



Decentralized Robust Optimal Control for Modular Robot Manipulators Based on Zero-Sum Game with ADP

Bo Dong^{1,2}(✉), Tianjiao An¹, Fan Zhou¹, Shenquan Wang¹, Yulian Jiang¹,
Keping Liu¹, Fu Liu³, Huiqiu Lu³, and Yuanchun Li¹

¹ Department of Control Science and Engineering,
Changchun University of Technology, Changchun, China
dongbo@ccut.edu.cn

² State Key Laboratory of Management and Control for Complex Systems,
Institute of Automation, Chinese Academy of Sciences, Beijing, China

³ Department of Control Science and Engineering,
Jilin University, Changchun, China

Abstract. In this paper, a decentralized robust zero-sum optimal control method is proposed for modular robot manipulators (MRMs) based on the adaptive dynamic programming (ADP) approach. The dynamic model of MRMs is formulated via joint torque feedback (JTF) technique that is deployed for each joint module, in which the local dynamic information is used to design the model compensation controller. An uncertainty decomposition-based robust control is developed to compensate the model uncertainties, and then the robust optimal control problem of MRMs with uncertain environments can be transformed into a two-player zero-sum optimal control one. According to the ADP algorithm, the Hamilton-Jacobi-Isaacs (HJI) equation is solved by constructing action-critic neural networks (NNs) and then the approximate optimal control policy derivation is possible. Experiments are conducted to verify the effectiveness of the proposed method.

Keywords: Modular robot manipulators · Adaptive dynamic programming · Optimal control · Zero-sum game

1 Introduction

Modular robot manipulators have attracted wildly attentions in robotics community since they are possessed of better structural adaptability and flexibility than conventional robot manipulators. An MRM is composed by standardized robotic modules, which consist of actuators, speed reducers, sensors and communication units. These modules can be assembled to desired configurations via standard mechanical interfaces according to the requirements of various tasks. Benefitting from the advantages above, MRMs are often utilized in dangerous and complex environments. Correspondingly, appropriate control systems are

required to guarantee the robustness and precision of MRMs with uncertain environments.

In order to seek out appropriate control for MRMs, joint torque feedback techniques have attracted attentions of both robotic researchers and industrial manufacturers. JTF technique can not only eliminate the necessity of contact model in motion control of robot-environment interactive tasks, it also facilitates in suppressing the effect of uncertain contact force created between the end-effector and payloads [1–3]. However, the existed JTF-based motion/force control methods have not considered the issue of optimal implementation between the error performance of joint motion and the output torque of actuator. Indeed, an ideal controllers for MRMs should be possessed of the properties that ensure the robustness of the robotic systems, and simultaneously take into account the optimal realization of the composite of error performance and output power.

Adaptive dynamic programming methodology is considered as one of the key directions to address the optimal control problems in complex systems. Some investigations reported the latest research progress of ADP-based optimal control methods for robot manipulators systems [4–6]. These strategies both follow the premise that the dynamic models of the robotic systems are completely unknown, thus the application of these methods are limited to address the optimal control problems of specific classes of robotic systems without implementing dynamic compensation. Dong et al. address the optimal control problems of MRMs by combining the model-based compensation control with ADP-based learning control [7], and their researches are further expanded to deal with the optimal tracking control issues of MRMs with uncertain environments [8]. However, the existing methods consider the disturbance torques, which are introduced by environmental contacts, as a class of dynamic uncertainties, and ignoring the intractability of decomposing the effects of model uncertainties and environmental contact disturbance targetedly. Indeed, there is less discussion on ADP-based robust optimal control for robot manipulators with both dynamic decomposition and optimal compensation, especially, for MRMs with uncertain environmental contact.

In this paper, a novel decentralized robust zero-sum neuro-optimal control method is presented for MRMs with uncertain environments. First, the dynamic model of the MRM systems is formulated via JTF technique, and a model-based compensation controller is designed by effectively utilizing the local dynamic information of each joint module. Second, a decentralized robust controller is proposed to deal with the compensation control issue of model uncertainties. Then, the robust optimal control problem of uncertain environment-contacted MRMs 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 disturbance can be approximated by constructing one critic NN and two action NNs, and then the decentralized robust zero-sum neuro-optimal control is developed. Finally, experiments are performed to verify the effectiveness of the proposed method.

2 Dynamic Model Formulation

We consider the MRMs comprise n joint modules installed in series, and each integrated with a DC motor, a speed reducer, two encoders and a joint torque sensor. By referencing the modelling methods of JTF-based robot manipulators with multi-degree of freedom (DOF) [9, 10], we formulate the dynamic model of the MRM system as a synthesis of interconnected robotic joint subsystems, in which the i th joint subsystem dynamic model is represented as:

$$I_{mi}\gamma_i\ddot{q}_i + f_i(q_i, \dot{q}_i) + z_i(q, \dot{q}, \ddot{q}) + \frac{\tau_{si}}{\gamma_i} + d_i(q_i) = \tau_i, \quad (1)$$

where the subscript “ i ” indicates the i th subsystem; I_{mi} is the moment of inertia of actuator rotor about the rotation axis; γ_i denotes the gear ratio; q_i , \dot{q}_i and \ddot{q}_i are the joint angular position, velocity and acceleration respectively; $f_i(q_i, \dot{q}_i)$ represents the joint lumped frictional torque; $z_i(q_i, \dot{q}_i, \ddot{q}_i)$ denotes the interconnected dynamic coupling (IDC) among the joint subsystems; τ_{si} represents the coupling torque at the torque sensor location; $d_i(q_i)$ is the disturbance torque and τ_i indicates the control torque.

(1) Joint friction

The friction term can be defined as a function with respect to joint angular position and velocity, which is given as:

$$f_i(q_i, \dot{q}_i) \approx \hat{f}_{bi}\dot{q}_i + \left(\hat{f}_{si}e^{-\hat{f}_{\tau i}\dot{q}_i^2} + \hat{f}_{ci} \right) \text{sgn}(\dot{q}_i) + f_{pi}(q_i, \dot{q}_i) + Y(\dot{q}_i)\tilde{F}_i, \quad (2)$$

where $\tilde{F}_i = [f_{bi} - \hat{f}_{bi} \ f_{ci} - \hat{f}_{ci} \ f_{si} - \hat{f}_{si} \ f_{\tau i} - \hat{f}_{\tau i}]^T$ indicates the parametric uncertainty of the friction; f_{bi} , f_{si} , $f_{\tau i}$ and f_{ci} represent the ideal parameters of the friction model; $f_{pi}(q_i, \dot{q}_i)$ denotes the position dependency of friction and the other friction modeling errors; $\text{sgn}(\dot{q})$ is a classical sign function; \hat{f}_{bi} , \hat{f}_{si} , $\hat{f}_{\tau i}$ and \hat{f}_{ci} are estimated parameters of the friction model and $Y(\dot{q}_i)$ is defined as:

$$Y(\dot{q}_i) = \left[\dot{q}_i, \text{sgn}(\dot{q}_i), e^{-\hat{f}_{\tau i}\dot{q}_i^2} \text{sgn}(\dot{q}_i), -\hat{f}_{si}\dot{q}_i^2 e^{-\hat{f}_{\tau i}\dot{q}_i^2} \text{sgn}(\dot{q}_i) \right]. \quad (3)$$

(2) Interconnected dynamic coupling

The IDC term $z_i(q, \dot{q}, \ddot{q})$, which is considered a complex nonlinear function, can be defined with respect to the positions, velocities and accelerations of the lower joint modules (from joint 1 to $i-1$), that is given as:

$$z_i(q, \dot{q}, \ddot{q}) = I_{mi} \sum_{j=1}^{i-1} D_j^i \ddot{q}_j + I_{mi} \sum_{j=2}^{i-1} \sum_{k=1}^{j-1} \Theta_{kj}^i \dot{q}_k \dot{q}_j = U_{zi} + V_{zi},$$

$$U_{zi} = \sum_{j=1}^{i-1} \left[I_{mi} \hat{D}_j^i \ I_{mi} \right] \left[\ddot{q}_j \ \tilde{D}_j^i \ddot{q}_j \right]^T, \quad V_{zi} = \sum_{j=2}^{i-1} \sum_{k=1}^{j-1} \left[I_{mi} \hat{\Theta}_{kj}^i \ I_{mi} \right] \left[\dot{q}_k \dot{q}_j \ \tilde{\Theta}_{kj}^i \dot{q}_k \dot{q}_j \right]^T, \quad (4)$$

where D_j^i and Θ_{kj}^i are given as $D_j^i = z_{mi}^T z_{lj}$ and $\Theta_{kj}^i = z_{mi}^T (z_{lk} \times z_{lj})$ respectively, z_{mi} , z_{lj} and z_{lk} are the unity vectors along the axis of rotation of the

i th robot, j th joint and k th joint respectively. Moreover, we have the relations $\tilde{D}_j^i = D_j^i - \tilde{D}_j^i$ and $\tilde{\Theta}_{kj}^i = \Theta_{kj}^i - \tilde{\Theta}_{kj}^i$, where \tilde{D}_j^i and $\tilde{\Theta}_{kj}^i$ are the alignment errors.

(3) Joint torque and disturbance torque

The coupled joint torques τ_{si} , which are measured by using the embedded torque sensors, can be represented as the following form:

$$\tau_{si} = \tau_{sri} + \tau_{sci}, \quad (5)$$

where τ_{sri} indicates the joint torque that is obtained in free space and τ_{sci} denotes the torque produced by environmental contact in constrained space. Besides, the disturbance torque term $d_i(q_i)$ can be defined as:

$$d_i(q_i) = d_{is}(q_i) + d_{ic}(q_i), \quad (6)$$

where $d_{is}(q_i)$ represents the disturbance of torque sensor and $d_{ic}(q_i)$ indicates unmeasured environmental contact torque in joint space.

Assumption 1. The uncertainty term \tilde{F}_i is bounded by $|\tilde{F}_i| \leq \rho_{Fin}$, the friction modeling error is bounded by $|Y(\dot{q}_i)\tilde{F}_i| \leq Y(\dot{q}_i)\rho_{Fin}$ and $f_{pi}(q_i, \dot{q}_i)$ is bounded by $|f_{pi}(q_i, \dot{q}_i)| \leq \rho_{fpi}$.

Assumption 2. The terms U_{zi} and V_{zi} are bounded by $|U_{zi}| \leq \rho_{Ui}$, $|V_{zi}| \leq \rho_{Vi}$.

Assumption 3. The disturbance $d_{is}(q_i)$ is bounded by $|d_{is}(q_i)| \leq \rho_{dsi}$ and the unmeasured environmental contact torque $d_{ic}(q_i)$ is bounded by $|d_{ic}(q_i)| \leq \rho_{dci}$.

Then, define $B_i = (I_{mi}\gamma_i)^{-1} \in R^+$ and the state vector $x_i = [x_{i1} \ x_{i2}]^T = [q_i \ \dot{q}_i]^T \in R^{2 \times 1}$ and the control input $u_i = \tau_i \in R^{1 \times 1}$, $i = 1, 2, \dots, n$, the state space of i th subsystem can be formulated as

$$S_i = \begin{cases} \dot{x}_{i1} = x_{i2} \\ \dot{x}_{i2} = \phi_i(x_i) + v_i(x) + p_i(x_i) + B_i u_i, \\ y = x_{i1} \end{cases}, \quad (7)$$

where $\phi_i(x_i)$, $v_i(x)$ and $p_i(x_i)$ are represented as follows:

$$\begin{aligned} \phi_i(x_i) &= B_i \begin{pmatrix} - \left(\hat{f}_{si} e^{(-\hat{f}_{\tau i} \dot{q}_i^2)} + \hat{f}_{ci} \right) \text{sgn}(\dot{q}_i) \\ - \hat{f}_{bi} \dot{q}_i - \frac{\tau_{si}}{\gamma_i} \end{pmatrix}, \\ v_i(x) &= B_i \left(-U_{zi} - V_{zi} - f_{pi}(q_i, \dot{q}_i) - Y(\dot{q}_i)\tilde{F}_i \right) \\ p_i(x_i) &= B_i (-d_{is}(q_i) - d_{ic}(q_i)) \end{aligned} \quad (8)$$

Next, a decentralized robust zero-sum optimal control method is presented for MRMs to ensure that both position and velocity of the closed-loop robotic systems are asymptotically stable.

3 Decentralized Robust Zero-Sum Neuro-Optimal Control Based on ADP

Consider the MRM systems with the dynamic model formulation (1) and the state space description (7), if the contact disturbance $p_i(x_i)$ is viewed as a system control input, the robust optimal control problem can be transformed into a two-player zero-sum optimal control issue, in which the continuously differentiable infinite horizon local cost function can be defined as follows:

$$J_i(s_{i0}(e_i)) = \int_0^\infty \left\{ \begin{array}{l} s_i(e_i(\tau))^T Q_i s_i(e_i(\tau)) + u_i(\tau)^T R_i u_i(\tau) \\ - \gamma_{pi}^2 p_i(x_i(\tau))^T p_i(x_i(\tau)) \end{array} \right\} d\tau, \quad (9)$$

where

$$s_i(e_i) = \dot{e}_i + \alpha_{ei} e_i \quad (10)$$

is a filtered error function with the initial filtered error $s_{i0}(e_i) = s_i(0)$; $e_i = x_{i1} - x_{id}$ and $\dot{e}_i = x_{i2} - \dot{x}_{id}$ indicate the position and velocity tracking error of the i th joint module respectively; x_{id} , \dot{x}_{id} and \ddot{x}_{id} represent the desired angular position, velocity and acceleration that are known and bounded reference states. $Q_i^T = Q_i$, $R_i^T = R_i$ denote the positive constant matrixes; α_{ei} and γ_{pi} are determined positive constants.

According to (9), one can define the local Hamiltonian as:

$$\begin{aligned} H_i(s_i, u_i, p_i, \nabla J_i) &= s_i^T Q_i s_i + u_i^T R_i u_i \\ &+ \nabla J_i^T(s_i) (\phi_i + v_i + p_i + \alpha_{ei} \dot{e}_i + B_i u_i - \ddot{x}_{id}) - \gamma_{pi}^2 p_i^T p_i \end{aligned} \quad (11)$$

where $\nabla J_i(s_i) = (\partial J_i(s_i)/\partial s_i)$ denotes the partial derivative of $J_i(s_i)$ with regard of s_i . Define the utility function as

$$\Omega_i(s_i, u_i, p_i) = s_i^T Q_i s_i + u_i^T R_i u_i - \gamma_{pi}^2 p_i^T p_i. \quad (12)$$

Then, one obtain that the optimal local cost function J_i^* satisfies:

$$J_i^*(s_i) = \min_{u_i} \max_{p_i} \int_0^\infty \Omega_i(s_i, u_i, p_i) d\tau = \max_{p_i} \min_{u_i} \int_0^\infty \Omega_i(s_i, u_i, p_i) d\tau. \quad (13)$$

According to (11) and (13), the local optimal control pair (u_i^*, p_i^*) satisfies the following HJI equation:

$$0 = \nabla J_i^{*T}(s_i) (\phi_i + v_i + p_i^* + \alpha_{ei} \dot{e}_i + B_i u_i^* - \ddot{x}_{id}) + \Omega_i(s_i, u_i^*, p_i^*). \quad (14)$$

On the basis of the principle of optimality, the local optimal control pair (u_i^*, p_i^*) can be represented as follows:

$$u_i^* = -\frac{1}{2} R_i^{-1} B_i^T \nabla J_i^*, \quad p_i^* = \frac{1}{2\gamma_{pi}^2} \nabla J_i^*. \quad (15)$$

Rewriting the optimal control law u_i^* of the i th subsystem as:

$$u_i^* = u_{i1} + u_{i2} + u_{i3} \quad (16)$$

to deal with modeled dynamic term ϕ_i , model uncertainty term v_i and contact disturbance term p_i respectively, then, the HJI equation can be modified as:

$$0 = \nabla J_i^* \left(\begin{array}{c} \phi_i + v_i + p_i^* + \alpha_{ei}\dot{e}_i - \ddot{x}_{id} \\ + B_i u_{i1} + B_i u_{i2} + B_i u_{i3}^* \end{array} \right) + \Omega_i(s_i, u_i^*, p_i^*). \quad (17)$$

Therefore, the zero-sum optimal control problem has been transformed into the one of obtaining the decentralized control laws u_{i1} , u_{i2} and u_{i3}^* respectively, and therefore, to realize the optimal compensation of model uncertainty and environmental contact disturbance of the MRM system.

First, we can design control law u_{i1} , that is given as:

$$u_{i1} = - \left(- \left(\hat{f}_{si} e^{(-\hat{f}_{\tau i} x_{i2}^2)} + \hat{f}_{ci} \right) \text{sgn}(x_{i2}) - \hat{f}_{bi} x_{i2} - B_i^{-1} \ddot{x}_{id} - \frac{\tau_{si}}{\gamma_i} + B_i^{-1} \alpha_{ei} \dot{e}_i \right). \quad (18)$$

Second, we consider that the MRM systems, which work in free space, is possessed of accurate joint torque sensing. Thus, substituting (18) into (17), one obtains:

$$0 = \Omega_i(s_i, u_i^*, p_i^*) + \nabla J_i^*(v_i + p_i^* + B_i u_{i2} + B_i u_{i3}^*). \quad (19)$$

Note that we have $p_i^* = u_{i3}^* = 0$ when the MRM system works in free space, therefore, combining the time derivative of (10) and the model-based compensation control (18), one obtains

$$\dot{s}_i = B_i \left(-f_{pi}(x_{i1}, x_{i2}) - U_{zi} - Y(x_{i2}) \tilde{F}_i - V_{zi} \right) + B_i u_{i2}. \quad (20)$$

Here, we introduce the decentralized robust control u_{i2} to compensate the effects of model uncertainties. For the unmodeled friction in (2), we decomposed \tilde{F}_i as the summation of constant part and variable part, which is given as:

$$\tilde{F}_i = \tilde{F}_{ic} + \tilde{F}_{iv}, \quad (21)$$

where \tilde{F}_{ic} indicates an unknown constant vector and \tilde{F}_{iv} denotes an unknown variable vector that is bounded as:

$$\left| \tilde{F}_{iv} \right| \leq \rho_{Fivn}, n = 1, 2, 3, 4, \quad (22)$$

in which $\rho_{Fivn} = [\rho_{Fiv1} \rho_{Fiv2} \rho_{Fiv3} \rho_{Fiv4}]^T$ is a determined constant vector. Then, one can design the robust control law u_{i2f} , which is given as:

$$u_{i2f} = u_{i2fu} + Y(x_{i2})(u_{i2fc} + u_{i2fv}), \quad (23)$$

where u_{i2fu} is designed to compensate friction term $f_{pi}(x_{i1}, x_{i2})$; u_{i2fc} , u_{i2fv} are used to compensate the effect of uncertainties \tilde{F}_{ic} and \tilde{F}_{iv} respectively:

$$u_{i2fu} = \begin{cases} -\rho_{fpi} \frac{s_i}{|s_i|} \dot{i}f & |s_i| > \varepsilon_{ifu} \\ -\rho_{fpi} \frac{s_i}{\varepsilon_{ifu}} \dot{i}f & |s_i| \leq \varepsilon_{ifu} \end{cases}, \quad (24)$$

$$u_{i2fc} = -\varepsilon_{ifc} \left(\int_0^t Y(x_{i2}) s_i d\tau + s_i \right), \quad (25)$$

$$u_{i2fv} = \begin{cases} -\rho_{Fivn} \frac{Y(x_{i2}) s_i}{|Y(x_{i2}) s_i|} \text{ if } |Y(x_{i2}) s_i| > \varepsilon_{ifvn} \\ -\rho_{Fivn} \frac{Y(x_{i2}) s_i}{\varepsilon_{ifvn}} \text{ if } |Y(x_{i2}) s_i| \leq \varepsilon_{ifvn} \end{cases}, n = 1, 2, 3, 4 \quad (26)$$

where ε_{ifu} , ε_{ifc} and ε_{ifvn} are determined positive parameters. Moreover, we decomposed the terms of U_{zi} and V_{zi} as:

$$U_{zi} = U_{zic} + U_{ziv}, \quad V_{zi} = V_{zic} + V_{ziv}, \quad (27)$$

where U_{zic} and V_{zic} are with known up-bounds; U_{ziv} and V_{ziv} are bounded as:

$$|U_{ziv}| \leq \rho_{Uiv}, \quad |V_{ziv}| \leq \rho_{Viv}, \quad (28)$$

where ρ_{Uiv} and ρ_{Viv} are constant up-bounds to be determined.

Then, by adopting the decomposition-based analysis of the uncertainties, we can design the robust control law u_{i2h} , which is formulated as:

$$u_{i2h} = u_{i2hU} + u_{i2hV}, \quad (29)$$

where u_{i2hU} is used to compensate the effect of the term U_{zi} , that is given as:

$$\begin{aligned} u_{i2hU} &= u_{i2hUc} + u_{i2hUv}, \\ u_{i2hUc} &= -\varepsilon_{iUs} \left(\int_0^t \varepsilon_{iUs} s_i d\tau + s_i \right), u_{i2hUv} = \begin{cases} -\rho_{Uiv} \frac{\varepsilon_{iUs} s_i}{|\varepsilon_{iUs} s_i|} \text{ if } |\varepsilon_{iUs} s_i| > \varepsilon_{iUB} \\ -\rho_{Uiv} \frac{\varepsilon_{iUs} s_i}{\varepsilon_{iUB}} \text{ if } |\varepsilon_{iUs} s_i| \leq \varepsilon_{iUB} \end{cases} \end{aligned} \quad (30)$$

where ε_{iUs} and ε_{iUB} are positive parameters to be determined. Moreover, one can design the control law u_{i2hV} to compensate the effect of the term V_{zi} :

$$\begin{aligned} u_{i2hV} &= u_{i2hVc} + u_{i2hVv} \\ u_{i2hVc} &= -\varepsilon_{iVs} \left(\int_0^t \varepsilon_{iVs} s_i d\tau + s_i \right), u_{i2hVv} = \begin{cases} -\rho_{Viv} \frac{\varepsilon_{iVs} s_i}{|\varepsilon_{iVs} s_i|} \text{ if } |\varepsilon_{iVs} s_i| > \varepsilon_{iVB} \\ -\rho_{Viv} \frac{\varepsilon_{iVs} s_i}{\varepsilon_{iVB}} \text{ if } |\varepsilon_{iVs} s_i| \leq \varepsilon_{iVB} \end{cases} \end{aligned} \quad (31)$$

where ε_{iVs} and ε_{iVB} are positive parameters to be determined. Then, by combining (23) with (29), one obtain the decentralized control law u_{i2} , which is given as follows:

$$u_{i2} = u_{i2f} + u_{i2h}. \quad (32)$$

Then, we introduce the critic NN, u -action NN and p -action NN to approximate the performance index function $J_i(s_i)$, the optimal control law u_{i3}^* and the worst contact disturbance p_i^* respectively.

By employing RBF NNs, the ideal critic NN is represented as:

$$J_i(s_i) = w_{ci}^T \sigma_{ci}(s_i) + \varepsilon_{ci}, \quad (33)$$

where w_{ci} is the unknown ideal NN weight, ε_{ci} represent the finite approximation error of the critic NN, $\sigma_{ci}(s_i)$ indicates a standard activation function. Besides, the gradient of the approximated cost function is given as:

$$\nabla J_i(s_i) = \nabla \sigma_{ci}(s_i)^T w_{ci} + \nabla \varepsilon_{ci}^T, \quad (34)$$

where $\nabla \sigma_{ci}(s_i) = \partial \sigma_{ci}(s_i) / \partial s_i$ indicates the gradients of the activation function and $\nabla \varepsilon_{ci}$ is the gradient error. Then, one can rewrite the Hamiltonian and obtain the following relation:

$$H_i(s_i, u_i, p_i, w_{ci}) = \Omega_i(s_i, u_i, p_i) + (w_{ci}^T \nabla \sigma_{ci}(s_i)) \dot{s}_i - e_{cHi} = 0, \quad (35)$$

where e_{cHi} denotes the approximation error of the critic NN:

$$e_{cHi} = \Omega_i + (w_{ci}^T \nabla \sigma_{ci}) \dot{s}_i. \quad (36)$$

Let \hat{w}_{ci} be the estimated weight vector of w_{ci} , therefore, one obtains:

$$\hat{J}_i(s_i) = \hat{w}_{ci}^T \sigma_{ci}(s_i). \quad (37)$$

Then, one can define the approximate Hamilton function as:

$$\hat{H}_i(s_i, u_i, p_i, \hat{w}_{ci}) = \Omega_i(s_i, u_i, p_i) + (\hat{w}_{ci}^T \nabla \sigma_{ci}(s_i)) \dot{s}_i. \quad (38)$$

Define the weight approximation error $\tilde{w}_{ci} = w_{ci} - \hat{w}_{ci}$, then, one obtains:

$$e_{ci} = e_{cHi} - \tilde{w}_{ci}^T \nabla \sigma_{ci}(s_i) \dot{s}_i. \quad (39)$$

Define the residual error function $E_{ci} = \frac{1}{2} e_{ci}^2$, which is minimized to adjust the weight vector of the critic NN, that is updated by

$$\dot{\hat{w}}_{ci} = -\alpha_{ci} \left(\frac{\partial E_{ci}}{\partial \hat{w}_{ci}} \right) = -\alpha_{ci} \frac{p_{ci} (p_{ci}^T \hat{w}_{ci} + \Omega_i)}{(p_{ci}^T p_{ci} + 1)^2}, \quad (40)$$

where $\alpha_{ci} > 0$ is the learning rate of the critic NN, and $p_{ci} = \nabla \sigma_{ci}(s_i) \dot{s}_i$.

Then, we employ the action NNs to approximate u_{i3}^* and p_i^* respectively. The ideal u - and p -action NNs are given as follows:

$$u_{i3}^* = w_{ai}^T \sigma_{ai}(s_i) + \varepsilon_{ai}, \quad p_i^* = w_{pi}^T \sigma_{pi}(s_i) + \varepsilon_{pi}, \quad (41)$$

where w_{ai} , w_{pi} represent the ideal weight vectors, σ_{ai} , σ_{pi} denote the activation functions and ε_{ai} , ε_{pi} are the finite approximation errors of u - and p -action respectively. Let \hat{w}_{ai} and \hat{w}_{pi} be the estimation weight vectors of w_{ai} and w_{pi} , then we have:

$$\hat{u}_{i3} = \hat{w}_{ai}^T \sigma_{ai}(s_i), \quad \hat{p}_i = \hat{w}_{pi}^T \sigma_{pi}(s_i). \quad (42)$$

Since $\left(\partial \hat{H}_i(s_i, \hat{u}_i, \hat{p}_i, \hat{w}_{ci}) / \partial \hat{u}_i \right) = 0$ and $\left(\partial \hat{H}_i(s_i, \hat{u}_i, \hat{p}_i, \hat{w}_{ci}) / \partial \hat{p}_i \right) = 0$, so that \hat{u}_{i3} and \hat{p}_i can be further described as follows:

$$\hat{u}_{i3} = -\frac{1}{2} R_i^{-1} B_i^T \nabla \sigma_{ci}^T(s_i) \hat{w}_{ci}, \quad \hat{p}_i = \frac{1}{2\gamma_{pi}^2} \nabla \sigma_{ci}^T(s_i) \hat{w}_{ci}, \quad (43)$$

where the update laws for u - and p -action NN weights are given as:

$$\dot{\hat{w}}_{ai} = -\alpha_{ai}\sigma_{ai}(s_i) \cdot \left(\hat{w}_{ai}^T \sigma_{ai}(s_i) + \frac{1}{2} R_i^{-1} B_i^T \nabla \sigma_{ci}^T(s_i) \hat{w}_{ci} \right)^T, \quad (44)$$

$$\dot{\hat{w}}_{pi} = -\alpha_{pi}\sigma_{pi}(s_i) \cdot \left(\hat{w}_{pi}^T \sigma_{pi}(s_i) - \frac{1}{2\gamma_{pi}^2} \nabla \sigma_{ci}^T(s_i) \hat{w}_{ci} \right)^T, \quad (45)$$

where α_{ai} and α_{pi} are the positive learning rates of u - and p -action NNs.

Then, by combining (18), (32) with (43), the proposed decentralized robust zero-sum neuro-optimal control law u_i^* is given as:

$$u_i^* = - \begin{pmatrix} -\hat{f}_{bi}x_{i2} - B_i^{-1}\ddot{x}_{id} - \frac{\tau_{si}}{\gamma_i} + B_i^{-1}\alpha_{ei}\dot{e}_i \\ -(\hat{f}_{si}e^{(-\hat{f}_{\tau_i}x_{i2}^2)} + \hat{f}_{ci}) \operatorname{sgn}(x_{i2}) \end{pmatrix} - (-u_{i2hU} - u_{i2hV}) \\ - (-u_{i2fu} - Y(x_{i2})(u_{i2fc} + u_{i2fv})) - \frac{1}{2} R_i^{-1} B_i^T \nabla \sigma_{ci}^T \hat{w}_{ci} \quad (46)$$

Theorem 1. Consider an n -DOF modular robot manipulator, with the dynamic model of the i th joint subsystem formulated in (1); the model uncertainties existed in (2), (4) and the environmental contact disturbance given in (6). The closed-loop robotic system is asymptotically stable under the decentralized robust zero-sum neuro-optimal control law proposed by (46).

Proof. A Lyapunov function candidate is selected as:

$$V_g(t) = \sum_{i=1}^n V_{gi}(t) = \sum_{i=1}^n \left(\delta_{si} s_i^T s_i + \delta_{\Omega_i} J_i(s_i) + \delta_{pi} \int_t^\infty \gamma_{pi}^2 p_i^T(\tau) p_i(\tau) d\tau \right), \quad (47)$$

where $\delta_{si} > 0$ and $\delta_{pi} \geq \delta_{\Omega_i} > 0$ are determined positive definite constants. Then, one obtains the time derivative of (47) that is given as:

$$\begin{aligned} \dot{V}_g(t) &\leq \sum_{i=1}^n \left(2\delta_{si} \left(\|s_i\|^2 + \frac{1}{2} \|B_i\|^2 \|u_{i3}^*\|^2 \right) \right) - \sum_{i=1}^n \left(-\delta_{si} - (\delta_{\Omega_i} - \delta_{pi}) \gamma_{pi}^2 \right) \|p_i\|^2 \\ &\quad - \sum_{i=1}^n \delta_{\Omega_i} \left(\lambda_{\min}(Q_i) \|s_i\|^2 + \lambda_{\min}(R_i) \|u_{i1}\|^2 + \lambda_{\min}(R_i) \|u_{i2}\|^2 + \lambda_{\min}(R_i) \|u_{i3}^*\|^2 \right) \\ &= - \sum_{i=1}^n \left(\delta_{\Omega_i} \lambda_{\min}(Q_i) - 2\delta_{si} \right) \|s_i\|^2 - \sum_{i=1}^n \left(\delta_{\Omega_i} \lambda_{\min}(R_i) \right) \|u_{i1}\|^2 \\ &\quad - \sum_{i=1}^n \left(\delta_{\Omega_i} \lambda_{\min}(R_i) \right) \|u_{i2}\|^2 - \sum_{i=1}^n \left(\delta_{\Omega_i} \lambda_{\min}(R_i) - \delta_{si} \|B_i\|^2 \right) \|u_{i3}^*\|^2 \\ &\quad - \sum_{i=1}^n \left(-\delta_{si} - (\delta_{\Omega_i} - \delta_{pi}) \gamma_{pi}^2 \right) \|p_i\|^2 \end{aligned} \quad (48)$$

Therefore, we know that $\dot{V}_g(t) \leq 0$ when the following condition holds:

$$\lambda_{\min}(Q_i) \geq \frac{2\delta_{si}}{\delta_{\Omega_i}}, \quad \lambda_{\min}(R_i) \geq \frac{\delta_{si}}{\delta_{\Omega_i} I_{mi}^2 \gamma_i^2}, \quad \delta_{pi} \geq \delta_{\Omega_i} + \frac{\delta_{si}}{\gamma_{pi}^2}. \quad (49)$$

Besides, from (48), one obtains that if (49) is satisfied, then $\dot{V}_g(t) < 0$ for any $s_i \neq 0$. Therefore, by the Lyapunov theory, the closed-loop robotic system is asymptotically stable under the proposed decentralized robust zero-sum neuro-optimal control (46). This completes the proof of the Theorem 1.

4 Experimental Results

In this section, the experimental results are presented based on the experimental platform as illustrated in [8], and the experimental results are collected to verify the effectiveness of the proposed control method. We consider the environmental contact as a kind of instantaneous collision between the MRM link and the collision object that may create instantaneous contact forces.

Table 1. Parameters setting

| Name | Value | Name | Value | Name | Value |
|---|-------------------------------------|--|--------|--|----------------------|
| \hat{f}_{bi} | 12 mNms/rad | \hat{f}_{ci} | 30 mNm | \hat{f}_{si} | 40 mNm |
| $\hat{f}_{\tau i}$ | 20 s ² /rad ² | γ_i | 100 | I_{mi} | 120 gcm ² |
| ρ_{Fiv1} | 30 mNms/rad | ρ_{Fiv2} | 60 mNm | ρ_{Fiv3} | 80 mNm |
| ρ_{Fiv4} | 50 s ² /rad ² | ρ_{dsi} | 10 mNm | ρ_{dci} | 50 mNm |
| ρ_{Viv} | 2.3 | ρ_{fpi} | 0.5 | ρ_{Uiv} | 2.6 |
| α_{ei} | 0.5 | ε_{ifu} | 0.1 | ε_{ifc} | 1 |
| $\alpha_{ci}, \alpha_{pi}, \alpha_{ai}$ | 0.8 | $\varepsilon_{ifv1}, \varepsilon_{ifv2}$ | 0.01 | $\varepsilon_{ifv3}, \varepsilon_{ifv4}$ | 0.01 |
| γ_{pi} | 1.87 | $\varepsilon_{iUs}, \varepsilon_{iVs}$ | 0.02 | $\varepsilon_{iVB}, \varepsilon_{iUB}$ | 0.02 |

The parameter definition is represented as follows: The critic NN is chosen as 2-5-1 with 2 input neurons, 5 hidden neurons and 1 output neuron, and the weight vectors are selected as $\hat{w}_{c1} = [\hat{w}_{c11} \hat{w}_{c12} \hat{w}_{c13} \hat{w}_{c14} \hat{w}_{c15}]^T$ for joint 1 and $\hat{w}_{c2} = [\hat{w}_{c21} \hat{w}_{c22} \hat{w}_{c23} \hat{w}_{c24} \hat{w}_{c25}]^T$ for joint 2, with the initial values $\hat{w}_{c1} = \hat{w}_{c2} = [0]$. The NN structures and initial weight values of the and action NNs are selected as the same of the critic NNs. The other model parameters, uncertainty up-bound parameters and control parameters are given in Table 1.

Figure 1 illustrate the position and velocity error curves under the situation of instantaneous collision contact. In this figure, obvious position and velocity deviations are captured, which are due to the effect of collision force. Moreover, we also observe that the error values decreased to normal ranges with very short time periods, which may attribute to the performance of the proposed zero-sum

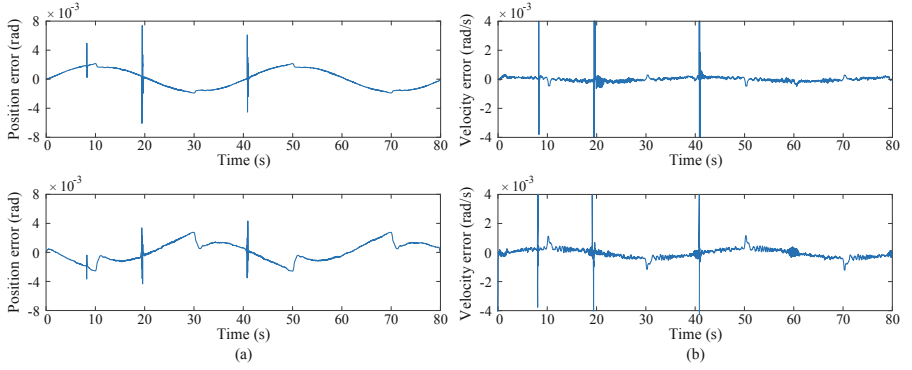


Fig. 1. Error curves under the collision contact, (a) position errors, (b) velocity errors.

neuro-optimal controller that realized the optimal compensation of unknown contact disturbance.

Figure 2(a) illustrate the control torque curves under the situation of instantaneous collision contact. From this figure, the reliable control torque outputs are obtained, in which the instant increase of control torques keep in safe limits. This is attribute the contribution of the proposed optimal control that realizes the optimization of tracking errors and output torques.

Figure 2(b) illustrate the variations of the estimated critic NN weights under the proposed method. From this figure, one obtain that the critic NN weights may converge in certain ranges for each isolated joint subsystem. With the critic NN weight estimations, the HJI equation and the optimal control law can be solved and derived respectively.

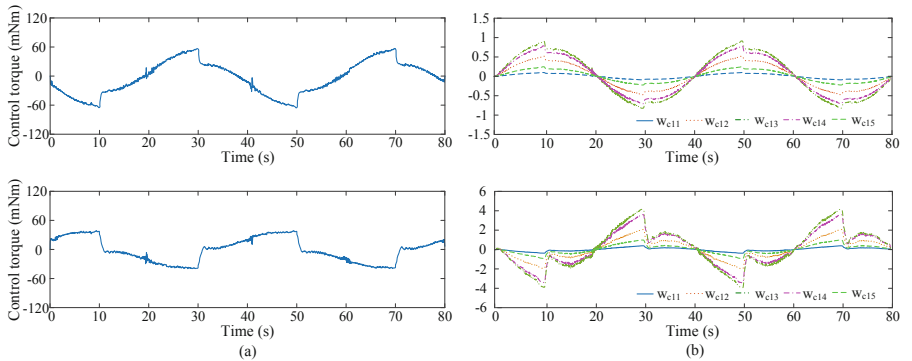


Fig. 2. Control torque and weight adjustment curves, (a) control torque, (b) weight adjustment.

5 Conclusion

In this paper, we develop a decentralized robust zero-sum neuro-optimal control scheme for MRMs. Based on local dynamic information of each joint module and the JTF technique, a model-based compensation controller and an uncertainty decomposition-based robust controller are designed for controlling MRM systems. Based on ADP algorithm, the HJI equation is solved by constructing the action-critic NNs, and then the decentralized robust zero-sum neuro-optimal control is developed. Finally, experimental results are illustrated to verify the effectiveness and advantages of the proposed method.

Acknowledgements. This work is supported by the National Natural Science Foundation of China (Grant nos. 61374051, 61773075 and 61703055) and the Scientific Technological Development Plan Project in Jilin Province of China (Grant nos. 20170204067GX, 20160520013JH and 20160414033GH).

References

1. Albu-Schaffer, A., Ott, C., Hirzinger, G.: A unified passivity-based control framework for position, torque, and impedance control of flexible joint robots. *Int. J. Robot. Res.* **26**(1), 23–39 (2007)
2. Liu, G., Liu, Y., Goldenberg, A.A.: Design, analysis, and control of a spring-assisted modular and reconfigurable robot. *IEEE/ASME Trans. Mech.* **16**(4), 695–706 (2011)
3. Lee, H., Kim, S., Chang, H., Kim, J.: Development of a compact optical torque sensor with decoupling axial-interference effects for pHRI. *Mechatronics* **52**, 90–101 (2018)
4. He, W., Dong, Y.: Adaptive fuzzy neural network control for a constrained robot using impedance learning. *IEEE Trans. Neural Netw. Learn. Syst.* **29**(4), 1174–1186 (2018)
5. Roveda, L., Pallucca, G., Pedrocchi, N., Braghin, F., Tosatti, L.M.: Iterative learning procedure with reinforcement for high-accuracy force tracking in robotized tasks. *IEEE Trans. Ind. Electron.* **14**(4), 1753–1763 (2018)
6. Leotta, D., Ruiz-del-Solar, J., Babuska, R.: Decentralized reinforcement learning of robot behaviors. *Artif. Intell.* **256**, 130–159 (2018)
7. Dong, B., Zhou, F., Liu, K., Li, Y.: Decentralized robust optimal control for modular robot manipulators via criticidentifier structure-based adaptive dynamic programming. *Neural Comput. Appl.* (2018). <https://doi.org/10.1007/s00521-018-3714-8>
8. Dong, B., Zhou, F., Liu, K., Li, Y.: Torque sensorless decentralized neuro-optimal control for modular and reconfigurable robots with uncertain environments. *Neurocomputing* **282**, 60–73 (2018)
9. Imura, J., Yokokohji, Y., Yoshikawa, T., Sugie, T.: Robust control of robot manipulators based on joint torque sensor information. *Int. J. Robot. Res.* **13**(5), 434–442 (1994)
10. Dong, B., Liu, K., Li, Y.: Decentralized control of harmonic drive based modular robot manipulator using only position measurements: theory and experimental verification. *J. Intell. Robot. Syst.* **88**, 3–18 (2017)