Research Article

Adaptive Reinforcement Learning-Enhanced Motion/Force Control Strategy for Multirobot Systems

Phuong Nam Dao , Duy Khanh Do, and Dinh Khue Nguyen

School of Electrical Engineering, Hanoi University of Science and Technology, Hanoi, Vietnam

Correspondence should be addressed to Phuong Nam Dao; nam.daophuong@hust.edu.vn

Received 17 January 2021; Revised 14 March 2021; Accepted 4 May 2021; Published 1 June 2021

Academic Editor: Hussein Abulkasim

Copyright © 2021 Phuong Nam Dao et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

This paper presents an adaptive reinforcement learning- (ARL-) based motion/force tracking control scheme consisting of the optimal motion dynamic control law and force control scheme for multimanipulator systems. Specifically, a new additional term and appropriate state vector are employed in designing the ARL technique for time-varying dynamical systems with online actor/critic algorithm to be established by minimizing the squared Bellman error. Additionally, the force control law is designed after obtaining the computation of constraint force coefficient by the Moore–Penrose pseudo-inverse matrix. The tracking effectiveness of the ARL-based optimal control is verified in the closed-loop system by theoretical analysis. Finally, simulation studies are conducted on a system of three manipulators to validate the physical realization of the proposed optimal tracking control design.

1. Introduction

In recent years, the tracking control problem for robotic systems has been remarkable, considered from not only academia but also industrial automation [1, 2]. Among different schemes, robotic systems can mainly be separated into two categories: (1) mobile robotics [3, 4] and (2) robot manipulators [5]. However, because of the growing requirements for high productivity, the networked robotic systems that are known as the collaboration of many robots can fulfill different tasks by mobile manipulators [6], tractor trailers [3], bilateral teleoperators [5], and cooperating mobile manipulators (CMMs) [7]. It is worth emphasizing that, unlike nonlinear multiagent systems [8, 9] studying the relation between agents, CMMs consider the interaction with the rigid object. In the literature of CMMs' control objectives, they can be classified into two categories, i.e., multiple mobile robot manipulators in cooperation carrying a common object with unknown parameters and disturbances [7, 10–18]; one of them tightly holds the object by the end effector, and the remaining mobile manipulators' end effector follows a trajectory on the surface of the object [19, 20]. Most of the existing control literature of networked

robotics mainly focuses on the implementation of nonlinear controllers. According to Lyapunov's direct technique, authors in [5] presented the robust controller for a bilateral teleoperation under time-varying delays and without relative motion. Recently, the backstepping technique of networked robotics has achieved much attention [3, 11, 21]. In [3], a backstepping method is utilized to implement an adaptive trajectory tracking control design for a tractor trailer, where the proposed control scheme can stabilize for cascade control systems. For the purpose of easily computing the backstepping method and handling the full-state constraint, authors in [21, 22] pointed out the appropriate term α in the asymmetric barrier Lyapunov function (ABLF). This backstepping method is extended in [11] to deal with unknown dead zone nonlinearity, time-varying term using the input-driven filter term and Nussbaum function, respectively. Another consideration addressed in the existing literature is to mention the holonomic-nonholonomic constraint [12, 19, 20, 23, 24]. Motivated by this approach, the corresponding motion dynamic model and constraint force equation are established for finding appropriate controllers [12, 19, 20, 23, 24]. As a result, the robust adaptive controller was realized by estimating the derivative of the Lyapunov

function and combining the constraint force control to obtain the motion/force control scheme for multimanipulator systems [12]. Due to dynamic uncertainties and external disturbance, fuzzy and neural network (NN) method are popularly considered to estimate dynamic uncertainties for guaranteeing system stability [7, 10, 20]. However, the work in [23] utilized an estimation in the absence of fuzzy and NNs for designing a decentralized controller for cooperating manipulators. In contrast to this technique, authors in [7, 12, 25] proposed the centralized control law based on the consideration of dynamics of m manipulators. The scheme in [7] extends the neural network-based centralized coordination control to achieve the tracking of each mobile manipulator by the consideration of additional actuators using servo motors. However, these nonlinear control solutions do not take into account the challenges, such as input-state saturation, and control objectives are different from the traditional tracking control problem. Therefore, we propose motion/force control with a progressive optimality principle to improve the tracking performance.

In the implementation of the optimality principle, an important aspect is developed for robotic systems with the consideration of the robot-environment interaction [26–28]. In control design for robotic systems, and so on, there always exists an interaction between the end-effector and the environment, which may lead to various challenges of system performance. Therefore, the control scheme should maintain a solution for considering the system interactions. Needless to say, the control systems of the robot-environment interaction have been paid much attention by many researchers for several decades [28]. Due to the effect of the interaction between the robot's end effector and the unknown environment, the desired trajectory in the joint space needs to be modified by the optimality approach [26–28]. In [26], the method to generate the virtual desired trajectory can be implemented by adaptive output feedback optimal control for the unknown impedance model. A different approach for finding the virtual reference trajectory can be developed by Q-learning [27]. Furthermore, based on ARL-based admittance control, authors in [28] proposed the completed control structure with the additional term of torque estimation and adaptive controller. An extension of the robot-environment interaction can be known as the physical human-robot interaction [15, 29]. In [29], the dynamic model of an n -dimensional manipulator system is established under the task space with the consideration of the end effector of the robot and human or environment. Moreover, the dynamic model established more a desired admittance model in the task space for obtaining virtual desired trajectory under constraint region by a soft saturation function [29]. A different discussion of the human-robot interaction with the estimation of human intention was mentioned in [15]. Furthermore, the motion/force control law was proposed using the computation of constraint force without the optimality problem [15]. An approach of optimization for the dual-arm manipulator with the consideration of relative Jacobian was mentioned in [16] by the QP solver. The extension from dual-arm manipulators to multiarm robots is discussed with the classical

backstepping structure and optimal control law being constructed to tackle optimal behaviors involving obstacle avoidance and energy saving [11]. Additionally, authors in [17] studied multiarm systems under time-varying communication delay by task-space synchronization with a fixed threshold. For different networked systems of bilateral teleoperation systems, the optimality principle was also considered for implementing control design [30]. However, the disadvantage of the control law in [30] is no consideration in solving the Riccati equation. Furthermore, the optimality principle in robotic systems has been almost focused on the interaction of the human-robot's end effector and the environment. This article proposes the consideration of optimal control in trajectory tracking in the motion/force control problem.

For finding an optimal control problem to minimize a given performance index in dynamical systems, it is necessary to solve the so-called Hamilton–Jacobi–Bellman (HJB) equation obtained from the dynamic principle. However, it is hard to analytically solve this HJB equation due to leading to a nonlinear partial differential equation. Among the numerical methods which have been considered to solve the HJB equation, the remarkable iterative structure has been developed to solve via online based on the reinforcement learning (RL) principle being inspired by machine learning [27, 31–39]. In order to implement the numerical algorithm for solving the HJB equation, there are two major directions, including online actor/critic [31, 32] and off-policy technique using integral reinforcement learning (IRL) [38, 39]. The first approach using online actor/critic is to develop based on the property of Hamiltonian $H(x, u^*, (\partial V^*/\partial x))$ being 0. As a result, the iterative algorithm can be considered by the Bellman error with solving the optimization problem. For designing the ARL scheme in linear dynamical systems [31] and in nonlinear systems [32, 40], the methods are realized to find the optimal control input being Kronecker product and approximating neural networks (NNs), respectively. Furthermore, this technique is extended for several situations, such as goal representation heuristic dynamic programming (GRHDP) with the multivariable tracking scheme [35] and uncertain discrete-time systems by using NNs [33]. The second approach using the off-policy technique is implemented by considering the deviation of the performance index in two different times with appropriate data collection [38, 39]. It is worth emphasizing that almost existing works discussed the ARL design in time-invariant systems [27, 33, 35, 36, 38, 39]. In contrast, for establishing the optimal control problem in robotic systems, it requires the consideration of ARL in time-varying systems because of the desired time-varying trajectory [31, 32]. Furthermore, the ARL-based optimal control with the actor/critic learning structure has just been discussed in [41] for time-invariant systems. The key idea is to establish simultaneously the learning of neural networks (NNs) in the actor and critic part [41]. The work in [32] developed the actor/critic learning structure for manipulators by combining the sliding mode control (SMC) technique in the absence of constraint force. Regarding the multimanipulator systems, the proposed control law in [12]

only implemented the robust adaptive control based on the traditional nonlinear control technique and model separation technique without considering the optimality principle.

In this paper, we investigate the ARL-based motion/force tracking control problem for a multimanipulator system in the presence of disturbance. The control goal is to obtain the unification of the optimality principle and tracking problem. The main contributions of this paper can be summarized as follows:

- (1) First, it is obviously different from [26, 27, 34, 41, 42] studying optimal control for time-invariant systems; we propose ARL-based optimal control design in the situation of the time-varying nonlinear dynamical system under the influence of the time-varying desired trajectory by utilizing the online actor/critic technique for the motion dynamic model of multimanipulator systems. Furthermore, this work is still proposed for the motion/force control problem by extending more the force control term in the control scheme.
- (2) Second, different from numerous existing nonlinear control approaches [10, 12, 13, 17, 20, 23–25, 43, 44], the proposed optimal control algorithm is investigated with the unification of the optimization and tracking problem to be discussed through theoretical analysis and simulation studies.

The rest of this article is organized as follows. Section 2 provides the problem statement of this paper. Section 3 describes the ARL-based motion/force control design for multimanipulator systems with tracking analysis. Simulation studies on the multirobots with 3 manipulators are presented in Section 4, and the conclusion remarks are pointed out in Section 5.

2. Preliminaries and Problem Statement

We investigate n two-link manipulators, as shown in Figure 1, which includes n robots interacted by each pair of manipulators, the unknown environment with the original Cartesian coordinate system $O(X, Y, Z)$, and the corresponding coordinate E_i of the end effector of each robot. The interaction between each pair of manipulators leads to a constraint condition as described in the following equation:

$$\Xi_{ij} = |\mathbf{OE}_i - \mathbf{OE}_j| - d_{ij} = 0 (\forall 1 \leq i < j \leq n), \quad (1)$$

where $\mathbf{OE}_i$ is a vector with the first point being the original point and the end effector E_i under the Cartesian coordinate system and d_{ij} is the distance between end effectors of the i th robot manipulator and the j th robot manipulator. According to (1), combining all the constraints in this multimanipulator, we can obtain the constraint vector as follows:

$$\Xi(\eta) = [\dots, \Xi_{ij}(\eta_i, \eta_j), \dots]^T = 0, \quad (2)$$

where $\Xi(\eta) \in \mathbb{R}^{m \times 1}$, η is the joint vector of n two-link manipulator systems. It can be seen that there are two main

FIGURE 1: The cooperating manipulator system.

cases in general robots. First, there exist no constraint forces due to no interaction between the robot and the rigid object. Second, the interaction leads to the existence of constraint force, and we only consider the constraint motion under the following assumption.

Assumption 1. The number of joints is more than the number of constraint conditions, i.e., $2n > m$.

In [12], the dynamic equation of n two-link robot manipulators with constraint force by the interaction between each couple of manipulators can be described as

$$M(\eta)\ddot{\eta} + C(\eta, \dot{\eta})\dot{\eta} + F(\eta, \dot{\eta}, t) = \tau - J^T(\eta)\lambda, \quad (3)$$

where $J(\eta) = (\partial \Xi / \partial \eta) \in \mathbb{R}^{m \times 2n}$ and $\lambda \in \mathbb{R}^{m \times 1}$ is the vector of Lagrange multiplier.

The control objective is to obtain not only the tracking of joints to desired trajectories but also guaranteeing the constant distance between end effectors of each pair of manipulators. Moreover, the motion/force control objective also mentions the tracking of Lagrangian coefficient λ in (3). On the contrary, due to the proposed optimality-based motion/force control design, it requires the unification of the optimal control and tracking problem.

Remark 1. It is worth emphasizing that different from classical nonlinear control approaches [7, 17, 19, 20, 23, 24, 43], the approach in this paper is developed by the optimal control technique for uncertain systems based on adaptive/approximate reinforcement learning (ARL). Moreover, regarding the mathematical model of multimanipulators, this work is the extension of Chen et al. [11] by considering the holonomic constraint force with separation of the joint vector.

3. Motion/Force Control Design with ARL

In this paper, a completed control structure in Figure 2 is investigated with 2 independent parts, including ARL-based optimal control for the motion dynamic model and the force control scheme for tracking the constraint coefficient. Moreover, the convergence and stability of the whole system

FIGURE 2: Control structure of the surface vessel.

(Figure 2) are discussed for determining the performance of the proposed motion/force controller.

3.1. ARL-Based Motion Controller. For the purpose of designing the motion controller in multimanipulator systems, the motion dynamic model is discussed after eliminating the constraint forces. Moreover, it is due to the description of time-varying systems in trajectory tracking control; the modification of the tracking error model needs to be implemented for developing the ARL-based optimal control design.

Because of $\Xi(\eta) \in \mathbb{R}^{m \times 1}$ with $m(m < 2n)$ independent constraint conditions, there exists an algebraic map $\varphi: \mathbb{R}^{2n-m} \mapsto \mathbb{R}^{2n}$ satisfying

$$\eta = \varphi(\xi), \quad (4)$$

where $\xi \in \mathbb{R}^{(2n-m) \times 1}$ is the vector of independent joint variables. Taking the derivative of (4), we imply that

$$\begin{aligned} \dot{\eta} &= \Lambda \dot{\xi}, \\ \ddot{\eta} &= \Lambda \ddot{\xi} + \dot{\Lambda} \dot{\xi}, \end{aligned} \quad (5)$$

where $\Lambda = (\partial\varphi/\partial\xi) \in \mathbb{R}^{2n \times (2n-m)}$. With the purpose of eliminating the constraint force in the motion dynamic model, we use the Moore–Penrose pseudo-inverse matrix of $J(\eta)$ being $J^+ = J^T(JJ^T)^{-1}$ to achieve the matrix $Q = (I - J^+J)$ satisfying $QJ^T = 0$. Multiplying both sides of (3) with matrix Q , it leads to the following equation after utilizing $QJ^T = 0$:

$$QM\ddot{\eta} + QC\dot{\eta} + QF = Q(\tau - J^T\lambda) = Q\tau - QJ^T\lambda = Q\tau = \tau^1. \quad (6)$$

According to (5) and (6), the motion dynamic can be expressed as

$$QM\Lambda\ddot{\xi} + (QM\dot{\Lambda} + QC\Lambda)\dot{\xi} + QF = \tau^1. \quad (7)$$

Regarding the independent joint coordinates, the dynamic model can be represented after multiplying with matrix $[I_{2n-m}, 0] \in \mathbb{R}^{(2n-m) \times 2n}$ on both sides of (6) as follows:

$$\begin{cases} \dot{\xi} = v, \\ \overline{M}\dot{v} + \overline{C}v + \overline{F} = \tau_{\text{ind}}^1, \end{cases} \quad (8)$$

where

$$\begin{aligned} \overline{M} &= [I_{2n-m}, 0]QM\Lambda, \\ \overline{C} &= [I_{2n-m}, 0](QM\dot{\Lambda} + QC\Lambda), \\ \overline{F} &= [I_{2n-m}, 0]QF, \\ \tau_{\text{ind}}^1 &= [I_{2n-m}, 0]\tau^1. \end{aligned} \quad (9)$$

Remark 2. This paper only discusses the situation that the number of constraint conditions is smaller than the number of joints ($2n > m$) as mentioned in Assumption 1. For the case of ($2n = m$), it leads to no existence of independent joints. As a result, the above modifications are not implemented, and the control method is developed with different approaches [45]. Additionally, no constraint forces are considered in the special case ($m = 0$).

The motion control design is proposed as

$$\tau_{\text{ind}}^1 = \tau_d + u, \quad (10)$$

where

$$\begin{aligned} \tau_d &= \overline{M}\dot{v}_d + \overline{C}v_d + \overline{F}, \\ v_d &= (\dot{\xi}_d - \beta z_\xi), \\ z_\xi &= (\xi - \xi_d), \\ z_v &= (v - v_d). \end{aligned} \quad (11)$$

Therefore, with the assumption that $(d/dt)\xi_d$ is the function of ξ_d as $(d/dt)\xi_d = h_1(\xi_d)$, we can obtain that

$$\left\{ \frac{d}{dt}z_v = \frac{d}{dt}(v - v_d) = \overline{M}^{-1}u - \overline{M}^{-1}\overline{C}(v - v_d), \frac{d}{dt}z_\xi = v - \frac{d}{dt}\xi_d, \frac{d}{dt}\xi_d = h_1(\xi_d). \right. \quad (12)$$

Furthermore, v_d is the function of (z_ξ, ξ_d) as $v_d = (\dot{\xi}_d - \beta z_\xi) = h_1(\xi_d) - \beta z_\xi \Rightarrow v = (z_v + v_d)$ is the function of

(z_v, z_ξ, ξ_d) . Combining with (12), a continuous-time affine system has been obtained with the typical form

$$\dot{Z} = F(Z) + G(Z)u, \quad (13)$$

where the vector of state variables is $Z = [z_v, z_\xi, \xi_d]^T \in \mathbb{R}^{12}$.

Remark 3. Due to the time-varying reference in trajectory tracking, the closed system needs to be known as the time-varying property. However, the framework of using term (11) and choosing the state variable vector Z is able to obtain time-invariant systems (12) and (13). Therefore, in contrast to the existing methods in [36], the proposed control easily handles the situation of time-varying systems by the elimination of the term $-(\partial V^*/\partial t)$.

3.1.1. The Optimal Control Problem. In this section, we consider the general class of time-varying systems:

$$\frac{d}{dt}Z = F(Z, u, t), \quad (14)$$

with the associated cost function to be defined as

$$J(Z, u) = \int_0^\infty r(Z(\tau), u(\tau))d\tau, \quad (15)$$

where $r(Z(\tau), u(\tau)) = (Z^T Q Z + u^T R u)$ under the weighting matrices $Q = Q^T \geq 0$ and $R = R^T > 0$ to be defined with appropriate dimensions. It can be seen that the optimal problem for the motion model of multimanipulators (12) and (13) is to realize an admissible control policy [31, 32] for the purpose of obtaining the minimal value. This is because that the control policy $\hat{u}(Z)$ needs to guarantee the existence of the estimated Bellman function $\hat{V}(Z)$ obtained from the HJB equation.

Definition 1. A control law $u(Z)$ is considered being “admissible” with respect to the performance index (15) on a compact set Ω , which is known as a continuous signal $u(Z) \in Y(\Omega)$, if the properties of the stabilization of (14) and the limitation of $J(Z, u)$ for every $Z \in \Omega$ are satisfied.

It is worth emphasizing that, in the general case of time-varying system (14), the optimal control strategy needs to be considered as a time-varying function $u^*(Z, t)$. Hence, the time-varying Bellman function with respect to arbitrary time $V^*(Z(t), t)$ can be given as

$$V^*(Z(t)) = \min_{u(Z) \in Y(\Omega)} J(Z(t), u(t)). \quad (16)$$

The HJB equation can be achieved by taking the time derivative of $V^*(X(t))$ under two different methods. Firstly, it is directly computed along system (14) as

$$\frac{d}{dt}V^*(Z(t), t) = \frac{\partial V^*}{\partial Z} \frac{dZ}{dt} + \frac{\partial V^*}{\partial t} = \frac{\partial V^*}{\partial Z} F(Z, u^*, t) + \frac{\partial V^*}{\partial t}. \quad (17)$$

Secondly, thanks to the dynamic programming principle, we also compute the derivative of $V^*(Z(t))$ with respect to time as

$$\frac{d}{dt}V^*(Z(t), t) = -r(Z(\tau), u^*(\tau)). \quad (18)$$

According to (17) and (18), the first part of the HJB problem can be given:

$$r(Z(\tau), u^*(\tau)) + \frac{\partial V^*}{\partial Z} F(Z, u^*, t) = -\frac{\partial V^*}{\partial t}. \quad (19)$$

Moreover, the second part of the HJB problem can be developed with the performance index to be formulated as

$$V^*(Z(t), t) = \min_{u(Z) \in Y(\Omega)} \int_t^\infty r(Z(\tau), u^*(\tau))d\tau, \quad (20)$$

$$V^*(Z(t), t) = \min_{u^*(Z) \in Y(\Omega)} \int_t^{t+\Delta t} r(Z, u^*)d\tau + \min_{u^*(Z) \in Y(\Omega)} \int_{t+\Delta t}^\infty r(Z, u^*)d\tau. \quad (21)$$

Because of utilizing the dynamic programming principle, the performance index is represented as

$$V^*(Z(t), t) = \min_{u(Z) \in Y(\Omega)} \int_t^{t+\Delta t} r(Z(\tau), u(\tau))d\tau + V^*(Z(t + \Delta t), t + \Delta t). \quad (22)$$

Therefore, we imply that the Bellman function can be rewritten as

$$0 = \min_{u(Z) \in Y(\Omega)} \left\{ \frac{1}{\Delta t} \int_t^{t+\Delta t} r(Z, u^*)d\tau + \frac{V^*(Z(t + \Delta t), t + \Delta t) - V^*(Z(t), t)}{\Delta t} \right\}. \quad (23)$$

As the convergence of $\Delta t \rightarrow 0^+$, one can derive that

$$\min_{u(Z) \in Y(\Omega)} \left\{ r(Z(t), u^*(t)) + \frac{\partial V^*}{\partial Z} F(Z, u^*, t) \right\} = -\frac{\partial V^*}{\partial t}. \quad (24)$$

Remark 4. It is worth noting here that the term $(\partial/\partial t)V^*(Z(t), t)$ in (19) and (24) is the disadvantage in developing the ARL algorithm, and there has been very little research considering the direct method for this challenge. Authors in [36] developed the solutions of the ARL-based time-varying optimal controller using the framework of the data-driven and Newton–Leibniz formula-based function approximation method. In this paper, we indirectly handle the time-varying tracking error model through the consideration of systems (12) and (13). Therefore, the two parts of the HJB problem can be modified as

$$r(Z, u^*) + \frac{\partial V^*}{\partial Z} (F(Z) + G(Z)u^*) = 0, \quad (25)$$

$$\min_{u(X) \in Y(\Omega)} \left[r(X, u) + \frac{\partial V^*}{\partial X} (F(X) + G(X)u) \right] = 0. \quad (26)$$

3.1.2. ARL-Based Control Design for Independent Joints. It is because of difficulties in directly solving HJB equation (25) to obtain corresponding optimal control algorithm. Hence, the approximate solution with a neural network (NN) is mentioned to develop the ARL design in the motion dynamic model of multimanipulators. Thanks to the smooth property of functions $V^*(Z)$ and $u^*(Z)$, they can be described over any compact domain $C \subseteq \mathbb{R}^{12}$:

$$V^*(Z) = W^T \psi(Z) + \epsilon(Z), \quad (27)$$

$$u^*(Z) = -\frac{1}{2} R^{-1} G^T(Z) \left(\left(\frac{\partial \psi}{\partial Z} \right)^T W + \left(\frac{\partial \epsilon(Z)}{\partial Z} \right)^T \right), \quad (28)$$

where $W \in \mathbb{R}^N$ is an uncertain ideal NN weight vector, N is chosen as the number of neurons, $\psi(Z) \in \mathbb{R}^N$ denotes an activation function vector satisfying $\psi_j(0) = 0$ and $(\partial \psi_j / \partial Z)|_{Z=0} = 0, \forall j = 1, \dots, N$, and $\epsilon(Z) \in \mathbb{R}$ is known as the estimation error of the Bellman function $V^*(Z)$. Due to the unknown ideal NN weight vector $W \in \mathbb{R}^N$, it is necessary to obtain the corresponding adjusting mechanism $\hat{W}_a, \hat{W}_c$ for estimating the actor/critic NNs to achieve the optimal control design in the absence of finding the HJB equation analytically (for more details, see [41]). Moreover, it is able to choose the smooth NN activation function vector $\psi(Z) \in \mathbb{R}^N$ based on the description of multimanipulators (see Section 4). In [41], thanks to the Weierstrass approximation theorem, we can uniformly approximate not only $V^*(Z)$ but also $(\partial V^*(Z)/\partial Z)$ with $\epsilon(Z), (\partial \epsilon(Z)/\partial Z) \rightarrow 0$ as $N \rightarrow \infty$. For a fixed number N , the approximated Bellman function of critic NN $\hat{V}(Z)$ and the approximated optimal control policy of actor NN $\hat{u}(Z)$ are represented as

$$\hat{V}(Z) = \hat{W}_c^T \psi(Z), \quad (29)$$

$$\hat{u}(Z) = -\frac{1}{2} R^{-1} G^T(Z) \left(\frac{\partial \psi}{\partial Z} \right)^T \hat{W}_a. \quad (30)$$

It should be noted that properties (19) and (24) of Hamiltonian

$H(Z, u, (\partial V/\partial Z)) = r(Z, u) + (\partial V/\partial Z)(F(Z) + G(Z)u)$ with the optimal controller $u^*(Z)$ and corresponding value function $V^*(Z)$ guarantee that the adjusting mechanisms of actor NN $\hat{W}_a$ and critic $\hat{W}_c$ weights are simultaneously tuned to minimize the squared Bellman error δ_{hjb} and the corresponding integral, respectively. According to the property of HJB problem (25), it leads to $H^*(Z, u^*, (\partial V^*/\partial Z)) = 0$. After obtaining the error between approximated functions $\hat{V}(Z), \hat{u}(Z)$ and optimal control input $V^*(Z), u^*(Z)$, the Bellman error δ_{hjb} is represented as

$$\delta_{hjb} = \hat{H}\left(Z, \hat{u}, \frac{\partial \hat{V}}{\partial Z}\right) - H^*\left(Z, u^*, \frac{\partial V^*}{\partial Z}\right) = \hat{H}\left(Z, \hat{u}, \frac{\partial \hat{V}}{\partial Z}\right) = \hat{W}_c^T \sigma(Z, \hat{u}) + Z^T Q_T Z + \hat{u}^T R \hat{u}, \quad (31)$$

where $\sigma(Z, \hat{u}) = (\partial \psi / \partial Z)(F(Z) + G(Z)\hat{u})$ is the regression vector of the critic part.

Based on the results in [31], the adjusting mechanism of the critic NN is computed:

$$\frac{d}{dt} \hat{W}_c = -k_c \lambda \frac{\sigma}{1 + \gamma \sigma^T \lambda \sigma} \delta_{hjb}, \quad (32)$$

where $\gamma, k_c \in \mathbb{R}$ are constant positive coefficients and $\lambda(t) \in \mathbb{R}^{N \times N}$ is an estimated symmetric matrix being the solution of the differential equation as

$$\frac{d}{dt} \lambda = -k_c \lambda \frac{\sigma \sigma^T}{1 + \gamma \sigma^T \lambda \sigma} \lambda, \quad (33)$$

$$\lambda(t_s^+) = \lambda(0) = \varphi_0 I,$$

where t_s^+ denotes the resetting time satisfying that $\lambda_{\min}\{\lambda(t)\} \leq \varphi_1, \varphi_0 > \varphi_1 > 0$. In [31], because of the condition that $\lambda(t)$ is positive definite and preventing the covariance wind-up problem, the covariance matrix $\lambda(t)$ is chosen as

$$\varphi_1 I \leq \lambda(t) \leq \varphi_0 I. \quad (34)$$

Additionally, the adjusting mechanism of the actor NN part is represented by minimizing the squared Bellman error:

$$\begin{aligned} \frac{d}{dt}\hat{W}_a &= -\frac{k_{a1}}{\sqrt{1+\sigma^T\sigma}} \frac{\partial\psi}{\partial Z} \text{GR}^{-1} G^T \frac{\partial\psi^T}{\partial Z} (\hat{W}_a - \hat{W}_c) \delta_{hjb} \\ &\quad -k_{a2}(\hat{W}_a - \hat{W}_c) \end{aligned} \quad (35)$$

Remark 5. ARL technique has been investigated with many different approaches, such as off-policy integral reinforcement learning (IRL) [46, 47] and Q-learning [42]. However, there is a wide difference between the online actor/critic technique in this paper and IRL and Q-learning. The off-policy IRL requires the complete data collection between the two sequential sampling times in computing the optimal law [46, 47]. Meanwhile, it is obviously different from the off-policy IRL technique; the online actor/critic method in (29), (30), (35), (33) requires the initial values to implement the computation. This could be exemplified by this article design control law under HJB problems (25) and (26), and the off-policy IRL method considers the relation between the deviation of integrals at two sampling times and the collected data under the dynamic programming law. In addition, the Q-learning method mentioned the Q-function in terms of not only state variables but also control inputs, being different from the Bellman function [42]. This property can result in the difficulties for implementing the Q-learning technique of the continuous-time system. Additionally, compared with the data-driven method in [48], this work provides a neural network-based technique to avoid the Kronecker product in estimating the actor/critic term. The actor/critic-based approaches have been discussed in [49, 50] for nonlinear affine systems using residual error δ_{hjb} . However, in view of the consideration of the identifier in [49, 50], it implies the difference in the computation of residual error δ_{hjb} and training laws in actor/critic weights between the proposed method and the work in [49, 50].

3.1.3. The Control Design of Dependent Joints. After completing the tracking control design of independent joints, we can achieve the remaining dependent joints' controller based on the relation between two groups of joint variables, including independent joints and dependent joints. According to (8), it leads to the following representation:

$$\ddot{\xi} = \overline{M}^{-1} \left(\tau_{\text{ind}}^1 - \overline{C}\dot{\xi} - \overline{F} \right). \quad (36)$$

On the contrary, thanks to the relation $\tau^1 = [\tau_{\text{ind}}^{1T}, \tau_{\text{dep}}^{1T}]^T$, multiplying both sides of (6) with $[0, I_m] \in \mathbb{R}^{(m) \times 2n}$, we achieve the dependent joints to be computed as

$$M'' \ddot{\xi} + C'' \dot{\xi} + F'' = \tau_{\text{dep}}^1, \quad (37)$$

where

$$\begin{aligned} M'' &= [0, I_m] Q M \Lambda, \\ C'' &= [0, I_m] (Q M \dot{\Lambda} + Q C \Lambda), \\ F'' &= [0, I_m] Q F, \\ \tau_{\text{dep}}^1 &= [0, I_m] \tau^1. \end{aligned} \quad (38)$$

Substituting (36) into (37), we obtain the control law of dependent joint variables to be implemented as follows:

$$\tau_{\text{dep}}^1 = M'' \left\{ \overline{M} \left(\tau_{\text{ind}}^1 - \overline{C}\dot{\xi} - \overline{F} \right) \right\} + C'' \dot{\xi} + F''. \quad (39)$$

It is worth noting that the problem of synchronization between multiple joints can be considered by the relation between dependent and independent joints. Therefore, this problem is able to obtain the desired trajectories of independent joints and the trajectories of remaining joints depending on these independent joints. Furthermore, it leads to the relationship between the control input of independent joints and dependent joints (39).

3.2. Constraint Force Control Design. For the purpose of designing the motion/force control scheme for n manipulators, the constraint force controller and motion control law are necessary to be established based on the computation of lambda coefficient in (3). Before demonstrating the motion/force control design, we need to compute the coefficient λ to be dependent on control input τ as follows. According to (3)–(5), we achieve the following equation:

$$J^{*T}(\xi) \lambda = \tau - \{M^*(\eta) \ddot{\xi} + C^*(\eta, \dot{\eta}) \dot{\xi} + F^*(\eta, \dot{\eta}, t)\}, \quad (40)$$

where

$$\begin{aligned} J^{*T}(\xi) &= J^T(\varphi(\xi)), \\ M^*(\eta) &= M(\eta) \Lambda, \\ C^*(\eta, \dot{\xi}) &= C(\eta, \dot{\eta}) \Lambda + M(\eta) \dot{\Lambda}, \\ F^*(\eta, \dot{\eta}, t) &= F(\eta, \dot{\eta}, t). \end{aligned} \quad (41)$$

Moreover, we find the method to compute coefficient λ after eliminating the term ξ by multiplying both sides with matrix N satisfying that

$$\begin{cases} N M(\eta) \Lambda = 0, \\ N J^{*T}(\xi) = I. \end{cases} \quad (42)$$

Therefore, we can obtain the following coefficient λ :

$$\lambda = N \tau - \{N C^*(\eta, \dot{\eta}) \dot{\xi} + N F^*(\eta, \dot{\eta}, t)\}, \quad (43)$$

with N to be chosen as

$$N = (A J^{*T}(\xi))^{-1} A, \quad (44)$$

where $A = J^*(\xi) M^{-1}(\eta)$.

Based on the proposed motion/force, control design is described in the following equation:

$$\tau = \begin{bmatrix} \tau_{\text{ind}}^{(1)} \\ \tau_{\text{dep}}^{(1)} \end{bmatrix} - J^T(\eta) \tau^{(2)}. \quad (45)$$

This leads to the following relations that

$$Q\tau^{(1)} = Q \begin{bmatrix} \tau_{\text{ind}}^{(1)} \\ \tau_{\text{dep}}^{(1)} \end{bmatrix} = \tau^1, \quad (46)$$

$$\lambda = N\tau^{(1)} - \tau^{(2)} - \{NC^*(\eta, \dot{\eta})\dot{\xi} + NF^*(\eta, \dot{\eta}, t)\}. \quad (47)$$

It should be noted that the constraint force coefficient λ depends on the ARL-based motion dynamic control $\tau^{(1)} = \begin{bmatrix} \tau_{\text{ind}}^{(1)} \\ \tau_{\text{dep}}^{(1)} \end{bmatrix}$ to be described in (47). Therefore, with the purpose of tracking the coefficient λ , we can propose the remaining term $\tau^{(2)}$ in (47) as given in the following:

$$\tau^{(2)} = K_f e_\lambda - \lambda_d - NC^*(\eta, \dot{\eta})\dot{\xi}_d. \quad (48)$$

The tracking problem will be proved in the next section.

3.3. Convergence and Stability Analysis. With the purpose of considering the stability and convergence of the closed system under the proposed motion-force control design, we need to implement several following steps. First, the Bellman error δ_{hjb} , critic estimated weight error $\tilde{W}_c = W - \hat{W}_c$, and the persistence of excitation (PE) condition need to be given. Second, thanks to the optimal Bellman function and PE condition, the corresponding Lyapunov function candidate is chosen. Third, the stability of the whole system (Figure 1) and the role of constraint force control and dependent joints' motion control are discussed.

According to (27)–(31), the Bellman error function δ_{hjb} is dependent on state variable vector $Z(t)$ as

$$\begin{aligned} \delta_{\text{hjb}} &= \tilde{W}_c^T \sigma(Z, \hat{u}) + Z^T Q_T Z + \hat{u}^T R \hat{u} - W^T \sigma(Z, u^*) - u^{*T} R u^* \\ &\quad - \frac{\partial \epsilon}{\partial Z} (F(Z) + G(Z)u^*) - Z^T Q_T Z \\ &= -\tilde{W}_c^T \sigma(Z, \hat{u}) + \frac{1}{4} \tilde{W}_a^T \left(\frac{\partial \psi}{\partial Z} \right) G R^{-1} G^T \left(\frac{\partial \psi}{\partial Z} \right)^T \tilde{W}_a \\ &\quad - \frac{1}{4} \left(\frac{\partial \epsilon}{\partial Z} \right) G R^{-1} G^T \left(\frac{\partial \epsilon}{\partial Z} \right)^T - \frac{\partial \epsilon}{\partial Z} (F(Z) + G(Z)u^*). \end{aligned} \quad (49)$$

Combining (49) with (32), we imply the dynamic equation of critic error $\tilde{W}_c = W - \hat{W}_c$ to be given as

$$\begin{aligned} \frac{d}{dt} \tilde{W}_c &= -k_c \lambda \Pi \Pi^T \tilde{W}_c + k_c \lambda \frac{\sigma(Z, \hat{u})}{1 + \nu \sigma^T(Z, \hat{u}) \lambda \sigma(Z, \hat{u})} \left(\frac{1}{4} \tilde{W}_a^T \left(\frac{\partial \psi}{\partial Z} \right) G R^{-1} G^T \left(\frac{\partial \psi}{\partial Z} \right)^T \tilde{W}_a \right. \\ &\quad \left. - \frac{1}{4} \left(\frac{\partial \epsilon}{\partial Z} \right) G R^{-1} G^T \left(\frac{\partial \epsilon}{\partial Z} \right)^T - \frac{\partial \epsilon}{\partial Z} (F(Z) + G(Z)u^*) \right), \end{aligned} \quad (50)$$

where $\tilde{W}_c = W - \hat{W}_c$ and $\Pi(t) = (\sigma(Z, \hat{u}) / \sqrt{1 + \nu \sigma^T(Z, \hat{u}) \lambda \sigma(Z, \hat{u})})$ is satisfied by

$$\|\Pi\| \leq \frac{1}{\sqrt{\nu} \varphi_1}. \quad (51)$$

The nominal system is given after eliminating the term of NN errors:

$$\frac{d}{dt} \tilde{W}_c = -k_c \lambda \Pi \Pi^T \tilde{W}_c. \quad (52)$$

Due to the purpose of exponential convergence of $\tilde{W}_c$, it is necessary to guarantee the PE condition as follows:

$$\mu_2 I \geq \int_{t_0}^{t_0 + \delta} \Pi(s) \Pi(s)^T ds \geq \mu_1 I, \quad \forall t_0 \geq 0. \quad (53)$$

The following main theorem is proposed with the assumptions being discussed in [41] to estimate the attraction region of tracking errors.

Theorem 1. Consider the cooperating manipulators in Figure 1 with the assumption of the bound conditions of the NN including ideal weights, activation function $\psi(\cdot)$, its derivative to be described in [41], and PE condition of the signal vector $\Pi(t)$, and the selection of parameters is implemented with the following condition:

$$\frac{c_3}{k_{a1}} > k_1 k_2, \quad (54)$$

where the coefficients in (54) are mentioned in (55), (56), and (35). Let us consider the control structure (Figure 2), (10) with ARL-based control scheme (46), (10), (39), and (30), the associated adjusting mechanisms (35) and (33) for the actual controller, and constraint force control vector (48); then, (1) the actor-critic weight errors $\tilde{W}_a$ and $\tilde{W}_c$ are UUB; (2) the tracking effectiveness of not only $\tilde{z}_v = (v - v_d)$ but also $\tilde{z}_\eta = \xi - \xi_d$ in the closed control system (Figure 1) is also UUB; (3) the tracking of constraint force coefficient vector λ and the remaining terms of joint variables' vector $\eta_{2n} = [\xi_{(2n-m)}, p_m]$ is also UUB.

Proof. First, the tracking effectiveness of motion tracking control is considered under the ARL-based optimal control scheme for corresponding tracking error motion models (12) and (13). Due to the satisfaction of PE condition (53), there exists a time-varying function $V_c: \mathbb{R}^N \times [0, \infty) \rightarrow \mathbb{R}$ satisfying several following inequalities:

$$c_1 \|\tilde{W}_c\|^2 \leq V_c(\tilde{W}_c, t) \leq c_2 \|\tilde{W}_c\|^2,$$

$$\frac{\partial V_c}{\partial t} + \frac{\partial V_c}{\partial \tilde{W}_c} (-k_c \lambda \Pi \Pi^T \tilde{W}_c) \leq -c_3 \|\tilde{W}_c\|^2, \quad (55)$$

$$\left\| \frac{\partial V_c}{\partial \tilde{W}_c} \right\| \leq c_4 \|\tilde{W}_c\|,$$

where $c_1, c_2, c_3, c_4 \in \mathbb{R}$ are positive constant coefficients. The function $V_c(\tilde{W}_c, t)$ is one part of the Lyapunov function candidate. Moreover, according to the bound conditions of the proposed NN [41], we can achieve the following results

with the purpose of estimating the time derivative of the Lyapunov function in the next steps:

$$\begin{aligned}
 & i. \|\tilde{W}_a\| \leq k_1, \\
 & i^2. \left\| \frac{\partial \psi}{\partial Z} GR^{-1} G^T \left(\frac{\partial \psi}{\partial Z} \right)^T \right\| \leq k_2, \\
 & i^3. \left\| \frac{1}{4} \tilde{W}_a^T \left(\frac{\partial \psi}{\partial Z} \right) GR^{-1} G^T \left(\frac{\partial \psi}{\partial Z} \right)^T \tilde{W}_a - \frac{1}{4} \left(\frac{\partial \epsilon}{\partial Z} \right) GR^{-1} G^T \left(\frac{\partial \epsilon}{\partial Z} \right)^T - \frac{\partial \epsilon}{\partial Z} (F(Z) + G(Z)u^*) \right\| \leq k_3, \\
 & i^4. \left\| \frac{1}{2} W^T \left(\frac{\partial \psi}{\partial Z} \right) GR^{-1} G^T \left(\frac{\partial \epsilon}{\partial Z} \right)^T + \frac{1}{2} \left(\frac{\partial \epsilon}{\partial Z} \right) GR^{-1} G^T \left(\frac{\partial \epsilon}{\partial Z} \right)^T \right. \\
 & \quad \left. + \frac{1}{2} W^T \left(\frac{\partial \psi}{\partial Z} \right) GR^{-1} G^T \left(\frac{\partial \psi}{\partial Z} \right)^T \tilde{W}_a + \frac{1}{2} \left(\frac{\partial \epsilon}{\partial Z} \right) GR^{-1} G^T \left(\frac{\partial \psi}{\partial Z} \right)^T \tilde{W}_a \right\| \leq k_4.
 \end{aligned} \tag{56}$$

Thanks to Bellman function $V^*(Z)$ (20) and time-varying $V_c(\tilde{W}_c, t)$ (55), a Lyapunov function candidate can be chosen to determine the stability of the whole of the cascade control system (Figure 2) and the tracking of the weights of the actor NN and critic NN:

$$V_L \triangleq \frac{1}{2} \rho z_\xi^T z_\xi + V^*(Z) + V_c(\tilde{W}_c, t) + \frac{1}{2} \tilde{W}_a^T \tilde{W}_a. \tag{57}$$

It is worth noting that the term $(1/2)\rho z_\xi^T z_\xi$ (ρ is a positive constant coefficient) is the additional term for the purpose of tracking of the whole motion control system. Because $V^*(Z)$ is a smooth function and positive definite, there exist two K -class functions γ_1 and γ_2 such that

$$\gamma_1(\|Z\|) \leq V^*(Z) \leq \gamma_2(\|Z\|). \tag{58}$$

Because of (55) and (8), we can determine the strict proper representation of Lyapunov function $V_L(Z)$:

$$\begin{aligned}
 & \frac{1}{2} \rho \|z_\xi\|^2 + \alpha_1(\|Z\|) + c_1 \|\tilde{W}_c\|^2 + \frac{1}{2} \|\tilde{W}_a\|^2 \leq V_L \\
 & \leq \frac{1}{2} \rho \|z_\xi\|^2 + \alpha_2(\|Z\|) \\
 & \quad + c_2 \|\tilde{W}_c\|^2 + \frac{1}{2} \|\tilde{W}_a\|^2.
 \end{aligned} \tag{59}$$

Implementing the derivative of V_L along motion system (13) under the control input $\hat{u}(Z)$, we imply that

$$\begin{aligned}
 & \frac{d}{dt} V_L = \rho z_\xi^T (z_v - \beta z_\xi) + \frac{\partial V^*}{\partial Z} (F(Z) + G(Z)\hat{u}) \\
 & \quad + \frac{\partial V_c}{\partial t} + \frac{\partial V_c}{\partial \tilde{W}_c} (\Theta_1 + \Theta_2) \\
 & \quad - \tilde{W}_a^T \dot{\tilde{W}}_a + \frac{\partial V^*}{\partial Z} G(Z)\Delta,
 \end{aligned} \tag{60}$$

where

$$\Theta_1 = -k_c \lambda \Pi \Pi^T \tilde{W}_c$$

$$\begin{aligned}
 \Theta_2 = k_c \lambda \frac{\sigma}{1 + \nu \sigma^T \lambda \sigma} & \left(\frac{1}{4} \tilde{W}_a^T \left(\frac{\partial \psi}{\partial Z} \right) GR^{-1} G^T \left(\frac{\partial \psi}{\partial Z} \right)^T \right. \\
 & \left. \tilde{W}_a - \frac{1}{4} \left(\frac{\partial \epsilon}{\partial Z} \right) GR^{-1} G^T \left(\frac{\partial \epsilon}{\partial Z} \right)^T - \frac{\partial \epsilon}{\partial Z} (F(Z) + G(Z)u^*) \right).
 \end{aligned} \tag{61}$$

According to (19), the relation between the Bellman function and the optimal control law can be represented as

$$\frac{\partial V^*}{\partial Z} F(Z) = -\frac{\partial V^*}{\partial Z} G(Z)u^*(Z) - Z^T Q_T Z - u^{*T}(Z) R u^*(Z). \tag{62}$$

According to (62) and optimal law (28), (35), and (55) in (60), we can conclude that

$$\begin{aligned}
 & \frac{d}{dt} V_L \leq -\rho \beta \|z_\xi\|^2 + \rho z_\xi^T z_v - z_v^T Q z_v - u^{*T} R u^* \\
 & \quad + 2u^{*T} R (u^* - \hat{u}) - c_3 \|\tilde{W}_c\|^2 + c_4 \|\tilde{W}_c\| \|\Theta_2\| + \\
 & \quad k_{a2} \tilde{W}_a^T (\hat{W}_a - \tilde{W}_c) - 2u^{*T} R \Delta + \frac{k_{a1}}{\sqrt{1 + \sigma^T \sigma}} \tilde{W}_a^T \frac{\partial \psi}{\partial Z} \\
 & \quad G(Z) R^{-1} G^T(Z) \left(\frac{\partial \psi}{\partial Z} \right)^T (\hat{W}_a - \tilde{W}_c) \delta_{\text{hjb}}.
 \end{aligned} \tag{63}$$

Using Young's inequality, one can obtain that

$$\rho z_\xi^T z_v \leq \frac{\rho}{2} \|z_\xi\|^2 + \frac{\rho}{2} z_v^T z_v = \frac{\rho}{2} \|z_\xi\|^2 + \frac{\rho}{2} \|z_v\|^2. \tag{64}$$

From (30), (19), and (56), the following inequality is given:

$$\begin{aligned} 2u^{*T}R(u^* - \hat{u}) &= \frac{1}{2}W^T \frac{\partial \phi}{\partial Z} GR^{-1}G^T \left(\frac{\partial \epsilon}{\partial Z} \right)^T + \frac{1}{2}W^T \frac{\partial \phi}{\partial Z} GR^{-1}G^T \left(\frac{\partial \phi}{\partial Z} \right)^T \tilde{W}_a \\ &+ \frac{1}{2} \frac{\partial \epsilon}{\partial Z} GR^{-1}G^T \left(\frac{\partial \phi}{\partial Z} \right)^T \tilde{W}_a + \frac{1}{2} \frac{\partial \epsilon}{\partial Z} GR^{-1}G^T \left(\frac{\partial \epsilon}{\partial Z} \right)^T \leq k_4, \end{aligned} \quad (65)$$

and the term Θ_2 is bounded by

$$\|\Theta_2\| \leq \frac{k_c \varphi_0}{2\sqrt{\gamma \varphi_1}} k_3. \quad (66)$$

Following (56) and (49), we have

$$\begin{aligned} &\frac{k_{a1}}{\sqrt{1+\sigma^T \sigma}} \tilde{W}_a^T \frac{\partial \psi}{\partial Z} G(Z) R^{-1} G^T(Z) \left(\frac{\partial \psi}{\partial Z} \right)^T (\hat{W}_a - \hat{W}_c) \delta_{\text{hjb}} = \\ &\frac{k_{a1}}{\sqrt{1+\sigma^T \sigma}} \tilde{W}_a^T \frac{\partial \psi}{\partial Z} G(Z) R^{-1} G^T(Z) \left(\frac{\partial \psi}{\partial Z} \right)^T (\tilde{W}_c - \tilde{W}_a) \\ &\times \left(-\tilde{W}_c^T \sigma + \frac{1}{4} \tilde{W}_a^T \left(\frac{\partial \psi}{\partial Z} \right) GR^{-1}G^T \left(\frac{\partial \psi}{\partial Z} \right)^T \tilde{W}_a \right. \\ &\left. - \frac{1}{4} \left(\frac{\partial \epsilon}{\partial Z} \right) GR^{-1}G^T \left(\frac{\partial \epsilon}{\partial Z} \right)^T - \frac{\partial \epsilon}{\partial Z} (F(Z) + G(Z)u^*) \right) \\ &\leq k_{a1} k_1 k_2 \|\tilde{W}_c\|^2 + k_{a1} k_1^2 k_2 \|\tilde{W}_c\| + k_{a1} k_1 k_2 k_3 \|\tilde{W}_c\| + k_{a1} k_1^2 k_2 k_3. \end{aligned} \quad (67)$$

It is obvious that

$$k_{a2} \tilde{W}_a^T (\hat{W}_a - \hat{W}_c) = k_{a2} \tilde{W}_a^T (\tilde{W}_c - \tilde{W}_a) \leq k_{a2} k_1 \|\tilde{W}_c\| - k_{a2} \|\tilde{W}_a\|^2. \quad (68)$$

Moreover, it can be seen that

$$-u^{*T}Ru^* - 2u^{*T}R\Delta \leq \Delta^T R\Delta \leq \lambda_{\max}(R)\bar{\Delta}^2. \quad (69)$$

Replacing (64) and (68) in (63), we can conclude that

$$\begin{aligned} \frac{d}{dt}V_L &\leq -\rho \left(\beta - \frac{1}{2} \right) \|z_\xi\|^2 - z_v^T (Q - 0.5\rho I) z_v - \\ &(c_3 - k_{a1}k_1k_2) \|\tilde{W}_c\|^2 - k_{a2} \|\tilde{W}_a\|^2 + k_{a1}k_1^2k_2k_3 + \\ &k_4 + \left(\frac{c_4 k_c \varphi_0}{2\sqrt{\gamma \varphi_1}} k_3 + k_{a1}k_1k_2k_3 + k_{a1}k_1^2k_2 + k_{a2}k_1 \right) \\ &\|\tilde{W}_c\| + \lambda_{\max}(R)\bar{\Delta}^2. \end{aligned} \quad (70)$$

Based on the estimation $ab \leq \gamma a^2 + (1/4\gamma)b^2$, we can obtain that

$$\begin{aligned} \frac{d}{dt}V_L &\leq -\rho \left(\beta_\xi - \frac{1}{2} \right) \|z_\eta\|^2 - z_v^T (Q - 0.5\rho I) z_v + \lambda_{\max}(R)\bar{\Delta}^2 \\ &- (1 - \theta)(c_3 - k_{a1}k_1k_2) \|\tilde{W}_c\|^2 - k_{a2} \|\tilde{W}_a\|^2 + k_{a1}k_1^2k_2k_3 + \\ &k_4 + \frac{1}{4\theta(c_3 - k_{a1}k_1k_2)} \left(\frac{c_4 \eta_c \varphi_0}{2\sqrt{\gamma \varphi_1}} k_3 + k_{a1}k_1k_2k_3 + k_{a2}k_1 + k_{a1}k_1^2k_2 \right)^2. \end{aligned} \quad (71)$$

Let us select the coefficients to be satisfied such that $\beta > 1/2$, $0 < \rho < 2\lambda_{\min}(Q)$, $0 < \theta < 1$, and $c_3 > k_{a1}k_1k_2$.

With the state vector $z = [z_\eta^T, z_v^T, \tilde{W}_c^T, \tilde{W}_a^T]^T$, it is able to determine that there exist two K -class functions α_3 and α_4 satisfying

$$\begin{aligned} \alpha_3(\|z\|) &\leq \rho \left(\beta_\eta - \frac{1}{2} \right) \|z_\eta\|^2 + z_v^T (Q - 0.5\rho I) z_v + (1 - \theta)(c_3 - k_{a1}k_1k_2) \|\tilde{W}_c\|^2 \\ &+ k_{a2} \|\tilde{W}_a\|^2 \leq \alpha_4(\|z\|). \end{aligned} \quad (72)$$

FIGURE 3: The constrained multimanipulator system [12].

FIGURE 4: The response of 6 joint variables.

Based on (72), inequality (71) can be written as

$$\begin{aligned} \frac{d}{dt}V_L \leq & -\alpha_3(\|z\|) + k_{a1}k_1^2k_2k_3 + k_4 + \lambda_{\max}(R)\bar{\Delta}^2 + \\ & \frac{1}{4\theta(c_3 - k_{a1}k_1k_2)} \left(\frac{c_4k_c\varphi_0}{2\sqrt{\gamma\varphi_1}}k_3 + k_{a1}k_1k_2k_3 + k_{a1}k_1^2ka_2 + k_{a2}k_1 \right)^2. \end{aligned} \quad (73)$$

FIGURE 5: The tracking errors of 6 joint variables.

FIGURE 6: Weights 1-6 of the actor NN.

It is obvious that $(d/dt)V_L$ is negative if $z(t)$ lies outside the attraction region as

$$\Omega_z \triangleq \left\{ z: \|z\| \leq \alpha_3^{-1} \left(\frac{1}{4\theta(c_3 - k_{a1}k_1k_2)} \left(\frac{c_4k_c\varphi_0}{2\sqrt{\gamma}\varphi_1} k_3 + k_{a1}k_1k_2k_3 + k_{a1}k_1^2k_2 + k_{a2}k_1 \right)^2 + k_{a1}k_1^2k_2k_3 + k_4 + \lambda_{\max}(R)\overline{\Delta}^2 \right) \right\}. \quad (74)$$

FIGURE 7: Weights 7–12 of the actor NN.

FIGURE 8: Weights 13–18 of the actor NN.

As a result of this, we can conclude that $\|z\|$ is UUB with attraction region (74) depending on the number of neurons in the critic NN.

Second, the tracking of constraint force is considered through the dynamic of $e_\lambda = \lambda - \lambda_d$ obtained from (47) and (48) as

$$(I + K_f)e_\lambda = N\tau^{(1)} - NC^*(\dot{\xi} - \dot{\xi}_d) - NF^*(\xi, \dot{\xi}, t). \quad (75)$$

According to $\dot{z}_\xi = \dot{\xi} - \dot{\xi}_d = z_v - \beta z_\xi$ and z_v, z_ξ are UUB, $\tau^{(1)}, N$, and $F^*(\eta, \dot{\eta}, t)$ are bounded; then, e_λ is UUB with the attraction region depending on K_f and the bound of ARL-based motion dynamic control $\tau^{(1)} = \begin{bmatrix} \tau_{\text{ind}}^{(1)} \\ \tau_{\text{dep}}^{(1)} \end{bmatrix}$. The proposed controller is absolutely implemented for uncertain multi-manipulators with attraction region (74) depending on model error $\bar{\Delta}$. This completes the proof. $\square$

4. Simulation Results

In this section, an example with 3 manipulators (Figure 3) is given to validate the effectiveness of the proposed control structure (Figure 2) by the *m*-file script in MATLAB software. This example has been considered in [12] by the classical robust nonlinear controller. However, it is obviously different from the existing method in [12], and the proposed motion/force control based on ARL is first structured to obtain optimality satisfaction. Similar to the model in [12], the constrained multimanipulator system is represented by dynamic equation (3) with the following constraints:

$$\begin{aligned} \phi_{12} &= [(x_{E1} - x_{E2})^2 + (z_{E1} - z_{E2})^2]^{(1/2)} - D_{12} = 0 \\ \phi_{13} &= [(x_{E1} - x_{E3})^2 + (y_{E1} - y_{E3})^2 + (z_{E1} - z_{E3})^2]^{(1/2)} - D_{13} = 0. \end{aligned} \quad (76)$$

In this example, the matrices in (3) can be represented as in [12]:

$$\begin{aligned} M &= \begin{bmatrix} M_1 & & \\ & M_2 & \\ & & M_3 \end{bmatrix}, \\ M_1 &= \begin{bmatrix} M_{11} & M_{12} \\ M_{21} & M_{22} \end{bmatrix}, \\ M_2 &= \begin{bmatrix} M_{33} & M_{34} \\ M_{43} & M_{44} \end{bmatrix}, \\ M_3 &= \begin{bmatrix} M_{55} & M_{56} \\ M_{65} & M_{66} \end{bmatrix}, \\ C &= \begin{bmatrix} C_1 & & \\ & C_2 & \\ & & C_3 \end{bmatrix}, \\ C_1 &= \begin{bmatrix} C_{11} & C_{12} \\ C_{21} & C_{22} \end{bmatrix}, \\ C_2 &= \begin{bmatrix} C_{33} & C_{34} \\ C_{43} & C_{44} \end{bmatrix}, \\ C_3 &= \begin{bmatrix} C_{55} & C_{56} \\ C_{65} & C_{66} \end{bmatrix}. \end{aligned} \quad (77)$$

FIGURE 9: Weights 1–6 of the critic NN.

FIGURE 10: Weights 7–12 of the critic NN.

Moreover, to easily compare with the existing work [12], the parameters of multimanipulators are given as $D_{12} = 7$ m, $D_{13} = 3.5$ m, $l_1 = l_2 = 1$ m, $m_1 = m_2 = 1$ kg, $r_1 = r_2 = 0.9$ m, $l_3 = l_4 = 1.5$ m, $m_3 = m_4 = 2$ kg, $r_3 = r_4 = 1.2$ m, $l_5 = l_6 = 1.2$ m, $m_5 = m_6 = 1.5$ kg, and $r_5 = r_6 = 1$ m. Because of two constraints (76), this multimanipulator has 4 independent joints η_1, η_2, η_3 , and η_5 with the desired trajectories $\eta_{1d} = (\pi/12)$, $\eta_{2d} = 1.91\pi + 0.2 \sin(t)$, $\eta_{3d} = 0.51\pi$, and $\eta_{5d} = 0.191\pi$, and the initial values of these independent joints are $\eta_1(0) = \pi/6$, $\eta_2(0) = 1.92\pi$, $\eta_3(0) = 2\pi/3$, $\eta_4(0) = -0.0565$, $\eta_5(0) = \pi/4$, and $\eta_6(0) = -0.0853$. According to the given multimanipulator and above parameters, we achieve the necessary models for investigating the control

design, including motion dynamic model (7), (37), and constraint force (43). In light of Theorem 1, the control parameters are given as follows: $k_1 = 0.5$, $\eta_c = 5$, $\eta_{a1} = 5$, $\eta_{a2} = 10$, and $\nu = 0.01$. The dynamic model of 3 manipulators with the above parameters and ARL-based motion/force control scheme is established by the *m*-file script and Simulink in MATLAB software. The purpose of simulations is to verify the tracking effectiveness of not only joint variables but also constraint force coefficients. Additionally, due to the implementation of optimal control for the motion dynamic model by the radial basis function (RBF) network-based actor/critic reinforcement learning algorithm, the convergence of the adjusting mechanism of

FIGURE 11: Weights 13–18 of the critic NN.

FIGURE 12: The response of constraint coefficients.

actor NN (35) and critic NN (33) should be verified, and the activation function is chosen depending on multi-manipulator systems:

$$\psi(Z) = [z_1^2, z_1 z_2, z_1 z_3, z_1 z_4, z_2^2, z_2 z_3, z_2 z_4, z_3^2, z_3 z_4, z_4^2, z_1^2 z_9, z_2^2 z_{10}, z_3^2 z_{11}, z_4^2 z_{12}, z_1^2 z_5^2, z_2^2 z_6^2, z_3^2 z_7^2, z_4^2 z_8^2], \quad (78)$$

where the vector of state variables is $Z = [z_v, z_\xi, \xi_d]^T \in \mathbb{R}^{12}$. Moreover, the adjusting mechanisms of weights in actor and critic are implemented as in (32 and 35). On the contrary, in order to complete the proposed motion/force control structure, the remaining terms need to be realized, (39) and (48). In order to illustrate the performance of the ARL scheme, we examine to obtain

the actor/critic weight response and the tracking effectiveness of joint variables. Referring to Figures 4 and 5, it is seen that the response of joint variables 1–6 with tracking errors converging to zero is 10 s, which leads to that the effectiveness of closed systems is highly precise under the ARL technique. Additionally, the convergence of adjusting mechanisms of the actor NN and critic NN to the ideal weight is also obtained (0.0006, -0.0055, 0.0045, 0.0038, 0.014, -0.018, -0.009, 0.019, 0.011, 0.0055, 0.000041, 0.00051, 0.0492, 0.002, 0.000015, 0.0024, 0.0047, 0.00019) (Figures 6–11). Finally, the high-performance responses of constraint force and control inputs are shown in Figures 12 and 13. It should be noted that these simulation results are conducted for three manipulators to be described in the work [12]. However, they are obviously different with [12];

FIGURE 13: The control inputs of multimanipulator systems.

due to the ARL-based approach, the proposed control scheme also considers the convergence of weights in actor/critic neural networks (Figures 6–11).

5. Conclusions

This paper presented an ARL-based motion/force tracking control for multimanipulator systems, which consists of the ARL-based motion control law and nonlinear force controller. Both the minimization of the performance index and closed-loop stability are guaranteed by theoretical analysis. Moreover, the proposed control structure is able to handle the time-varying model using the additional term. The simulation studies are conducted to illustrate the efficacy of the proposed control structure. Future work is to develop completely uncertain multimanipulator systems by integrating model-free reinforcement learning.

Data Availability

This publication is supported by multiple datasets, which are available at locations cited in the reference section.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This research was supported by the Ministry of Education and Training, Vietnam, under Grant B2020-BKA-05.

References

- [1] J. Gallardo-Alvarado and J. H. Tinajero-Campos, "A parallel manipulator with planar configurable platform and three end-effectors," *Mathematical Problems in Engineering*, vol. 2019, Article ID 7972837, 12 pages, 2019.
- [2] J. Shi, Y. Mao, P. Li et al., "Hybrid mutation fruit fly optimization algorithm for solving the inverse kinematics of a redundant robot manipulator," *Mathematical Problems in Engineering*, vol. 2020, Article ID 6315675, 13 pages, 2020.
- [3] N. T. Binh, N. A. Tung, D. P. Nam, and N. H. Quang, "An adaptive backstepping trajectory tracking control of a tractor trailer wheeled mobile robot," *International Journal of Control, Automation and Systems*, vol. 17, no. 2, pp. 465–473, 2019.
- [4] T. Nguyena, T. Hoang, M. Pham, and N. Dao, "A Gaussian wavelet network-based robust adaptive tracking controller for a wheeled mobile robot with unknown wheel slips," *International Journal of Control*, vol. 92, no. 11, pp. 2681–2692, 2019.
- [5] Y.-C. Liu, P. N. Dao, and K. Y. Zhao, "On robust control of nonlinear teleoperators under dynamic uncertainties with variable time delays and without relative velocity," *IEEE Transactions on Industrial Informatics*, vol. 16, no. 2, pp. 1272–1280, 2020d.
- [6] Z. P. Wang, T. Zhou, Y. Mao, and Q. J. Chen, "Adaptive recurrent neural network control of uncertain constrained nonholonomic mobile manipulators," *International Journal of Systems Science*, vol. 45, no. 2, pp. 133–144, 2014.
- [7] M. Rani, N. Kumar, and H. P. Singh, "Motion/force control scheme for electrically driven cooperative multiple mobile manipulators," *Control Engineering Practice*, vol. 88, pp. 52–64, 2019.
- [8] J. Han, X. Liu, X. Gao, and X. Wei, "Intermediate observer-based robust distributed fault estimation for nonlinear multiagent systems with directed graphs," *IEEE Transactions on Industrial Informatics*, vol. 16, no. 12, pp. 7426–7436, 2020.
- [9] X. Liu, X. Gao, and J. Han, "Distributed fault estimation for a class of nonlinear multiagent systems," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 50, no. 9, pp. 3382–3390, 2020a.
- [10] B. Baigzadehnoe, Z. Rahmani, A. Khosravi, and B. Rezaie, "On position/force tracking control problem of cooperative robot manipulators using adaptive fuzzy backstepping approach," *ISA Transactions*, vol. 70, pp. 432–446, 2017.
- [11] C. Chen, Z. Liu, Y. Zhang, and S. Xie, "Coordinated motion/force control of multiarm robot with unknown sensor nonlinearity and manipulated object's uncertainty," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 47, pp. 1123–1134, 2016.
- [12] H. Dou and S. Wang, "Robust adaptive motion/force control for motion synchronization of multiple uncertain two-link manipulators," *Mechanism and Machine Theory*, vol. 67, pp. 77–93, 2013.
- [13] L. Kong, W. He, C. Yang, Z. Li, and C. Sun, "Adaptive fuzzy control for coordinated multiple robots with constraint using impedance learning," *IEEE Transactions on Cybernetics*, vol. 49, no. 8, pp. 3052–3063, 2019.
- [14] Y. Li, C. Yang, W. Yan, R. Cui, and A. Annamalai, "Admittance-based adaptive cooperative control for multiple manipulators with output constraints," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 30, no. 12, pp. 3621–3632, 2019.
- [15] Z. Li, B. Huang, A. Ajoudani, C. Yang, C. Y. Su, and A. Bicchi, "Asymmetric bimanual control of dual-arm exoskeletons for human-cooperative manipulations," *IEEE Transactions on Robotics*, vol. 34, pp. 264–271, 2017b.
- [16] Z. Li, W. Yuan, S. Zhao, Z. Yu, Y. Kang, and C. P. Chen, "Brain-actuated control of dual-arm robot manipulation with relative motion," *IEEE Transactions on Cognitive and Developmental Systems*, vol. 11, pp. 51–62, 2017c.

- [17] V. T. Ngo and Y. C. Liu, "Object transportation with force-sensorless control and event-triggered synchronization for networked uncertain manipulators," *IEEE Transactions on Industrial Electronics*, vol. 68, pp. 1–10, 2020.
- [18] T. Zhao, Y. Liu, Z. Li, C. Y. Su, and Y. Feng, "Adaptive control and optimization of mobile manipulation subject to input saturation and switching constraints," *IEEE Transactions on Automation Science and Engineering*, vol. 16, pp. 1543–1555, 2018.
- [19] Z. Li, P. Y. Tao, S. S. Ge, M. Adams, and W. S. Wijesoma, "Robust adaptive control of cooperating mobile manipulators with relative motion," *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, vol. 39, pp. 103–116, 2008b.
- [20] Z. Li, C. Yang, and Y. Tang, "Decentralised adaptive fuzzy control of coordinated multiple mobile manipulators interacting with non-rigid environments," *IET Control Theory & Applications*, vol. 7, no. 3, pp. 397–410, 2013.
- [21] M. Deng, Z. Li, Y. Kang, C. P. Chen, and X. Chu, "A learning-based hierarchical control scheme for an exoskeleton robot in human–robot cooperative manipulation," *IEEE Transactions on Cybernetics*, vol. 50, pp. 112–125, 2018.
- [22] X. Yu, W. He, H. Li, and J. Sun, "Adaptive fuzzy full-state and output-feedback control for uncertain robots with output constraint," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 99, pp. 1–14, 2020.
- [23] Z. Li, S. Deng, C. Y. Su et al., "Decentralised adaptive control of cooperating robotic manipulators with disturbance observers," *IET Control Theory & Applications*, vol. 8, no. 7, pp. 515–521, 2014a.
- [24] Z. Li, S. S. Ge, and A. Ming, "Adaptive robust motion/force control of holonomic-constrained nonholonomic mobile manipulators," *IEEE Transactions on Systems, Man and Cybernetics, Part B (Cybernetics)*, vol. 37, no. 3, pp. 607–616, 2007.
- [25] Z. Li, W. Chen, and J. Luo, "Adaptive compliant force-motion control of coordinated non-holonomic mobile manipulators interacting with unknown non-rigid environments," *Neuro-computing*, vol. 71, no. 7–9, pp. 1330–1344, 2008a.
- [26] X. Liu, S. S. Ge, F. Zhao, and X. Mei, "Optimized impedance adaptation of robot manipulator interacting with unknown environment," *IEEE Transactions on Control Systems Technology*, vol. 29, pp. 411–419, 2020b.
- [27] X. Liu, S. S. Ge, F. Zhao, and X. Mei, "Optimized interaction control for robot manipulator interacting with flexible environment," *IEEE/ASME Transactions on Mechatronics, Early Access*, vol. 99, p. 1, 2020c.
- [28] C. Yang, G. Peng, Y. Li, R. Cui, L. Cheng, and Z. Li, "Neural networks enhanced adaptive admittance control of optimized robot–environment interaction," *IEEE Transactions on Cybernetics*, vol. 49, pp. 2568–2579, 2018.
- [29] W. He, C. Xue, X. Yu, Z. Li, and C. Yang, "Admittance-based controller design for physical human-robot interaction in the constrained task space," *IEEE Transactions on Automation Science and Engineering*, vol. 62, 2020b.
- [30] S. Ganjefar, S. Najibi, and H. Momeni, "A novel structure for the optimal control of bilateral teleoperation systems with variable time delay," *Journal of the Franklin Institute*, vol. 348, no. 7, pp. 1537–1555, 2011.
- [31] P. N. Dao and Y. C. Liu, "Adaptive reinforcement learning strategy with sliding mode control for unknown and disturbed wheeled inverted pendulum," *International Journal of Control, Automation and Systems*, vol. 60, pp. 1–12, 2021.
- [32] P. N. Dao, P. T. Loc, T. Q. Huy et al., "Sliding variable-based online adaptive reinforcement learning of uncertain/disturbed nonlinear mechanical systems," *Journal of Control, Automation and Electrical Systems*, vol. 32, no. 2, pp. 1–10, 2021.
- [33] J. Li, J. Ding, T. Chai, F. L. Lewis, and S. Jagannathan, "Adaptive interleaved reinforcement learning: robust stability of affine nonlinear systems with unknown uncertainty," *IEEE Transactions on Neural Networks and Learning Systems, Early Access*, vol. 34, no. 1, pp. 1–11, 2021.
- [34] Y. Lv, X. Ren, and J. Na, "Adaptive optimal tracking controls of unknown multi-input systems based on nonzero-sum game theory," *Journal of the Franklin Institute*, vol. 356, no. 15, pp. 8255–8277, 2019.
- [35] C. Mu, Z. Ni, C. Sun, and H. He, "Data-driven tracking control with adaptive dynamic programming for a class of continuous-time nonlinear systems," *IEEE Transactions on Cybernetics*, vol. 47, pp. 1460–1470, 2016.
- [36] T. Sun and X. M. Sun, "An adaptive dynamic programming scheme for nonlinear optimal control with unknown dynamics and its application to turbofan engines," *IEEE Transactions on Industrial Informatics*, vol. 17, p. 14, 2020.
- [37] Z. Wang, H. Li, H. Wu, and Z. Wu, "Improving maneuver strategy in air combat by alternate freeze games with a deep reinforcement learning algorithm," *Mathematical Problems in Engineering*, vol. 2020, Article ID 7180639, 17 pages, 2020.
- [38] H. N. Wu and Z. Y. Liu, "Data-driven guaranteed cost control design via reinforcement learning for linear systems with parameter uncertainties," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 50, 2019.
- [39] X. Yang, H. He, and X. Zhong, "Approximate dynamic programming for nonlinear-constrained optimizations," *IEEE Transactions on Cybernetics*, vol. 99, pp. 1–14, 2019.
- [40] W. He, H. Gao, C. Zhou, C. Yang, and Z. Li, "Reinforcement learning control of a flexible two-link manipulator: an experimental investigation," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 11, 2020a.
- [41] S. Bhasin, R. Kamalapurkar, M. Johnson, K. G. Vamvoudakis, F. L. Lewis, and W. E. Dixon, "A novel actor-critic-identifier architecture for approximate optimal control of uncertain nonlinear systems," *Automatica*, vol. 49, no. 1, pp. 82–92, 2013.
- [42] K. G. Vamvoudakis, "Q-learning for continuous-time linear systems: a model-free infinite horizon optimal control approach," *Systems & Control Letters*, vol. 100, pp. 14–20, 2017.
- [43] Z. Li and S. S. Ge, *Fundamentals in Modeling and Control of Mobile Manipulators*, CRC Press, Boca Raton, FL, USA, 2013.
- [44] Z. Li, C. Yang, C. Y. Su, S. Deng, F. Sun, and W. Zhang, "Decentralized fuzzy control of multiple cooperating robotic manipulators with impedance interaction," *IEEE Transactions on Fuzzy Systems*, vol. 23, pp. 1044–1056, 2014b.
- [45] W. He, Y. Chen, and Z. Yin, "Adaptive neural network control of an uncertain robot with full-state constraints," *IEEE Transactions on Cybernetics*, vol. 46, pp. 620–629, 2015.
- [46] Y. Zhu, D. Zhao, and X. Li, "Using reinforcement learning techniques to solve continuous-time non-linear optimal tracking problem without system dynamics," *IET Control Theory & Applications*, vol. 10, no. 12, pp. 1339–1347, 2016.
- [47] J. Li, B. Kiumarsi, T. Chai, F. L. Lewis, and J. Fan, "Off-policy reinforcement learning: optimal operational control for two-time-scale industrial processes," *IEEE Transactions on Cybernetics*, vol. 47, no. 12, pp. 4547–4558, 2017a.
- [48] Y. Peng, Q. Chen, and W. Sun, "Reinforcement q-learning algorithm for h tracking control of unknown discrete-time

- linear systems," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 50, 2019.
- [49] Y. Lv, X. Ren, and J. Na, "Online nash-optimization tracking control of multi-motor driven load system with simplified rl scheme," *ISA Transactions*, vol. 98, pp. 251–262, 2020.
- [50] J. Na, Y. Lv, K. Zhang, and J. Zhao, "Adaptive identifier-critic-based optimal tracking control for nonlinear systems with experimental validation," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 60, no. 1, pp. 1–14, 2020.