

A Learning-Based Decentralized Optimal Control Method for Modular and Reconfigurable Robots with Uncertain Environment

Bo Dong^{1,2}, Keping Liu^{1,2}, Hui Li^{1,2}, and Yuanchun Li^{1,2}(✉)

¹ Department of Control Science and Engineering,
Changchun University of Technology, Yan'an Avenue. 2055,
Changchun 130012, China

{dongbo,liukeping,lihui,liyc}@ccut.edu.cn

² State Key Laboratory of Management and Control for Complex Systems,
Institute of Automation, Chinese Academy of Sciences,
Zhongguancun East Road. 95, Beijing 100190, China

Abstract. This paper presents a novel decentralized control approach for modular and reconfigurable robots (MRRs) with uncertain environment contact under a learning-based optimal compensation strategy. Unlike the known optimal control methods that are merely suitable for specific classes of robotic systems without implementing dynamic compensations, in this investigation, the dynamic model of the MRR system is described as a synthesis of interconnected subsystems, in which the obtainable local dynamic information is utilized effectively to construct the feedback controller, thus making the decentralized optimal control problem of the MRR system be formulated as an optimal compensation issue of the model uncertainty. A policy iteration algorithm is employed to solve the Hamilton-Jacobi-Bellman (HJB) equation with a modified cost function, which is approximated by constructing a critic neural network, and then the approximate optimal control policy can be derived. The asymptotic stability of the closed-loop MRR system is proved by using the Lyapunov theory. At last, simulations are performed to verify the effectiveness of the proposed decentralized optimal control approach.

Keywords: Modular and reconfigurable robots · Decentralized control · Adaptive dynamic programming · Optimal control · Neural networks

1 Introduction

Modular and reconfigurable robots (MRRs) are comprised of the robot modules, which contain power supplies, processing systems, actuators and sensors. These modules are assembled to desirable configurations with standard electromechanical interfaces to satisfy the requirements of various tasks with complex working environments. Furthermore, MRRs need appropriate control systems that taking into consideration of both control precision and power consumption.

To ensure the stability and accuracy of trajectory tracking of the robotic systems, and simultaneously taking into account the optimal realization of the composite of control performance and power consumption have attached widespread attention in the robotics community. As an effective tool to address the optimal control problems in nonlinear systems, the adaptive dynamic programming (ADP) methodology has been considered as one of the key directions for the researches on designing discrete-time, continuous-time and data driven-based intelligent systems. In the past few years, numerous studies have been carried out on analytical description of robot manipulator systems under the ADP-based optimal control [1–4]. However, these methods are concentrated on centralized control, indeed, a centralized controller designed on the basis of an entire system may hardly be applicable for controlling MRRs. To avoid these problems, Liu et al. presented an online learning-based decentralized stabilization method [5,6] to deal with the decentralized optimal control problems of the classical nonlinear systems. However, the application of these methods are limited to address the optimal control problems of specific classes of robotic systems without implementing optimal dynamic compensation. Therefore, it is meaningful to investigate the decentralized optimal control approach by combining the model-based compensation control method and ADP-based optimal control policy for MRRs.

In this paper, a novel learning-based optimal control method is constructed to attain the decentralized controller design for MRRs with uncertain environment contact. The dynamic model of MRRs is described as a synthesis of interconnected subsystems, and the decentralized optimal control problem of the whole robotic system is reformulated as an optimal compensation issue of the model uncertainty. Moreover, a policy iteration-based learning algorithm is employed to solve the HJB equation with a modified cost function, and then a critic neural network is used to approximate the cost function, so that the approximate optimal control policy can be derived. Based on the Lyapunov theory, the asymptotic stability of the closed-loop robotic system are proved. Finally, simulations are conducted for 2-DOF MRRs with different configurations to investigate the effectiveness of the proposed decentralized optimal control approach.

2 Dynamic Model Formulation

By referencing the dynamic model of n -DOF MRR, which is proposed in our previous investigation [7], and the modeling approach for the robot manipulator with torque sensing [8], the dynamic model of the MRR system is described as a synthesis of interconnected subsystems, in which the subsystem dynamic model is formulated as:

$$I_{mi}\gamma_i\ddot{\theta}_i + f_i(\theta_i, \dot{\theta}_i) + Z_i(\theta, \dot{\theta}, \ddot{\theta}) + \frac{\tau_{fi}}{\gamma_i} = \tau_i, \quad (1)$$

where the subscript “ i ” represents the i th module, I_{mi} is the moment of inertia of the rotor about the axis of rotation, γ_i denotes the gear ratio, θ , $\dot{\theta}$ and $\ddot{\theta}$ represent the angular position, velocity and acceleration respectively, $f_i(\theta_i, \dot{\theta}_i)$ represents the frictional torque, $Z_i(\theta, \dot{\theta}, \ddot{\theta})$ indicates the interconnected joint coupling, τ_{fi} denotes the joint torque that including the dynamic information of the load torque and the external environment contact torque, and τ_i is the motor output torque. The friction term $f_i(\theta_i, \dot{\theta}_i)$ in (1), which is considered as a function of the joint position and velocity, is defined as:

$$f_i(\theta_i, \dot{\theta}_i) = \hat{b}_{fi}\dot{\theta}_i + \left(\hat{f}_{ci} + \hat{f}_{si}e^{(-\hat{f}_{\tau i}\dot{\theta}_i^2)} \right) \text{sgn}(\dot{\theta}_i) + f_{pi}(\theta_i, \dot{\theta}_i) + Y(\dot{\theta}_i)\tilde{F}_i, \quad (2)$$

where b_{fi} , f_{ci} , f_{si} , $f_{\tau i}$ and $f_{pi}(\theta_i, \dot{\theta}_i)$ are the nominal values of the friction model parameters, $\tilde{F}_i = [b_{fi} - \hat{b}_{fi}, f_{ci} - \hat{f}_{ci}, f_{si} - \hat{f}_{si}, f_{\tau i} - \hat{f}_{\tau i}]^T$ indicates the parametric uncertainty vector of the friction, \hat{b}_{fi} , \hat{f}_{ci} , \hat{f}_{si} and $\hat{f}_{\tau i}$ represent the estimated values of the friction parameters, and the vector $Y(\dot{\theta}_i)$ is defined as

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

where $|\tilde{F}_i| \leq \rho_{F_{il}}$ ($l = 1, 2, 3, 4$) and $|f_{pi}(\theta_i, \dot{\theta}_i)| \leq \rho_{f_{pi}}$ are the known up-bounds. Moreover, according to the torque estimation method proposed in [9], one can estimate the joint torque τ_{fi} by substituting the position measurements into a control-oriented harmonic drive model, which is represented as

$$\tau_{fi} = \frac{1}{c_f} \tan \left(c_f k_{f0} \left(\Delta\theta_i - \frac{\text{sgn}(\tau_{wi})(1 - e^{-c_w|\tau_{wi}|})}{\gamma_i c_w k_{w0}} \right) \right), \quad (4)$$

where $\Delta\theta_i = \theta_{fOi} - \theta_{wIi}/\gamma_i$ is the harmonic drive torsional angle θ_{wIi} and θ_{fOi} denote the motor-side angular position and the link-side angular position, which are measured by using the motor-side and the link-side encoders respectively, τ_{wi} denotes the wave generator torque, which can be obtained by using the motor torque command, c_f , c_w , k_{f0} and k_{w0} are positive constants to be determined. Additionally, the interconnected joint coupling term $Z_i(\theta, \dot{\theta}, \ddot{\theta})$ in (1) is defined as follows:

$$Z_i(\theta, \dot{\theta}, \ddot{\theta}) = I_{mi} \sum_{j=1}^{i-1} z_{mi}^T z_{lj} \ddot{\theta}_j + I_{mi} \sum_{j=2}^{i-1} \sum_{k=1}^{j-1} z_{mi}^T (z_{lk} \times z_{lj}) \dot{\theta}_k \dot{\theta}_j, \quad (5)$$

where z_{mi} , z_{lj} and z_{lk} are the unity vectors along the axis of rotation of the i th rotor, j th joint and k th joint respectively. In order to facilitate the analysis of the interconnected joint couplings, rewriting $I_{mi} \sum_{j=2}^{i-1} \sum_{k=1}^{j-1} z_{mi}^T (z_{lk} \times z_{lj}) \dot{\theta}_k \dot{\theta}_j$ and $I_{mi} \sum_{j=1}^{i-1} z_{mi}^T z_{lj} \ddot{\theta}_j$ as

$$I_{mi} \sum_{j=1}^{i-1} z_{mi}^T z_{lj} \ddot{\theta}_j = \sum_{j=1}^{i-1} U_j^i, \quad (6)$$

$$I_{mi} \sum_{j=2}^{i-1} \sum_{k=1}^{j-1} z_{mi}^T (z_{lk} \times z_{lj}) \dot{\theta}_k \dot{\theta}_j = \sum_{j=2}^{i-1} \sum_{k=1}^{j-1} V_{kj}^i. \quad (7)$$

where $\left| \sum_{j=1}^{i-1} U_j^i \right| \leq \rho_{Uj}$ and $\left| \sum_{j=2}^{i-1} \sum_{k=1}^{j-1} V_{kj}^i \right| \leq \rho_{Vj}$ are the known up-bounds.

Define the system state vector $x_i = [x_{i1}, x_{i2}]^T = [\theta_i, \dot{\theta}_i]^T \in R^{2 \times 1}$, and the control input $u_i = \tau_i \in R^{1 \times 1}$, $i = 1, 2, \dots, n$. Then, the state space of i th subsystem is formulated as follows:

$$S_i \begin{cases} \dot{x}_{i1} = x_{i2} \\ \dot{x}_{i2} = -(\phi_i(x_i, \dot{x}_i) + h_i(x, \dot{x}, \ddot{x})) + B_i u_i \\ y = x_{i1} \end{cases} \quad (8)$$

where $\phi_i(\theta_i, \dot{\theta}_i) = B_i \left(\hat{b}_{fi} \dot{\theta}_i + \left(\hat{f}_{ci} + \hat{f}_{si} e^{(-\hat{f}_{\tau i} \dot{\theta}_i^2)} \right) \text{sgn}(\dot{\theta}_i) + \frac{\tau_{fi}}{\gamma_i} \right)$ represents the modeled and estimated part of the dynamic model, $B_i = (I_{mi} \gamma_i)^{-1} \in R^+$, and $h_i(\theta, \dot{\theta}, \ddot{\theta}) = B_i \left(Y(\dot{\theta}_i) \tilde{F}_i + f_{pi}(\theta_i, \dot{\theta}_i) + \sum_{j=1}^{i-1} U_j^i + \sum_{j=2}^{i-1} \sum_{k=1}^{j-1} V_{kj}^i \right)$ is the model uncertainty term.

3 Learning-Based Decentralized Optimal Control Method

3.1 Problem Transformation

Let the desired position, velocity and acceleration of the i th joint be x_{id} , \dot{x}_{id} and \ddot{x}_{id} respectively. Then, consider the MRR system (8) with an continuously differentiable infinite horizon cost function written as:

$$J_i(s_i(e_i)) = \int_0^\infty \{U_i(s_i(e_i(\tau)), u_i(\tau)) + D_i^T D_i\} d\tau, \quad (9)$$

where $s_i(e_i)$ is defined as $s_i(e_i) = \alpha_{ei} e_i + \dot{e}_i$, in which $e_i = x_{i1} - x_{id}$ and $\dot{e}_i = \dot{x}_{i1} - \dot{x}_{id}$ denote the position and velocity tracking error of the i th joint respectively, α_{ei} is a determined constant, $U_i(s_i(e_i), u_i) = s_i^T Q_i s_i + u_i^T R_i u_i$ represents the utility function, in which $Q_i = Q_i^T$ and $R_i = R_i^T$ are determined positive constant matrixes, $D_i \in R^+$ denotes the up-bound function, then we can give a specifies form for the term D_i as:

$$D_i = B_i (|Y(\dot{x}_i)| \rho_{Fil} + \rho_{Ui} + \rho_{Vi} + \rho_{fpi}) \quad l = 1, 2, 3, 4, \quad (10)$$

Obviously, the model uncertainty term h_i and the up-bound function D_i satisfy the relation $h_i^T h_i \leq D_i^T D_i$. Then, for the MRR system (8) with the cost function (9), one can define the Hamiltonian function and the optimal cost function as

$$H_i(s_i, u_i, \nabla J_i) = U_i(s_i, u_i) + \nabla J_i(s_i)^T (-\phi_i - h_i + B_i u_i + \alpha_{ei} \dot{e}_i - \ddot{x}_{id}) + D_i^T D_i, \quad (11)$$

$$J_i^*(s_i) = \min_{u_i} \int_0^\infty \{U_i(s_i(e_i(\tau)), u_i(\tau)) + D_i^T D_i\} d\tau, \quad (12)$$

where $\nabla J_i(s_i) = \partial J_i(s_i) / \partial s_i$.

If the solution of J_i^* is existent and continuously differentiable, the optimal control law of the MRR system (8) can be computed as:

$$u_i^* = -\frac{1}{2} R_i^{-1} B_i^T \nabla J_i^*(s_i). \quad (13)$$

Rewriting the decentralized optimal control law u_i^* as the form of $u_i^* = u_{i1} + u_{i2}^*$ to deal with the terms of ϕ_i and h_i in (8) respectively, then one can modify the HJB equation as follows:

$$0 = U_i(s_i, u_i^*) + \nabla J_i^*(s_i)^T (-\phi_i - h_i + B_i u_{i1} + B_i u_{i2}^* + \alpha_{ei} \dot{e}_i - \ddot{x}_{id}) + D_i^T D_i. \quad (14)$$

Note that the terms $\alpha_{ei} \dot{e}_i$ and \ddot{x}_{id} are measurable and known, as well as the term ϕ_i includes the certain part of the dynamic model, which is directly obtainable, so that the feedback control law u_{i1} can be designed as

$$u_{i1} = \hat{b}_{fi} \dot{x}_i + \left(\hat{f}_{ci} + \hat{f}_{si} e^{(-\hat{f}_{\tau i} \dot{x}_i^2)} \right) \text{sgn}(\dot{x}_i) + \frac{\tau f_i}{\gamma_i} - B_i^{-1} (\alpha_{ei} \dot{e}_i) + B_i^{-1} \ddot{x}_{id}, \quad (15)$$

to compensate the modeled and estimated terms of the dynamic model.

3.2 Policy Iteration-Based Learning Algorithm

In this part, the online policy iteration-based learning algorithm is implemented to derive the solution of the HJB equation. The policy iteration algorithm consists of policy evaluation based on (13) and policy improvement based on (14). Specifically, the iterative procedure of the policy iteration algorithm with cost function (9) can be described in [10].

3.3 Neural Network Implementation

Neural network is a well-known tool for approximating nonlinear functions. Since the cost function is highly nonanalytic and nonlinear, in this part, the cost function $J_i(s_i)$ is approximated by using a single hidden layer neural network, which is defined as follows:

$$J_i(s_i) = W_{ci}^T \sigma_{ci}(s_i) + \varepsilon_{ci}, \quad (16)$$

where W_{ci} is the ideal weight vector, $\sigma_{ci}(s_i)$ denotes the activation function, and ε_{ci} is the approximation error of neural network. Then, the gradient of $\nabla J_i(s_i)$ is given as:

$$\nabla J_i(s_i) = (\nabla \sigma_{ci}(s_i))^T W_{ci} + \nabla \varepsilon_{ci}, \quad (17)$$

where $\nabla \sigma_{ci}(s_i) = \partial \sigma_{ci}(s_i) / \partial s_i$ and $\nabla \varepsilon_{ci}$ are the gradients of the activation function and the approximation error respectively. Since the ideal weight W_{ci} is always unknown, a critic neural network is built with approximated weight \hat{W}_{ci} to estimate the cost function as:

$$\hat{J}_i(s_i) = \hat{W}_{ci}^T \sigma_{ci}(s_i). \quad (18)$$

According to the definition of Hamiltonian (11) and the HJB equation (14), the Hamiltonian is further expressed by

$$H_i(s_i, u_i, W_{ci}) = U_i(s_i, u_i) + D_i^T D_i + (W_{ci}^T \sigma_{ci}(s_i)) \cdot (-\phi_i - h_i + B_i u_{i1} + B_i u_{i2} + \alpha_{ei} \dot{e}_i - \ddot{x}_{id}) - e_{cHi}, \quad (19)$$

where e_{cHi} is the residual error that is brought from the neural network approximation error, and defined as follows:

$$e_{cHi} = -\nabla \varepsilon_{ci}^T (-\phi_i - h_i + B_i u_{i1} + B_i u_{i2} + \alpha_{ei} \dot{e}_i - \ddot{x}_{id}). \quad (20)$$

The approximate Hamiltonian function, in the same manner, is given as:

$$\hat{H}_i(s_i, u_i, \hat{W}_{ci}) = U_i(s_i, u_i) + D_i^T D_i + (\hat{W}_{ci}^T \sigma_{ci}(s_i)) \cdot (-\phi_i - h_i + B_i u_{i1} + B_i u_{i2} + \alpha_{ei} \dot{e}_i - \ddot{x}_{id}). \quad (21)$$

Define the error function $e_{ci} = \hat{H}_i(s_i, u_i, \hat{W}_{ci}) - H_i(s_i, u_i, W_{ci})$, and the weight estimation error $\tilde{W}_{ci} = W_{ci} - \hat{W}_{ci}$, by combining (19) and (21), one obtain the expression of e_{ci} in terms of \tilde{W}_{ci} as

$$e_{ci} = e_{cHi} - \tilde{W}_{ci}^T \nabla \sigma_{ci}(s_i) \cdot (-\phi_i - h_i + B_i u_{i1} + B_i u_{i2} + \alpha_{ei} \dot{e}_i - \ddot{x}_{id}). \quad (22)$$

For the purpose of training and adjusting the weight information of the critic neural network, we employ the objective function $E_{ci} = \frac{1}{2} e_{ci}^T e_{ci}$, which is minimized by \hat{W}_{ci} . Moreover, the neural network weight is updated by using

$$\dot{\hat{W}}_{ci} = -\alpha_{ci} \left(\frac{\partial E_{ci}}{\partial \hat{W}_{ci}} \right), \quad (23)$$

where $\alpha_{ci} > 0$ denotes the learning rate of the critic neural network.

When implementing the online policy iteration algorithm to accomplish the policy improvement, one obtain the approximate optimal control law \hat{u}_{i2}^* as:

$$\hat{u}_{i2}^* = -\frac{1}{2} R_i^{-1} B_i^T (\nabla \sigma_{ci}(s_i))^T \hat{W}_{ci}. \quad (24)$$

From (24), one concludes that the optimal control law is derived depending on only critic neural network, unlike the conventional method that also relay on training of action neural network. Then, combining (15) and (24), the proposed decentralized optimal control law u_i^* is given as

$$u_i^* = \hat{b}_{fi} \dot{x}_i + \left(\hat{f}_{ci} + \hat{f}_{si} e^{(-\hat{f}_{\tau_i} \dot{x}_i^2)} \right) \text{sgn}(\dot{x}_i) + \frac{\tau_{fi}}{\gamma_i} - B_i^{-1} (\alpha_{ei} \dot{e}_i) + B_i^{-1} \ddot{x}_{id} - \frac{1}{2} R_i^{-1} B_i^T (\nabla \sigma_{ci}(s_i))^T \hat{W}_{ci}. \quad (25)$$

Theorem. *Given an environmental contacted modular and reconfigurable robot comprised of n modules, with the joint dynamic model as formulated in (1), and the model uncertainties that exist in (2), (6) and (7) with the up-bound function (10). The closed-loop robotic system is asymptotically stable under the decentralized optimal control law designed by (25), with the weight update law given by (23).*

Proof. Select the Lyapunov function candidate as

$$V(t) = \sum_{i=1}^n V_i(t) = \sum_{i=1}^n \left(\frac{1}{2} s_i^T s_i + J_i^*(s_i) \right). \quad (26)$$

The time derivative of (26) is obtained as

$$\dot{V}(t) = \sum_{i=1}^n s_i^T \begin{pmatrix} -\phi_i - h_i + B_i u_{i1} - \ddot{x}_{id} \\ + B_i u_{i2}^* + \alpha_{ei} \dot{e}_i \end{pmatrix} + \sum_{i=1}^n \nabla J_i^*(s_i)^T \begin{pmatrix} -\phi_i - h_i + B_i u_{i1} \\ + B_i u_{i2}^* + \alpha_{ei} \dot{e}_i - \ddot{x}_{id} \end{pmatrix}. \quad (27)$$

It is noted that the control law u_{i1} is designed as (15) for the purpose of compensating the certain terms ϕ_i , $\alpha_{ei} \dot{e}_i$ and \ddot{x}_{id} in the HJB equation (14), then, one can rewrite $\dot{V}(t)$ as

$$\dot{V}(t) = \sum_{i=1}^n (s_i^T (-h_i + B_i u_{i2}^*) - s_i^T Q_i s_i - u_i^{*T} R_i u_i^* - D_i^T D_i). \quad (28)$$

By Young's inequation, we know the (28) can be reformulated as:

$$\begin{aligned} \dot{V}(t) \leq & \sum_{i=1}^n \left(\frac{1}{2} \|s_i\|^2 + \frac{1}{2} \|h_i\|^2 + \frac{1}{2} \|s_i\|^2 + \frac{1}{2} \|B_i\|^2 \|u_{i2}^*\|^2 \right) \\ & - \sum_{i=1}^n \left(\lambda_{\min}(Q_i) \|s_i\|^2 + \lambda_{\min}(R_i) \|u_{i1}\|^2 \right) - \sum_{i=1}^n \left(\lambda_{\min}(R_i) \|u_{i2}^*\|^2 + \|D_i\|^2 \right), \end{aligned} \quad (29)$$

where $\lambda_{\min}(Q_i)$ and $\lambda_{\min}(R_i)$ denotes the minimum eigenvalue of Q_i and R_i respectively. Since h_i and D_i satisfy the relation $h_i^T h_i \leq D_i^T D_i$, then one obtains

$$\begin{aligned} \dot{V}(t) \leq & - \sum_{i=1}^n (\lambda_{\min}(Q_i) - 1) \|s_i\|^2 - \sum_{i=1}^n \frac{1}{2} \|D_i\|^2 \\ & - \sum_{i=1}^n \left(\lambda_{\min}(R_i) - \frac{1}{2} \|B_i\|^2 \right) \|u_{i2}^*\|^2 - \sum_{i=1}^n \lambda_{\min}(R_i) \|u_{i1}\|^2. \end{aligned} \quad (30)$$

Therefore, one concludes that $\dot{V}(t) \leq 0$ if the following condition holds

$$\left\{ \begin{array}{l} \lambda_{\min}(Q_i) \geq 1 \quad \lambda_{\min}(R_i) \geq \frac{1}{2I_{mi}^2 \gamma_i^2} \end{array} \right. \quad (31)$$

Besides, (30) also implies that $\dot{V}(t) < 0$ for any $s_i \neq 0$ when the condition (31) is hold, therefore, according to the Lyapunov theory, we conclude that the closed-loop MRR system is asymptotically stable under the proposed decentralized optimal control in (25). This concludes the proof of the Theorem.

4 Simulations

In order to verify the effectiveness of the proposed decentralized optimal control method, in this section, two 2-DOF MRRs with uncertain environment contact is used to conduct the simulations. The dynamic model, friction model parameters and desired trajectories are adopted by referring our previous investigation [11], and the parameters of the controller are given in Table 1. Let $Q_i = R_i = I$ (identity matrix), expressing the weigh vector \hat{W}_{ci} ($i = 1, 2$) as $\hat{W}_{c1} = [\hat{W}_{c11}, \hat{W}_{c12}, \hat{W}_{c13}]^T$ and $\hat{W}_{c2} = [\hat{W}_{c21}, \hat{W}_{c22}, \hat{W}_{c23}]^T$, and setting the activation function $\sigma_{ci}(s_i)$ ($i = 1, 2$) as the form of $\sigma_{c1}(s_1) = [e_1^2, e_1 s_1, s_1^2]$ and $\sigma_{c2}(s_2) = [e_2^2, e_2 s_2, s_2^2]$. Two types of external environment contacts are considered in the simulations that including continuous time-varying environment constraint (configuration A) and collision at random time point (configuration B). The time-varying constraint force and the the constant collision force are still follow our previous study [12].

Table 1. Parameters of the controller

Name	Value	Name	Value	Name	Value	Name	Value	Name	Value	Name	Value
I_{mi}	120	\hat{f}_{si}	4.0	k_{w0}	1.33	ρ_{Ui}	2.37	c_w	83.5	$\hat{f}_{\tau i}$	80
k_{f0}	8.3e+3	ρ_{Vi}	2.2519	c_f	8.9e-2	\hat{f}_{ci}	3.0	\hat{b}_{fi}	1.2	ρ_{fpi}	0.32
ρ_{Fi1}	0.3	ρ_{Fi2}	1.0	α_{ei}	0.5	ρ_{Fi3}	0.7	ρ_{Fi4}	20	α_{ci}	0.8

Figure 1 illustrated the position tracking error curves. For configuration A, in the first 10s, the tracking errors of both situations are relatively obvious due to the decentralized optimal controllers require a period of time for training the critic neural network, after that, the tracking errors may converge to a small range (less than $10e - 2rad$) since the model uncertainty has been compensated accurately. For configuration B, one observes that the instantaneous position deviations are occurred at the time points of 30s and 45s, which can be attributed to the influence of the environmental collision, after this, the tracking errors are converged rapidly under the action of the decentralized optimal control.

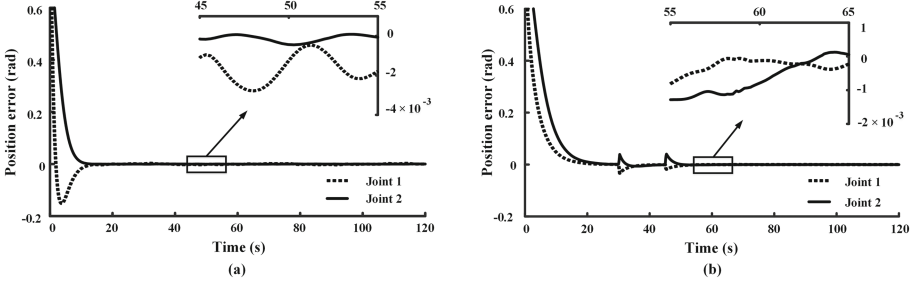


Fig. 1. Position tracking error curves. (a) Configuration A. (b) Configuration B

Figure 2 shows the control torque curves of the MRRs of configuration A and B. From this figure, one concludes that the control torques, which are continuous and smooth motor output torques, are available for implementing in the actual MRR systems. Besides, benefit from the proposed optimal control strategy, the torque consumptions are optimized in a suitable range for matching the output power of the motors in each joint module. Note that the decentralized optimal controller are suitable for different configurations of MRRs without the needs of readjusting the control parameters.

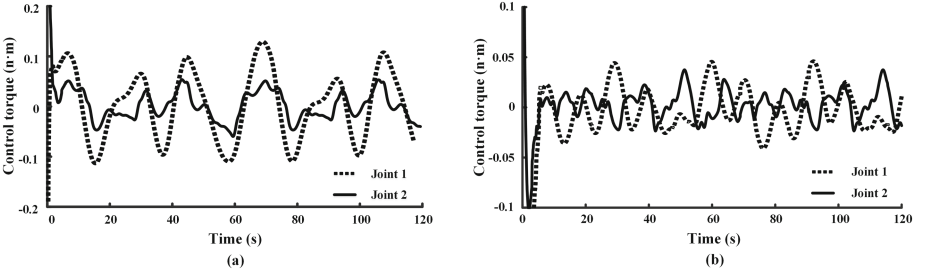


Fig. 2. Control torque curves. (a) Configuration A. (b) Configuration B.

During the implementation process of the online policy iteration algorithm and critic neural network training, for each isolated subsystem, we obtain that convergence results of the weights have occurred after two seconds for each situation. Actually, the weights of the critic neural networks converge to

$$\begin{cases} \hat{W}_{c1A} = [33.34 & 52.99 & 59.43] \\ \hat{W}_{c2A} = [45.66 & 50.16 & 46.78] \end{cases}, \begin{cases} \hat{W}_{c1B} = [23.20 & 4.70 & 43.96] \\ \hat{W}_{c2B} = [30.65 & 18.33 & 34.37] \end{cases}$$

for configuration A and configuration B respectively.

From the simulation results above, we conclude that the proposed decentralized optimal control method can provide accuracy and stability for MRRs to satisfy the requirements of various tasks with complex working environment.

5 Conclusions

This paper focus on investigating of MRRs with uncertain environment contact, and addresses the problem of decentralized control with a learning-based optimal compensation strategy. The dynamic model of MRRs are formulated as a synthesis of interconnected subsystems, and the optimal control problem for the whole robotic system is reformulated as an optimal compensation issue of the model uncertainty. The policy iteration algorithm is developed to solve the HJB equation by constructing a critic neural network, and then the approximate optimal control policy can be derived directly. The Lyapunov theory is used to prove the asymptotic stability of the closed-loop MRR systems. Finally, simulations are performed for two 2-DOF MRRs with uncertain environment contact to verify the effectiveness of the proposed decentralized optimal control method.

Acknowledgments. This work is supported by the National Natural Science Foundation of China (Grant no. 61374051), the State Key Laboratory of Management and Control for Complex Systems (Grant no. 20150102), the Scientific Technological Development Plan Project in Jilin Province of China (Grant nos. 20160520013JH, 20160414033GH and 20150520112JH) and the Science and Technology project of Jilin Provincial Education Department of China during the 13th Five-Year Plan Period (JJKH20170569KJ).

References

1. Patchaikani, P., Behera, L., Prasad, G.: A single network adaptive critic-based redundancy resolution scheme for robot manipulators. *IEEE Trans. Ind. Electron.* **59**, 3241–3253 (2012)
2. Tang, L., Liu, Y., Tong, S.: Adaptive neural control using reinforcement learning for a class of robot manipulator. *Neural Comput. Appl.* **25**, 135–141 (2014)
3. Li, Y., Chen, L., Tee, K., Li, Q.: Reinforcement learning control for coordinated manipulation of multi-robots. *Neurocomputing* **170**, 168–175 (2015)
4. Nagesh Rao, S., Lopes, G., Jeltsema, D., Babuska, R.: Passivity-based reinforcement learning control of a 2-DOF manipulator arm. *Mechatronics* **24**, 1001–1007 (2014)
5. Liu, D., Wang, D., Li, H.: Decentralized stabilization for a class of continuous-time nonlinear interconnected systems using online learning optimal control approach. *IEEE Trans. Neural Netw. Learn. Syst.* **25**, 418–428 (2014)
6. Wang, D., Liu, D., Mu, C., Ma, H.: Decentralized guaranteed cost control of interconnected systems with uncertainties: a learning-based optimal control strategy. *Neurocomputing* **214**, 297–306 (2016)
7. Dong, B., Li, Y.: Decentralized integral nested sliding mode control for time varying constrained modular and reconfigurable robot. *Adv. Mech. Eng.* **7**, 1–15 (2015)
8. Imura, J., Yokokohji, Y., Yoshikawa, T., Sugie, T.: Robust control of robot manipulators based on joint torque sensor information. *Int. J. Robot. Res.* **13**, 434–442 (1994)
9. Zhang, H., Ahmad, S., Liu, G.: Torque estimation for robotic joint with harmonic drive transmission based on position measurements. *IEEE Trans. Robot.* **31**, 322–330 (2015)

10. Zhao, B., Liu, D., Li, Y.: Online fault compensation control based on policy iteration algorithm for a class of affine non-linear systems with actuator failures. *IET Control Theory Appl.* **10**, 1816–1823 (2016)
11. Dong, B., Li, Y.: Decentralized reinforcement learning robust optimal tracking control for time varying constrained reconfigurable modular robot based on ACI and Q-function. *Math. Probl. Eng.* **2013**, 1–16 (2013)
12. Dong, B., Li, Y., Liu, K.: Decentralized control for harmonic drive-based modular and reconfigurable robots with uncertain environment contact. *Adv. Mech. Eng.* **9**, 1–14 (2017)

Neural Information Processing

24th International Conference, ICONIP 2017,

Guangzhou, China, November 14–18, 2017,

Proceedings, Part VI

Liu, D.; Xie, S.; Li, Y.; Zhao, D.; El-Alfy, E.-S.M. (Eds.)

2017, XVIII, 912 p. 415 illus., Softcover

ISBN: 978-3-319-70135-6