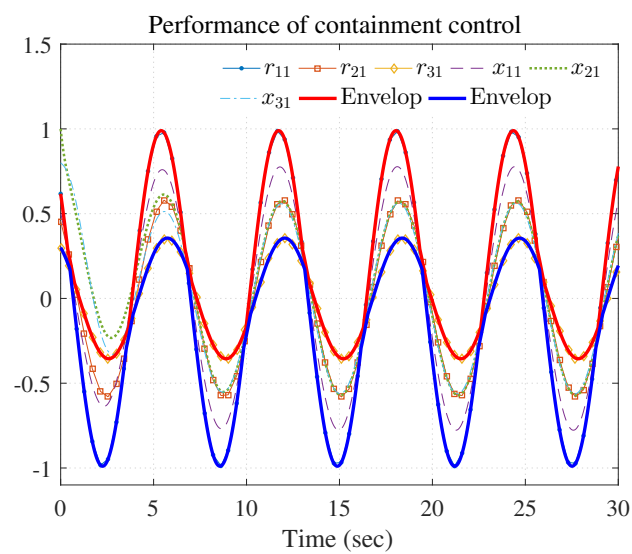


Figure 4. Performance of containment control (r_{j1} and x_{i1}).

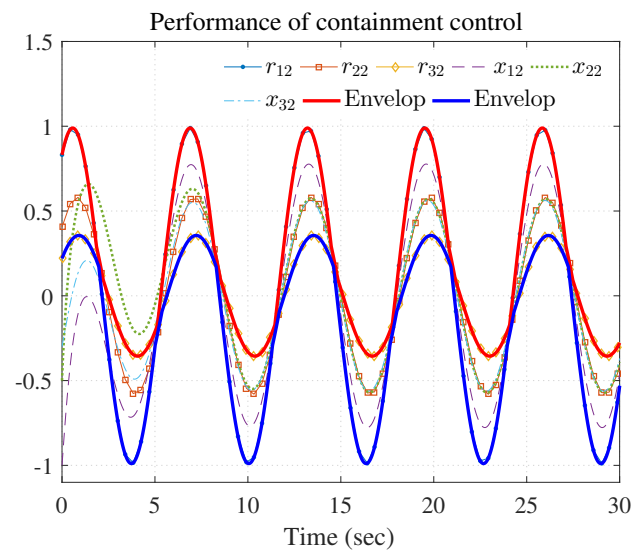


Figure 5. Performance of containment control (r_{j2} and x_{i2}).

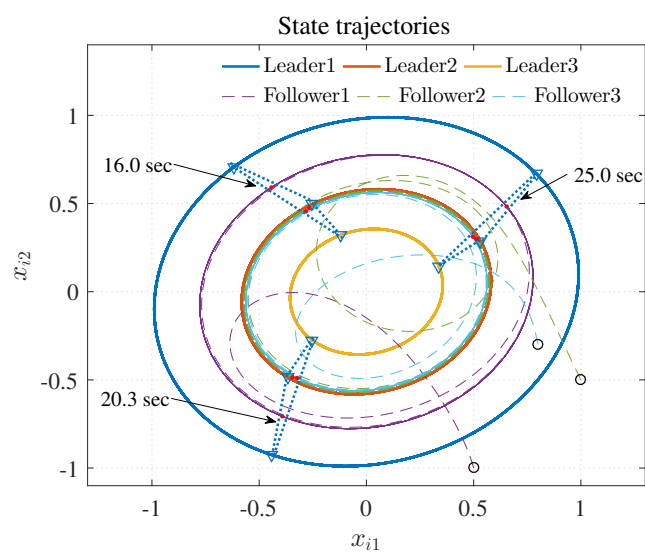


Figure 6. State trajectories.

Table 1. Notations of the single-link robot arm.

Symbol	Notation
z_i	Link angle
\dot{z}_i	Angular velocity of the link
\mathcal{J}	Total rotational inertia of the link and motor
\bar{B}	Overall damping coefficient
\bar{M}	Total mass of the link
l	Distant from joint axis to mass center of the link
u_i	Command generator

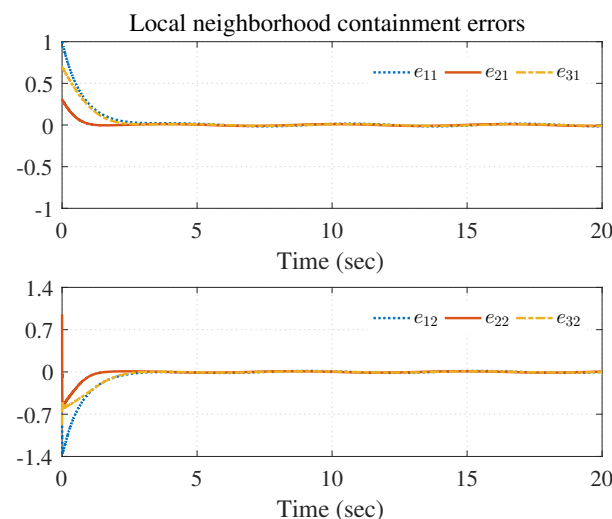
Define $x_i = [x_{i1}, x_{i2}]^T = [z_i, \dot{z}_i]^T \in \mathbb{R}^2$ and $\mu_i = u_i$. For the i th follower, the model (36) can be rewritten as

$$\begin{bmatrix} \dot{x}_{i1} \\ \dot{x}_{i2} \end{bmatrix} = \begin{bmatrix} x_{i2} \\ -\frac{\bar{M}gl}{\mathcal{J}} \sin(x_{i1}) - \frac{\bar{B}}{\mathcal{J}} x_{i2} \end{bmatrix} + \begin{bmatrix} 0 \\ \frac{1}{\mathcal{J}} \end{bmatrix} \mu_i. \quad (37)$$

Similar to Example Section 4.1, the local neighborhood containment error vector was given as $e_i = [e_{i1}, e_{i2}]^T \in \mathbb{R}^2$.

The critic NN structures and the related activation functions were initialized as in Example Section 4.1. The critic NN weights were initialized as the random values within $(0, 36)$ and the parameters of initialization and control were chosen as $r_1(0) = [0, 0.6]^T$, $r_2(0) = [0.4 \sin(\frac{\pi}{6}), 0.4 \cos(\frac{\pi}{6})]^T$, $r_3(0) = [0.2 \sin(-\frac{\pi}{6}), 0.2 \cos(-\frac{\pi}{6})]^T$, $x_1(0) = [0.8, 0.1]^T$, $x_2(0) = [0.6, 0.5]^T$, $x_3(0) = [0.7, -0.3]^T$, $Q_{ip} = 18I_n$, $R_{ip} = 5$, $t_\tau = 0.1$ s, $l_{ci} = 0.1$ and $l_{si} = 0.1$.

Figures 7–11 show the simulation results. The local neighborhood containment errors converged to a small region around zero as depicted in Figure 7, which shows that the containment control of the nonlinear MAS was achieved. In Figures 8 and 9, it can be found that the state trajectories of single-link robot arms (36) entered and stayed within the region enveloped by the leader nodes as the time progressed, which indicated the satisfactory performance of the developed scheme. The evolution curves of all agents are illustrated as the 2-D phase plane plot in Figure 10. We can see that the convex hull formed by the leaders v_1, v_2 and v_3 contains the followers at the time instants $t = 5.0$ s, 10.0 s, 14.5 s and 26.0 s, which implies that the followers converged to the convex hull. Figure 11 describes the curves of the containment control inputs, which shows the regulation process of the containment error system.

**Figure 7.** Local neighborhood containment errors of triple followers.

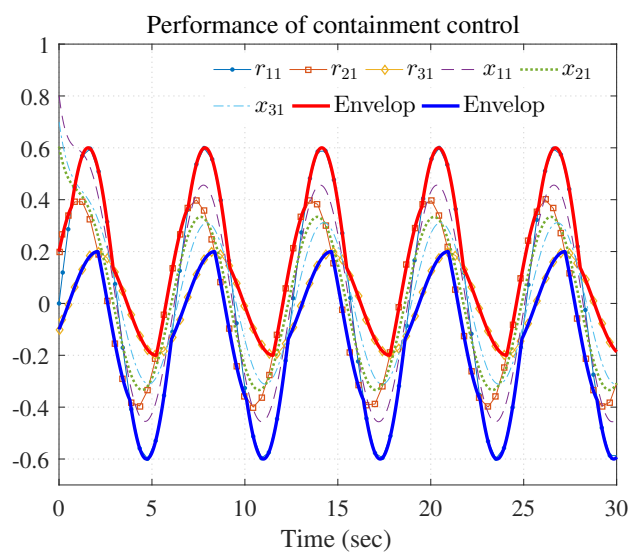


Figure 8. Performance of containment control (r_{j1} and x_{i1}).

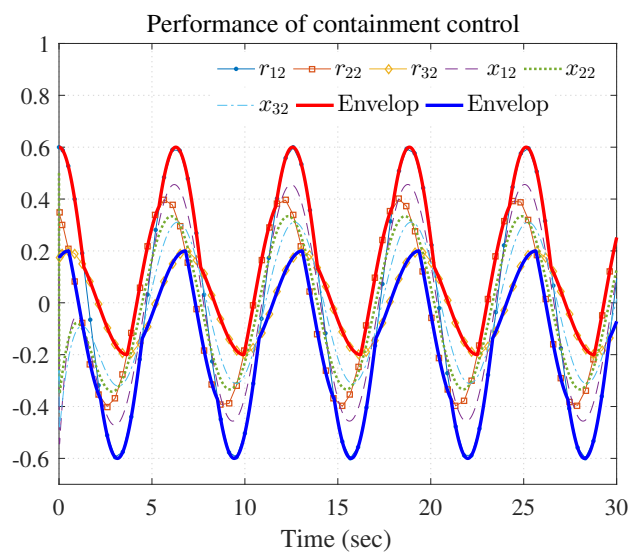


Figure 9. Performance of containment control (r_{j2} and x_{i2}).

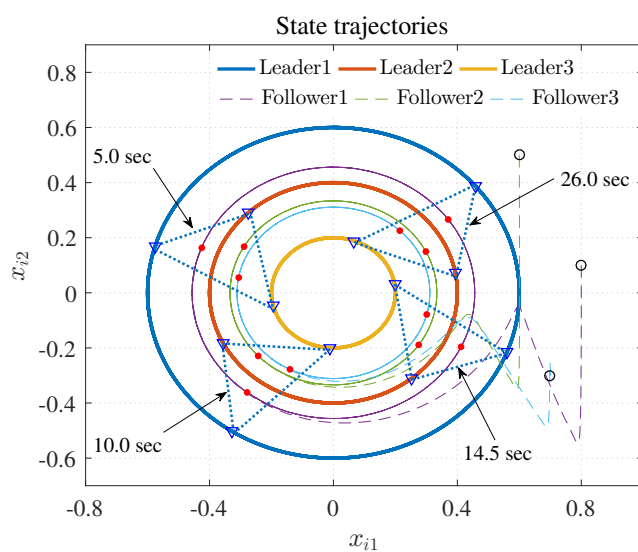


Figure 10. State trajectories.

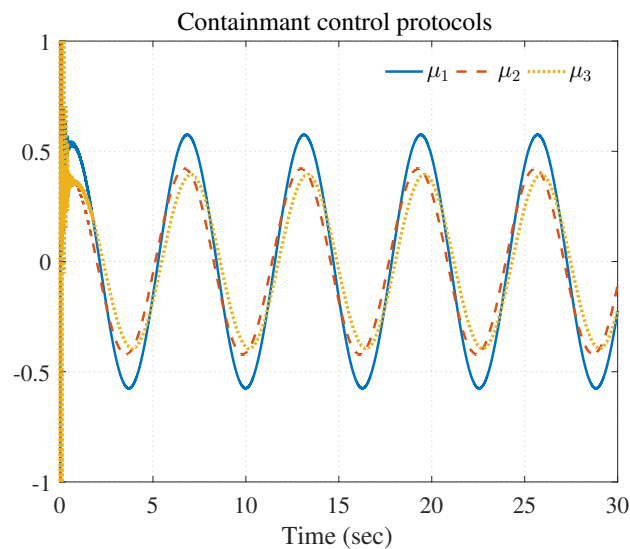


Figure 11. Containment control inputs of triple followers.

5. Conclusions

This paper investigated the OCC problem of nonlinear MASs with partially unknown dynamics via the IRL method. Based on the IRL method, the integral Bellman equation was constructed to relax the requirement of the drift dynamics. The proposed control algorithm was guaranteed to converge by analyzing the convergence of IRL. With the aid of the universal approximation capability of the NN, the solution of the HJBE was acquired by a critic NN with a modified weight-updating law which guaranteed the asymptotical stability of the weight error dynamics. By using the Lyapunov stability theorem, we showed that the closed-loop containment error system was UUB. From the simulation results of two examples, the effectiveness of the proposed IRL-based OCC scheme was illustrated. In the considered MASs, the information among all agents was transmitted by a desired communication network, which is always confronted with some security issues, such as attacks and packet dropouts. The focus of our future work is to develop a novel distributed resilient containment control for the MASs subjected to attacks and packet dropouts.

Author Contributions: Conceptualization, Q.W. and Y.W. (Yonghua Wang); methodology, Q.W.; software, Q.W.; investigation, Q.W.; writing—original draft preparation, Q.W.; writing—review and editing, Y.W. (Yongheng Wu); visualization, Y.W. (Yongheng Wu); supervision, Y.W. (Yonghua Wang); funding acquisition, Y.W. (Yonghua Wang). All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by Open Research Fund of The State Key Laboratory for Management and Control of Complex Systems under grant no. 20220118.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data is contained within this manuscript.

Acknowledgments: We appreciate all the authors for their contributions and the support of the foundation.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Jimenez, A.F.; Cardenas, P.F.; Jimenez, F. Intelligent IoT-multiagent precision irrigation approach for improving water use efficiency in irrigation systems at farm and district scales. *Comp. Electr. Agric.* **2022**, *192*, 106635. [[CrossRef](#)]
2. Vallejo, D.; Castro-Schez J.; Glez-Morcillo C.; Albusac, J. Multi-agent architecture for information retrieval and intelligent monitoring by UAVs in known environments affected by catastrophes. *Eng. Appl. Artif. Intell.* **2020**, *87*, 103243. [[CrossRef](#)]

3. Liu, Y.; Wang, Y.; Li, Y.; Gooi, H.B.; Xin, H. Multi-agent based optimal scheduling and trading for multi-microgrids integrated with urban transportation networks. *IEEE Trans. Power. Syst.* **2021**, *36*, 2197–2210.
4. Deng, Q.; Peng, Y.; Qu, D.; Han, T.; Zhan, X. Neuro-adaptive containment control of unmanned surface vehicles with disturbance observer and collision-free. *ISA Trans.* **2022**, *129*, 150–156. [[CrossRef](#)] [[PubMed](#)]
5. Hamani, N.; Jamont, J.P.; Occello, M.; Ben-Yelles, C.B.; Lagreze, A.; Koudil, M. A multi-cooperative-based approach to manage communication in wireless instrumentation systems. *IEEE Syst. J.* **2018**, *12*, 2174–2185.
6. Ren, W.; Beard, R.W. Consensus seeking in multiagent systems under dynamically changing interaction topologies. *IEEE Trans. Autom. Control* **2005**, *50*, 655–661.
7. Luo, K.; Guan, Z.H.; Cai, C.X.; Zhang, D.X.; Lai, Q.; Xiao, J.W. Coordination of nonholonomic mobile robots for diffusive threat defense. *J. Frankl. Inst.* **2019**, *356*, 4690–4715. [[CrossRef](#)]
8. Yu, Z.; Liu, Z.; Zhang, Y.; Qu, Y.; Su, C.Y. Distributed finite-time fault-tolerant containment control for multiple unmanned aerial vehicles. *IEEE Trans. Neural Netw. Learn. Syst.* **2020**, *31*, 2077–2091. [[CrossRef](#)]
9. Li, Y.; Qu, F.; Tong, S. Observer-based fuzzy adaptive finite-time containment control of nonlinear multiagent systems with input delay. *IEEE Trans. Cybern.* **2021**, *51*, 126–137. [[CrossRef](#)]
10. Li, Z.; Xue, H.; Pan, Y.; Liang, H. Distributed adaptive event-triggered containment control for multi-agent systems under a funnel function. *Int. J. Robust Nonlinear Control* **2022**. [[CrossRef](#)]
11. Li, Y.; Liu, M.; Lian, J.; Guo, Y. Collaborative optimal formation control for heterogeneous multi-agent systems. *Entropy* **2022**, *24*, 1440. [[CrossRef](#)]
12. Zhao, L.; Yu, J.; Shi, P. Command filtered backstepping-based attitude containment control for spacecraft formation. *IEEE Trans. Syst. Man Cybern. Syst.* **2021**, *51*, 1278–1287.
13. Liu, D.; Wei, Q.; Wang, D.; Yang, X.; Li, H. *Adaptive Dynamic Programming With Applications in Optimal Control*; Springer: Cham, Switzerland, 2017.
14. Bellman, R.E. *Dynamic Programming*; Princeton Univ. Press: Trenton, NJ, USA, 1957.
15. Abu-Khalaf, M.; Lewis, F.L. Nearly optimal control laws for nonlinear systems with saturating actuators using a neural network HJB approach. *Automatica* **2005**, *41*, 779–791. [[CrossRef](#)]
16. Liu, D.; Xue, S.; Zhao, B.; Luo, B.; Wei, Q. Adaptive dynamic programming for control: A survey and recent advances. *IEEE Trans. Syst. Man. Cybern. Syst.* **2021**, *51*, 142–160.
17. Vamvoudakis, K.G.; Lewis, F.L.; Hudas, G.R. Multi-agent differential graphical games: Online adaptive learning solution for synchronization with optimality. *Automatica* **2012**, *48*, 1598–1611. [[CrossRef](#)]
18. 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*, 152–163. [[CrossRef](#)]
19. Zhao, W.; Zhang, H. Distributed optimal coordination control for nonlinear multi-agent systems using event-triggered adaptive dynamic programming method. *ISA Trans.* **2019**, *91*, 184–195. [[CrossRef](#)]
20. Cui, J.; Pan, Y.; Xue, H.; Tan, L. Simplified optimized finite-time containment control for a class of multi-agent systems with actuator faults. *Nonlinear Dyn.* **2022**, *109*, 2799–2816. [[CrossRef](#)]
21. Xu, J.; Wang, L.; Liu, Y.; Xue, H. Event-triggered optimal containment control for multi-agent systems subject to state constraints via reinforcement learning. *Nonlinear Dyn.* **2022**, *109*, 1651–1670. [[CrossRef](#)]
22. Xiao, W.; Zhou, Q.; Liu, Y.; Li, H.; Lu, R. Distributed reinforcement learning containment control for multiple nonholonomic mobile robots. *IEEE Trans. Circuits Syst. I Reg. Papers* **2022**, *69*, 896–907. [[CrossRef](#)]
23. Chen, C.; Lewis, F.L.; Xie, K.; Xie, S.; Liu, Y. Off-policy learning for adaptive optimal output synchronization of heterogeneous multi-agent systems. *Automatica* **2020**, *119*, 109081. [[CrossRef](#)]
24. Yu, D.; Ge, S.S.; Li, D.; Wang, P. Finite-horizon robust formation-containment control of multi-agent networks with unknown dynamics. *Neurocomputing* **2021**, *458*, 403–415. [[CrossRef](#)]
25. Zuo, S.; Song, Y.; Lewis, F.L.; Davoudi, A. Optimal robust output containment of unknown heterogeneous multiagent system using off-policy reinforcement learning. *IEEE Trans. Cybern.* **2018**, *48*, 3197–3207. [[CrossRef](#)] [[PubMed](#)]
26. Mazouchi, M.; Naghibi-Sistani, M.B.; Hosseini Sani, S.K.; Tatari, F.; Modares, H. Observer-based adaptive optimal output containment control problem of linear heterogeneous Multiagent systems with relative output measurements. *Int. J. Adapt. Control Signal Process.* **2019**, *33*, 262–284.
27. Yang, Y.; Modares, H.; Wunsch, D.C.; Yin, Y. Optimal containment control of unknown heterogeneous systems with active leaders. *IEEE Trans. Control Syst. Technol.* **2019**, *27*, 1228–1236. [[CrossRef](#)]
28. Zhang, H.; Lewis, F.L.; Qu, Z. Lyapunov, adaptive, and optimal design techniques for cooperative systems on directed communication graphs. *IEEE Trans. Ind. Electron.* **2012**, *59*, 3026–3041. [[CrossRef](#)]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.