IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received June 25, 2019, accepted July 3, 2019, date of publication July 9, 2019, date of current version August 2, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2927511

Actor-Critic-Identifier Structure-Based Decentralized Neuro-Optimal Control of Modular Robot Manipulators With Environmental Collisions

BO DONG[1,2], TIANJIAO AN[1], FAN ZHOU[1], KEPING LIU[1], WEIBO YU[1], AND YUANCHUN LI[1]
[1]Department of Control Science and Engineering, Changchun University of Technology, Changchun 130012, China
[2]The State Key Laboratory of Management and Control for Complex Systems, Institute of Automation, Chinese Academy of Sciences, Beijing 100190, China

Corresponding author: Yuanchun Li (liyc@ccut.edu.cn)

This work was supported in part by the National Natural Science Foundation of China under Grant 61773075 and Grant 61703055, in part by the Scientific Technological Development Plan Project in Jilin Province of China under Grant 20190103004JH, Grant 20170204067GX, Grant 20160520013JH, and Grant 20160414033GH, and in part by the Project of Engineering Laboratory of Intelligent Robot and Vision Measurement and Control Technology in Jilin Province under Grant 2019C010.

ABSTRACT This paper presents a decentralized zero-sum optimal control method for MRMs with environmental collisions via an actor-critic-identifier (ACI) structure-based adaptive dynamic programming (ADP) algorithm. The dynamic model of the MRMs is formulated via a novel collision identification method that is deployed for each joint module, in which the local position and torque information are used to design the model compensation controller. A neural network (NN) identifier is developed to compensate the model uncertainties and then, the optimal control problem of the MRMs with environmental collisions can be transformed into a two-player zero-sum optimal control one. Based on the ADP algorithm, the Hamilton-Jacobi-Isaacs (HJI) equation is solved by constructing the actor-critic NNs, thus making the derivation of the approximate optimal control policy feasible. Based on the Lyapunov theory, the closed-loop robotic system is proved to be asymptotically stable. Finally, the experiments are conducted to verify the effectiveness and advantages of the proposed method.

INDEX TERMS Adaptive dynamic programming, collision identification, decentralized optimal control, modular robot manipulators, zero-sum game.

I. INTRODUCTION

Modular robot manipulators have attracted extensive attentions in robotics community since they have better structural adaptability and flexibility than conventional robot manipulators. The standardized robotic modules, as the basic element to constitute an MRM, consist of actuators, speed reducers, sensors and communication units. These modules can be assembled to desired configurations via standard mechanical connectors according to the requirements of different tasks. Profiting from the advantages above, MRMs are always employed in dangerous and complex environments, such as space exploration, hazard survey, rescue operation and sports activities et al., in which the collisions from external environments may occur at any time. Therefore, we need to develop appropriate control systems to guarantee the stability and precision of MRMs even with external environmental collisions.

To address the problems of enhancing the stability and control precision of the robot manipulator systems in the face of environmental collisions, collision identification, which aims at obtaining the collision joint torques, is considered an efficient method to implement the joint torque feedback and to facilitate the controller design of the robotic systems with collisions. Several researches [1]–[3] report direct torque sensing-based techniques, in which the joint torque information are measured by embedding torque sensors into the joint modules and thus completing the joint torque feedback. However, the joint torques may increase significantly while unexpected environmental collisions occurred intensely. Unfortunately, this may cause irreversible damage to the strain gauges, which are the core components of joint

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

torque sensors. Luca et al. propose a sensorless collision identification method [4], in which the environmental collisions are considered the faulty behaviors of the robot actuating systems, and then the research results are further applied to address the problems of collision detection and safe reaction with lightweight robot manipulators [5]. However, when the actuator faults and environmental collisions exist at the same time, these methods can hardly distinguish their effects, which may leave potential safety hazard for the robotic systems. Some other feasible solutions of implementing collision identification include that measure the collision force/torque information by using the sensing fusion of the tactile together with exteroceptive information [6]–[8]. Nevertheless, the tactile and exteroceptive sensors are terribly expensive and in a general view, difficult to be generalized to multiple types of robotic systems. Some forward-looking researches, which investigated the synchronization [9] and convergence [10] of the NNs, are presented for improving the noise tolerance [11] of multi-agent systems [12] and robotic systems [13], but these methods still can hardly complete the optimal implementation of the controllers. Therefore, an ideal collision identification method should simultaneously take into account both the safety, accuracy and economy of robotic systems, especially, satisfy the requirements of modularity and reconfiguration of the MRM systems.

Besides the requirement of the reliable collision identification, MRM systems also need appropriate optimal controllers, which are with the properties that guarantee the stability of robotic systems and simultaneously consider the optimality of composite of error characteristics and output energy-efficiency. Adaptive dynamic programming methodology, which was proposed by Werbos [14], is recognized as one of the key directions to address the optimal control problems of complex systems. There are several synonyms adopted for ADP, including ''approximate dynamic programming'' [15], ''neural dynamic programming'' [16], ''adaptive critic designs'' [17] and ''reinforcement learning'' (RL) [18], [19] etc. Recently, ADP-based approaches have been utilized to deal with the optimal control issues of discrete-time systems [20]–[22], continuous-time systems [23]–[25], data driven-based systems [26]–[28], two-player zero-sum games [29]–[31], and further employed to develop the intractability of nonlinear optimal control with input/output constraints [32], [33], uncertainties and/or external disturbance [34]–[36], and actuator failures [37], etc. Some investigations report the latest research progress of ADP-based optimal control methods for robot manipulators systems. In [38], an adaptive fuzzy NN control method is developed for constrained robots via ADP-based impedance learning. An online adaptive tracking control algorithm is proposed to solve the tracking problem of wheeled mobile robots with uncertain resistance by using the ADP algorithm [39]. Reference [40] investigated the behaviors of robot manipulators that use an RL-based decentralized control scheme. Reference [41] presented a deep RL framework to handle the learning control issue in physical human-robot interaction (pHRI) tasks. An RL optimal control method is reported in [42] to address the problems of global dynamic compensation and force tracking control in industrial interaction robotics tasks. The mentioned strategies above both follow the premise that the dynamic models of the robotic systems are completely unknown, thus the implementation of these methods are restricted in solving the optimal control problems of special type of robotic systems without local dynamic compensation. Dong et al. address the optimal control problems of MRMs by combining the model-based compensation control with ADP-based learning control [43], and their researches are further expanded to deal with the optimal tracking control issues of MRMs with uncertain environments [44]. However, the existing methods consider the disturbance torques, which are introduced by environmental contacts and collisions, a class of dynamic uncertainties, and ignore the intractability of explicitly compensating the effects of model uncertainties and environmental collision disturbance. To the best of the authors' knowledge, there is less discussion on ADP-based optimal control for robot manipulators in the face of environmental collisions with both the implementation of reliable collision identification and optimal disturbance compensation, especially, for the MRM systems with theoretical and experimental investigation.

In this paper, an ACI structure-based decentralized zero-sum neuro-optimal control approach is presented for MRMs with environmental collisions. First, the dynamic model of the MRM systems is formulated via a novel harmonic drive compliance model-based collision identification method, and then a model-based compensation controller is developed by effectively utilizing the local position and torque information of each joint module. Second, an NN identifier, which is established to approximate the dynamics of the model uncertainty, is used to design the learning-based compensation controller, then, the optimal control problem of environment-collided MRM systems is transformed into a two-player zero-sum optimal control one. The ADP algorithm is employed to solve the HJI equation, in which the cost function, optimal control policy and worst collision disturbance can be approximated by constructing one critic NN and two actor NNs, and then the decentralized zero-sum neuro-optimal control is developed. Based on the Lyapunov theory, the joint position/velocity errors and NN weight approximate errors are proved to be uniform ultimate bounded (UUB), and the closed-loop MRM systems are also proved to be asymptotically stable. Finally, experiments are performed to clarify the effectiveness and advantage of the proposed method.

The main contributions of this work are briefly summarized as follows:

1) It is a novel idea that develops the collision identification method based on a nonlinear harmonic drive compliance model. Note that the compliance characteristics of the harmonic drive components, which are considered the uncertain perturbance in some existing researches, are effectively utilized in this paper to

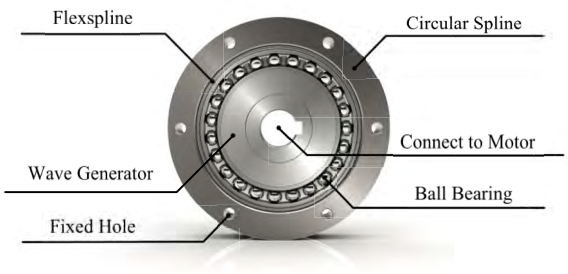


FIGURE 1. Schematic illustration of the harmonic drive device.

estimate the joint torques of the MRM system along with the environmental collisions.

2) Unlike the existing methods that leave the intractability of optimal compensation of environmental collision disturbance out of consideration. In our works, a decentralized optimal control method is developed via a novel ACI structure-based zero-sum neuro-optimal control scheme, and the advantages of the proposed method is verified by using the established experimental platform.

The remainder of this paper is organized as follows: Section 2 analyses the collision identification method and the dynamic model formulation. The decentralized optimal control method is developed in Section 3. Experiments are performed in Section 4. Finally, a brief conclusion is derived in Section 5.

II. COLLISION IDENTIFICATION AND DYNAMIC MODEL FORMULATION

A. COLLISION IDENTIFICATION

Collision identification aims at estimating the coupled torques of each robotic joint, while the uncertain environmental collisions are acting along the robot structure [45]. In this part, a novel collision identification method is presented based on a nonlinear harmonic drive compliance model as well as the position and current measurements of each joint module.

1) HARMONIC DRIVE COMPLIANCE MODEL

As illustrated in Figure 1, the adopted harmonic drive device includes a wave generator, a flexspline and a circular spline. The wave generator consists of an elliptical wave generator plug, which is assembly inserted into the ball-bearing, thus providing an elliptical shape for the bearing too. The flexspline, which is a thin cylindrical cup with external teeth, is designed with a slightly smaller pitch diameter than the circular spline with internal teeth. Since the flexspline has two less teeth than the circular spline, so that this may lead to a small phase deviation between the corresponding teeth in engagement. After assembled, the flexspline fits tightly over the wave generator, accordingly, when the wave generator plug is rotating, the flexspline deforms and molds into the shape of the rotating ellipse. When the harmonic drive device is equipped into a joint module, the wave generator is connected to a motor, the circular spline is fixed with joint shell and the flexspline, which is sandwiched between the wave generator and the circular spline, is connected to the joint output. Then, for the purpose of identifying the coupled joint torque that using only position and current measurement, here, we need to formulate a harmonic drive model, in which the nonlinear compliance features of each harmonic drive component can be clearly reflected.

According to the kinematic relation and the compliance behavior of harmonic drive components in [46], [47], one can represent the torsional angular variables of wave generator and flexspline, which are given as follows:

$$\begin{aligned} \Delta q_{iF} &= q_{iFout} - q_{iFin}, \\ \Delta q_{iW} &= q_{iWout} - q_{iWin}, \end{aligned} \tag{1}$$

where the first subscript letter "i" means the ith joint module, q_{iWout} and q_{iWin} indicate the wave generator angular positions at the outside part and the center part respectively; q_{iFout} and q_{iFin} denote the angular positions of load-side and gear-side of the flexspline respectively. Note that it is impossible to embed any sensors into a harmonic drive device, so that only the flexspline output position (joint's angular position) and the wave generator input position can be measured by the link-side and motor-side encoders respectively. These angular position measurements can be used to calculate the harmonic drive torsional angle Δq_i with the following relation:

$$\Delta q_i = q_{iFout} - \frac{q_{iWin}}{\gamma_i}, \tag{2}$$

where γ_i denotes the gear ratio of the harmonic drive device. By adding and subtracting the terms q_{iFin} and $\frac{q_{iWout}}{\gamma_i}$ to (2), one may get

$$\begin{aligned} \Delta q_i = q_{iFout} - q_{iFin} \\ + \left(q_{iFin} - \frac{q_{iWout}}{\gamma_i}\right) + \left(\frac{q_{iWout}}{\gamma_i} - \frac{q_{iWin}}{\gamma_i}\right). \end{aligned} \tag{3}$$

Assuming that there is no relation displacement between the flexspline input and the wave generator output, which means that $\gamma_i q_{iFin} - q_{iWout} = 0$, then, substituting (1) into (3), one obtains

$$\Delta q_i = \Delta q_{iF} + \frac{\Delta q_{iW}}{\gamma_i}. \tag{4}$$

According to the hysteresis behaviors of a harmonic drive device that is shown in Figure 2, one may conclude that the local elastic coefficient e_{iFl} increases along with the increase of flexspline output torque τ_{iF}, so that we can define the local elastic coefficient e_{iFl} as the form of

$$e_{iFl} = \frac{d\tau_{iF}}{d\Delta q_{iF}}. \tag{5}$$

Considering the symmetric property of the harmonic drive stiffness and using Taylor expansion, the local elastic coefficient can be approximated as

$$e_{iFl} = e_{iFl0}\left(1 + (c_{iF}\tau_{iF})^2\right), \tag{6}$$

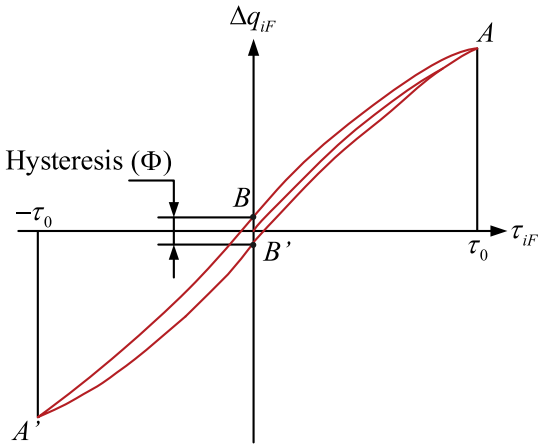


FIGURE 2. Typical hysteresis curve of a harmonic drive.

where e_{iFl0} and c_{iF} are determined position constants. If $e_{iFl0} \neq 0$, one can obtain the flexspline torsion which is expressed as follows:

$$\Delta q_{iF} = \int_0^{\tau_{iF}} \frac{d\tau_{iF}}{e_{iFl}}. \tag{7}$$

Substituting (6) into (7), the flexspline torsion can be calculated as

$$\Delta q_{iF} = \frac{\arctan(c_{iF}\tau_{iF})}{c_{iF}e_{iFl0}}, \tag{8}$$

where arctan(·) represents the arctangent function. Besides, from Figure 2, one obtains that the torsional angle Δq_{iF} may range from $-\Phi/2$ to $\Phi/2$, while τ_{iF} equals to zero; moreover, the torsional angle turns to zero, while τ_{iF} reaches to the nominal torque, where Φ represents the hysteresis loss. To replicate the shape of the hysteresis curve, we formulate the local elastic coefficient of wave generator as follows:

$$e_{iWl} = e_{iWl0}e^{c_{iW}|\tau_{iF}|}, \tag{9}$$

where e_{iWl0} and c_{iW} are determined positive constants. If $e_{iWl0} \neq 0$, one may calculate the torsional angle of wave generator Δq_{iW} by using the relation

$$\Delta q_{iW} = \int_0^{\tau_{iW}} \frac{d\tau_{iW}}{e_{iWl}}, \tag{10}$$

where τ_{iW} is the torque at the wave generator location. By substituting the expression of e_{iWl} in (9) into (10), one obtains

$$\Delta q_{iW} = \frac{\mathrm{sgn}(\tau_{iW})}{c_{iW}e_{iWl0}}(1 - e^{-c_{iW}|\tau_{iW}|}), \tag{11}$$

where sgn(·) is a typical sign function, expressing as

$$\mathrm{sgn}(\cdot) = \begin{cases} 1 & for \cdot > 0 \\ 0 & for \cdot = 0 \\ -1 & for \cdot < 0. \end{cases} \tag{12}$$

2) COLLISION IDENTIFICATION BASED ON THE HARMONIC DRIVE COMPLIANCE MODEL

Based on the nonlinear harmonic drive compliance model, in this part, we focus on estimating the coupled joint torque, which equal to the flexspline output torque, when the environmental collision happened.

Combining the flexspline torsional angle (8) with the wave generator torsional angle (11), the coupled joint torque can be estimated by:

$$\begin{aligned}\tau_{iF} &= \frac{\tan(\Delta q_{iF}c_{iF}e_{iFl0})}{c_{iF}} \\ &= \frac{1}{c_{iF}}\tan\left(c_{iF}e_{iFl0}\left(\Delta q_i - \frac{\mathrm{sgn}(\tau_{iW})\left(1-e^{-c_{iW}|\tau_{iW}|}\right)}{\gamma_i c_{iW}e_{iWl0}}\right)\right),\end{aligned} \tag{13}$$

where the torsional angle Δq_i is calculated by (2) and the wave generator torque τ_{iW} can be obtain by using the following relation:

$$\tau_{iW} = \tau_{im} - I_{im}\ddot{q}_{iW}, \tag{14}$$

where τ_{im} is the motor output torque that can be calculate by using the measured motor current; I_{im} denotes a coupled moment of inertia including the inertia the inertia of the motor's rotor, the shaft and the wave generator. Note that the magnitude of $I_{im}\ddot{q}_{iW}$ is very small, so that one can approximately consider the wave generator torque τ_{iW} is equal to the motor output torque τ_{im}.

Remark 1: From (13), one obtains that the coupled joint torque can be estimated by substituting the measurements of motor-side and link-side position as well as the motor current into the harmonic drive compliance model. Note that the position and current information can also be accurately measured, when he undesired environmental collision happened, so that the collision identification is realized without the need of embedding joint torque sensors into joint modules.

In the following part, the coupled joint torque is utilized in dynamic model formulation of MRMs.

B. DYNAMIC MODEL FORMULATION

We consider an MRM system that is comprised of n joint modules assembled in series and each module constitute a direct current (DC) motor as actuator, a harmonic drive device as speed reducer, a motor-side incremental encoder as well as a link-side absolute encoder as position and velocity sensors. Then, by referencing the modeling methods of joint torque feedback-based robot manipulators with multi-degree of freedom (DOF) [48], [49], the dynamic model of the MRM can be formulated as an integration of independent joint subsystems with interconnected dynamic coupling (IDC) effects, where the ith subsystem model is formulated as:

$$I_{im}\gamma_i\ddot{q}_i + \frac{\tau_{iF}}{\gamma_i} + f_{ri}(q_i, \dot{q}_i) + h_i(q, \dot{q}, \ddot{q}) + d_i(q_i) = \tau_i, \tag{15}$$

where the subscript "i" means the ith subsystem; $q_i, \dot{q}_i$, $\ddot{q}_i$ represent the joint position, velocity and acceleration

variables respectively; τ_{iF} is the coupled joint torque; $f_{ri}(q_i, \dot{q}_i)$ indicates the lumped friction torque of the joint module; $h_i(q, \dot{q}, \ddot{q})$ denotes the IDC among the joint subsystems; $d_i(q_i)$ is the disturbance torque and τ_i is the control torque.

1) JOINT FRICTION TORQUE

We consider the lumped friction torque $f_{ri}(q_i, \dot{q}_i)$ that includes the motor friction and the harmonic drive friction in each joint module. According to [50], [51], a friction function, which is regarding of the joint position and velocity information, is defined as follows:

$$\begin{aligned} f_{ri}(q_i, \dot{q}_i) &= \hat{f}_{ib}\dot{q}_i + \left(\hat{f}_{is}e^{(-\hat{f}_{i\tau}\dot{q}_i^2)} + \hat{f}_{ic}\right)\mathrm{sgn}(\dot{q}_i) \\ &\quad + f_{ip}(q_i, \dot{q}_i) + Y(\dot{q}_i)\tilde{F}_i, \\ Y(\dot{q}_i) &= \Big[\dot{q}_i, \quad \mathrm{sgn}(\dot{q}_i), e^{(-\hat{f}_{i\tau}\dot{q}_i^2)}\mathrm{sgn}(\dot{q}_i), \\ &\quad -\hat{f}_{is}\dot{q}_i^2 e^{(-\hat{f}_{i\tau}\dot{q}_i^2)}\mathrm{sgn}(\dot{q}_i)\Big] \\ \tilde{F}_i &= \left[f_{ib} - \hat{f}_{ib}, f_{ic} - \hat{f}_{ic}, f_{is} - \hat{f}_{is}, f_{i\tau} - \hat{f}_{i\tau}\right]^T, \end{aligned} \tag{16}$$

where f_{ib} denotes the viscous friction parameter, f_{is} is the static friction coefficient, $f_{i\tau}$ indicates the Stribect effect parameter, f_{ic} is the Coulomb friction coefficient and $f_{ip}(q_i, \dot{q}_i)$ represents the position dependency friction term. Moreover, $\hat{f}_{ib}, \hat{f}_{is}, \hat{f}_{i\tau}$ and $\hat{f}_{ic}$ are estimated parameters of f_{ib}, $f_{is}, f_{i\tau}$ and f_{ic} respectively.

2) INTERCONNECTED DYNAMIC COUPLING

The IDC term, which is considered a complex nonlinear function, is defined as:

$$\begin{aligned} h_i(q, \dot{q}, \ddot{q}) &= I_{im}\sum_{j=1}^{i-1} z_{mi}^T z_{lj}\ddot{q}_j + I_{im}\sum_{j=2}^{i-1}\sum_{k=1}^{j-1} z_{mi}^T\left(z_{lk}\times z_{lj}\right)\dot{q}_k\dot{q}_j \\ &= I_{im}\sum_{j=1}^{i-1} D_j^i\ddot{q}_j + I_{im}\sum_{j=2}^{i-1}\sum_{k=1}^{j-1}\Theta_{kj}^i\dot{q}_k\dot{q}_j, \end{aligned} \tag{17}$$

where z_{mi}, z_{lj} and z_{lk} denote the unity vectors along with the rotation axis of the ith, jth and kth joint respectively. Moreover, we also have the relations $D_j^i = z_{mi}^T z_{lj}$ and $\Theta_{kj}^i = z_{mi}^T\left(z_{lk}\times z_{lj}\right)$. Reformulating (17), one obtains

$$\begin{aligned} h_i(q, \dot{q}, \ddot{q}) &= U_{iz} + V_{iz}, \\ U_{iz} &= \sum_{j=1}^{i-1}\left[I_{im}\hat{D}_j^i \quad I_{im}\right]\left[\ddot{q}_j \quad \tilde{D}_j^i\ddot{q}_j\right]^T, \\ V_{iz} &= \sum_{j=2}^{i-1}\sum_{k=1}^{j-1}\left[I_{im}\hat{\Theta}_{kj}^i \quad I_{im}\right]\left[\dot{q}_k\dot{q}_j \quad \tilde{\Theta}_{kj}^i\dot{q}_k\dot{q}_j\right]^T, \end{aligned} \tag{18}$$

where $\hat{D}_j^i = D_j^i - \tilde{D}_j^i$ denotes the dot product of unit vector z_{mi} and z_{lj}, $\tilde{D}_j^i$ is the alignment error. Similarly, $\hat{\Theta}_{kj}^i = \Theta_{kj}^i - \tilde{\Theta}_{kj}^i$ indicates the dot product of z_{mi} and $z_{lk}\times z_{lj}$, $\tilde{\Theta}_{kj}^i$ is the alignment error.

3) DISTURBANCE TORQUE

Define the disturbance torque term $d_i(q_i)$ as:

$$d_i(q_i) = d_{ie}(q_i) + d_{ic}(q_i), \tag{19}$$

where $d_{ie}(q_i)$ indicates the joint torque estimation error, $d_{ic}(q_i)$ represents the unacquirable perturbed torque, which is caused by undesired environmental collision. According to the model formulation in (15), (16), (18) and (19), we know that there exist many uncertainties, which include the friction modeling error, the IDC effect, the joint torque estimation error and the unobtainable collision disturbance. It is noted that the uncertainties are with the following properties.

Property 1: For the frictional model (16), since the friction coefficients f_{ib}, f_{is}, $f_{i\tau}$, f_{ic} and their approximate ones are bounded, the approximation error vector $\tilde{F}_i$ is also bounded, where the up-bound can be given as $\left|\tilde{F}_i\right| \le \rho_{Fil}$, where $l = 1, 2, 3, 4$ and ρ_{Fil} is a known positive constant vector. Accordingly, the friction modeling error term $Y(\dot{q}_i)\tilde{F}_i$ is bounded by $\left|Y(\dot{q}_i)\tilde{F}_i\right| \le Y(\dot{q}_i)\rho_{Fil}$. Moreover, the position dependency friction term $f_{ip}(q_i, \dot{q}_i)$ is bounded by $\left|f_{ip}(q_i, \dot{q}_i)\right| \le \rho_{fip}$, where ρ_{fip} is a known positive constant.

Property 2: For the IDC terms that represented in (17) and (18), we know that the vector products among the terms z_{mi}, z_{lj} and z_{lk} are bounded by $\left|D_j^i\right| = \left|z_{mi}^T z_{lj}\right| \le 1$ and $\left|\Theta_{kj}^i\right| = \left|z_{mi}^T\left(z_{lk}\times z_{lj}\right)\right| \le 1$ respectively. Moreover, if the lower j and k joints ($1 < j, k < i-1$) are stabilized, then the terms U_{iz} and V_{iz} are bounded, and satisfying the following relations:

$$\begin{aligned} |U_{iz}| &\le \rho_{iU} \\ |V_{iz}| &\le \rho_{iV}, \end{aligned} \tag{20}$$

where ρ_{iU} and ρ_{iV} are known constants.

Property 3: The joint torque estimation error $d_{ie}(q_i)$ is bounded and the known up-bound $|d_{ie}(q_i)| \le \rho_{die}$ can be calculated by the product of the kinematic error ratio of harmonic drive transmission and the joint torque. Moreover, the perturbed torque $d_{ic}(q_i)$, which is attributed to the unobtainable collision disturbance, is also bounded, since the joint external torque is bounded. Here, the up-bound can be given by $|d_{ic}(q_i)| \le \rho_{dic}$, where ρ_{dic} is a known position constant.

4) STATE SPACE DESCRIPTION

Then, according to the dynamic model of the ith joint subsystem (15), define the system state $x_i = \left[x_{i1}\ x_{i2}\right]^T = \left[q_i\ \dot{q}_i\right]^T \in R^{2\times 1}$, the control torque $u_i = \tau_i \in R^{1\times 1}$ and the positive coefficient $B_i = (I_{im}\gamma_i)^{-1} \in R^+$, one can obtain the state space of the ith subsystem that is formulated as follows:

$$S_i: \begin{cases} \dot{x}_{i1} = x_{i2} \\ \dot{x}_{i2} = \phi_i(x_i) + \upsilon_i(x) + p_i(x_i) + B_i u_i \\ y = x_{i1} \end{cases}, \tag{21}$$

where $\phi_i(x_i)$ represented the accurately modeled and estimated parts of the subsystem dynamics, $\upsilon_i(x)$ indicates a

global model uncertainty term of the subsystem and $p_i(x_i)$ is the environmental collision disturbance term. Here, we can represent these terms as follows:

$$\begin{aligned}\phi_i(x_i) &= B_i\begin{pmatrix}-\left(\hat{f}_{is}e^{\left(-\hat{f}_{i\tau}\dot{q}_i^2\right)}+\hat{f}_{ic}\right)\operatorname{sgn}(\dot{q}_i)\\ -\hat{f}_{ib}\dot{q}_i-\dfrac{\tau_{iF}}{\gamma_i}\end{pmatrix}\\ \upsilon_i(x) &= B_i\left(-U_{iz}-V_{iz}-f_{ip}(q_i,\dot{q}_i)-Y(\dot{q}_i)\tilde{F}_i\right)-d_{ie}(q_i)\\ p_i(x_i) &= B_i(-d_{ic}(q_i)).\end{aligned}\tag{22}$$

In the following section, based on ADP algorithm and ACI structure, a decentralized zero-sum optimal control method is presented for MRMs in face of environmental collisions to ensure that both position and velocity of MRM systems are asymptotically stable.

III. ACI STRUCTURE-BASED DECENTRALIZED ZERO-SUM OPTIMAL CONTROL VIA ADP ALGORITHM

A. ZERO-SUM OPTIMAL CONTROL AND PROBLEM TRANSFORMATION

For an MRM system with the dynamic model formulation (15) and state space description (21), we consider the environmental collision disturbance $p_i(x_i)$ as a kind of system control input, then the optimal control problem can be transformed into a two-player zero-sum optimal control issue, in which the continuously differentiable infinite horizon local performance index function can be defined as follows:

$$J_i(l_{i0}(e_i)) = \int_0^\infty \begin{Bmatrix} l_i(e_i(\tau))^T Q_i l_i(e_i(\tau)) \\ +u_i(\tau)^T R_i u_i(\tau) \\ -\gamma_{ip}^2 p_i(x_i(\tau))^T p_i(x_i(\tau)) \end{Bmatrix} d\tau, \tag{23}$$

where $l_{i0}(e_i)$, which indicates a filtered error function with respect to the joint position and velocity errors, is defined as

$$l_i(e_i) = \dot{e}_i + \alpha_{ie} e_i, \tag{24}$$

where $e_i = x_{i1} - x_{id}$ and $\dot{e}_i = x_{i2} - \dot{x}_{id}$ represent the position and velocity tracking error of the ith robotic joint respectively; x_{id}, $\dot{x}_{id}$, $\ddot{x}_{id}$ indicate the determined reference joint variables; $Q_i^T = Q_i$, $R_i^T = R_i$ are positive constant matrixes and α_{ie}, γ_{ip} are the positive constants to be determined. Then, according to the local performance index function $J_i(l_{i0})$ in (23), we can define the local Hamiltonian as follows:

$$\begin{aligned}&H_i(l_i, u_i, p_i, \nabla J_i)\\ &\quad= l_i^T Q_i l_i + u_i^T R_i u_i - \gamma_{ip}^2 p_i^T p_i + \nabla J_i^T(l_i)\dot{l}_i\\ &\quad= l_i^T Q_i l_i + u_i^T R_i u_i - \gamma_{ip}^2 p_i^T p_i\\ &\qquad+ \nabla J_i^T(l_i)(\phi_i + \upsilon_i + p_i + \alpha_{ie}\dot{e}_i + B_i u_i - \ddot{x}_{id}),\end{aligned}\tag{25}$$

where $\nabla J_i(l_i) = \partial J_i(l_i)/\partial l_i$ represents the partial derivative of $J_i(l_i)$ with respect to l_i, and $\dot{l}_i$ denotes the time derivative of l_i. To ensure the existence of the optimal performance index function, here, we assume that the L_2-gain [29] of the robotic system (21) is less than or equal to γ_{ip}.

Define the utility function as:

$$U_i(l_i, u_i, p_i) = l_i^T Q_i l_i + u_i^T R_i u_i - \gamma_{ip}^2 p_i^T p_i, \tag{26}$$

then, according to [52], [53], the optimal performance index function can be given as:

$$\begin{aligned}J_i^*(l_i) &= \min_{u_i}\max_{p_i}\int_0^\infty \begin{pmatrix} l_i(e_i(\tau))^T Q_i l_i(e_i(\tau)) \\ +u_i(\tau)^T R_i u_i(\tau) \\ -\gamma_{ip}^2 p_i(x_i(\tau))^T p_i(x_i(\tau)) \end{pmatrix} d\tau\\ &= \min_{u_i}\max_{p_i}\int_0^\infty U_i(l_i, u_i, p_i)\, d\tau\\ &= \max_{p_i}\min_{u_i}\int_0^\infty U_i(l_i, u_i, p_i)\, d\tau.\end{aligned}\tag{27}$$

From (25) and (27), we conclude that the optimal control pair (u_i^*, p_i^*) satisfies the following HJI equation:

$$\begin{aligned}0 = {}&\nabla J_i^{*T}(l_i)\left(\phi_i + \upsilon_i + p_i^* + \alpha_{ie}\dot{e}_i + B_i u_i^* - \ddot{x}_{id}\right)\\ &+ U_i\left(s_i, u_i^*, p_i^*\right).\end{aligned}\tag{28}$$

According to the optimality principle, if $J_i^*(l_i)$ is the positive definite solution of the HJI equation (27), then one can represent the optimal control pair (u_i^*, p_i^*) that is given as follows:

$$\begin{aligned}u_i^* &= -\frac{1}{2}R_i^{-1}B_i^T\nabla J_i^*\\ p_i^* &= \frac{1}{2\gamma_{ip}^2}\nabla J_i^*.\end{aligned}\tag{29}$$

In order to deal with the subsystem dynamic term ϕ_i, model uncertainty term υ_i and the collision disturbance term p_i respectively, here, we can rewrite the optimal control law u_i^* as the form of

$$u_i^* = u_{i1} + u_{i2} + u_{i3}^*, \tag{30}$$

so that the HJI equation can be modified as:

$$0 = \nabla J_i^*\begin{pmatrix}\phi_i + \upsilon_i + p_i^* + \alpha_{ie}\dot{e}_i - \ddot{x}_{id}\\ +B_i u_{i1} + B_i u_{i2} + B_i u_{i3}^*\end{pmatrix} + U_i\left(l_i, u_i^*, p_i^*\right). \tag{31}$$

From (31), we conclude that the optimal control problem has been transformed into the one of obtaining decentralized compensation control laws u_{i1}, u_{i2} and decentralized zero-sum optimal control law u_{i3}^*, thus, to realize optimal compensation of model uncertainty and collision disturbance of MRM systems.

In (31), we know that the dynamic model term ϕ_i, the desired joint acceleration term $\ddot{x}_{id}$ and the joint velocity error term $\dot{e}_i$ are all known. In order to compensate these accurately modeled and estimated dynamic model terms, we can design the decentralized compensation control law u_{i1} as follows:

$$u_{i1} = -\begin{pmatrix}-\left(\hat{f}_{is}e^{\left(-\hat{f}_{i\tau}x_{i2}^2\right)}+\hat{f}_{ic}\right)\operatorname{sgn}(x_{i2}) - \frac{\tau_{iF}}{\gamma_i}\\ -\hat{f}_{ib}x_{i2} - B_i^{-1}\ddot{x}_{id} + B_i^{-1}\alpha_{ie}\dot{e}_i\end{pmatrix}. \tag{32}$$

Next, an ACI-structured-based decentralized zero-sum optimal control method is presented to deal with the optimal control problem of environmental collided MRM systems via NN implementation.

B. ACI-STRUCTURE-BASED DECENTRALIZED ZERO-SUM OPTIMAL CONTROL VIA NN IMPLEMENTATION

Neural networks, which possess of excellent capability for approximating unknown nonlinearities, are widely used in community of robotic control. In this part, we employ RBF-NNs to deal with the decentralized zero-sum optimal control problem via an actor-critic-identifier structure.

1) IDENTIFIER FOR MODEL UNCERTAINTY

In this part, we establish an RBF-NN identifier to approximate the model uncertainty. First, the following assumption should be considered:

Assumption 1 [54]: The NN approximation error $\varepsilon(\cdot)$ is upper-bounded and the NN activation function $\sigma(\cdot)$ as well as its derivative with respect of its arguments $\sigma'(\cdot)$ are also upper-bounded.

According to the Assumption 1, the model uncertainty term $\upsilon_i(x)$ in (22) can be approximated by employing an RBF-NN, which is given as:

$$\upsilon_i = w_{i\upsilon}^T \sigma_{i\upsilon}(x_\upsilon) + \varepsilon_{i\upsilon}(x_{i\upsilon}), \tag{33}$$

where $x_{i\upsilon}$ denotes a determined NN state; $x_\upsilon = [x_d, x_{i\upsilon}]^T = [x_{1d}, x_{2d} \cdots x_{md}, x_{i\upsilon}]^T$, $m < i$ is as defined state vector that is composed of the NN state as well as the known and bounded reference robotic system states; $w_{i\upsilon}$ is the unknown ideal NN weight; $\varepsilon_{i\upsilon}(x_{i\upsilon})$ represents the finite NN approximation error; $\sigma_{i\upsilon}(x_\upsilon)$ indicates the activation function that is selected as the following Gaussian function:

$$\sigma_{i\upsilon}(x_\upsilon) = \exp\left(\frac{-(x_\upsilon - \gamma_\upsilon)^T (x_\upsilon - \gamma_\upsilon)}{\beta_\upsilon}\right), \tag{34}$$

in which the constant γ_υ is the center of the activation function and the positive constant β_υ indicates the width of the activation function. Besides, on the basis of the representation of model uncertainly in (33), we consider the following nonlinear dynamic system, with a bounded control input $u_{i\upsilon}$, as follows:

$$\dot{x}_{i\upsilon} = \upsilon_i + u_{i\upsilon} = w_{i\upsilon}^T \sigma_{i\upsilon}(x_\upsilon) + \varepsilon_{i\upsilon}(x_{i\upsilon}) + u_{i\upsilon}. \tag{35}$$

Then, we can define the following NN identifier to approximate (35):

$$\dot{\hat{x}}_{i\upsilon} = \hat{\upsilon}_i + u_{i\upsilon} = \hat{w}_{i\upsilon}^T \sigma_{i\upsilon}(x_\upsilon) + u_{i\upsilon} + \delta_{i\upsilon}, \tag{36}$$

where $\hat{x}_{i\upsilon}$ indicates the identified state of $x_{i\upsilon}$; $\hat{\upsilon}_i$ denotes the estimation of υ_i; $\hat{w}_{i\upsilon}$ is the approximated weight and $\delta_{i\upsilon}$ denotes a robust integral of sign of the error (RISE [55]) feedback term, defined as:

$$\delta_{i\upsilon} = \alpha_{i\upsilon} e_{i\upsilon} + \Xi_{i\upsilon}, \tag{37}$$

where $e_{i\upsilon}$ is the state identification error that is represented as $e_{i\upsilon} = x_{i\upsilon} - \hat{x}_{i\upsilon}$, $\Xi_{i\upsilon}$ is the generalized solution of

$$\dot{\Xi}_{i\upsilon} = (\alpha_{i\upsilon}\beta_{i\upsilon} + \gamma_{i\upsilon}) e_{i\upsilon} + \eta_{i\upsilon 1} \operatorname{sgn}(e_{i\upsilon}), \tag{38}$$

in which the terms $\alpha_{i\upsilon}$, $\beta_{i\upsilon}$, $\gamma_{i\upsilon}$ and $\eta_{i\upsilon 1}$ are positive control parameters to be determined. By combining (35) with (36), the state identification error dynamic can be represented as follows:

$$\dot{e}_{i\upsilon} = \tilde{\upsilon}_i = w_{i\upsilon}^T \sigma_{i\upsilon} - \hat{w}_{i\upsilon}^T \sigma_{i\upsilon} + \varepsilon_{i\upsilon} - \delta_{i\upsilon}, \tag{39}$$

where $\tilde{\upsilon}_i = \upsilon_i - \hat{\upsilon}_i$. On this basis, define the following identification error function

$$l_{i\upsilon} = \dot{e}_{i\upsilon} + \beta_{i\upsilon} e_{i\upsilon}. \tag{40}$$

Then, one can easily obtain the time derivative of (40), which is given as:

$$\begin{aligned} \dot{l}_{i\upsilon} = {} & w_{i\upsilon}^T \sigma'_{i\upsilon} \dot{x}_{i\upsilon} - \dot{\hat{w}}_{i\upsilon}^T \sigma_{i\upsilon} - \hat{w}_{i\upsilon}^T \sigma'_{i\upsilon} \dot{\hat{x}}_{i\upsilon} + \dot{\varepsilon}_{i\upsilon} \\ & - \alpha_{i\upsilon} l_{i\upsilon} - \gamma_{i\upsilon} e_{i\upsilon} - \eta_{i\upsilon 1} \operatorname{sgn}(e_{i\upsilon}) + \beta_{i\upsilon} \dot{e}_{i\upsilon}, \end{aligned} \tag{41}$$

in which the weight update law of the NN is given as:

$$\dot{\hat{w}}_{i\upsilon} = proj\left(P_{i\upsilon} \sigma'_{i\upsilon} \dot{\hat{x}}_{i\upsilon} e_{i\upsilon}^T\right), \tag{42}$$

where $proj(\cdot)$ is a smooth projection operation [56], $P_{i\upsilon}$ indicates a positive constant gain matric. Rewriting (41), one obtains:

$$\begin{aligned} \dot{l}_{i\upsilon} = {} & \tilde{M}_{i\upsilon 1}\left(e_{i\upsilon}, w_{i\upsilon}, \hat{w}_{i\upsilon}\right) + M_{i\upsilon 2}\left(x_{i\upsilon}, w_{i\upsilon}, \hat{w}_{i\upsilon}\right) \\ & + \hat{M}_{i\upsilon 3}\left(\hat{x}_{i\upsilon}, !\tilde{w}_{i\upsilon}\right) - \alpha_{i\upsilon} l_{i\upsilon} - \gamma_{i\upsilon} e_{i\upsilon} - \eta_{i\upsilon 1} \operatorname{sgn}(e_{i\upsilon}), \end{aligned} \tag{43}$$

in which $\tilde{M}_{i\upsilon 1}$, $M_{i\upsilon 2}$ and $\hat{M}_{i\upsilon 3}$ are defined as follows:

$$\begin{aligned} \tilde{M}_{i\upsilon 1} &= \beta_{i\upsilon} \dot{e}_{i\upsilon} - \dot{\hat{w}}_{i\upsilon}^T \sigma_{i\upsilon} + \frac{1}{2} w_{i\upsilon}^T \sigma'_{i\upsilon} \dot{e}_{i\upsilon} + \frac{1}{2} \hat{w}_{i\upsilon}^T \sigma'_{i\upsilon} \dot{e}_{i\upsilon} \\ M_{i\upsilon 2} &= \frac{1}{2} w_{i\upsilon}^T \sigma'_{i\upsilon} \dot{x}_{i\upsilon} - \frac{1}{2} \hat{w}_{i\upsilon}^T \sigma'_{i\upsilon} \dot{x}_{i\upsilon} + \dot{\varepsilon}_{i\upsilon} \\ \hat{M}_{i\upsilon 3} &= \frac{1}{2} \tilde{w}_{i\upsilon}^T \sigma'_{i\upsilon} \dot{\hat{x}}_{i\upsilon}, \end{aligned} \tag{44}$$

where $\tilde{w}_{i\upsilon} = w_{i\upsilon} - \hat{w}_{i\upsilon}$. Then, define an auxiliary $M_{i\upsilon 3}(x_{i\upsilon}, \tilde{w}_{i\upsilon})$ that replace $\hat{x}_{i\upsilon}$ in $\hat{M}_{i\upsilon 3}$ with $\dot{x}_{i\upsilon}$, moreover, we can also define the relations $\tilde{M}_{i\upsilon 3} = \hat{M}_{i\upsilon 3} - M_{i\upsilon 3}$ and $M_{i\upsilon 23} = M_{i\upsilon 2} + M_{i\upsilon 3}$ to facilitate the subsequent stability analysis. Therefore, according to the Assumption 1 and using (40), (42) and (44), we may have the following bounds:

$$\begin{aligned} \left\|\tilde{M}_{i\upsilon 1}\right\| &\le \varphi_{i\upsilon}(\|b_{i\upsilon}\|)\|b_{i\upsilon}\|, \\ \|M_{i\upsilon 2}\| &\le m_{i\upsilon 1}, \|M_{i\upsilon 3}\| \le m_{i\upsilon 2}, \\ \|M_{i\upsilon 23}\| &\le m_{i\upsilon 3} + m_{i\upsilon 4}\varphi_{i\upsilon 2}(\|b_{i\upsilon}\|)\|b_{i\upsilon}\|, \\ \left\|\dot{e}_{i\upsilon}^T \tilde{M}_{i\upsilon 3}\right\| &\le m_{i\upsilon 5}\|e_{i\upsilon}\|^2 + m_{i6}\|l_{i\upsilon}\|^2, \end{aligned} \tag{45}$$

where $b_{i\upsilon} = \left[e_{i\upsilon}^T \; l_{i\upsilon}^T\right]^T$, $\varphi_{i\upsilon 1}(\cdot)$, $\varphi_{i\upsilon 2}(\cdot)$ are positive, non-decreasing, globally invertible functions, and $m_{i\upsilon f}, f = 1, \ldots, 6$, are positive constants. In order to facilitate the subsequent stability analysis, we define

$$T_{i\upsilon} = \left[e_{i\upsilon}^T \quad l_{i\upsilon}^T \quad \sqrt{Q_{i\upsilon}} \quad \sqrt{R_{i\upsilon}}\right]^T, \tag{46}$$

in which the auxiliary function $Q_{i\upsilon}$ is the generalized solution of the following differential equation:

$$\dot{Q}_{i\upsilon} = -l_{i\upsilon}^T \left(M_{i\upsilon 2} - \eta_{i\upsilon 1}\,\mathrm{sgn}(e_{i\upsilon})\right) + \dot{e}_{i\upsilon}^T M_{i\upsilon 3} - \eta_{i\upsilon 2}\varphi_{i\upsilon 2}\left(\|b_{i\upsilon}\|\right)\|b_{i\upsilon}\|\,\|e_{i\upsilon}\|, \quad (47)$$

where $\eta_{i\upsilon 1}$, $\eta_{i\upsilon 2}$ can be selected according to the following conditions to ensure that $Q_{i\upsilon}$ is nonnegative definite:

$$\varphi_{i\upsilon 1} > \max\left(m_{i\upsilon 1} + \frac{m_{i\upsilon 3}}{\beta_{i\upsilon}}, m_{i\upsilon 1} + m_{i\upsilon 2}\right), \quad \varphi_{i\upsilon 2} > \varphi_{i\upsilon 4}. \quad (48)$$

Moreover, the auxiliary function $R_{i\upsilon}$ in (46) is defined as:

$$R_{i\upsilon} = \frac{1}{4}\beta_{i\upsilon}\left[tr\left(\tilde{w}_{i\upsilon}^T P_{i\upsilon}^{-1}\tilde{w}_{i\upsilon}\right)\right], \quad (49)$$

where $tr(\cdot)$ denotes the trace of matrix.

Theorem 1: Consider the global model uncertainty of an MRM system that is represented as (33), and the dynamic system formulated in (35). The RBF-NN identifier presented in (36) along with the weight update law given in (42) can ensure the model uncertainty term to be asymptotically identified, in the sense that

$$\lim_{t\to\infty}\|\tilde{\upsilon}_i\| = 0,$$

afforded the parameters

$$\alpha_{i\upsilon} > m_{i\upsilon 6}, \gamma_{i\upsilon} > \frac{m_{i\upsilon 5}}{\beta_{i\upsilon}},$$

in which $\alpha_{i\upsilon}$, $\beta_{i\upsilon}$, $\gamma_{i\upsilon}$, $m_{i\upsilon 5}$ and $m_{i\upsilon 6}$ are introduced in (38) and (45) respectively.

Proof: See Appendix A.

Then, according to (31), (36), (37) and (38), one can design the model uncertainty identifier-based learning control law u_{i2}, which is given as:

$$u_{i2} = -B_i^{-1}\left(\begin{array}{c}\hat{w}_{i\upsilon}^T\sigma_{i\upsilon} + \alpha_{i\upsilon}e_{i\upsilon} \\ + \int_0^t \left(\begin{array}{c}(\alpha_{i\upsilon}\beta_{i\upsilon} + \gamma_{i\upsilon})e_{i\upsilon} \\ +\eta_{i\upsilon 1}\,\mathrm{sgn}(e_{i\upsilon})\end{array}\right)dt\end{array}\right), \quad (50)$$

where the NN weight $\hat{w}_{i\upsilon}$ is updated by (42).

From (32) and (50), we know that the control law u_{i1} is developed to address the effect of modeled and estimated part of the dynamic model, as well as u_{i2} is designed to deal with the model uncertainty of the MRM system. Next, we focus on finding a zero-sum neuro-optimal control u_{i3}^* to realize the optimal compensation of the effect of disturbance that is caused by environmental collisions.

2) CRITIC AND ACTOR NNS

In this part, we employ the critic NN, $u-$actor NN and $p-$actor NN to estimate the performance index function $J_i(l_i)$, the zero-sum optimal control law u_{i3}^* and the worst environmental collision disturbance p_i^* respectively.

a: CRITIC NN

By using RBF NNs, the ideal critic NN can be given as

$$J_i(l_i) = w_{ic}^T\sigma_{ic}(l_i) + \varepsilon_{ic}, \quad (51)$$

where w_{ic} denotes the unknown ideal NN weight, ε_{ic} is the finite estimation error of the critic NN, $\sigma_{ic}(s_i)$ indicates the activation function that is similar to the form of (34). Then, the gradient of the estimated performance index function is given as:

$$\nabla J_i(l_i) = \nabla\sigma_{ic}(l_i)^T w_{ic} + \nabla\varepsilon_{ic}^T, \quad (52)$$

where $\nabla\sigma_{ic}(l_i)$ is the activation function gradient that is given as $\nabla\sigma_{ic}(l_i) = \partial\sigma_{ic}(l_i)/\partial l_i$ and $\nabla\varepsilon_{ic}$ is the gradient error. Then, we can rewrite the Hamiltonian in (25) and obtain the following relation:

$$H_i(l_i, u_i, p_i, w_{ic}) = \begin{array}{c}U_i(l_i, u_i, p_i) \\ +\left(w_{ic}^T\nabla\sigma_{ic}(l_i)\right)\dot{l}_i - e_{icH}\end{array} = 0, \quad (53)$$

where e_{icH}, which is the estimation error of the critic NN, can be represented by:

$$e_{icH} = U_i + \left(w_{ic}^T\nabla\sigma_{ic}\right)\dot{l}_i. \quad (54)$$

Let $\hat{w}_{ic}$ be the approximated weight vector of w_{ic}, so that the actual output of the critic NN can be given as:

$$\hat{J}_i(l_i) = \hat{w}_{ic}^T\sigma_{ic}(l_i). \quad (55)$$

Then, one obtains the approximated Hamilton function that is given as:

$$\hat{H}_i\left(l_i, u_i, p_i, \hat{w}_{ic}\right) = U_i(l_i, u_i, p_i) + \left(\hat{w}_{ic}^T\nabla\sigma_{ic}(l_i)\right)\dot{l}_i. \quad (56)$$

Define the error function for adjusting the critic NN as $e_{ic} = \hat{H}_i - H_i$, which is in fact $e_{ic} = \hat{H}_i$. Let the weight estimation error to be $\tilde{w}_{ic} = w_{ic} - \hat{w}_{ic}$, then, combining (53) with (56), we obtain that $e_{ic} = e_{icH} - \tilde{w}_{ic}^T\nabla\sigma_{ic}(l_i)\dot{l}_i$. Then, based on the gradient descent method, define the residual error function $E_{ic} = \frac{1}{2}e_{ic}^2$ that is minimized to adjust the critic NN (55), which is updated by

$$\dot{\hat{w}}_{ic} = -\alpha_{ic}\left(\frac{\partial E_{ic}}{\partial\hat{w}_{ic}}\right) = -\alpha_{ic}\frac{p_{ic}\left(p_{ic}^T\hat{w}_{ic} + U_i\right)}{\left(p_{ic}^T p_{ic} + 1\right)^2}, \quad (57)$$

where p_{ic} is defined by $p_{ic} = \nabla\sigma_{ic}(l_i)\dot{l}_i$ and α_{ic} is a positive learning rate of the critic NN. Moreover, define $m_{ic} = p_{ic}^T p_{ic} + 1$ and $n_{ic} = \frac{p_{ic}}{m_{ic}}$, so that the critic NN error dynamic can be given as:

$$\dot{\tilde{w}}_{ic} = -\alpha_{ic}n_{ic}n_{ic}^T\tilde{w}_{ic} + \alpha_{ic}n_{ic}\frac{e_{icH}}{m_{ic}}. \quad (58)$$

b: ACTOR NNs

Here, we employ the actor NNs that include u-actor NN and p-actor NN to approximate the optimal control law u_{i3}^* and the worst environmental collision disturbance. Consider that the ideal $u-$ and $p-$ actor NNs are given as:

$$u_{i3}^* = w_{ia}^T\sigma_{ia}(l_i) + \varepsilon_{ia}, \quad (59)$$

$$p_i^* = w_{ip}^T\sigma_{ip}(l_i) + \varepsilon_{ip}, \quad (60)$$

where w_{ia}, w_{ip} are the ideal weight vectors; σ_{ia}, σ_{ip}, which are selected as the similar form of (34), represent the activation functions and $\varepsilon_{ia}, \varepsilon_{ip}$ denote the finite approximation errors. Let $\hat{w}_{ia}$ and $\hat{w}_{ip}$ be the approximated weight vectors of w_{ia} and w_{ip}, then, the actual output of the $u-$ and $p-$actor NNs are given as:

$$\hat{u}_{i3} = \hat{w}_{ia}^T \sigma_{ia}(l_i), \tag{61}$$

$$\hat{p}_i = \hat{w}_{ip}^T \sigma_{ip}(l_i). \tag{62}$$

Substituting (61) and (62) into (56), one obtains the relations:

$$\left(\partial \hat{H}_i\left(l_i, \hat{u}_i, \hat{p}_i, \hat{w}_{ic}\right)/\partial \hat{u}_i\right) = \partial \hat{H}_i\left(l_i, \hat{u}_i, \hat{p}_i, \hat{w}_{ic}\right)/\partial \hat{p}_i = 0. \tag{63}$$

Therefore, we can rewrite the approximated optimal control and the worst environmental collision disturbance as follows:

$$\hat{u}_{i3} = -\frac{1}{2} R_i^{-1} B_i^T \nabla \sigma_{ic}^T(l_i) \hat{w}_{ic}, \tag{64}$$

$$\hat{p}_i = \frac{1}{2\gamma_{ip}^2} \nabla \sigma_{ic}^T(l_i) \hat{w}_{ic}. \tag{65}$$

Then, the approximation error of the $u-$ and $p-$ actor NNs can be expressed as:

$$\begin{aligned} e_{ia} &= \hat{w}_{ia}^T \sigma_{ia}(l_i) + \frac{1}{2} R_i^{-1} B_i^T \nabla \sigma_{ic}^T(l_i) \hat{w}_{ic} \\ e_{ip} &= \hat{w}_{ip}^T \sigma_{ip}(l_i) - \frac{1}{2\gamma_{ip}^2} \nabla \sigma_{ic}^T(l_i) \hat{w}_{ic}. \end{aligned} \tag{66}$$

We know that the objective of the $u-$ and $p-$ actor NNs are to select the weight estimation $\hat{w}_{ia}$ and $\hat{w}_{ip}$ that minimize the residual error functions $E_{ia} = \frac{1}{2} e_{ia}^2$ and $E_{ip} = \frac{1}{2} e_{ip}^2$. Then, based on the gradient descent rule, the NN weight update laws can be given as follows:

$$\dot{\hat{w}}_{ia} = -\alpha_{ia} \sigma_{ia}(l_i) \cdot \begin{pmatrix} \hat{w}_{ia}^T \sigma_{ia}(l_i) \\ +\frac{1}{2} R_i^{-1} B_i^T \nabla \sigma_{ic}^T(l_i) \hat{w}_{ic} \end{pmatrix}^T, \tag{67}$$

$$\dot{\hat{w}}_{ip} = -\alpha_{ip} \sigma_{ip}(l_i) \cdot \begin{pmatrix} \hat{w}_{ip}^T \sigma_{ip}(l_i) \\ -\frac{1}{2\gamma_{ip}^2} \nabla \sigma_{ic}^T(l_i) \hat{w}_{ic} \end{pmatrix}^T, \tag{68}$$

where α_{ia} and α_{ip} are the positive learning rates to be determined. Define the following weight estimation errors:

$$\begin{aligned} \tilde{w}_{ia} &= w_{ia} - \hat{w}_{ia} \\ \tilde{w}_{ip} &= w_{ip} - \hat{w}_{ip}. \end{aligned} \tag{69}$$

Then, one can define the error dynamics of the $u-$ and $p-$ actor NNs, which are given as follows:

$$\dot{\tilde{w}}_{ia} = \alpha_{ia} \sigma_{ia}(l_i) \cdot \begin{pmatrix} -\tilde{w}_{ia}^T \sigma_{ia}(l_i) + w_{ia}^T \sigma_{ia}(l_i) \\ -\frac{1}{2} R_i^{-1} B_i^T \nabla \sigma_{ic}^T(l_i) \tilde{w}_{ic} \\ +\frac{1}{2} R_i^{-1} B_i^T \nabla \sigma_{ic}^T(l_i) w_{ic} \end{pmatrix}^T$$

$$\dot{\tilde{w}}_{ip} = \alpha_{ip} \sigma_{ip}(l_i) \cdot \begin{pmatrix} -\tilde{w}_{ip}^T \sigma_{ip}(l_i) + w_{ip}^T \sigma_{ip}(l_i) \\ +\frac{1}{2\gamma_{ip}^2} \nabla \sigma_{ic}^T(l_i) \tilde{w}_{ic} \\ -\frac{1}{2\gamma_{ip}^2} \nabla \sigma_{ic}^T(l_i) w_{ic} \end{pmatrix}^T. \tag{70}$$

Remark 2: According to the Hamiltonian (25) and the HJI equation (31), we know the fact that $H_i\left(l_i, u_i^*, p_i^*, \nabla J_i^*\right) = 0$. Moreover, from (52), (61) and (62), we also conclude that ∇J_i^*, u_i^* and p_i^* can be rewritten with respect to the ideal weights of critic and actor NNs. In this sense, the Hamiltonian (53) can be further expressed as the relation of: $H_i(l_i, u_i, p_i, w_{ic}) = H_i\left(l_i, w_{ia}, w_{ip}, w_{ic}\right) = 0$.

Theorem 2: Let the performance index function and the optimal control pair be approximated by the critic NN (51) and the actor NNs (59), (60) with the ideal weight w_{ic}, w_{ia} and w_{ip}, respectively. If the estimated performance index function and optimal control pair are expressed by (55), (61) and (62) that are built with approximated weight $\hat{w}_{ic}$, $\hat{w}_{ia}$ and $\hat{w}_{ip}$ respectively, as well as the NN weights are updated by (57), (67) and (68) respectively, then, the weight approximation errors $\tilde{w}_{ic}$, $\tilde{w}_{ia}$ and $\tilde{w}_{ip}$ are UUB.

Proof: See Appendix B.

By combining (32), (50) with (64), we can formulated the completed decentralized zero-sum neuro-optimal control law u_i^* that is given as:

$$\begin{aligned} u_i^* &= u_{i1} + u_{i2} + u_{i3}^* \\ &= -\begin{pmatrix} -\left(\hat{f}_{is} e^{\left(-\hat{f}_{i\tau} x_{i2}^2\right)} + \hat{f}_{ic}\right) \operatorname{sgn}(x_{i2}) \\ -\hat{f}_{ib} x_{i2} - B_i^{-1} \ddot{x}_{id} - \frac{\tau_{iF}}{\gamma_i} + B_i^{-1} \alpha_{ie} \dot{e}_i \end{pmatrix} \\ &\quad - B_i^{-1} \begin{pmatrix} \hat{w}_{iv}^T \sigma_{iv} + \alpha_{iv} e_{iv} \\ + \int_0^t \begin{pmatrix} (\alpha_{iv} \beta_{iv} + \gamma_{iv}) e_{iv} \\ + \eta_{iv1} \operatorname{sgn}(e_{iv}) \, dt \end{pmatrix} \end{pmatrix} \\ &\quad - \frac{1}{2} R_i^{-1} B_i^T \nabla \sigma_{ic}^T \hat{w}_{ic}. \end{aligned} \tag{71}$$

C. STABILITY ANALYSIS OF THE CLOSED-LOOP ROBOTIC SYSTEM

In this part, we focus on investigating the stability problems of the closed-loop MRM systems under the proposed decentralized control scheme (71). The theorem is given as follows:

Theorem 3: Consider a modular robot manipulator system that subjects to environmental collisions, with the subsystem dynamic model formulated in (15), the model uncertainties and the collision disturbance existed in (22). The closed-loop robotic system is asymptotically stable under the proposed decentralized zero-sum neuro-optimal control given in (71).

Proof: See Appendix C.

IV. EXPERIMENTS

Based on the established experimental platform and the collected experimental results, in this section, we present the experimental verifications to analysis the advantages and effectiveness of the proposed method.

A. EXPERIMENTAL SETUP

As illustrated in Figure 3, the details of the experimental setup are given as follows: An MRM is consist of two modular robotic joints, each joint composing a DC motor,

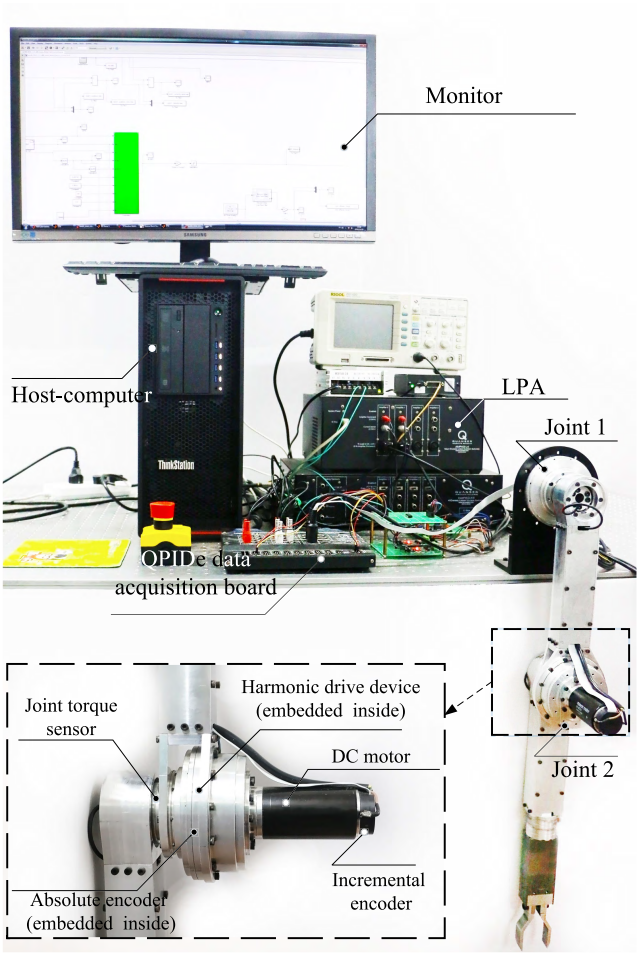


FIGURE 3. Experimental setup.

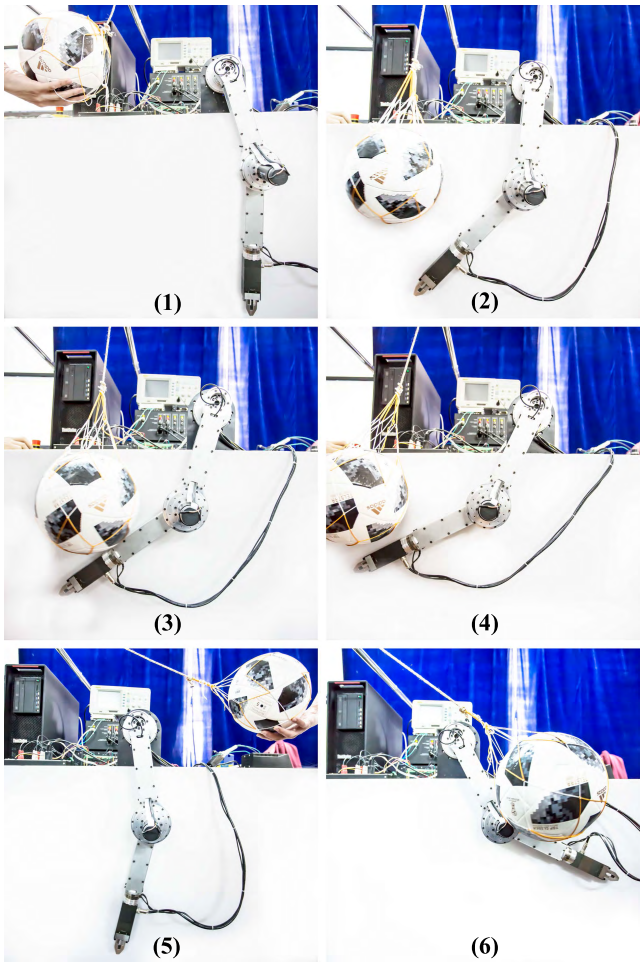


FIGURE 4. Experiments of the MRM system with environmental collisions, (1)–(6) Collision example.

an incremental encoder, an absolute encoder, a harmonic drive device and a joint torque sensor. The DC motors, with type no. 218014 from Maxon Inc., possess the nominal torque 0.2 Nm and nominal voltage 48V. The harmonic drive devices, which are with the reduction ratio of 100:1, are coupled with the DC motor. The motor-side position variables are measured by the 500-line incremental encoder from Maxon Inc. and the link-side position variables are obtained by using the absolute encoder with 19-bit resolution from Netzer Inc.. The joint torque sensors, which are with nominal torque 20 Nm are equipped at the end of each joint module, are used to verify the joint torque estimation performance. A linear power amplifier (LPA) and a QPIDe data acquisition board from Quanser Inc. are adopted to drive the motor and to collect the experimental data. The LPA is seamless embedded into Simulink from MathWorks Inc. by using QUARC from Quanser Inc., and these softwares are installed in a host-computer, which can communicate with the QPIDe data acquisition board to process the experimental data in Simulink.

Remark 3: It is noted that the proposed control algorithm, which is in the form of continuous time, needs to be realized discretely, when it is implemented in experiments. Fortunately, the control system, which is constructed under the Simulink environment, may complete the discrete realization automatically and adjust the sampling period adaptively.

We consider that the environmental collisions occur at random place of the MRM links. As illustrated in Figure 4, the collisions between the collision objects and the MRM links may create instantaneous contact forces. Note that the collision external torques are estimated and calculated in advance to avoid exceeding the instantaneous maximum range of the joint torque sensors. Moreover, to match up the nominal values of the DC motors and the torque sensors, the robotic joints follow the desired trajectories of

$$q_{1d} = \frac{-30\pi}{180}\sin\left(\frac{\pi}{20}\right)t, \quad q_{2d} = \frac{45\pi}{180}\sin\left(\frac{\pi}{20}\right)t,$$

in which $0 \leq t \leq 80\ s$.

The MRM system parameters, which include the model parameters, the uncertainty up-bound parameters and the control parameters, are represented in Table 1. Besides, the activation function for the identifier is selected as a symmetric sigmoid function, the critic NN is chosen as the structure of 2-5-1 with 2 input neurons, 5 hidden neurons and 1 output neuron, and the weight vectors are selected as

TABLE 1. Parameters setting.

Parameter type	Name	Value	Name	Value	Name	Value
Model parameters	$\hat{f}_{ib}$	$12mNms/rad$	$\hat{f}_{ic}$	$30mNm$	$\hat{f}_{is}$	$40mNm$
	$\hat{f}_{i\tau}$	$20s^2/rad^2$	γ_i	100	I_{im}	$120gcm^2$
	c_{iF}	$83.5Nm^{-1}$	c_{iW}	$89mNm^{-1}$		
	e_{iFl0}	$1.33Nm/rad$	e_{iWl0}	8.3e+3Nm/rad		
Uncertainty up-bound parameters	ρ_{Fi1}	$30mNms/rad$	ρ_{Fi2}	$60mNm$	ρ_{Fi3}	$80mNm$
	ρ_{Fi4}	$50s^2/rad^2$	ρ_{die}	$0.2Nm$	ρ_{dic}	$0.5Nm$
	ρ_{iV}	2.3	ρ_{iU}	2.6	ρ_{fip}	0.3
Control parameters	α_{ie}	0.5	α_{ic}	0.85	α_{ip}	0.85
	α_{ia}	0.85	γ_{ip}	1.9	α_{iv}	800
	β_{iv}	350	γ_{iv}	5	η_{iv1}	$0.1I$

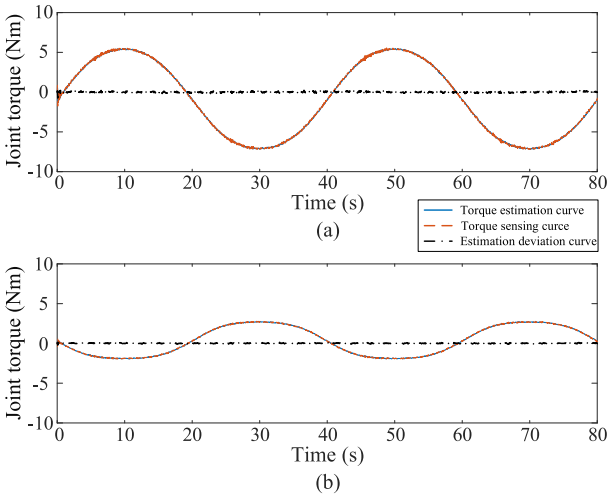


FIGURE 5. Estimated joint torque without environmental collisions, (a) Joint 1, (b) Joint 2.

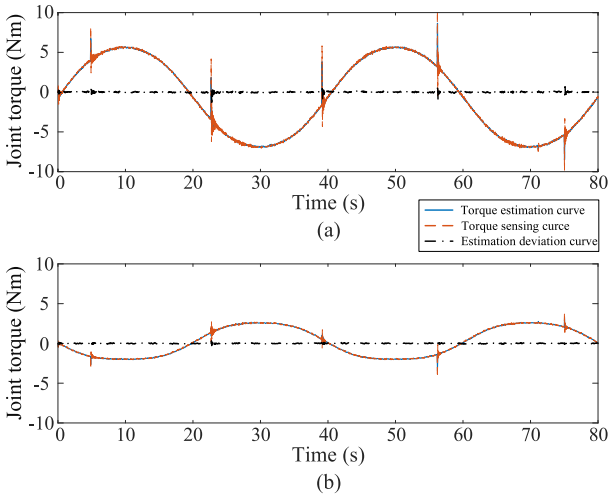


FIGURE 6. Estimated joint torque with environmental collisions, (a) Joint 1, (b) Joint 2.

$\hat{w}_{c1} = \left[\hat{w}_{c11}\ \hat{w}_{c12}\ \hat{w}_{c13}\ \hat{w}_{c14}\ \hat{w}_{c15}\right]^T$ for joint 1 and $\hat{w}_{c2} = \left[\hat{w}_{c21}\ \hat{w}_{c22}\ \hat{w}_{c23}\ \hat{w}_{c24}\ \hat{w}_{c25}\right]^T$ for joint 2, with the initial values $\hat{w}_{c1} = \hat{w}_{c2} = [0]$. Moreover, the NN structures and initial weight values of the actor NNs are selected as the same of the critic NNs.

B. EXPERIMENTAL RESULTS

In this part, the experimental results are proposed to analyze the variations of joint torque estimation performance, trajectory tracking performance, position and velocity tracking errors, control torques and the network weights. In order to compare the advantages between the existing methods, e.g. [57]–[59], and the proposed method, in this paper, two different control schemes are considered in the experiments, which include the existing learning-based decentralized control method and the proposed decentralized zero-sum neuro-optimal control method. The performance comparisons with each control method are listed in Table 2.

1) JOINT TORQUE ESTIMATION

Figures 5 and 6 illustrate the joint torque curves under the situation of collision free and uncertain collision environments by using the torque sensing technique and the proposed harmonic drive model-based joint torque estimation method, respectively. In Figure 5, the joint torques are obtained under the collision free environment and we observe that the torque estimations are highly consistent with the torque sensing measurements as well as the torque estimation deviation keeps in a very small range. Besides, Figure 5 also verifies that the torque estimation method is suitable for multiple joint trajectories. Figure 6 illustrates the joint torque curves under the situation of uncertain collision environment. From this figure, one observes that there exists significant jump on the torque curves while the environmental collision occurred. Fortunately, the proposed joint torque estimation method, which is rely on the position and current measurements that are sensitive to the collisions, may effectively capture the

TABLE 2. Performance comparisons under the existing and the proposed control methods.

Control method selection	Joint No.	Mean absolute value of position error	Mean absolute value of velocity error	Mean absolute value of control torque consumption
The existing method	Joint 1	$1.32\text{e}-3/rad$	$3.95\text{e}-4/rad/s$	$95.65/mNm$
	Joint 2	$1.64\text{e}-3/rad$	$6.86\text{e}-4/rad/s$	$64.72/mNm$
The proposed method	Joint 1	$0.74\text{e}-3/rad$	$2.98\text{e}-4/rad/s$	$85.27/mNm$
	Joint 2	$0.80\text{e}-3/rad$	$4.20\text{e}-4/rad/s$	$57.36/mNm$

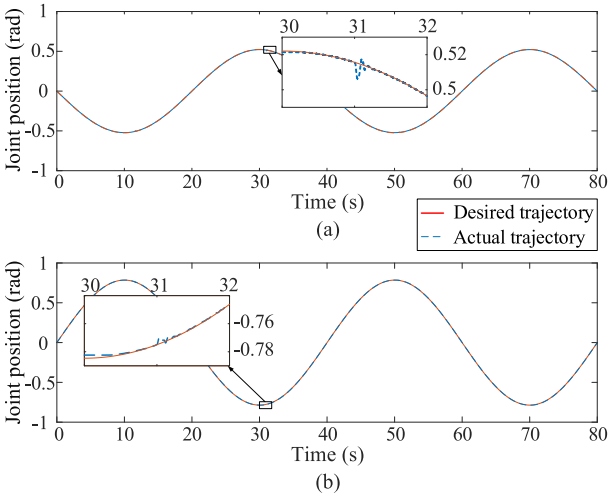


FIGURE 7. Trajectory tracking curves by using the existing methods, (a) Joint 1, (b) Joint 2.

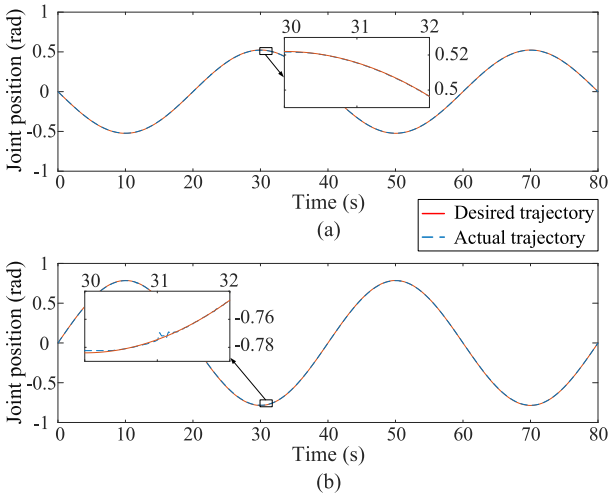


FIGURE 8. Trajectory tracking curves by using the proposed method, (a) Joint 1, (b) Joint 2.

instantaneous changing trend of the joint torque, therefore, torque estimation deviation still keeps in an acceptable range when the collision occurred. Note that the proposed joint torque estimation method is independent of the selected control methods.

2) TRAJECTORY TRACKING PERFORMANCE

Figures 7 and 8 illustrate the trajectory tracking curves of the MRM system under the existing and the proposed control methods with uncertain environmental collisions respectively. In these figures, the desired trajectories are tracked appropriately by the actual joint trajectories, which means the trajectory tracking tasks are completed successfully under both of the two control methods. Besides, the initial joint positions of the MRM system are equated to the initial locations of the trajectory tracking tasks, that is attribute to the effect of the initialization localization programs.

3) POSITION AND VELOCITY TRACKING ERRORS

The position and velocity tracking error curves under the situation of environmental collisions are shown in Figures 9–12. In Figures 9 and 10, the steady state position and velocity errors are less than 4e-3 rad and 2e-3 rad/s respectively for each joint module under the existing control methods, and the errors keep in an acceptable range.

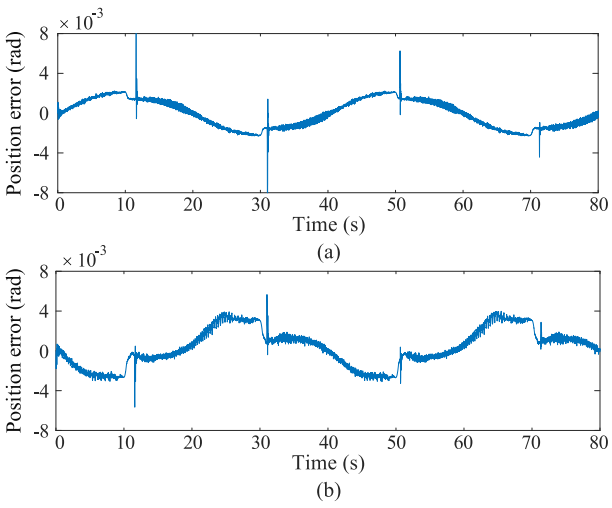


FIGURE 9. Position error curves by using the existing methods, (a) Joint 1, (b) Joint 2.

Besides, obvious instantaneous increasing of the position and velocity errors are captured in these figures, which are due to the effect of collision force. However, since the dynamic identification and optimal compensation of the model uncertainties and environmental collision disturbance

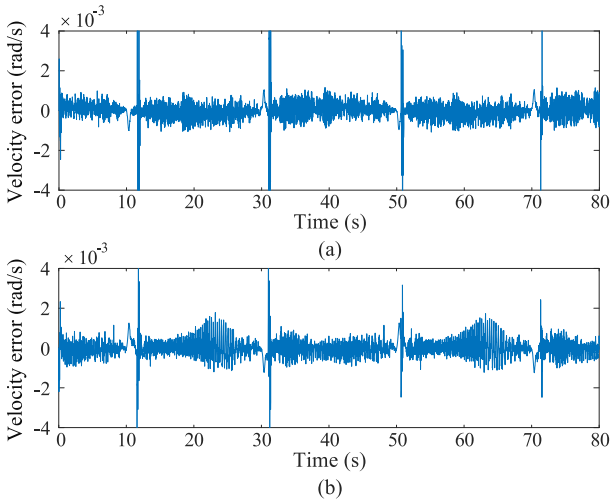


FIGURE 10. Velocity error curves by using the existing methods, (a) Joint 1, (b) Joint 2.

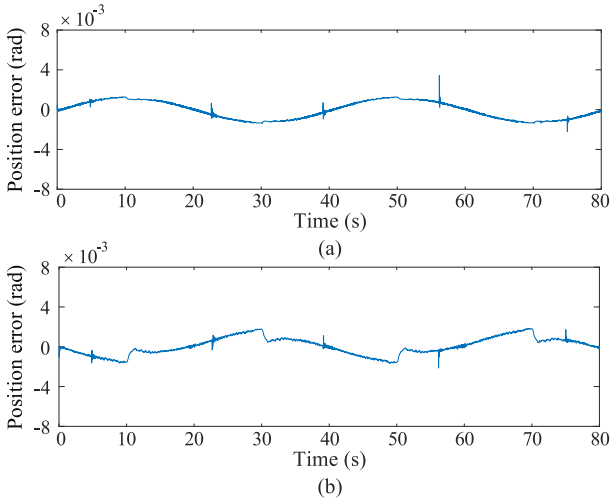


FIGURE 11. Position error curves by using the proposed method, (a) Joint 1, (b) Joint 2.

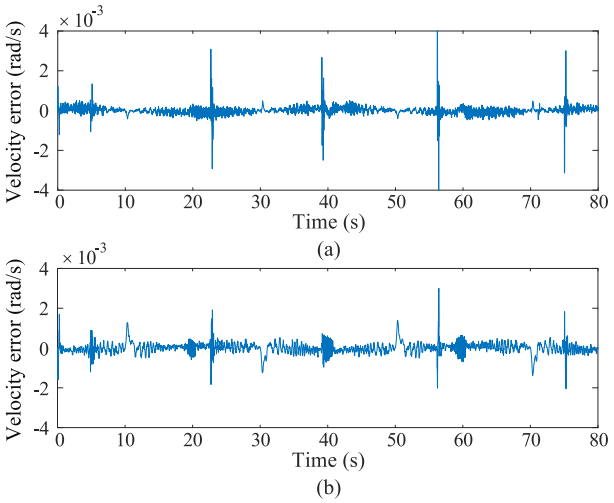


FIGURE 12. Velocity error curves by using the proposed method, (a) Joint 1, (b) Joint 2.

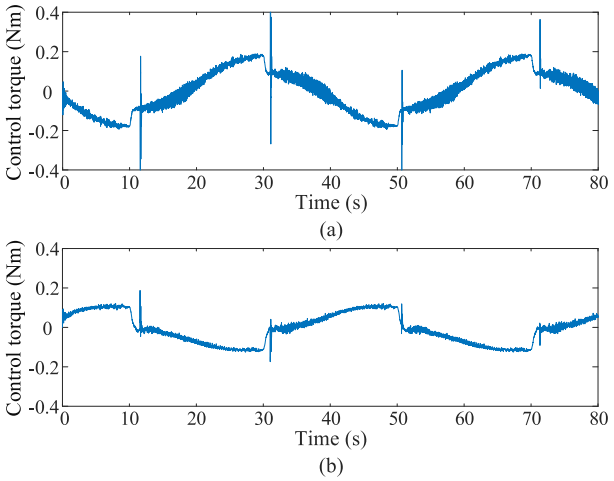


FIGURE 13. Control torque curves by using the existing methods, (a) Joint 1, (b) Joint 2.

are not implemented in the existing control methods, the error curves are featured with obvious chattering and noise effects. In Figures 11 and 12, the steady state position and velocity errors are less than 2e-3 rad and 1e-3 rad/s respectively, which are with better performance than the former ones, and the main reason of this phenomenon is that the proposed control method has compensated the effects of frictional modeling error, IDCs and environmental collision disturbance targeted and effectively. Moreover, we also observe that the instantaneous increasing of position and velocity errors, which are caused by the environmental collisions, return to normal ranges within very short time periods, which is attributed to the performance of the proposed zero-sum neuro-optimal controller that realizes the optimal compensation of uncertain collision disturbance.

4) CONTROL TORQUE

In Figure 13, the control torque curves, which is plotted under the existing control methods, are with serious chattering effect that may decrease the precision of joint trajectory tracking and affect the durability of DC motors. Moreover, we observe that the control torques may sharply increase when the collision contact occurred instantly, note that this may trigger the circuit break and protect mechanism of the power amplifier and thus making the robotic system out of control. Smooth and healthy control torque curves are represented in Figure 14 that profits from the proposed control method, in which the output torques have been optimized with an appropriate behavior in accordance with the output power of motors. Besides, we also observe that the instant increase of control torques are kept within safe limits while the collision occurred. This is attribute to the proposed optimal control that realizes the optimization of tracking errors and output torques.

5) CRITIC NN WEIGHT ESTIMATIONS

Figure 15 illustrates the variations of the estimated critic NN weights under the proposed method with environmental

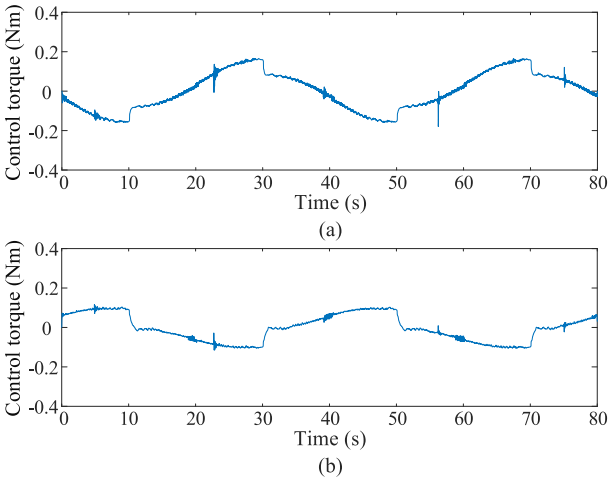


FIGURE 14. Control torque curves by using the proposed method, (a) Joint 1, (b) Joint 2.

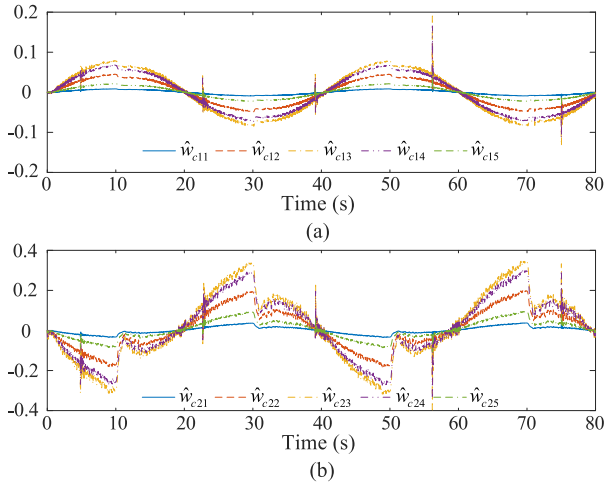


FIGURE 15. Estimated critic NN weight adjustment curves by using the proposed method, (a) Joint 1, (b) Joint 2.

collisions. From this figure, one observes that the weight estimation curves, which can effectively reflect the influence of the environmental collisions, are updated regularly within certain boundaries. Consequently, with the update of the weight estimations, the critic NN can learn the performance index function in real time, as well as the HJI equation and the optimal control law can be solved and derived respectively. Besides, the weight estimations of the u- and p-actor NNs are with the same RBF NN structures of the critic NNs.

From the experimental comparative cases, we conclude that the proposed control method is with better control performance in the aspect of motion control accuracy and power consumption than the existing ones. Moreover, we can also conclude that both position and velocity variables of the MRM systems are asymptotically stable under the proposed decentralized zero-sum neuro-optimal control method, and the control torques are smooth and healthy in the tasks with environmental collisions. All the experiments reported in this section are repeatable, and the experimental results are consistent.

V. CONCLUSION

In this paper, we propose a decentralized zero-sum neuro-optimal control scheme for MRMs with environmental collisions. Based on the collision identification method and the local dynamic information, we formulate the dynamic model of the MRM system and then a model-based compensation controller is designed. An NN identifier, which is established to approximate the model uncertainties, is adopted to develop the learning controller, then, we can transform the problem of optimal control for MRMs with environmental collisions into a two-player zero-sum optimal control one. Based on ADP algorithm, the HJI equation is solved by establishing the actor-critic NNs that include one critic NN and two actor NNs, and then the decentralized zero-sum neuro-optimal control is developed. Based on the Lyapunov theory, the closed-loop robotic system is proved to be asymptotically stable by the implementation of a set of proposed decentralized controllers. At last, experimental results are illustrated to verify the effectiveness and advantages of the proposed method.

APPENDIX A
PROOF OF THEOREM 1

We define a Lyapunov function $V_{i\upsilon}(T_{i\upsilon}) : \Gamma_{i\upsilon} \to K_{i\upsilon}$ as:

$$V_{i\upsilon}(T_{i\upsilon}) = \frac{1}{2} l_{i\upsilon}^T l_{i\upsilon} + \frac{1}{2} \gamma_{i\upsilon} e_{i\upsilon}^T e_{i\upsilon} + Q_{i\upsilon} + R_{i\upsilon}, \quad \text{(A1)}$$

which satisfies the following relations:

$$U_1(T_{i\upsilon}) \le V_{i\upsilon}(T_{i\upsilon}) \le U_2(T_{i\upsilon}), \quad \text{(A2)}$$

where $U_1(T_{i\upsilon})$, $U_2(T_{i\upsilon})$ represent continuous positive definite functions:

$$\begin{aligned} U_1(T_{i\upsilon}) &= \frac{1}{2} \min(1, \gamma_{i\upsilon}) \|T_{i\upsilon}\|^2, \\ U_2(T_{i\upsilon}) &= \max(1, \gamma_{i\upsilon}) \|T_{i\upsilon}\|^2. \end{aligned} \quad \text{(A3)}$$

Under Filippov' s framework, we obtain the time derivative of (A1) as:

$$\begin{aligned} \dot{V}_{i\upsilon} < {} & l_{i\upsilon}^T \begin{pmatrix} \tilde{M}_{i\upsilon 1} + M_{i\upsilon 2} + \hat{M}_{i\upsilon 3} - \alpha_{i\upsilon} l_{i\upsilon} \\ -\eta_{i\upsilon 1} \operatorname{sgn}(e_{i\upsilon}) - \gamma_{i\upsilon} e_{i\upsilon} \end{pmatrix} \\ & + \gamma_{i\upsilon} e_{i\upsilon}^T (l_{i\upsilon} - \beta_{i\upsilon} e_{i\upsilon}) \\ & - l_{i\upsilon}^T (M_{i\upsilon 2} - \eta_{i\upsilon 1} \operatorname{sgn}(e_{i\upsilon})) - \dot{e}_{i\upsilon}^T M_{i\upsilon 3} \\ & + \eta_{i\upsilon 2} \varphi_{i\upsilon 2}(\|b_{i\upsilon}\|) \|b_{i\upsilon}\| \|e_{i\upsilon}\| \\ & - \frac{1}{2} \beta_{i\upsilon} \left[tr\left(\tilde{w}_{i\upsilon}^T P_{i\upsilon}^{-1} \dot{\hat{w}}_{i\upsilon} \right) \right]. \end{aligned} \quad \text{(A4)}$$

Decomposing for $\alpha_{i\upsilon} = \alpha_{i\upsilon 1} + \alpha_{i\upsilon 2}$ and $\gamma_{i\upsilon} = \gamma_{i\upsilon 1} + \gamma_{i\upsilon 2}$, using (44) and completing the squares, we obtain the following upper bound:

$$\begin{aligned} \dot{V}_{i\upsilon} \le & -(\beta_{i\upsilon}\gamma_{i\upsilon 1} - m_{i\upsilon 5}) \|e_{i\upsilon}\|^2 - (\alpha_{i\upsilon 1} - m_{i\upsilon 6}) \|l_{i\upsilon}\|^2 \\ & + \frac{\varphi_{i\upsilon 1}(\|b_{i\upsilon}\|)^2}{4\alpha_{i\upsilon 2}} \|b_{i\upsilon}\|^2 + \frac{\eta_{i\upsilon 2}^2 \varphi_{i\upsilon 2}(\|b_{i\upsilon}\|)^2}{4\beta_{i\upsilon}\gamma_{i\upsilon 2}} \|b_{i\upsilon}\|^2. \end{aligned} \quad \text{(A5)}$$

If the inequality conditions in Theorem 1 can be satisfied, then (A5) can be rewritten as:

$$\dot{V}_{iv} \leq -\lambda_{iv1}\|b_{iv}\|^2 + \frac{\varphi_{iv}(\|b_{iv}\|)^2}{4\lambda_{iv2}}\|b_{iv}\|^2 \leq -U_{ivc}(T_{iv}), \quad \forall T_{iv} \in \Gamma_{iv}, \tag{A6}$$

where $\lambda_{iv1} = \min\{\alpha_{iv1} - m_{iv6}, \beta_{iv}\gamma_{iv1} - m_{iv5}\}$, $\lambda_{iv2} = \min\left\{\frac{\beta_{iv}\gamma_{iv2}}{\eta_{iv2}^2}, \alpha_{iv2}\right\}$ and $\varphi_{iv}(\|b_{iv}\|)^2 = \varphi_{iv1}(\|b_{iv}\|)^2 + \varphi_{iv2}(\|b_{iv}\|)^2$. Moreover, $U_{ivc}(T_{iv})$, which is considered a continuous positive semidefinite function $U_{ivc}(T_{iv}) = \alpha_{ivc}\|b_{iv}\|^2$, is defined on the domain $\Gamma_{iv} = \{T_{iv} | \|T_{iv}\| \leq \varphi_{iv}^{-1}\left(2\sqrt{\lambda_{iv1}\lambda_{iv2}}\right)\}$, and α_{ivc} is a positive constant.

Let $K_{iv} \subset \Gamma_{iv}$ represents a set, defined as:

$$K_{iv} = \left\{T_{iv} \subset \Gamma_{iv} | \, U_{iv2}(T_{iv}) < \frac{1}{2}\left(\varphi_{iv}^{-1}\left(2\sqrt{\lambda_{iv1}\lambda_{iv2}}\right)\right)^2\right\}. \tag{A7}$$

Then, one can adjust the region of attraction in (A7) to be arbitrarily large that includes any initial conditions, thus, one may have $\alpha_{ivc}\|b_{iv}\|^2 \to 0$, while $\forall T_{iv}(0) \in K_{iv}, t \to \infty$.

Per the definition of b_{iv}, l_{iv} and υ_i, we conclude that $\|e_{iv}\|, \|\dot{e}_{iv}\|, \|l_{iv}\| \to 0$, while $\forall T_{iv}(0) \in K_{iv}, t \to \infty$. Therefore, according to the representation in (39), one obtains the conclusion that $\|\tilde{\upsilon}_i\| \to 0$ while $t \to \infty$. This completes the proof of the Theorem 1.

APPENDIX B
PROOF OF THEOREM 2

We select the Lyapunov function candidate as:

$$V_{iN} = \frac{1}{2\alpha_{ic}}\tilde{w}_{ic}^T\tilde{w}_{ic} + \frac{z_{ia}}{2\alpha_{ia}}tr\left(\tilde{w}_{ia}^T\tilde{w}_{ia}\right) + \frac{z_{ip}}{2\alpha_{ip}}tr\left(\tilde{w}_{ip}^T\tilde{w}_{ip}\right), \tag{B1}$$

where z_{ia}, z_{ip} are positive constants to be determined and $tr(\cdot)$ denotes the trace of a matrix. Then, one can formulate the time derivative of (B1) that is given as:

$$\begin{aligned}
\dot{V}_{iN} = {} & \frac{1}{\alpha_{ic}}\tilde{w}_{ic}^T\dot{\tilde{w}}_{ic} \\
& - z_{ia}tr\left\{\tilde{w}_{ia}^T\sigma_{ia}\begin{pmatrix}\tilde{w}_{ia}^T\sigma_{ia} - w_{ia}^T\sigma_{ia} \\ -\frac{1}{2}R_i^{-1}B_i^T\nabla\sigma_{ic}^T w_{ic} \\ +\frac{1}{2}R_i^{-1}B_i^T\nabla\sigma_{ic}^T\tilde{w}_{ic}\end{pmatrix}^T\right\} \\
& - z_{ip}tr\left\{\tilde{w}_{ip}^T\sigma_{ip}\begin{pmatrix}\tilde{w}_{ip}^T\sigma_{ip} - \frac{1}{2\gamma_{ip}^2}\nabla\sigma_{ic}^T\tilde{w}_{ic} \\ -w_{ip}^T\sigma_{ip} + \frac{1}{2\gamma_{ip}^2}\nabla\sigma_{ic}^T w_{ic}\end{pmatrix}^T\right\} \\
\leq {} & \begin{pmatrix}\frac{z_{ia}}{8}\left\|R_i^{-1}B_i^T\nabla\sigma_{ic}^T\right\|^2 \\ +\frac{z_{ip}}{8\gamma_{ip}^2}\left\|\nabla\sigma_{ic}^T\right\|^2 - n_{ic}n_{ic}^T\end{pmatrix}\tilde{w}_{ic}^T\tilde{w}_{ic} \\
& - \frac{1}{2}z_{ia}\left(\tilde{w}_{ia}^T\sigma_{ia}\right)^T\left(\tilde{w}_{ia}^T\sigma_{ia}\right) \\
& - \frac{1}{2}z_{ip}\left(\tilde{w}_{ip}^T\sigma_{ip}\right)^T\left(\tilde{w}_{ip}^T\sigma_{ip}\right) + \tilde{w}_{ic}^T n_{ic}\frac{e_{icH}}{m_{ic}} \\
& + z_{ia}\left(\tilde{w}_{ia}^T\sigma_{ia}\right)^T\cdot\left(w_{ia}^T\sigma_{ia} + \frac{1}{2}R_i^{-1}B_i^T\nabla\sigma_{ic}^T w_{ic}\right) \\
& + z_{ip}\left(\tilde{w}_{ip}^T\sigma_{ip}\right)^T\cdot\left(w_{ip}^T\sigma_{ip} - \frac{1}{2\gamma_{ip}^2}\nabla\sigma_{ic}^T w_{ic}\right).
\end{aligned} \tag{B2}$$

Besides, we define the matrices ψ_{N1}, ψ_{N2} and ψ_{N3} as:

$$\begin{aligned}
\psi_{N1} &= n_{ic}n_{ic}^T - \frac{z_{ia}}{8}\left\|R_i^{-1}B_i^T\nabla\sigma_{ic}^T\right\|^2 - \frac{z_{ip}}{8\gamma_{ip}^2}\left\|\nabla\sigma_{ic}^T\right\|^2, \\
\psi_{N2} &= \frac{1}{2}z_{ia}, \quad \psi_{N3} = \frac{1}{2}z_{ip}.
\end{aligned} \tag{B3}$$

Define a matrix ψ_{Ng} as $\psi_{Ng} = diag(\psi_{N1}, \psi_{N2}, \psi_{N3})$ and select the constants z_{ia} and z_{ip} as the following form to make sure that ψ_{Ng} is positive definite:

$$z_{ia} < \frac{4n_{ic}n_{ic}^T}{\left\|R_i^{-1}B_i^T\nabla\sigma_{ic}\right\|^2}, \tag{B4}$$

$$z_{ip} < \frac{4\gamma_{ip}^2 n_{ic}n_{ic}^T}{\|\nabla\sigma_{ic}\|^2}. \tag{B5}$$

Consider the vectors M_n and N_n as:

$$\begin{aligned}
M_n &= \left[\tilde{w}_{ic}^T, \left(\tilde{w}_{ia}^T\sigma_{ia}\right)^T, \left(\tilde{w}_{ip}^T\sigma_{ip}\right)^T\right]^T, \\
N_n &= \begin{bmatrix}\left(\frac{e_{icH}}{m_{ic}}\right)^T, z_{ia}\left(w_{ia}^T\sigma_{ia} + \frac{1}{2}R_i^{-1}B_i^T\nabla\sigma_{ic}^T w_{ic}\right), \\ z_{ip}\left(w_{ip}^T\sigma_{ip} - \frac{1}{2\gamma_{ip}^2}\nabla\sigma_{ic}^T w_{ic}\right)\end{bmatrix}^T.
\end{aligned} \tag{B6}$$

Then, one obtains that (B2) satisfies the following relation:

$$\begin{aligned}
\dot{V}_{iN} &\leq -M_n^T\psi_{Ng}M_n + M_n^T N_n \\
&\leq -\|M_n\|^2\lambda_{\min}(k_n) + \|M_n\|\cdot\|N_n\|,
\end{aligned} \tag{B7}$$

where $\lambda_{\min}(\cdot)$ denotes the minimum eigenvalue of a matrix. According to (B7), let $\|M_n\| > M_{nB}$, in which

$$M_{nB} = \frac{\|N_n\|}{\lambda_{\min}\left(\psi_{Ng}\right)}. \tag{B8}$$

From (B7) and (B8), we know that $\dot{V}_{iN}$ is negative, therefore, one can obtain the conclusion that the weigh approximation errors $\tilde{w}_{ic}$, $\tilde{w}_{ia}$ and $\tilde{w}_{ip}$ are guaranteed to be UUB. This concludes the proof of Theorem 2.

APPENDIX C
PROOF OF THEOREM 3

We choose a Lyapunov function candidate as:

$$\begin{aligned}
V_M(t) &= \sum_{i=1}^{n} V_{iM}(t) \\
&= \sum_{i=1}^{n}\begin{pmatrix}\omega_{il}l_i^T l_i + \omega_{iR}J_i(l_i) \\ +\omega_{ip}\int_t^\infty \gamma_{ip}^2 p_i^T(\tau)p_i(\tau)d\tau\end{pmatrix},
\end{aligned} \tag{C1}$$

where ω_{il} and $\omega_{ip} > \omega_{iR}$ are positive definite constants. According to (24) and (31), we obtain the time derivative of (C1) that is given as:

$$\begin{aligned}\dot{V}_M(t) = &\sum_{i=1}^{n} 2\omega_{il}\left(l_i^T\begin{pmatrix}\phi_i+\upsilon_i+p_i^*+\alpha_{ie}\dot{e}_i\\ -\ddot{x}_{id}+B_iu_{i1}\\ +B_iu_{i2}+B_iu_{i3}^*\end{pmatrix}\right)\\ &-\sum_{i=1}^{n}\omega_{iR}\left(l_i^TQ_il_i+u_i^TR_iu_i-\gamma_{ip}^2p_i^Tp_i\right)\\ &-\sum_{i=1}^{n}\omega_{iR}\left(\gamma_{ip}^2p_i^Tp_i\right). \end{aligned}\tag{C2}$$

In (C2), we know that the control law u_{i1} is proposed as (32) to compensate the dynamic model term ϕ_i, the control law u_{i2} is given as (50) to due with the effect of the model uncertain term υ_i, in which the asymptotic convergence of identification error has been proved. Then, by substituting (32), (50) into (C2), $\dot{V}_M(t)$ can be approximated by:

$$\begin{aligned}\dot{V}_M(t) \le &\sum_{i=1}^{n}\left(2\omega_{il}\left(\|l_i\|^2+\frac{1}{2}\|B_i\|^2\left\|u_{i3}^*\right\|^2\right)\right)\\ &-\sum_{i=1}^{n}\omega_{iR}\begin{pmatrix}\lambda_{\min}(Q_i)\|l_i\|^2+\lambda_{\min}(R_i)\|u_{i1}\|^2\\ +\lambda_{\min}(R_i)\|u_{i2}\|^2+\lambda_{\min}(R_i)\left\|u_{i3}^*\right\|^2\end{pmatrix}\\ &-\sum_{i=1}^{n}\left(-\omega_{il}-\left(\omega_{iR}-\omega_{ip}\right)\gamma_{ip}^2\right)\|p_i\|^2\\ =&-\sum_{i=1}^{n}\left(\omega_{iR}\lambda_{\min}(Q_i)-2\omega_{il}\right)\|l_i\|^2\\ &-\sum_{i=1}^{n}\left(\omega_{iR}\lambda_{\min}(R_i)\right)\|u_{i1}\|^2\\ &-\sum_{i=1}^{n}\left(\omega_{iR}\lambda_{\min}(R_i)\right)\|u_{i2}\|^2\\ &-\sum_{i=1}^{n}\left(\omega_{iR}\lambda_{\min}(R_i)-\omega_{il}\|B_i\|^2\right)\left\|u_{i3}^*\right\|^2\\ &-\sum_{i=1}^{n}\left(-\omega_{il}-\left(\omega_{iR}-\omega_{ip}\right)\gamma_{ip}^2\right)\|p_i\|^2. \end{aligned}\tag{C3}$$

From (C3), we obtain that $\dot{V}_M(t) \le 0$, when the following condition satisfied:

$$\begin{cases}\lambda_{\min}(Q_i) \ge \dfrac{2\omega_{il}}{\omega_{iR}},\\ \lambda_{\min}(R_i) \ge \dfrac{\omega_{il}}{\omega_{iR}B_i^2},\\ \omega_{ip} \ge \omega_{iR}+\dfrac{\omega_{il}}{\gamma_{ip}^2}.\end{cases}\tag{C4}$$

If the conditions in (C4) are all hold, then for any $l_i \neq 0$, one concludes $\dot{V}_M(t) < 0$. So that according to the Lyapunov theory, the closed-loop robotic systems is asymptotically stable under the decentralized control law (71). This concludes the proof of the Theorem 3.

REFERENCES

[1] A. Albu-Schäffer, C. Ott, and G. Hirzinger, "A unified passivity-based control framework for position, torque and impedance control of flexible joint robots," *Int. J. Robot. Res.*, vol. 26, no. 1, pp. 23–39, Jan. 2007.

[2] G. Liu, S. Abdul, and A. A. Goldenberg, "Distributed control of modular and reconfigurable robot with torque sensing," *Robotica*, vol. 26, no. 1, pp. 75–84, Jan. 2008.

[3] G. Liu, Y. Liu, and A. A. Goldenberg, "Design, analysis, and control of a spring-assisted modular and reconfigurable robot," *IEEE/ASME Trans. Mechatronics*, vol. 16, no. 4, pp. 695–706, Aug. 2011.

[4] A. D. Luca and R. Mattone, "Sensorless robot collision detection and hybrid force/motion control," in *Proc. IEEE Int. Conf. Robot. Automat.*, Apr. 2005, pp. 999–1004.

[5] A. D. Luca, A. Albu-Schaffer, S. Haddadin, and G. Hirzinger, "Collision detection and safe reaction with the DLR-III lightweight manipulator arm," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, Oct. 2006, pp. 1623–1630.

[6] L. Manuelli and R. Tedrake, "Localizing external contact using proprioceptive sensors: The contact particle filter," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst. (IROS)*, Oct. 2016, pp. 5062–5069.

[7] M. Fumagalli, S. Ivaldi, M. Randazzo, L. Natale, G. Metta, G. Sandini, and F. Nori, "Force feedback exploiting tactile and proximal force/torque sensing," *Auton. Robots*, vol. 33, pp. 381–398, Nov. 2012.

[8] E. Magrini, F. Flacco, and A. D. Luca, "Estimation of contact forces using a virtual force sensor," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, Sep. 2014, pp. 2126–2133.

[9] J. Qu, Z. Ji, C. Lin, and H. Yu, "Fast consensus seeking on networks with antagonistic interactions," *Complexity*, vol. 15, Dec. 2018, Art. no. 7831317.

[10] H. Ma, X. Jia, N. Cai, and J. Xi, "Adaptive guaranteed-performance consensus control for multiagent systems with an adjustable convergence speed," *Discrete Dyn. Nature Soc.*, vol. 9, May 2019, Art. no. 5190301.

[11] N. Cai, M. He, Q. Wu, and M. J. Khan, "On almost controllability of dynamical complex networks with noises," *J. Syst. Sci. Complex.*, vol. 45, pp. 1–19, Jun. 2017.

[12] J. Xi, C. Wang, H. Liu, and Z. Wang, "Dynamic output feedback guaranteed-cost synchronization for multiagent networks with given cost budgets," *IEEE Access*, vol. 6, pp. 28923–28935, 2018.

[13] W. He, Z. Yan, Y. Sun, Y. Ou, and C. Sun, "Neural-learning-based control for a constrained robotic manipulator with flexible joints," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 29, no. 12, pp. 5993–6003, Dec. 2018.

[14] P. J. Werbos, "Approximate dynamic programming for real-time control and neural modeling," in *Handbook of Intelligent Control: Neural, Fuzzy and Adaptive Approaches*, D. A. White and D. A. Sofge, Eds. New York, NY, USA: Van Nostrand, 1992, ch. 13.

[15] A. Al-Tamimi, F. L. Lewis, and M. Abu-Khalaf, "Discrete-time nonlinear HJB solution using approximate dynamic programming: Convergence proof," *IEEE Trans. Syst., Man, Cybern. B. Cybern.*, vol. 38, no. 4, pp. 943–949, Aug. 2008.

[16] D. Wang, C. Mu, and D. Liu, "Data-driven nonlinear near-optimal regulation based on iterative neural dynamic programming," *Acta Automatica Sinica*, vol. 43, no. 3, pp. 366–375, Mar. 2017.

[17] D. V. Prokhorov and D. C. Wunsch, "Adaptive critic designs," *IEEE Trans. Neural Netw.*, vol. 8, no. 5, pp. 997–1007, Sep. 1997.

[18] L. P. Kaelbling, M. L. Littman, and A. W. Moore, "Reinforcement learning: A survey," *J. Artif. Intell. Res.*, vol. 4, no. 1, pp. 237–285, Jan. 1996.

[19] Z. Chi, Z. Wei, C. Ningbo, and G. Junshan, "Trajectory tracking control for rotary steerable systems using interval type-2 fuzzy logic and reinforcement learning," *J. Franklin Inst.*, vol. 355, pp. 803–826, Jan. 2018.

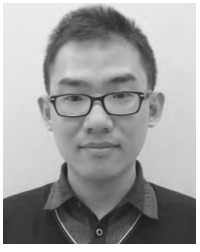
[20] N. Dong and Z. Chen, "A novel ADP based model-free predictive control," *Nonlinear Dyn.*, vol. 69, pp. 89–97, Jul. 2012.

[21] B. Luo, D. Liu, and H. Wu, "Adaptive constrained optimal control design for data-based nonlinear discrete-time systems with critic-only structure," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 29, no. 6, pp. 2099–2111, Jun. 2018.

[22] Q. Wei, B. Li, and R. Song, "Discrete-time stable generalized self-learning optimal control with approximation errors," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 29, no. 4, pp. 1226–1238, Apr. 2018.

[23] B. Zhao, D. Wang, G. Shi, D. Liu, and Y. Li, "Decentralized control for large-scale nonlinear systems with unknown mismatched interconnections via policy iteration," *IEEE Trans. Syst., Man, Cybern. Syst.*, vol. 48, no. 10, pp. 1725–1735, Oct. 2018.

[24] B. Zhao, L. Jia, H. Xia, and Y. Li, "Adaptive dynamic programming-based stabilization of nonlinear systems with unknown actuator saturation," *Nonlinear Dyn.*, vol. 93, pp. 2089–2103, Sep. 2018.
[25] Q. Zhang, D. Zhao, and D. Wang, "Event-based robust control for uncertain nonlinear systems using adaptive dynamic programming," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 29, no. 1, pp. 37–50, Jan. 2018.
[26] Q. Wei and D. Liu, "Data-driven neuro-optimal temperature control of water—Gas shift reaction using stable iterative adaptive dynamic programming," *IEEE Trans. Ind. Electron.*, vol. 61, no. 11, pp. 6399–6408, Nov. 2014.
[27] B. Luo, H.-N. Wu, T. Huang, and D. Liu, "Data-based approximate policy iteration for affine nonlinear continuous-time optimal control design," *Automatica*, vol. 50, no. 12, pp. 3281–3290, 2014.
[28] C. Li, D. Liu, and D. Wang, "Data-based optimal control for weakly coupled nonlinear systems using policy iteration," *IEEE Trans. Syst., Man, Cybern. Syst.*, vol. 48, no. 4, pp. 511–521, Apr. 2018.
[29] Q. Wei, R. Song, and P. Yan, "Data-driven zero-sum neuro-optimal control for a class of continuous-time unknown nonlinear systems with disturbance using ADP," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 27, no. 2, pp. 444–458, Feb. 2016.
[30] Q. Wei, D. Liu, Q. Lin, and R. Song, "Adaptive dynamic programming for discrete-time zero-sum games," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 29, no. 4, pp. 957–969, Apr. 2018.
[31] J. Sun, C. Liu, and X. Zhao, "Backstepping-based zero-sum differential games for missile-target interception systems with input and output constraints," *IET Control Theory Appl.*, vol. 12, no. 2, pp. 243–253, Nov. 2017.
[32] H. Zhang, Y. Luo, and D. Liu, "Neural-network-based near-optimal control for a class of discrete-time affine nonlinear systems with control constraints," *IEEE Trans. Neural Netw.*, vol. 20, no. 9, pp. 1490–1503, Sep. 2009.
[33] X. Yang, D. Liu, and D. Wang, "Reinforcement learning for adaptive optimal control of unknown continuous-time nonlinear systems with input constraints," *Int. J. Control*, vol. 87, no. 3, pp. 553–566, Mar. 2014.
[34] Y. Jiang and Z.-P. Jiang, "Robust adaptive dynamic programming for large-scale systems with an application to multimachine power systems," *IEEE Trans. Circuits Syst., II, Exp. Briefs*, vol. 59, no. 10, pp. 693–697, Oct. 2012.
[35] D. Wang, D. Liu, and H. Li, "Policy iteration algorithm for online design of robust control for a class of continuous-time nonlinear systems," *IEEE Trans. Autom. Sci. Eng.*, vol. 11, no. 2, pp. 627–632, Apr. 2014.
[36] D. Wang, D. Liu, H. Li, B. Luo, and H. Ma, "An approximate optimal control approach for robust stabilization of a class of discrete-time nonlinear systems with uncertainties," *IEEE Trans. Syst., Man, Cybern., Syst.*, vol. 46, no. 5, pp. 713–717, May 2016.
[37] B. Zhao, D. Liu, and Y. Li, "Online fault compensation control based on policy iteration algorithm for a class of affine non-linear systems with actuator failures," *IET Control Theory Appl.*, vol. 10, no. 15, pp. 1816–1823, Jun. 2016.
[38] W. He and Y. Dong, "Adaptive fuzzy neural network control for a constrained robot using impedance learning," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 29, no. 4, pp. 1174–1186, Apr. 2018.
[39] S. Li, L. Ding, H. Gao, Y.-J. Liu, L. Huang, and Z. Deng, "ADP-based online tracking control of partially uncertain time-delayed nonlinear system and application to wheeled mobile robots," *IEEE Trans. Cybern.*, to be published.
[40] D. L. Leottau, J. Ruiz-del-Solar, and R. Babuška, "Decentralized reinforcement learning of robot behaviors," *Artif. Intell.*, vol. 256, pp. 130–159, Mar. 2018.
[41] A. H. Qureshi, Y. Nakamura, Y. Yoshikawa, and H. Ishiguro, "Intrinsically motivated reinforcement learning for human—Robot interaction in the real-world," *Neural Netw.*, vol. 107, pp. 23–33, Nov. 2018.
[42] L. Roveda, G. Pallucca, N. Pedrocchi, F. Braghin, L. M. Tosatti, "Iterative learning procedure with reinforcement for high-accuracy force tracking in robotized tasks," *IEEE Trans. Ind. Informat.*, vol. 14, no. 4, pp. 1753–1763, Apr. 2018.
[43] B. Dong, F. Zhou, K. Liu, and Y. Li, "Decentralized robust optimal control for modular robot manipulators via criticidentifier structure-based adaptive dynamic programming," *Neural Comput. Appl.*, vol. 21, pp. 1–18, Sep. 2018.
[44] B. Dong, F. Zhou, K. Liu, and Y. Li, "Torque sensorless decentralized neuro-optimal control for modular and reconfigurable robots with uncertain environments," *Neurocomputing*, vol. 282, pp. 60–73, Mar. 2018.
[45] S. Haddadin, A. De Luca, and A. Albu-Schäffer, "Robot collisions: A survey on detection, isolation, and identification," *IEEE Trans. Robot.*, vol. 33, no. 6, pp. 1292–1312, Dec. 2017.
[46] H. Zhang, S. Ahmad, and G. Liu, "Modeling of torsional compliance and hysteresis behaviors in harmonic drives," *IEEE/ASME Trans. Mechatronics*, vol. 20, no. 1, pp. 178–185, Feb. 2015.
[47] H. Zhang, S. Ahmad, and G. Liu, "Torque estimation for robotic joint with harmonic drive transmission based on position measurements," *IEEE Trans. Robot.*, vol. 31, no. 2, pp. 322–330, Apr. 2015.
[48] J.-I. Imura, Y. Yokokohji, T. Yoshikawa, and T. Sugie, "Robust control of robot manipulators based on joint torque sensor information," *Int. J. Robot. Res.*, vol. 13, no. 5, pp. 434–442, Oct. 1994.
[49] B. Dong, K. Liu, and Y. Li, "Decentralized control of harmonic drive based modular robot manipulator using only position measurements: Theory and experimental verification," *J. Intell. Robotic Syst.*, vol. 88, pp. 3–18, Oct. 2017.
[50] B. Armstrong-Hélouvry, P. Dupont, and C. C. De Wit, "A survey of models, analysis tools and compensation methods for the control of machines with friction," *Automatica*, vol. 30, pp. 1083–1138, Jul. 1994.
[51] G. Liu, A. A. Goldenberg, and Y. Zhang, "Precise slow motion control of a direct-drive robot arm with velocity estimation and friction compensation," *Mechatronics*, vol. 14, no. 7, pp. 821–834, Sep. 2004.
[52] T. Basar and P. Bernhard, *H1-Optimal Control and Related Minimax Design Problems: A Dynamic Game Approach* Boston, MA, USA: Birkhauser, 1995.
[53] T. Basar and G. J. Olsder, *Dynamic Noncooperative Game Theory*. 2nd ed. Philadelphia, Philadelphia, PA, USA: SIAM, 1999.
[54] 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, Jan. 2013.
[55] P. M. Patre, W. MacKunis, K. Kaiser, and W. E. Dixon, "Asymptotic tracking for uncertain dynamic systems via a multilayer neural network feedforward and RISE feedback control structure," *IEEE Trans. Autom. Control*, vol. 53, no. 9, pp. 2180–2185, Oct. 2008.
[56] W. E. Dixon, A. Behal, D. M. Dawson, and S. Nagarkatti, *Nonlinear Control of Engineering Systems: A Lyapunov-Based Approach*. Birkhauser, Boston, MA, USA: 2003.
[57] D. Liu, D. Wang, and H. Li, "Decentralized stabilization for a class of continuous-time nonlinear interconnected systems using online learning optimal control approach," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 25, no. 2, pp. 418–428, Feb. 2014.
[58] S. Tong, K. Sun, and S. Sui, "Observer-based adaptive fuzzy decentralized optimal control design for strict-feedback nonlinear large-scale systems," *IEEE Trans. Fuzzy Syst.*, vol. 26, no. 2, pp. 569–584, Apr. 2018.
[59] K. Sun, S. Sui, and S. Tong, "Fuzzy adaptive decentralized optimal control for strict feedback nonlinear large-scale systems," *IEEE Trans. Cybern.*, vol. 48, no. 4, pp. 1326–1339, Apr. 2018.



BO DONG received the M.S. and Ph.D. degrees from Jilin University, China, in 2012 and 2015, respectively. He is currently an Associate Professor with the Changchun University of Technology, China. He is also a Postdoctoral Fellow with the State Key Laboratory of Management and Control for Complex Systems, Institute of Automation, Chinese Academy of Sciences, Beijing, China. His research interests include intelligent mechanical and robot control.

TIANJIAO AN received the B.S. degree from the Changchun University of Technology, China, in 2017, where he is currently pursuing the M.S. degree with the Department of Control Science and Engineering. His research interests include robot control and adaptive dynamic programming.

FAN ZHOU received the M.S. and Ph.D. degrees from the Changchun University of Technology, China in 2015 and 2018, respectively, where she is currently a Lecturer. Her research interests include intelligent mechanical, robot control, and robust control.

KEPING LIU received the M.S. and Ph.D. degrees from Jilin University, China, in 1999 and 2002, respectively. He is currently a Professor with the Changchun University of Technology, China. His research interests include intelligent mechanical and robot control.

WEIBO YU received the B.S. and M.S. degrees from the Changchun University of Technology, China, in 1993 and 1998, respectively, where she is currently a Professor. Her research interests include intelligent control and image processing.

YUANCHUN LI received the M.S. and Ph.D. degrees from the Harbin Institute of Technology, China, in 1987 and 1990, respectively. He is currently a Professor with the Changchun University of Technology, China. His research interests include complex system modeling, intelligent mechanical, and robot control.