Received July 18, 2019, accepted July 21, 2019, date of publication July 25, 2019, date of current version August 9, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2930989

Model-Free Optimal Consensus Control of Networked Euler-Lagrange Systems

HUAIPIN ZHANG[1,2], **JU H. PARK**[2], **AND WEI ZHAO**[3]
[1]Institute of Advanced Technology, Nanjing University of Posts and Telecommunications, Nanjing 210023, China
[2]Department of Electrical Engineering, Yeungnam University, Gyongsan 38541, South Korea
[3]School of Mathematics, Southeast University, Nanjing 210096, China

Corresponding author: Ju H. Park (jessie@ynu.ac.kr)

This work was supported in part by the Nanjing University of Posts and Telecommunications Funding (NUPTSF) under Grant NY219037, and in part by the Basic Science Research Programs of the National Research Foundation of Korea funded by the Ministry of Education under Grant NRF-2017R1A2B2004671.

ABSTRACT This paper considers the model-free optimal consensus problem of networked Euler–Lagrange systems without velocity measurements. By employing the position information, a novel neural network-based velocity observer is built for each agent to estimate the unmeasurable velocity vector and unknown system model. Based on the estimated velocity information, we propose distributed optimal control policies depended on the solutions to the coupled Hamilton–Jacobi–Bellman (HJB) equations. Then, a model-free policy iteration (PI) algorithm is provided to learn the coupled HJB equations online. To implement the PI algorithm, the critic-action neural networks are built and their weights are updated based on the gradient descent method. The uniform ultimate boundedness of the integrated observer estimation errors, the integrated consensus errors, and the weight estimation errors for the observer-critic-action neural networks is demonstrated by the Lyapunov technique. Finally, the numerical simulation on a directed network with six nonlinear manipulators is presented to validate the theoretical results.

INDEX TERMS Euler-Lagrange systems, adaptive dynamic programming, consensus control, adaptive observer.

I. INTRODUCTION

Distributed control of networked Euler-Lagrange systems (NELSs) has attracted lots of attention due to Euler-Lagrange equations are usually utilized to describe the dynamical behaviors of many practical physical systems, such as robotic manipulators, power electronic systems and actuated autonomous vehicles. Extensive results have been reported on distributed cooperation problems of NELSs in [1]–[6]. The study of the consensus problem for NELSs over an ideal and fixed communication network has been considered in [7], which was extended to communication delays or switching networks in [3], [8]. Later, by introducing an auxiliary system, Hu *et al.* investigated the swarming behavior of NELSs with cooperation-competition interactions in [4]. It has been revealed that, under a stochastic sampling communication setting, distributed consensus for NELSs can be achieved in [9]. Most of these works assume that the complete state information of the system is available. Unfortunately, this assumption is rather restrictive and not realistic so as to stunt the application of the existing results to practical systems.

It is usually difficult to access velocity information in some mechanical systems since these systems may be either not equipped with velocity sensors or those with velocity sensors are often contaminated by noises. Hence, how to avoid using the velocity information in addressing the consensus problem of NELSs become a crucial and challenging issue. To deal with this issue, many works on partial state feedback control have been presented in [10]–[14]. By adopting the damping injecting principles of passivity-based control, a position feedback consensus control protocol for NELSs without velocity measurement was designed in [10]. For the finite-time tracking problem, Zhao *et al.* [11] built a sliding-mode observer-based distributed consensus control policy which only depends on the the position information.

In the process of achieving consensus, it is desirable to optimize the system performance [15]–[18]. So far, adaptive dynamic programming (ADP) [19], [20] combined reinforcement learning with adaptive control is an efficient and promising method to address optimal consensus problem forward in time. From the perspective of game theory, ADP

The associate editor coordinating the review of this manuscript and approving it for publication was Heng Zhang.

technique was employed in [21], [22] to investigate the optimal regulation problem of differential graph games, in which all agents can achieve optimal consensus while they are in Nash equilibrium. Extension to deal with the unknown system dynamics can be found in [17], [23], where off-policy reinforcement learning approach is introduced to address the optimal synchronization problem for model-free homogeneous and heterogeneous MASs. By equipping with omnidirectional vision sensors, [24] proposed the distributed optimal consensus controllers for uncertain mobile multi-robot systems with external disturbances.

In this paper, we address the distributed optimal consensus problem for NELSs without velocity measurements. Combined with the above results, the main contributions of this paper are given as follows.

1) By using only the position information of each agent, we design a novel neural network-based observer to estimate the velocity information and the system dynamic model.

2) It is noted that NELSs can be transformed into second-order multi-agent systems (MASs) based on coordinate transformation. Then we will employ ADP technique to solve the optimal consensus problem for second-order MASs.

3) Compared with the existing model-based adaptive consensus approach of NELSs [6], [25], a distributed model-free optimal consensus algorithm is proposed in which the critic-action network framework is built to approximate the optimal performance and optimal consensus policies.

The rest of this paper is organized as follows. Section II presents the graph theory and problem formulation. Section III designs a neural network-based observer for each agent, while optimal control policy is obtained in Section IV. The critic-action network structure is built to implement the proposed algorithm in Section V. A numerical simulation is provided in Section VI. Finally, the conclusion is drawn in Section VII.

II. PRELIMINARIES

A. GRAPH THEORY

Let $\mathcal{G} = \{\mathcal{V}, \mathcal{E}\}$ be a digraph in which $\mathcal{V} = \{1, 2, N\}$ is the set of nodes and $\mathcal{E} \subseteq \mathcal{V} \times \mathcal{V}$ is the set of directed edges. A directed edge $(i, j) \in \mathcal{E}$ exists if and only if node i can get the information from node j. And $\mathcal{N}_i = \{j : (i, j) \in \mathcal{E}, j \neq i\}$ denotes the neighbor set of node i and $\bar{\mathcal{N}_i} = \{\mathcal{N}_i, i\}$. A directed path is a directed edge sequence in the form of $(i_1, i_2), (i_2, i_3), \ldots, (i_{k-1}, i_k)$. A digraph is said to be strongly connected if there is a directed path from one node to any other node. Graph $\mathcal{G}$ has a directed spanning tree if there is a root node that has a directed path to any other node. $A = [a_{ij}] \in R^{N\times N}$ denotes the nonnegative adjacency matrix of digraph $\mathcal{G}$, where $a_{ij} > 0$ if $(i, j) \in \mathcal{E}$, or else $a_{ij} = 0$. And $L = [l_{ij}] \in R^{N\times N}$ is the Laplacian matrix of digraph $\mathcal{G}$ with $l_{ii} = \sum_{j\in N_i} a_{ij}$ and $l_{ij} = -a_{ij}, i \neq j$.

B. PROBLEM STATEMENT

Consider a group of N agents governed by the Euler-Lagrange equations, the dynamics of each agent can be described as

$$M_i(q_i)\ddot{q}_i + C_i(q_i, \dot{q}_i)\dot{q}_i + g_i(q_i) = \tau_i, \quad i = 1, \ldots, N \tag{1}$$

where $q_i \in \mathbb{R}^m$ is the generalized configuration coordinate, $M_i(q_i) \in \mathbb{R}^{m\times m}$ denotes the inertia matrix, $C_i(q_i, \dot{q}_i) \in \mathbb{R}^{m\times m}$ denotes the Coriolis and centrifugal force matrix, $g_i(q_i) \in \mathbb{R}^m$ is the gravitational torque vector, and $\tau_i \in \mathbb{R}^m$ is the vector of control input torque exerted by the actuators.

Based on the structure of Euler-Lagrange systems, the following assumption is given.

Assumption 1 [4], [8]: The inertia matrices $M_i(q_i)$ are symmetric positive definite and bounded, i.e., $m_{mi} \leq \|M_i(q_i)\| \leq m_{Mi}$ for all agents in which $m_{mi} = \lambda_{min}(M_i(q_i))$ and $m_{Mi} = \lambda_{max}(M_i(q_i))$.

To simplify the dynamics model, $v_i = \dot{q}_i$ is defined as the velocity vector of agent i. Thus the EL dynamics systems (1) can be transformed into the second-order dynamic system as follows

$$\begin{aligned} \dot{q}_i &= v_i \\ \dot{v}_i &= -F_i(q_i, v_i) + M_i^{-1}(q_i)\tau_i \end{aligned} \tag{2}$$

where $F_i(q_i, v_i) = M_i^{-1}(q_i)(C_i(q_i, v_i)v_i + g_i(q_i))$.

Now the problem is formulated as follows.

Problem 1: Consider the NELSs given by (1), the following distributed control protocol is designed as

$$u_i = k_i(q_{\bar{\mathcal{N}_i}}, v_{\bar{\mathcal{N}_i}}) \tag{3}$$

where $q_{\bar{\mathcal{N}_i}} = \{q_j : j \in \bar{\mathcal{N}_i}\}$ and $v_{\bar{\mathcal{N}_i}} = \{v_j : j \in \bar{\mathcal{N}_i}\}$ such that all the agents can achieve consensus, i.e., $\|q_i - q_j\| \to 0$, $\|v_i - v_j\| \to 0, \forall i, j \in \mathcal{V}$.

The designed distributed controllers (3) rely on the local position and velocity information. In practice, yet, the Euler-Lagrange systems can only obtain the position information. To address this issue, an adaptive observer will be designed for each agent to estimate the unknown velocity information and the unknown system model.

III. NEURAL NETWORK-BASED OBSERVER DESIGN

According to the approximation property of neural networks (NNs) [26], [27], the velocity dynamics of the ith agent can be presented as

$$\begin{aligned} \dot{v}_i &= -F_i(q_i, v_i) + M_i^{-1}(q_i)\tau_i \\ &= w_{F,i}^T\psi_{F,i}(q_i, v_i) + \varepsilon_{F,i}(q_i, v_i) + w_{M,i}^T\psi_{M,i}(q_i)\tau_i \\ &\quad + \varepsilon_{M,i}(q_i)\tau_i \\ &= \begin{bmatrix} w_{F,i}^T & w_{M,i}^T \end{bmatrix} \begin{bmatrix} \psi_{F,i}(q_i, v_i) & 0 \\ 0 & \psi_{M,i}(q_i) \end{bmatrix} \begin{bmatrix} 1 \\ \tau_i \end{bmatrix} \\ &\quad + \begin{bmatrix} \varepsilon_{F,i}(q_i, v_i) & \varepsilon_{M,i}(q_i) \end{bmatrix} \begin{bmatrix} 1 \\ \tau_i \end{bmatrix} \\ &= w_{o,i}^T\psi_i(q_i, v_i) + \varepsilon_{o,i}(q_i, v_i) \end{aligned} \tag{4}$$

where $w_{o,i} = \left[w_{F,i}^T \; w_{M,i}^T \right]^T$ denotes the unknown ideal weight of the observer NN,

$$\psi_i(q_i, v_i) = \begin{bmatrix} \psi_{F,i}(q_i, v_i) & 0 \\ 0 & \psi_{M,i}(q_i) \end{bmatrix} \begin{bmatrix} 1 \\ \tau_i \end{bmatrix}$$

is the activation function vector and $\varepsilon_{o,i}(q_i, v_i) = \left[\varepsilon_{F,i}(q_i, v_i) \; \varepsilon_{M,i}(q_i) \right] \begin{bmatrix} 1 \\ \tau_i \end{bmatrix}$ is the estimation error. Thus the velocity of agent i can be identified by updating its observer NN weights.

Throughout this section, the following standard assumption resulting from the PE condition [28] is introduced for the observer NNs.

Assumption 2: The ideal weight matrices, the activation functions and the estimation errors of the observer NNs are bounded, i.e., there exist positive constants $w_{o,M}$, ψ_M and $\varepsilon_{o,M}$ such that $\|w_{o,i}\| \leq w_{o,M}$, $\|\psi_i\| \leq \psi_M$ and $\|\varepsilon_{o,i}\| \leq \varepsilon_{o,M}$.

When using NNs to estimate the velocity vectors, the dynamics (4) can be approximated by

$$\dot{\hat{v}}_i = \hat{w}_{o,i}^T \psi_i(\hat{q}_i, \hat{v}_i) \tag{5}$$

where $\hat{w}_{o,i}$ is the estimated weight matrix, $\hat{q}_i$ and $\hat{v}_i$ are the estimated values of q_i and v_i, respectively.

Due to the velocities are unavailable, the passive observers are designed by using the available position information to estimate the velocity vectors and the system model. For each agent, the following NN-based observer is designed as

$$\begin{aligned} \dot{\hat{q}}_i &= \hat{v}_i + c_{i1} l_i (q_i - \hat{q}_i) \\ \dot{\hat{v}}_i &= \hat{w}_{o,i}^T \psi_i(\hat{q}_i, \hat{v}_i) + c_{i2} l_i^2 (q_i - \hat{q}_i) \end{aligned} \tag{6}$$

where $l_i > 0$ is the observer gain to be determined later. c_{i1}, c_{i2} are coefficients of the Hurwitz polynomial $f_i(x) = x^2 + c_{i1}x + c_{i2}$.

Remark 1: The proposed observers (6) are designed based on the position information of itself. In comparison to distributed observers in [13], [29], the proposed observers (6) can reduce communication and computation load.

From (2) and (6), the estimation error dynamics for the position and velocity vectors can be derived as

$$\begin{aligned} \dot{\tilde{q}}_i &= \tilde{v}_i - c_{i1} l_i \tilde{q}_i \\ \dot{\tilde{v}}_i &= \tilde{w}_{o,i}^T \psi_i(\hat{q}_i, \hat{v}_i) + \zeta_i - c_{i2} l_i^2 \tilde{q}_i \end{aligned} \tag{7}$$

where $\tilde{q}_i = q_i - \hat{q}_i$ and $\tilde{v}_i = v_i - \hat{v}_i$ denote the position and velocity estimation error, respectively, $\tilde{w}_{o,i} = w_{o,i} - \hat{w}_{o,i}$ is the weight estimation error, and $\zeta_i = w_{o,i}^T(\psi_i(q_i, v_i) - \psi_i(\hat{q}_i, \hat{v}_i)) + \varepsilon_{o,i}(q_i, v_i)$ is the bounded term based on Assumption 2.

Inspired by [30], [31], the weight update law of the ith observer NN is given by

$$\dot{\hat{w}}_{o,i} = \sigma_i \psi_i(\hat{q}_i, \hat{v}_i) \tilde{q}_i^T - \gamma_i \hat{w}_{o,i} \tag{8}$$

where $\sigma_i > 0$ is the learning rate, $\gamma_i > 0$ is a small constant.

Furthermore, the integrated observer estimation error for each agent is defined as $\eta_i = col(\tilde{q}_i, \tilde{v}_i / l_i)$. Then one has

$$\dot{\eta}_i = l_i C_i \eta_i + \epsilon_i \tag{9}$$

where $C_i = \begin{bmatrix} -c_{i1} & 1 \\ -l_i c_{i2} & 0 \end{bmatrix}$ and $\epsilon_i = \begin{bmatrix} 0 \\ \tilde{w}_{o,i}^T \psi_i(\hat{q}_i, \hat{v}_i) + \zeta_i \end{bmatrix}$.

Given a positive definite matrix Q_i, then there exists a positive definite matrix P_i such that

$$C_i^T P_i + P_i C_i = -Q_i \tag{10}$$

Definition 1 [32]: The error ξ is uniformly ultimately bounded (UUB) with respect to a closed ball Ω if for $\forall \mathcal{B} > 0$, there exists $t_f(\mathcal{B})$ such that if $\|\xi(t_0)\| \leq \mathcal{B}$, then $\xi(t) \in \Omega$, $\forall\, t \geq t_0 + t_f(\mathcal{B})$.

Theorem 1: Consider the NELSs (1), the NN-based observers and the weight updated laws of NNs are provided as (6) and (8), then the integrated observer estimation errors η_i and the weight estimation errors $\tilde{w}_{o,i}$ are UUB.

Proof: Choose the Lyapunov candidate function of agent i as follows

$$L_{o,i} = L_{\eta,i} + L_{w,i} \tag{11}$$

where $L_{\eta,i} = \frac{1}{2}\eta_i^T P_i \eta_i$ and $L_{w,i} = \frac{1}{2} tr(\tilde{w}_{o,i}^T \tilde{w}_{o,i})$.

Taking the derivative of $L_{\eta,i}$ along the dynamics of the integrated observer estimation error (9) yields

$$\begin{aligned} \dot{L}_{\eta,i} &= \frac{l_i}{2}\eta_i^T(C_i^T P_i + P_i C_i)\eta_i + \eta_i^T P_i \epsilon_i \\ &= -\frac{l_i}{2}\eta_i^T Q_i \eta_i + \eta_i^T P_i \epsilon_i \\ &\leq -\frac{l_i}{2}\lambda_{min}(Q_i)\|\eta_i\|^2 + \|\eta_i\|\|P_i\|\|\epsilon_i\| \\ &\leq -\frac{l_i}{2}\lambda_{min}(Q_i)\|\eta_i\|^2 + \frac{1}{2}\|\eta_i\|^2\|P_i\|^2 \\ &\quad + \frac{1}{2}(\|\tilde{w}_{o,i}^T \psi_i(\hat{q}_i, \hat{v}_i)\|^2 + \|\zeta_i\|^2) \end{aligned} \tag{12}$$

Based on Assumption 2, it is known that

$$\begin{aligned} \|\zeta_i\| &\leq \|w_{o,i}^T(\psi_i(q_i, v_i) - \psi_i(\hat{q}_i, \hat{v}_i))\| + \|\varepsilon_{o,i}(q_i, v_i)\| \\ &\leq 2w_{o,M}\psi_M + \varepsilon_{o,M} \\ &\leq \bar{\zeta} \end{aligned} \tag{13}$$

where $\bar{\zeta} = 2w_{o,M}\psi_M + \varepsilon_{o,M}$.

Combined (12) with (13), one has

$$\begin{aligned} \dot{L}_{\eta,i} &\leq -\frac{1}{2}(l_i\lambda_{min}(Q_i) - \|P_i\|^2)\|\eta_i\|^2 \\ &\quad + \frac{1}{2}\psi_M^2\|\tilde{w}_{o,i}\|^2 + \frac{1}{2}\bar{\zeta}^2 \end{aligned} \tag{14}$$

Then taking the derivative of $L_{w,i}$ along (8) yields

$$\begin{aligned} \dot{L}_{w,i} &= -tr(\tilde{w}_{o,i}^T \dot{\hat{w}}_{o,i}) \\ &= -tr(\tilde{w}_{o,i}^T(\sigma_i \psi_i(\hat{q}_i, \hat{v}_i)\tilde{q}_i^T - \gamma_i \hat{w}_{o,i})) \\ &= -tr(\sigma_i \tilde{w}_{o,i}^T \psi_i(\hat{q}_i, \hat{v}_i)\tilde{q}_i^T) + tr(\gamma_i \tilde{w}_{o,i}^T(w_{o,i} - \tilde{w}_{o,i})) \\ &\leq \frac{\sigma_i \psi_M}{2}(\|\tilde{w}_{o,i}\|^2 + \|\tilde{q}_i\|^2) + \frac{\gamma_i}{2}(\|\tilde{w}_{o,i}\|^2 + w_{o,M}^2) \\ &\quad - \gamma_i\|\tilde{w}_{o,i}\|^2 \\ &\leq \frac{\sigma_i\psi_M - \gamma_i}{2}\|\tilde{w}_{o,i}\|^2 + \frac{\sigma_i\psi_M}{2}\|\eta_i\|^2 + \frac{\gamma_i}{2}w_{o,M}^2 \end{aligned} \tag{15}$$

Substituting (14) and (15) into (11), one can obtain

$$\begin{aligned}\dot{L}_{o,i} &= \dot{L}_{\eta,i} + \dot{L}_{w,i}\\ &\leq -\frac{1}{2}(l_i\lambda_{min}(Q_i) - \|P_i\|^2)\|\eta_i\|^2 + \frac{1}{2}\psi_M\|\tilde{w}_{o,i}\|^2 + \frac{1}{2}\bar{\xi}^2\\ &\quad + \frac{\sigma_i\psi_M - \gamma_i}{2}\|\tilde{w}_{o,i}\|^2 + \frac{\sigma_i\psi_M}{2}\|\eta_i\|^2 + \frac{\gamma_i}{2}w_{o,M}^2\\ &\leq -\frac{1}{2}(l_i\lambda_{min}(Q_i) - \|P_i\|^2 - \sigma_i\psi_M)\|\eta_i\|^2\\ &\quad - \frac{1}{2}(\gamma_i - (\sigma_i+1)\psi_M)\|\tilde{w}_{o,i}\|^2 + \frac{1}{2}\bar{\xi}^2 + \frac{\gamma_i}{2}w_{o,M}^2\end{aligned}$$

Setting $l_i > \frac{\|P_i\|^2+\sigma_i\psi_M}{\lambda_{min}(Q_i)}$ and $\gamma_i > (\sigma_i+1)\psi_M$, if the following inequalities hold

$$\|\eta_i\| > \sqrt{\frac{\bar{\xi}^2 + \gamma_i w_{o,M}^2}{l_i\lambda_{min}(Q_i) - \|P_i\|^2 - \sigma_i\psi_M}} = \Theta_{\eta,i}$$

or

$$\|\tilde{w}_{o,i}\| > \sqrt{\frac{\bar{\xi}^2 + \gamma_i w_{o,M}^2}{\gamma_i - (\sigma_i+1)\psi_M}} = \Theta_{o,i}$$

then we can obtain $\dot{L}_{o,i} < 0$. This means that the integrated observer estimation error η_i and the weight estimate error $\tilde{w}_{o,i}$ for each agent are UUB.

Remark 2: It is noted that $\dot{L}_{o,i}$ is negative definite outside the ball $\Theta_i = \{(\eta_i, \tilde{w}_{o,i}) | \|\eta_i\| > \Theta_{\eta,i} \text{ and } \|\tilde{w}_{o,i}\| > \Theta_{o,i}\}$. The size of the error bound $\Theta_{\eta,i}$ and $\Theta_{o,i}$ can be arbitrarily small by selecting the suitable parameters l_i, σ_i and γ_i.

IV. ADP-BASED OPTIMAL CONSENSUS CONTROLLER DESIGN

To synthesize the system performance, the optimal consensus problem of NELSs will be transformed into a distributed optimal regulation of second-order MASs in this section. Then we employ ADP technique to design the optimal consensus control policies for second-order MASs.

A. OPTIMAL CONSENSUS CONTROLLER DESIGN

The local neighbor position and velocity consensus errors for agent i are defined as follows

$$\begin{aligned}e_i^q &= \sum_{j\in\mathcal{N}_i} a_{ij}(q_i - q_j)\\ e_i^v &= \sum_{j\in\mathcal{N}_i} a_{ij}(v_i - v_j)\end{aligned} \tag{16}$$

Then the dynamics of the local neighbor position and velocity consensus errors are derived as

$$\begin{aligned}\dot{e}_i^q &= \sum_{j\in\bar{\mathcal{N}}_i} l_{ij}v_j\\ \dot{e}_i^v &= \sum_{j\in\bar{\mathcal{N}}_i} l_{ij}(-F_j(q_j,v_j) + M_j^{-1}(q_j)\tau_j)\end{aligned} \tag{17}$$

For each agent in digraph $\mathcal{G}$, the dynamics of integrated local neighbor consensus error are given by

$$\begin{aligned}\dot{e}_i &= \sum_{j\in\bar{\mathcal{N}}_i} l_{ij}(\begin{bmatrix} v_j \\ -F_j(q_j,v_j)\end{bmatrix} + \begin{bmatrix} 0 \\ M_j^{-1}(q_j)\end{bmatrix}\tau_j)\\ &= \sum_{j\in\bar{\mathcal{N}}_i} l_{ij}(\bar{F}_j(q_j,v_j) + \bar{M}_j(q_j)\tau_j)\\ &= \mathcal{L}_i(\bar{F}(q,v) + \bar{M}(q)\tau)\end{aligned} \tag{18}$$

where $e_i = col(e_i^q, e_i^v)$ is the integrated local neighbor consensus error, $q = col(q_1,\ldots,q_N)$ and $v = col(v_1,\ldots,v_N)$ denote the global position and velocity vectors, $\mathcal{L}_i = L_i \otimes I_{2m}$, $\bar{F}(q,v) = col(\bar{F}_1(q_1,v_1),\ldots,\bar{F}_N(q_N,v_N))$ and $\bar{M}(q,v) = col(\bar{M}_1(q_1,v_1),\ldots,\bar{M}_N(q_N,v_N))$ in which $\bar{F}_j(q_j,v_j) = col(v_j, -F_j(q_j,v_j))$ and $\bar{M}_j(q_j) = col(0, M_j^{-1}(q_j))$, $\tau = col(\tau_1,\ldots,\tau_N)$ is the global control input torque.

In order to find an optimal control policy for agent i, we define a value function which depends on the local neighbor consensus error e_i and the control input torques $\tau_{\bar{\mathcal{N}}_i}$ of agent i and its neighbors as follows,

$$V_i(e_i, \tau_{\bar{\mathcal{N}}_i}) = \int_t^\infty (e_i^T S_{ii} e_i + \sum_{j\in\bar{\mathcal{N}}_i} \tau_j^T R_{ij}\tau_j) d\varsigma \tag{19}$$

where $\tau_{\bar{\mathcal{N}}_i} = \{\tau_j : j \in \bar{\mathcal{N}}_i\}$, S_{ii} and R_{ij} are symmetric positive definite matrices. In practical, $V_i(e_i, \tau_{\bar{\mathcal{N}}_i})$ denotes the local performance on energy consumption of agent i during consensus evolution.

Based on Leibniz's formula, the coupled Hamiltonian equation of agent i is given by

$$\begin{aligned}H_i(e_i, V_{e_i}, \tau_{\bar{\mathcal{N}}_i}) &= e_i^T S_{ii} e_i + \sum_{j\in\bar{\mathcal{N}}_i} \tau_j^T R_{ij}\tau_j\\ &\quad + V_{e_i}^T \mathcal{L}_i(\bar{F}(q,v) + \bar{M}(q)\tau)\end{aligned} \tag{20}$$

where $V_{e_i} = \partial V_i(e_i)/\partial e_i$ denotes the partial derivative of the value function $V_i(e_i)$ with respect to e_i.

Using the stationarity conditions, the optimal control policy for each agent can minimize the value function (19), thus one has

$$\tau_i^* = -\frac{1}{2} d_i R_{ii}^{-1} \bar{M}_i^T(q_i) V_{e_i}^* \tag{21}$$

where $V_i^*(e_i)$ denote the optimal value function of agent i.

Inserting (21) into (20), the coupled HJB equation for each agent is rewritten as

$$\begin{aligned}0 &= H_i(e_i, V_{e_i}^*, \tau_{\bar{\mathcal{N}}_i}^*)\\ &= e_i^T S_{ii} e_i - \frac{1}{4} d_i^2 V_{e_i}^{*T} \bar{M}_i(q_i) R_{ii}^{-1} \bar{M}_i^T(q_i) V_{e_i}^*\\ &\quad + \frac{1}{4}\sum_{j\in\mathcal{N}_i} d_j^2 V_{e_j}^{*T} \bar{M}_j(q_j) R_{jj}^{-T} R_{ij} R_{jj}^{-1} \bar{M}_j^T(q_j) V_{e_j}^*\\ &\quad - \frac{1}{2}\sum_{j\in\mathcal{N}_i} d_j l_{ij} V_{e_i}^{*T} \bar{M}_j(q_j) R_{jj}^{-1} \bar{M}_j^T(q_j) V_{e_j}^*\\ &\quad + \sum_{j\in\bar{\mathcal{N}}_i} l_{ij} V_{e_i}^{*T} \bar{F}_j(q_j,v_j)\end{aligned} \tag{22}$$

To find the optimal consensus control policies, it requires to obtain the solutions of (22). Yet, it is difficult to find the solutions due to the nonlinear nature of the coupled HJB equations.

B. MODEL-FREE PI ALGORITHM

Here, a model-free PI algorithm working together with NN-based observers is developed to solve the coupled HJB equations (22). The NN-based observers (6) and the weight update laws (8) which only depend on the measurable position information provide the estimation of the system model and the velocity vectors. Thus based on the estimations, we develop the model-free PI algorithm for the integrated error system (18) as shown in the following.

Algorithm 1 guarantees that the value functions and control policies can converge to their optimal values, i.e., $V_i^{(l)} \to V_i^*$ and $u_i^{(l)} \to u_i^*$ as $l \to \infty$. The convergence analysis of Algorithm 1 is similar to [18], [21], thus it is omitted here.

Algorithm 1 PI Solutions for Second-Order MAS

Step 1: Initialization
Start with the admissible control policies $u_i^{(0)}, i \in \mathcal{V}$.
Step 2: Neural network observation
Update the weight law (8), estimate the inertia matrix $\hat{M}_i(q_i)$ and compute the estimated values of $\hat{q}_i$ and $\hat{v}_i$ according to (6).
Step 3: Policy evaluation
Employing the estimation information of $\hat{M}_i(q_i)$, $\hat{q}_i$ and $\hat{v}_i$, solve the Hamilton equations
$H_i(e_i, V_{e_i}^{(l)}, \tau_{\bar{\mathcal{N}}_i}^{(l)}) = 0$
Step 4: Policy improvement
Seek the control inputs using
$\tau_i^{(l+1)} = -\frac{1}{2} d_i R_{ii}^{-1} \hat{\bar{M}}_i^T(q_i) V_{e_i}^{(l)}$
Step 5: Stop Criterion
If $\|V_i^{(l+1)} - V_i^{(l)}\| \le \epsilon$ (ϵ is a small positive constant), stop and acquire the approximated optimal control policies u_i^*; Otherwise, set $l = l + 1$ and go to Step 3.

V. IMPLEMENTATION OF CRITIC-ACTION NN STRUCTURE

In this section, the critic-action network framework based on model-free PI algorithm will be introduced to approximate the solutions to the coupled HJB equations (22).

A. CRITIC NN DESIGN

Similar to observer NNs, the value functions can be approximated by the critic NNs as

$$V_i(\hat{e}_i) = w_{c,i}^T \phi_{c,i}(\hat{e}_i) + \varepsilon_{c,i}(\hat{e}_i)$$

where $w_{c,i} \in \mathbb{R}^{h_{vi}}$ is the ideal weight vector with h_{vi} being the number of neurons in the hidden layer, $\phi_{c,i}(\hat{e}_i) \in \mathbb{R}^{h_{vi}}$ is the activation function, $\varepsilon_{c,i}(\hat{e}_i)$ is the approximation error and $\hat{e}_i = \mathcal{L}_i col(\hat{q}, \hat{v})$ is the estimation of the local neighbor consensus error e_i.

Taking the derivative of V_i with respect to $\hat{e}_i$ yields

$$V_{\hat{e}_i} = \nabla \phi_{c,i}^T w_{c,i} + \nabla \varepsilon_{c,i}$$

where $\nabla \phi_{c,i} = \partial \phi_{c,i}(\hat{e}_i)/\partial \hat{e}_i$ and $\nabla \varepsilon_{c,i} = \partial \varepsilon_{c,i}(\hat{e}_i)/\partial \hat{e}_i$.

Suppose that $\hat{w}_{c,i}$ is the weight estimation of $w_{c,i}$, then the output of the critic NN and its derivative are given by

$$\hat{V}_i(\hat{e}_i) = \hat{w}_{c,i}^T \phi_{c,i}(\hat{e}_i) \tag{23}$$
$$\hat{V}_{\hat{e}_i} = \nabla \phi_{c,i}^T \hat{w}_{c,i} \tag{24}$$

Then, the approximated Hamiltonian equation of agent i corresponding to (20) is rewritten as

$$H_i(\hat{e}_i, \hat{V}_{e_i}, \tau_{\bar{\mathcal{N}}_i}) = \hat{e}_i^T S_{ii} \hat{e}_i + \sum_{j \in \bar{\mathcal{N}}_i} \tau_j^T R_{ij} \tau_j + \hat{w}_{c,i}^T \nabla \phi_{c,i} \mathcal{L}_i (\bar{F}(\hat{q}, \hat{v}) + \bar{M}(\hat{q}) \tau) \tag{25}$$

Based on (25), we select the weight update law of the critic NN for agent i as

$$\begin{aligned} \dot{\hat{w}}_{c,i} &= -\alpha_i \frac{\partial E_i}{\partial \hat{w}_{c,i}} \\ &= -\alpha_i \frac{\rho_i(\rho_i^T \hat{w}_{c,i} + r_i(\hat{e}_i, \tau_{\bar{\mathcal{N}}_i}))}{\rho_i^T \rho_i + 1} \end{aligned} \tag{26}$$

where $\alpha_i > 0$ is the learning gain, $\rho_i = \nabla \phi_{c,i} \dot{\hat{e}}_i$, and $r_i(\hat{e}_i, \tau_{\bar{\mathcal{N}}_i}) = \hat{e}_i^T S_{ii} \hat{e}_i + \sum_{j \in \bar{\mathcal{N}}_i} \tau_j^T R_{ij} \tau_j$.

Let $\tilde{w}_{c,i} = \hat{w}_{c,i} - w_{c,i}$, then we have

$$\begin{aligned} \dot{\tilde{w}}_{c,i} &= \dot{\hat{w}}_{c,i} - \dot{w}_{c,i} \\ &= -\alpha_i \frac{\rho_i(\rho_i^T \hat{w}_{c,i} + r_i(\hat{e}_i, \hat{\tau}_{\bar{\mathcal{N}}_i}))}{\rho_i^T \rho_i + 1} \\ &= -\alpha_i \frac{\rho_i(\rho_i^T \tilde{w}_{c,i} + \varepsilon_{HJ,i})}{\rho_i^T \rho_i + 1} \end{aligned} \tag{27}$$

where $\varepsilon_{HJ,i} = w_{c,i}^T \rho_i + r_i(\hat{e}_i, \tau_{\bar{\mathcal{N}}_i})$.

B. ACTION NN DESIGN

Here, an action NN is employed to approximate the control policy for agent i as follows

$$\tau_i(\hat{e}_i) = w_{a,i}^T \phi_{a,i}(\hat{e}_i) + \varepsilon_{a,i}(\hat{e}_i)$$

where $w_{a,i} \in \mathbb{R}^{h_{ci} \times m}$ is the ideal weight matrix with h_{ci} being the number of neurons in the hidden layer, $\phi_{a,i}(\hat{e}_i) \in \mathbb{R}^{h_{ci}}$ is the activation function, and $\varepsilon_{a,i}(\hat{e}_i)$ is the approximation error.

Let $\hat{w}_{a,i}$ be the weight estimation of $w_{a,i}$, then the estimated control policy for agent i can be expressed as

$$\hat{\tau}_i(\hat{e}_i) = \hat{w}_{a,i}^T \phi_{a,i}(\hat{e}_i) \tag{28}$$

Furthermore, the optimal control policy of agent i using the gradient of the value function (30) can be given as

$$\tau_i = -\frac{1}{2} d_i R_{ii}^{-1} \bar{M}_i^T(\hat{q}_i) \nabla \phi_{c,i}^T \hat{w}_{c,i}$$

Then the approximation error of the ith action NN is defined as

$$\nu_{a,i} := \hat{w}_{a,i}^T \phi_{a,i}(\hat{e}_i) + \frac{1}{2} d_i R_{ii}^{-1} \bar{M}_i^T(\hat{q}_i) \nabla \phi_{c,i}^T \hat{w}_{c,i} \quad (29)$$

Based on (29), the weight update law of the action NN for agent i is selected as

$$\hat{w}_{a,i} = -\frac{\beta_i \phi_{a,i}(\hat{e}_i) \nu_{a,i}^T}{\phi_{a,i}^T(\hat{e}_i)\phi_{a,i}(\hat{e}_i) + 1} \quad (30)$$

Let $\tilde{w}_{a,i} = \hat{w}_{a,i} - w_{a,i}$, then one can obtain

$$\begin{aligned}\dot{\tilde{w}}_{a,i} &= \dot{\hat{w}}_{a,i} - \dot{w}_{a,i} \\ &= -\frac{\beta_i \phi_{a,i}(\hat{e}_i) \nu_{a,i}^T}{(\phi_{a,i}^T(\hat{e}_i)\phi_{a,i}(\hat{e}_i) + 1)^2} \\ &= -\frac{\beta_i \phi_{a,i}(\hat{e}_i)(\tilde{w}_{a,i}^T \phi_{a,i}(\hat{e}_i) + D_i \tilde{w}_{c,i} + \Upsilon_i)^T}{\phi_{a,i}^T(\hat{e}_i)\phi_{a,i}(\hat{e}_i) + 1}\end{aligned}$$

where $D_i = \frac{1}{2} d_i R_{ii}^{-1} \bar{M}_i^T(\hat{q}_i) \nabla \phi_{c,i}^T$ and $\Upsilon_i = D_i w_{c,i} + w_{a,i}^T \phi_{a,i}(\hat{e}_i)$.

Assumption 3: The ideal weight matrices, the activation functions and the reconstruction errors of the critic and action NNs are bounded, that is, there exist positive constants $w_{c,M}$, $w_{a,M}$, $\phi_{c,M}$, $\phi_{a,m}$, $\phi_{a,M}$, $\varepsilon_{c,M}$ and $\varepsilon_{a,M}$ such that $\|w_{c,i}\| \leq w_{c,M}$, $\|w_{a,i}\| \leq w_{a,M}$ $\|\phi_{c,i}\| \leq \phi_{c,M}$, $\phi_{a,m} \leq \|\phi_{a,i}\| \leq \phi_{a,M}$, $\|\varepsilon_{c,i}\| \leq \varepsilon_{c,M}$ and $\|\varepsilon_{a,i}\| \leq \varepsilon_{a,M}$.

Theorem 2: Consider the error dynamics be given by (18). Let the critic-action NN for agent i be given by (23) and (28), the update laws for the two NNs are provided by (26) and (30). Then the integrated consensus error e_i, the critic and action NN estimation errors $\tilde{w}_{c,i}$ and $\tilde{w}_{a,i}$ are UUB.

Proof: Choose a local Lyapunov function candidate as

$$L_{s,i} = L_{i,1} + L_{i,2} + L_{i,3} + L_{i,4} \quad (31)$$

where $L_{i,1} = tr(\tilde{w}_{c,i}^T \tilde{w}_{c,i})/2\alpha_i$, $L_{i,2} = tr(\tilde{w}_{a,i}^T \tilde{w}_{a,i})/2\beta_i$, $L_{i,3} = tr(\tilde{w}_{o,i}^T \tilde{w}_{o,i})$, $L_{i,4} = e_i^T e_i + 2\Gamma_i V_i$.

Based on Assumption 1 and 3, it is easily known that $\varepsilon_{HJ,i}$, D_i and Υ_i are bounded, i.e., there exist positive constants $\varepsilon_{HJ,M}$, D_M and Υ_M such that $\|\varepsilon_{HJ,i}\| \leq \varepsilon_{HJ,M}$, $\|D_i\| \leq D_M$ and $\|\Upsilon_i\| \leq \Upsilon_M$.

According to (27), we take the derivation of $L_{i,1}$ as

$$\begin{aligned}\dot{L}_{i,1} &= \frac{1}{\alpha_i} tr(\tilde{w}_{c,i}^T \dot{\tilde{w}}_{c,i}) \\ &= -tr(\frac{\tilde{w}_{c,i}^T \rho_i(\rho_i^T \tilde{w}_{c,i} + \varepsilon_{HJ,i})}{\rho_i^T \rho_i + 1}) \\ &= \frac{-\tilde{w}_{c,i}^T \rho_i (\tilde{w}_{c,i}^T \rho_i)^T - \tilde{w}_{c,i}^T \rho_i \varepsilon_{HJ,i}}{\rho_i^T \rho_i + 1}\end{aligned} \quad (32)$$

Due to $\tilde{w}_{c,i}^T \rho_i (\tilde{w}_{c,i}^T \rho_i)^T > 0$, then there exists a constant $\Lambda_i > 0$ such that $\Lambda_i \|\tilde{w}_{c,i}\|^2 \leq \tilde{w}_{c,i}^T \rho_i (\tilde{w}_{c,i}^T \rho_i)^T$. And it can observe that $\frac{\rho_i^T \rho_i}{\rho_i^T \rho_i + 1} \leq 1$, then one has

$$\begin{aligned}\dot{L}_{i,1} &\leq -\frac{\Lambda_i}{\rho_i^T \rho_i + 1} \|\tilde{w}_{c,i}\|^2 + \frac{1}{2(\rho_i^T \rho_i + 1)} (\|\tilde{w}_{c,i}^T \rho_i\|^2 + \varepsilon_{HJ,i}^2) \\ &\leq -(\frac{\Lambda_i}{\rho_M^2 + 1} - \frac{1}{2}) \|\tilde{w}_{c,i}\|^2 + \frac{1}{2\rho_m^2 + 1} \varepsilon_{HJ,M}^2\end{aligned} \quad (33)$$

Taking the derivative of $L_{i,2}$ yields

$$\begin{aligned}\dot{L}_{i,2} &= \frac{1}{\beta_i} tr(\tilde{w}_{a,i}^T \dot{\tilde{w}}_{a,i}) \\ &= -tr(\frac{\tilde{w}_{a,i}^T \phi_{a,i}(\tilde{w}_{a,i}^T \phi_{a,i} + D_i \tilde{w}_{c,i} + \Upsilon_i)^T}{\phi_{a,i}^T \phi_{a,i} + 1}) \\ &= -\frac{(\tilde{w}_{a,i}^T \phi_{a,i})^T \tilde{w}_{a,i}^T \phi_{a,i}}{\phi_{a,i}^T \phi_{a,i} + 1} - \frac{(\tilde{w}_{a,i}^T \phi_{a,i})^T D_i \tilde{w}_{c,i}}{\phi_{a,i}^T \phi_{a,i} + 1} \\ &\quad - \frac{(\tilde{w}_{a,i}^T \phi_{a,i})^T \Upsilon_{a,i}}{\phi_{a,i}^T \phi_{a,i} + 1}\end{aligned} \quad (34)$$

Since $(\tilde{w}_{a,i}^T \phi_{a,i})^T \tilde{w}_{a,i}^T \phi_{a,i} > 0$, there exists a constant $\Pi_i > 0$ such that $\Pi_i \|\tilde{w}_{a,i}\|^2 \leq (\tilde{w}_{a,i}^T \phi_{a,i})^T \tilde{w}_{a,i}^T \phi_{a,i}$. And it is known that $\frac{\phi_{a,i}^T \phi_{a,i}}{\phi_{a,i}^T \phi_{a,i} + 1} \leq 1$, then (34) can be rewritten as

$$\begin{aligned}\dot{L}_{i,2} &\leq -\frac{\Pi_i \|\tilde{w}_{a,i}\|^2}{\phi_{a,M}^2 + 1} + \frac{\|\phi_{a,i}\|^2 \|\tilde{w}_{a,i}\|^2}{2(\phi_{a,i}^T \phi_{a,i} + 1)} + \frac{\|D_i\|^2 \|\tilde{w}_{c,i}\|^2}{2(\phi_{a,i}^T \phi_{a,i} + 1)} \\ &\quad + \frac{\|\phi_{a,i}\|^2 \|\tilde{w}_{a,i}\|^2}{2(\phi_{a,i}^T \phi_{a,i} + 1)} + \frac{\|\Upsilon_i\|^2}{2(\phi_{a,i}^T \phi_{a,i} + 1)} \\ &\leq -(\frac{\Pi_i}{\phi_{a,M}^2 + 1} - 1) \|\tilde{w}_{a,i}\|^2 + \frac{D_M^2}{2(\phi_{a,m}^2 + 1)} \|\tilde{w}_{c,i}\|^2 \\ &\quad + \frac{\Upsilon_M^2}{2(\phi_{a,m}^2 + 1)}\end{aligned} \quad (35)$$

Subsequently, consider the third term, $L_{i,3}$, the derivative remains the same as in (15), so it yields

$$\dot{L}_{i,3} \leq (\sigma_i \psi_M - \gamma_i) \|\tilde{w}_{o,i}\|^2 + \sigma_i \psi_M \|\tilde{\eta}_i\|^2 + \gamma_i w_{o,M}^2 \quad (36)$$

Next, consider the derivative of $L_{i,4}$ with respect to time, then one has

$$\begin{aligned}\dot{L}_{i,4} &= 2e_i^T \dot{e}_i + 2\Gamma_i \dot{V}_i \\ &= 2e_i^T \mathcal{L}_i(\bar{F}(q,v) + \bar{M}(q)\tau) - 2\Gamma_i r_i(e_i, \tau_{\bar{\mathcal{N}}_i}) \\ &= \sum_{j \in \bar{\mathcal{N}}_i} 2 l_{ij} e_i^T (\bar{F}_j(q_j, v_j) + \bar{M}_j(q_j)\tau_j) \\ &\qquad - 2\Gamma_i e_i^T S_{ii} e_i - 2\Gamma_i \sum_{j \in \bar{\mathcal{N}}_i} \tau_j^T R_{ij} \tau_j \\ &\leq \sum_{j \in \bar{\mathcal{N}}_i} (\|l_{ij} m_{Mi}\|^2 - 2\Gamma_i \lambda_{min}(R_{ij})) \|\tau_j\|^2 + 2(\|\bar{\mathcal{N}}_i\| \\ &\qquad - \Gamma_i \lambda_{min}(S_{ii})) \|e_i\|^2 + \sum_{j \in \bar{\mathcal{N}}_i} \|l_{ij} \bar{F}_j(q_j, v_j)\|^2\end{aligned} \quad (37)$$

where $\|\bar{\mathcal{N}}_i\|$ denotes the number of agent i and its neighborhood.

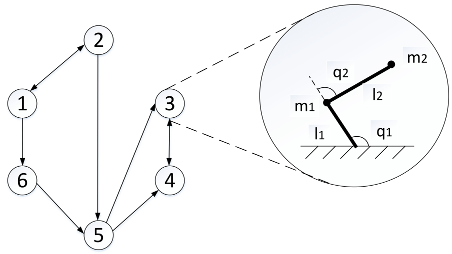
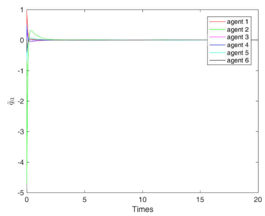
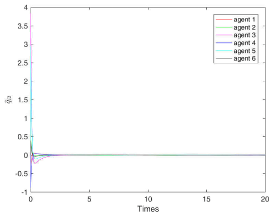


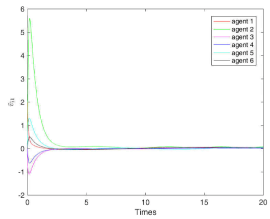
FIGURE 1. Communication graph of six 2-DOF nonlinear manipulators with revolute joints.



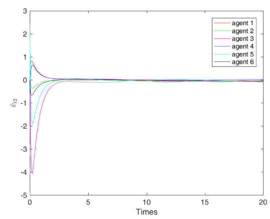
(a)



(b)



(c)

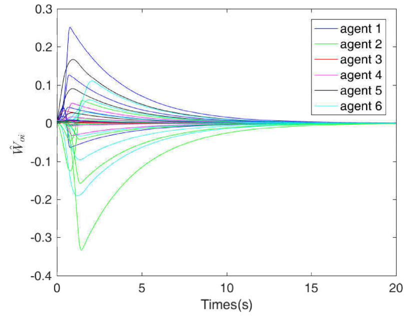


(d)

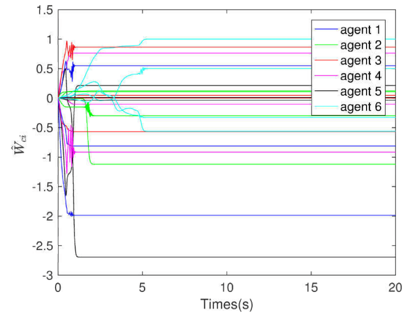
FIGURE 2. The estimation errors of the position and velocity vectors.

Finally, based on (33), (34), (36), and (37), the derivative of L_{si} is written as

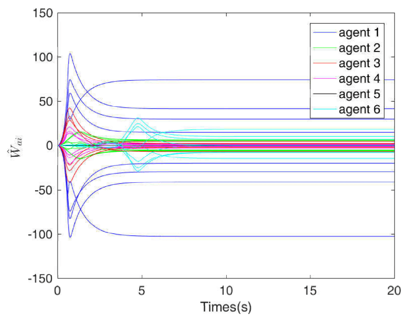
$$
\begin{aligned}
\dot{L}_{s,i} \leq& -(\frac{\Lambda_i}{\rho_M^2+1}-\frac{1}{2})\|\tilde{w}_{c,i}\|^2+\frac{1}{2\rho_m^2+1}\varepsilon_{HJ,M}^2 \\
&-(\frac{\Pi_i}{\phi_{a,M}^2+1}-1)\|\tilde{w}_{a,i}\|^2+\frac{D_M^2}{2(\phi_{a,m}^2+1)}\|\tilde{w}_{c,i}\|^2 \\
&+\frac{\Upsilon_M^2}{2(\phi_{a,m}^2+1)}+(\sigma_i\psi_M-\gamma_i)\|\tilde{w}_{o,i}\|^2+\sigma_i\psi_M\|\tilde{\eta}_i\|^2 \\
&+\gamma_i w_{o,M}^2+\sum_{j\in\bar{\mathcal{N}}_i}(\|l_{ij}m_{Mi}\|^2-2\Gamma_i\lambda_{min}(R_{ij}))\|\tau_j\|^2 \\
&+2(\|\bar{\mathcal{N}}_i\|-\Gamma_i\lambda_{min}(S_{ii}))\|e_i\|^2+\sum_{j\in\bar{\mathcal{N}}_i}\|l_{ij}\bar{F}_j(q_j,v_j)\|^2 \\
\leq& -(\gamma_i-\sigma_i\psi_M)\|\tilde{w}_{o,i}\|^2-(\frac{\Lambda_i}{\rho_M^2+1}-\frac{D_M^2}{2(\phi_{a,m}^2+1)} \\
&-\frac{1}{2})\|\tilde{w}_{c,i}\|^2-(\frac{\Pi_i}{\phi_{a,M}^2+1}-1)\|\tilde{w}_{a,i}\|^2+2(\|\bar{\mathcal{N}}_i\| \\
&-\Gamma_i\lambda_{min}(S_{ii}))\|e_i\|^2+\sum_{j\in\bar{\mathcal{N}}_i}(\|l_{ij}m_{Mi}\|^2-2\Gamma_i \\
&\times\lambda_{min}(R_{ij}))\|\tau_j\|^2+\Xi_i
\end{aligned}
\tag{38}
$$



(a)



(b)



(c)

FIGURE 3. The NN weight estimations: (a) Observer NNs; (b) Critic NNs; (c) Action NNs.

where $\Xi_i = \frac{1}{2\rho_m^2+1}\varepsilon_{HJ,M}^2+\frac{\Upsilon_M^2}{2(\phi_{a,m}^2+1)}+\sigma_i\psi_M\Theta_{\eta,i}^2+\gamma_i w_{o,M}^2+\sum_{j\in\bar{\mathcal{N}}_i}\|l_{ij}\bar{F}_j(q_j,v_j)\|^2$.

For simplicity, the following new variables is defined as

$$
\begin{aligned}
\varpi_{i,1} &= \gamma_i-\sigma_i\psi_M \\
\varpi_{i,2} &= \frac{\Lambda_i}{\rho_M^2+1}-\frac{D_M^2}{2(\phi_{a,m}^2+1)}-\frac{1}{2} \\
\varpi_{i,3} &= \frac{\Pi_i}{\phi_{a,M}^2+1}-1 \\
\varpi_{i,4} &= 2(\Gamma_i\lambda_{min}(S_{ii})-\|\bar{\mathcal{N}}_i\|)
\end{aligned}
$$

If Γ_i satisfies $\Gamma_i > \max\{\frac{2\|\bar{\mathcal{N}}_i\|}{\lambda_{min}(S_{ii})}, \frac{\|l_{ij}m_{Mi}\|^2}{2\lambda_{min}(R_{ij})}\}$ and the inequalities $\|\tilde{w}_{o,i}\| > \sqrt{\frac{\Xi_i}{\varpi_{i,1}}}$ or $\|\tilde{w}_{c,i}\| > \sqrt{\frac{\Xi_i}{\varpi_{i,2}}}$ or $\|\tilde{w}_{a,i}\| > \sqrt{\frac{\Xi_i}{\varpi_{i,3}}}$ or $\|e_i\| > \sqrt{\frac{\Xi_i}{\varpi_{i,4}}}$ hold, then $\dot{L}_{s,i} < 0$. This implies that the integrated error, the weight estimation errors of the observer-critic-action NNs are UUB.

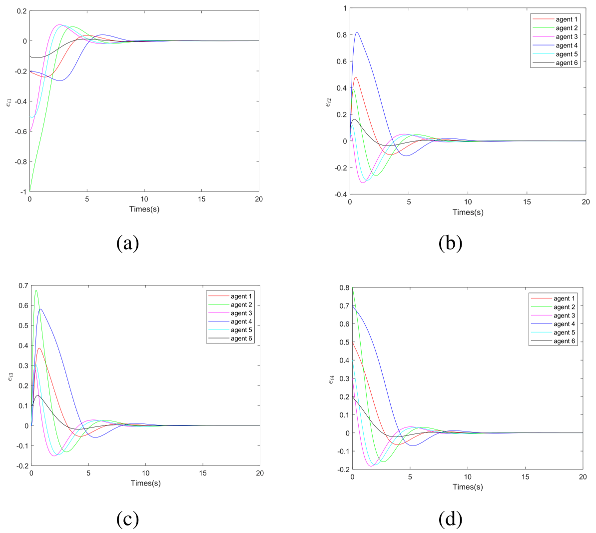


FIGURE 4. **The neighbor consensus errors of the position and velocity vectors.**

VI. SIMULATION

This section presents a numerical simulation to verify the resulting optimal consensus control algorithm. The NELS is a connection of six 2-DoF nonlinear manipulators with revolute joints as shown in Fig. 1. Refer to [25], [33], each agent is modeled as a two-link robotic manipulator with the the inertia and Coriolis matrices given by

$$M_i(q_i) = \begin{bmatrix} \delta_{i1} + \delta_{i2} + 2\delta_{i3}cos(q_{i2}) & \delta_{i2} + \delta_{i3}cos(q_{i2}) \\ \delta_{i2} + \delta_{i3}cos(q_{i2}) & \delta_{i2} \end{bmatrix}$$

$$C_i(q_i, \dot{q}_i) = \begin{bmatrix} -\delta_{i2}sin(q_{i2})\dot{q}_{i2} & -\delta_{i2}sin(q_{i2})(\dot{q}_{i1} + \dot{q}_{i2}) \\ \delta_{i2}sin(q_{i2})\dot{q}_{i1} & 0 \end{bmatrix}$$

and the gravity vector given by

$$g_i(q_i) = \begin{bmatrix} \delta_{i4}gcos(q_{i1}) + \delta_{i5}gcos(q_{i1} + q_{i2}) \\ \delta_{i5}gcos(q_{i1} + q_{i2}) \end{bmatrix}$$

where $\delta_{i1} = J_{i1} + m_{i2}l_{i1}^2$, $\delta_{i2} = J_{i2} + 0.25m_{i2}l_{i2}^2$, $\delta_{i3} = 0.5m_{i2}l_{i1}l_{i2}$, $\delta_{i4} = (0.5m_{i1} + m_{i2})l_{i1}$, $\delta_{i5} = 0.5m_{i2}l_{i2}$.

In the simulation, the parameters of six robot manipulators are: $m_{i1} = 4kg, m_{i2} = 2kg$ and $l_{i1} = 0.4m, l_{i2} = 0.4m$ for agent $i = 1, 2, 3$; $m_{i1} = 1.5kg, m_{i2} = 3kg$ and $l_{i1} = 0.3m, l_{i2} = 0.3m$ for agent $i = 4, 5, 6$, the weight matrices $S_{ii} = 5I_4, R_{ii} = I_2, R_{ij} = 0.5I_2$, and the gains $\sigma_i = 2, \gamma_i = 5$, $\alpha_i = 3 \times 10^{-2}$, $\beta_i = 5 \times 10^{-2}$. The observer-critic-actor NN activation functions are chosen as

$$\psi_i(\hat{q}_i, \hat{v}_i) = \begin{bmatrix} \hat{q}_i^T\hat{q}_i & 0 & 0 \\ \hat{v}_i^T\hat{v}_i & 0 & 0 \\ 0 & \hat{q}_{i1}^2 & \hat{q}_{i1}\hat{q}_{i2} \\ 0 & \hat{q}_{i1}\hat{q}_{i2} & \hat{q}_{i2}^2 \end{bmatrix} \begin{bmatrix} 1 \\ \tau_i \end{bmatrix}$$

$$\phi_{c,i}(e_i) = [tanh(e_{i1}^2)\ tanh(e_{i1}e_{i2})\ tanh(e_{i1}e_{i3})\ tanh(e_{i1}e_{i4})\ tanh(e_{i2}^2)\ tanh(e_{i2}e_{i3})\ tanh(e_{i2}e_{i4})\ tanh(e_{i3}^2)\ tanh(e_{i3}e_{i4})\ tanh(e_{i4}^2)]^T$$

$$\phi_{a,i}(e_i) = \begin{bmatrix} tanh(e_{i1}) & tanh(e_{i2}) & tanh(e_{i3}) & tanh(e_{i4}) \end{bmatrix}^T$$

Fig. 2 shows the estimation errors of the position and velocity for all agents, from which it can be observed that the estimated positions and velocities using the designed NN-based observers can converge to their real values well. In Fig. 3, the evolutions of the weight estimation for the observer-critic-action networks are depicted which shows that they are UUB. Moreover, Fig. 4 plots the local consensus error result for the positions and velocities of the manipulators, it is easily observed that the NELS can reach consensus under the proposed optimal controllers.

VII. CONCLUSION

In this paper, for the NELSs without velocity measurements, we built a distributed optimal control scheme using only the position information. For each agent, an NN-based observer was designed to estimate the unknown system dynamic model and unmeasured velocity vectors. Besides, the UUB of the integrated observer estimation errors and the weight estimation errors for observer NNs was proven. From the differential game point of view, the optimal consensus problem of NELSs was transformed into an optimal regulation problem of second-order MASs. To solve the optimal consensus issue online, a model-free PI algorithm was introduced to obtain the solutions of the coupled HJB equations. Then, the update rules for the critic-action network weight were designed and the stability of the closed-loop system was demonstrated. Finally, the simulation was provided to verify the effectiveness of the proposed approach. The proposed approach was based on local cooperation but not competition, so one future challenging problem is how to address the distributed coordination and competition problem.

REFERENCES

[1] Y. Sun, D. Dong, H. Qin, N. Wang, and X. Li, "Distributed coordinated tracking control for multiple uncertain Euler–Lagrange systems with time-varying communication delays," *IEEE Access*, vol. 7, pp. 12598–12609, 2019.

[2] W. He, C. Xu, Q.-L. Han, F. Qian, and Z. Lang, "Finite-time L_2 leader–follower consensus of networked Euler–Lagrange systems with external disturbances," *IEEE Trans. Syst., Man, Cybern., Syst.*, vol. 48, no. 11, pp. 1920–1928, Nov. 2018.

[3] M. Lu and L. Liu, "Leader-following consensus of multiple uncertain Euler–Lagrange systems subject to communication delays and switching networks," *IEEE Trans. Autom. Control*, vol. 63, no. 8, pp. 2604–2611, Aug. 2018.

[4] H.-X. Hu, G. Wen, W. Yu, Q. Xuan, and G. Chen, "Swarming behavior of multiple Euler–Lagrange systems with cooperation–competition interactions: An auxiliary system approach," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 29, no. 11, pp. 5726–5737, Nov. 2018.

[5] R. Machuca, C. I. Aldana, R. Munguía, and E. Nuño, "Cartesian space consensus of heterogeneous and uncertain Euler-Lagrange systems using artificial neural networks," *Int. J. Control, Automat. Syst.*, vol. 15, no. 3, pp. 1447–1455, Jun. 2017.

[6] Y. Liu and Y. Jia, "Adaptive consensus control for multiple Euler-Lagrange systems with external disturbance," *Int. J. Control, Automat. Syst.*, vol. 15, no. 1, pp. 205–211, Feb. 2017.

[7] H. Cai and J. Huang, "The leader-following consensus for multiple uncertain Euler-Lagrange systems with an adaptive distributed observer," *IEEE Trans. Autom. Control*, vol. 61, no. 10, pp. 3152–3157, Oct. 2016.

[8] J. R. Klotz, S. Obuz, Z. Kan, and W. E. Dixon, "Synchronization of uncertain Euler–Lagrange systems with uncertain time-varying communication delays," *IEEE Trans. Cybern.*, vol. 48, no. 2, pp. 807–817, Feb. 2018.

[9] X. Zhao, C. Ma, X. Xing, and X. Zheng, "A stochastic sampling consensus protocol of networked Euler-Lagrange systems with application to two-link manipulator," *IEEE Trans. Ind. Informat.*, vol. 11, no. 4, pp. 907–914, Aug. 2015.

[10] E. Nuño and R. Ortega, "Achieving consensus of Euler–Lagrange agents with interconnecting delays and without velocity measurements via passivity-based control," *IEEE Trans. Control Syst. Technol.*, vol. 26, no. 1, pp. 222–232, Jan. 2018.

[11] Y. Zhao, Z. Duan, and G. Wen, "Distributed finite-time tracking of multiple Euler–Lagrange systems without velocity measurements," *Int. J. Robust Nonlinear Control*, vol. 25, no. 11, pp. 1688–1703, 2015.

[12] Y. H. Kim and F. L. Lewis, "Neural network output feedback control of robot manipulators," *IEEE Trans. Robot. Autom.*, vol. 15, no. 2, pp. 301–309, Apr. 1999.

[13] X. Peng, Z. Geng, and J. Sun, "The specified finite-time distributed observers-based velocity-free attitude synchronization for rigid bodies on SO(3)," *IEEE Trans. Syst., Man, Cybern., Syst.*, to be published. doi: 10.1109/TSMC.2018.2831778.

[14] Q. Yang, H. Fang, J. Chen, Z.-P. Jiang, and M. Cao, "Distributed global output-feedback control for a class of Euler–Lagrange systems," *IEEE Trans. Autom. Control*, vol. 62, no. 9, pp. 4855–4861, Sep. 2017.

[15] H. A. Hashim, S. El-Ferik, and F. L. Lewis, "Adaptive synchronisation of unknown nonlinear networked systems with prescribed performance," *Int. J. Syst. Sci.*, vol. 48, no. 4, pp. 885–898, 2017.

[16] H. A. Hashim, S. El-Ferik, and F. L. Lewis, "Neuro-adaptive cooperative tracking control with prescribed performance of unknown higher-order nonlinear multi-agent systems," *Int. J. Control*, vol. 92, no. 2, pp. 445–460, 2019.

[17] H. Modares, F. L. Lewis, W. Kang, and A. Davoudi, "Optimal synchronization of heterogeneous nonlinear systems with unknown dynamics," *IEEE Trans. Autom. Control*, vol. 63, no. 1, pp. 117–131, Jan. 2018.

[18] H. Zhang, D. Yue, C. Dou, W. Zhao, and X. Xie, "Data-driven distributed optimal consensus control for unknown multiagent systems with input-delay," *IEEE Trans. Cybern.*, vol. 49, no. 6, pp. 2095–2105, Jun. 2019.

[19] F.-Y. Wang, H. Zhang, and D. Liu, "Adaptive dynamic programming: An introduction," *IEEE Comput. Intell. Mag.*, vol. 4, no. 2, pp. 39–47, May 2009.

[20] L. Zhang, Y. Zhu, H. Shi, and K.-S. Hwang, "Adaptive dynamic programming approach on optimal control for affinely pseudo-linearized nonlinear system," *IEEE Access*, vol. 7, pp. 75132–75142, 2019.

[21] K. G. Vamvoudakis, F. L. Lewis, and G. R. Hudas, "Multi-agent differential graphical games: Online adaptive learning solution for synchronization with optimality," *Automatica*, vol. 48, no. 8, pp. 1598–1611, 2012.

[22] M. I. Abouheaf, F. L. Lewis, K. G. Vamvoudakis, S. Haesaert, and R. Babuska, "Multi-agent discrete-time graphical games and reinforcement learning solutions," *Automatica*, vol. 50, no. 12, pp. 3038–3053, 2014.

[23] J. Li, H. Modares, T. Chai, F. L. Lewis, and L. Xie, "Off-policy reinforcement learning for synchronization in multiagent graphical games," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 28, no. 10, pp. 2434–2445, Oct. 2017.

[24] L. N. Tan, "Omnidirectional-vision-based distributed optimal tracking control for mobile multirobot systems with kinematic and dynamic disturbance rejection," *IEEE Trans. Ind. Electron.*, vol. 65, no. 7, pp. 5693–5703, Jul. 2018.

[25] Y. Dong and J. Chen, "Adaptive control for rendezvous problem of networked uncertain Euler–Lagrange systems," *IEEE Trans. Cybern.*, vol. 49, no. 6, pp. 2190–2199, Jun. 2019.

[26] Z. Wang, L. Liu, and H. Zhang, "Neural network-based model-free adaptive fault-tolerant control for discrete-time nonlinear systems with sensor fault," *IEEE Trans. Syst., Man, Cybern. Syst.*, vol. 47, no. 8, pp. 2351–2362, Apr. 2017.

[27] G. Wen, W. Yu, Z. Li, X. Yu, and J. Cao, "Neuro-adaptive consensus tracking of multiagent systems with a high-dimensional leader," *IEEE Trans. Cybern.*, vol. 47, no. 7, pp. 1730–1742, Jul. 2017.

[28] F. Lewis, S. Jagannathan, and A. Yesildirak, *Neural Network Control Of Robot Manipulators And Non-Linear Systems*. London, U.K.: Taylor & Francis, 1998.

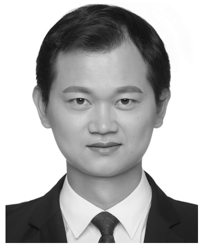
[29] Y. Hong, G. Chen, and L. Bushnell, "Distributed observers design for leader-following control of multi-agent networks," *Automatica*, vol. 44, no. 3, pp. 846–850, Mar. 2008.

[30] F. Abdollahi, H. A. Talebi, and R. V. Patel, "A stable neural network-based observer with application to flexible-joint manipulators," *IEEE Trans. Neural Netw.*, vol. 17, no. 1, pp. 118–129, Jan. 2006.

[31] D. Liu, Y. Huang, D. Wang, and Q. Wei, "Neural-network-observer-based optimal control for unknown nonlinear systems using adaptive dynamic programming," *Int. J. Control*, vol. 86, no. 9, pp. 1554–1566, Sep. 2013.

[32] Y. Lee and S. H. Zak, "Uniformly ultimately bounded fuzzy adaptive tracking controllers for uncertain systems," *IEEE Trans. Fuzzy Syst.*, vol. 12, no. 6, pp. 797–811, Dec. 2004.

[33] E. Nuño, "Consensus of Euler-Lagrange systems using only position measurements," *IEEE Trans. Control Netw. Syst.*, vol. 5, no. 1, pp. 489–498, Mar. 2018.



HUAIPIN ZHANG received the Ph.D. degree in control science and engineering from the Huazhong University of Science and Technology, Wuhan, China.

From 2017 to 2018, he was a Postdoctoral Research Fellow with the Department of Electrical Engineering, Yeungnam University, Gyeongsan, South Korea. He joined the Nanjing University of Posts and Telecommunications, Nanjing, China, in 2018. His research interests include nonlinear systems, cooperative control of multi-agent systems, adaptive dynamic programming, and event-triggered control scheme.

JU H. PARK received the Ph.D. degree in electronics and electrical engineering from the Pohang University of Science and Technology (POSTECH), Pohang, South Korea, in 1997.

In 2000, he joined Yeungnam University, Gyeongsan, South Korea, where he is currently the Chuma Chair Professor. From 2006 to 2007, he was a Visiting Professor with the Department of Mechanical Engineering, Georgia Institute of Technology. His research interests include neural networks, complex networks, multi-agent systems, and chaotic systems. He has published a number of papers in these areas. He is currently a Fellow of the Korean Academy of Science and Technology (KAST). He serves as an Editor for the *International Journal of Control, Automation, and Systems*. He is also a Subject Editor/an Associate Editor/an Editorial Board Member for several international journals, including *IET Control Theory and Applications*, *Applied Mathematics and Computation*, *Journal of The Franklin Institute*, *Nonlinear Dynamics*, *Journal of Applied Mathematics and Computing*, *Cogent Engineering*, and the IEEE TRANSACTIONS ON FUZZY SYSTEMS.

WEI ZHAO received the B.S. degree in mathematics and applied mathematics from Henan University, Kaifeng, China, in 2011, and the Ph.D. degree in mechanical and electronic engineering from the Huazhong University of Science and Technology, Wuhan, China, in 2017.

Since 2018, she has been a Postdoctoral Research Fellow with the School of Mathematics, Southeast University, Nanjing, China. Her current research interests include multi-agent systems, tracking control, adaptive dynamic programming, event-triggered control, and optimal control.