Model-free Adaptive Dynamic Programming Based Near-optimal Decentralized Tracking Control of Reconfigurable Manipulators

Bo Zhao* and Yuanchun Li

International Journal of Control, Automation and Systems 16(2) (2018) 478-490

ISSN:1598-6446 (print version)
eISSN:2005-4092 (electronic version)

To link to this article:
http://dx.doi.org/10.1007/s12555-016-0711-5

International Journal of Control, Automation and Systems 16(2) (2018) 478-490
http://dx.doi.org/10.1007/s12555-016-0711-5

ISSN:1598-6446 eISSN:2005-4092
http://www.springer.com/12555

Model-free Adaptive Dynamic Programming Based Near-optimal Decentralized Tracking Control of Reconfigurable Manipulators

Bo Zhao* and Yuanchun Li

Abstract: In this paper, a model-free near-optimal decentralized tracking control (DTC) scheme is developed for reconfigurable manipulators via adaptive dynamic programming algorithm. The proposed controller can be divided into two parts, namely local desired controller and local tracking error controller. In order to remove the norm-boundedness assumption of interconnections, desired states of coupled subsystems are employed to substitute their actual states. Using the local input/output data, the unknown subsystem dynamics of reconfigurable manipulators can be identified by constructing local neural network (NN) identifiers. With the help of the identified dynamics, the local desired control can be derived directly with corresponding desired states. Then, for tracking error subsystems, the local tracking error control is investigated by the approximate improved local cost function via local critic NN and the identified input gain matrix. To overcome the overall error caused by the substitution, identification and critic NN approximation, a robust compensation is added to construct the improved local cost function that reflects the overall error, regulation and control simultaneously. Therefore, the closed-loop tracking system can be guaranteed to be asymptotically stable via Lyapunov stability theorem. Two 2-degree of freedom reconfigurable manipulators with different configurations are employed to demonstrate the effectiveness of the proposed model-free near-optimal DTC scheme.

Keywords: Adaptive dynamic programming, decentralized tracking control, model-free, near-optimal, neural networks, reconfigurable manipulators.

1. INTRODUCTION

Reconfigurable manipulators [1], that consist of standard joint and link modules, are always considered as a set of subsystems interconnected by coupling torques. They can change their shapes to undertake different tasks by adding or removing modules. Hence, they have wide potential in industries, space explorations, smart manufacturing, high risk operations, and so on. However, it implies that modeling and controller design for each configuration are unacceptable to human operators. Therefore, to find a reasonable modeling and control strategy for reconfigurable manipulators become urgent problems.

In recent years, many efforts have been made to design controllers for reconfigurable manipulators. From the literature, control methods can be categorized into three strategies, namely centralized control [2–6], distributed control [7–10] and decentralized control [11, 12]. Decentralized control strategy can simplify the design complexity by using the local information of corresponding subsystems. Thus, the superiority leads it is well suited for reconfigurable manipulators that are assembled in modular structures with different degree of freedom (DOF). The main challenge in designing decentralized control lies in that how to get rid of negative effects from interconnections on control performance. Ababsa *et al.* [13] presented a decentralized control with two layers. In the first layer, based on current information perceived from environment, a genetic algorithm was used to generate well suited target configurations. While, in the second layer, a PacMan-like algorithm was used to plan the movement of modules to the target pattern emerged in the first layer. Butler *et al.* [14] presented generic locomotion algorithms inspired by cellular automata and geometric rules. It can be instantiated onto a variety of particular systems.

Inspired from that power efficiency reflects the operation health condition, Yuan *et al.* [15] developed a power efficiency estimation-based decentralized health monitoring and fault detection technique for modular and reconfigurable robot (MRR) with a joint torque sensor. Zhu

Manuscript received November 14, 2016; revised April 27, 2017 and June 13, 2017; accepted June 29, 2017. Recommended by Associate Editor Yongping Pan under the direction of Editor Jessie (Ju H.) Park. This work was supported in part by the National Natural Science Foundation of China under Grants 61603387, 61773075, and 61374051, and in part by the Scientific and Technological Development Plan Project in Jilin Province of China under Grants 20150520112JH and 20160414033GH.

Bo Zhao is with the State Key Laboratory of Management and Control for Complex Systems, Institute of Automation, Chinese Academy of Sciences, Beijing 100190, China (e-mail: zhaobo@ia.ac.cn). Yuanchun Li is with the Department of Control Science and Engineering, Changchun University of Technology, Changchun 130012, China (e-mail: liyc@ccut.edu.cn).
* Corresponding author.

et al. [16] developed a decentralized control based on virtue decomposition control with embedded field programmable gate array for MRR. This scheme led a precise control without joint torque sensing. Li *et al.* [17] proposed a decentralized robust control algorithm for MRRs based on backstepping techniques and a harmonic drive transmission system. By using proportional derivative controller with a saturated robust control, the uncertainties were compensated. We can see that the aforementioned methods all focused on controlling reconfigurable manipulators with available dynamics. However, modeling for changing configurations with different DOF is unexpected to operators. Therefore, decentralized control schemes without a prior knowledge of dynamics received more attention. Zhu *et al.* [18] used Takagi-Sugeno fuzzy logic systems to approximate the unknown dynamics of subsystems. The interconnection and fuzzy approximation error were removed by an adaptive sliding mode controller. Li *et al.* [19] identified the joint parameters of a modular robot by a genetic algorithm with fuzzy logic optimization. It can avoid converging to local optimal solutions and release dynamic control. Zhao *et al.* [20] proposed a neural network (NN) adaptive control scheme based on local joint information for fault-free reconfigurable manipulators. Then, fault compensation was implemented by employing a nonlinear velocity observer and a fault identifier to obtain acceptable control performance when it suffered from actuator faults. Despite of these methods have achieved excellent control performance, it is often desirable to design a controller which not only keeps systems stable, but also guarantees an adequate level of performance [21].

It is well known that the control objective can be achieved by optimal control, which is solved via dynamic programming for nonlinear systems. However, it is hardly to obtain the optimal feedback control since time-varying Hamilton-Jacobi-Bellman (HJB) equations are difficult to be solved. As an effective approximation approach, the adaptive dynamic programming (ADP) [22] which is aided by NNs [23] has consequently attracted much attention. There are several synonyms used for ADP, such as approximate dynamic programming [22], adaptive critic designs [24], neural dynamic programming [25], and reinforcement learning (RL) [26]. Werbos [22] classified the ADP schemes into heuristic dynamic programming (HDP), dual heuristic dynamic programming (DHP), action dependent HDP (ADHDP), and action dependent DHP (ADDHP). After that, other two approaches as globalized DHP (GDHP) and ADGDHP were proposed in [24]. In recent few years, ADP algorithms were developed further to solve control problems of continuous-time systems [27, 28], discrete-time systems [29], external disturbances and uncertainties [30], stochastic systems [31], trajectory tracking [21, 32, 33], control input saturation [34], fault tolerant [35–37], time-delay [38], zero-sum games [39, 40], etc. Meanwhile, ADP technique has been implemented further to tackle decentralized control problems. For linear interconnected systems, Jiang *et al.* [30] and Bian *et al.* [41] proposed robust ADP and policy iteration (PI) technique based decentralized controls for systems with dynamic uncertainties. Gao *et al.* [42] developed a data-driven output-feedback control policy based on PI and value iteration (VI) methods. Hioe *et al.* [43] presented decentralized control for dissipativity shaping problem via linear partial differential Hamilton-Jacobi equation. For nonlinear interconnected systems with available dynamics, Liu *et al.* [27] developed a decentralized control strategy with cost functions for subsystems, which were constructed by assumed upper bounded interconnections. Wang *et al.* [44] developed the decentralized guaranteed cost control by solving the modified HJB equation. The cost function was constructed for the overall plant. For unknown nonlinear interconnected systems, Liu *et al.* [45] proposed the decentralized stabilization scheme via actor-critic based online model-free integral PI algorithm. By assuming the input gain matrix was known and the unknown interconnection was weak, Mehraeen *et al.* [46] presented a decentralized near-optimal tracking control by using online tuned action NN and critic NN. In the previous ADP-based optimal tracking control methods, they were generally required to know the dynamics of plants, and the optimal tracking control can be designed directly. To the best of our knowledge, the result on the ADP-based decentralized tracking control (DTC) is rare for interconnected systems with unknown dynamics. Thus, it is a key problem to design the decentralized tracking controller via ADP for practical large-scale systems with unknown or unavailable dynamics, such as reconfigurable manipulators with changing shapes. This motivates our research.

In this paper, a near-optimal DTC is proposed for reconfigurable manipulators with unknown dynamics. By substituting the actual states of interconnections by their desired states, the common assumption on boundedness of interconnections can be removed. Then, the unknown subsystems of reconfigurable manipulators can be identified by using NNs and local input/output data. Therefore, the local desired control can be derived directly with the predefined desired trajectories. By employing a proper local cost function, the decentralized tracking control problem can be transformed into a near-optimal control problem. By using ADP technique, the approximate local tracking error control can be obtained by combining a critic NN with the identified control input matrix. Consider the substitution error, identification error and critic NN approximation error as the overall error, the local cost function is improved to reflect the overall error, regulation and control simultaneously. Therefore, the near-optimal DTC, which consists of the local desired control and local tracking error control, ensures the closed-loop reconfigurable manipulator system to be asymptotically stable. Simulations of two 2-DOF reconfigurable manipulators demonstrate the effectiveness of the developed model-free ADP based near-optimal DTC scheme.

The main contribution of this work have the following three aspects:

(i) The scheme extends the ADP technique to DTC problem for reconfigurable manipulators with unknown dynamics. The local input/output data is employed to identify the unknown subsystem model via local NN identifier. The local tracking control can be obtained by the identified dynamics. It implies that no matter how the configuration changes, system remodeling and controller redesign are not required any more. Thus, the proposed DTC strategy is more suitable for practice.

(ii) Interconnections are commonly assumed to be norm-bounded in existing methods. Different from them, the actual states of coupled subsystems in the interconnection are substituted by their desired states. It can remove the strict assumption, whose feasibility is difficult to be proved or ensured.

(iii) The improved local cost function reflects the overall error, regulation and control simultaneously. The overall error contains the substitution error, identification error and approximation error such that the closed-loop reconfigurable manipulator system can be guaranteed to be asymptotically stable by using the proposed DTC scheme, rather than ultimately uniformly bounded (UUB).

The rest of this paper is organized as follows: In Section 2, the problem statement is presented. In Section 3, the unknown dynamics of reconfigurable manipulators are identified by local NN identifiers, and the DTC is designed in detail. Then, the convergence and stability are discussed. In Section 4, two different 2-DOF reconfigurable manipulators are employed in simulation. In Section 5, the conclusion is drawn.

2. PROBLEM STATEMENT

The dynamic model of reconfigurable manipulator system with n-DOF can be formulated by Newton-Euler approach as

$$M(q)\ddot{q}+C(q,\dot{q})\dot{q}+G(q)=u, \tag{1}$$

where $q\in\mathbb{R}^n$ is the vector of joint displacements, $M(q)\in\mathbb{R}^{n\times n}$ is the inertia matrix, $C(q,\dot{q})\in\mathbb{R}^{n\times n}$ is the Coriolis and centripetal force, $G(q)\in\mathbb{R}^n$ is the gravity term, and $u\in\mathbb{R}^n$ is the applied joint torque.

In engineering practice, such as space manipulation and disaster rescue, reconfigurable manipulators consist of uncertain or large number of module joints. It brings complex control structure and heavy computational burden in traditional centralized control, which is difficult to be implemented. In order to relax the limitation, each joint is considered as a subsystem of the entire reconfigurable manipulator system interconnected by coupling torque. By separating terms only depending on local variables $(q_i,\dot{q}_i,\ddot{q}_i)$ from those terms of other joint variables, each subsystem dynamical model can be formulated in joint space as

$$M_i(q_i)\ddot{q}_i+C_i(q_i,\dot{q}_i)\dot{q}_i+G_i(q_i)+Z_i(q,\dot{q},\ddot{q})=u_i \tag{2}$$

with

$$\begin{aligned}
&Z_i(q,\dot{q},\ddot{q})\\
&=\left\{\sum_{j=1,j\neq i}^{n}M_{ij}(q)\ddot{q}_j+[M_{ii}(q)-M_i(q_i)]\ddot{q}_i\right\}\\
&\quad+\left\{\sum_{j=1,j\neq i}^{n}C_{ij}(q,\dot{q})\dot{q}_j+[C_{ii}(q,\dot{q})-C_i(q_i,\dot{q}_i)]\dot{q}_i\right\}\\
&\quad+\left[\bar{G}_i(q)-G_i(q_i)\right],
\end{aligned}$$

where $q_i,\dot{q}_i,\ddot{q}_i,\overline{G}_i(q)$ and $\bar{u}_i$ are the ith element of the vectors $q,\dot{q},\ddot{q},G(q)$ and u, $M_{ij}(q)$ and $C_{ij}(q,\dot{q})$ are the ijth element of the matrices $M(q)$ and $C(q,\dot{q})$, respectively. $Z_i(q,\dot{q},\ddot{q})$ is the coupling torque.

Let $x_i=[x_{i1},x_{i2}]^\top=[q_i,\dot{q}_i]^\top$, (2) can be expressed as

$$\begin{cases}\dot{x}_{i1}=x_{i2},\\ \dot{x}_{i2}=f_i(x_i)+g_i(x_i)u_i(x_i)+h_i(x),\\ y_i=x_i,\end{cases} \tag{3}$$

where x_i is the state of the ith subsystem, and

$$\begin{aligned}
f_i(x_i)&=M_i^{-1}(q_i)\left[-C_i(q_i,\dot{q}_i)\dot{q}_i-G_i(q_i)\right],\\
g_i(x_i)&=M_i^{-1}(q_i),\\
h_i(x)&=-M_i^{-1}(q_i)Z_i(q,\dot{q},\ddot{q}),
\end{aligned}$$

where $h_i(x)$ is the interconnection term, $x=[x_1,\ldots,x_n]^\top$ is the state vector of the entire reconfigurable manipulator.

Assumption 1: The nonlinear functions $f_i(x_i)$, $g_i(x_i)$ and $h_i(x)$ are Lipschitz and continuous in their arguments with $f_i(0)=0$, and the subsystem (3) is controllable.

Assumption 2: The desired trajectories q_{id} are twice differentiable and bounded as

$$\left\|\begin{matrix}q_{id}\\ \dot{q}_{id}\\ \ddot{q}_{id}\end{matrix}\right\|\leq q_{iA},$$

where $q_{iA}>0$ is a known constant.

In order to remove the norm-boundedness assumption on interconnections, the desired states of coupled subsystems are employed to substitute their actual ones. Thus, the interconnection term can be expressed as

$$h_i(x)=h_i(x_i,x_{jd})+\Delta h_i(x,x_{jd}),$$

where x_{jd} denotes the desired states of the coupled subsystems with $j=1,\ldots,i-1,i+1,\ldots N$, and $\Delta h_i(x,x_{jd}) = h_i(x)-h_i(x_i,x_{jd})$ denotes the substitution error. Thus, (3) becomes

$$\begin{cases} \dot{x}_{i1} = x_{i2}, \\ \dot{x}_{i2} = F_i(x_i,x_{jd}) + g_i(x_i)u_i(x_i) + \Delta h_i(x,x_{jd}), \\ y_i = x_i. \end{cases} \tag{4}$$

According to Assumption 1, $F_i(x_i,x_{jd}) = f_i(x_i) + h_i(x_i,x_{jd})$ is still Lipschitz continuous on a set $\Omega_i \in \mathbb{R}$. Since the interconnection satisfies the global Lipschitz condition, it thus implies

$$\|\Delta h_i(x,x_{jd})\| \leq \sum_{j=1,j\neq i}^{n} d_{ij}E_j, \tag{5}$$

where $E_j = \|x_j - x_{jd}\|$, and $d_{ij} \geq 0$ is an unknown global Lipschitz constant.

The objective of this paper is to find a DTC policy $u_i(x_i)$ such that the actual state of subsystem follows its desired trajectory. Furthermore, the set of DTC policies $u_1(x_1),\ldots,u_i(x_i),\ldots,u_n(x_n)$ ensure the closed-loop system of entire reconfigurable manipulator to be asymptotically stable in a near-optimal manner.

For ith subsystem, define the tracking error as

$$e_i = x_i - x_{id}.$$

where $e_i = [e_{i1}, e_{i2}]^{\mathsf{T}}$. From the dynamics of subsystem (4), the local tracking error dynamics can be expressed as

$$\dot{e}_i = \dot{x}_i - \dot{x}_{id}. \tag{6}$$

Therefore, associated with the local tracking error dynamics (6), the local tracking error control policy should minimize the following cost function

$$J_i(e_i(t)) = \int_t^{\infty} \left(\hat{\delta}_i \|e_i(\tau)\| + U_i(e_i(\tau), u_{ie}(\tau))\right) \mathrm{d}\tau, \tag{7}$$

where $U_i(e_i(t),u_{ie}(t)) = e_i^{\mathsf{T}}(t)Q_ie_i(t) + u_{ie}^{\mathsf{T}}(t)R_iu_{ie}(t)$ is the utility function, $U_i(0,0)=0$, and $U_i(e_i,u_{ie}) \geq 0$ for all e_i and u_{ie}, in which $Q_i \in \mathbb{R}^{2\times 2}$ and $R_i \in \mathbb{R}$ are positive definite matrices, $u_{ie} = u_i(x_i) - u_{id}(x_{id})$ is the local control input error, $u_{id}(x_{id})$ is the local desired control input, and $\hat{\delta}_i$ is the estimation of upper bound of the later defined overall error. It can be updated by

$$\dot{\hat{\delta}}_i = \Gamma_{i\delta}\|e_i\|. \tag{8}$$

From (7), we can see that it reflects the overall error, regulation and control simultaneously.

Remark 1: Actually, $\hat{\delta}_i\|e_i\|$ in the improved cost function (7) is a robust term, which is utilized to overcome the affection of the overall error caused by the substitution, identification and critic NN approximation. In this case, the closed-loop system can be guaranteed to be asymptotically stable, rather than UUB in many previous work.

Remark 2: Some existing works [27, 45] have constructed the cost function in similar form as (7), but they were always required to assume the interconnection to be upper bounded with a known scalar. However, the upper bound is difficult to or cannot be known in practice. Different from them, in our scheme, on one hand, it is not required to assume the interconnection to be upper bounded by substituting the actual states of coupling subsystems with their desired ones. On the other hand, the improved cost function (7) is updated adaptively to overcome the affection of the overall error, which contains the substitution error, identification error and critic NN approximation error.

3. MODEL-FREE ADP BASED DECENTRALIZED TRACKING CONTROLLER DESIGN

In this section, the detailed design procedure of model-free ADP based DTC in a near-optimal manner for reconfigurable manipulators is described.

3.1. Subsystem identification of reconfigurable manipulators

The system model cannot be available since the changing shape and uncertain DOF. Thus, according to the modularized property, the model-free strategy is an effective way to be taken into account. In this subsection, input/output data of the corresponding subsystems can be used to identify their dynamics.

Since the dynamics of subsystem (4) is unavailable, we use radial basis function (RBF) NNs to approximate the unknown terms as

$$F_i(x_i,x_{jd}) = W_{if}^{\mathsf{T}}\sigma_{if}(x_i,x_{jd}) + \varepsilon_{if}, \|\varepsilon_{if}\| \leq \varepsilon_{i1}, \tag{9}$$

$$g_i(x_i) = W_{ig}^{\mathsf{T}}\sigma_{ig}(x_i) + \varepsilon_{ig}, \|\varepsilon_{ig}\| \leq \varepsilon_{i2}, \tag{10}$$

where W_{if} and W_{ig} are ideal weight vectors from the hidden layer to the output layer, $\sigma_{if}(x_i,x_{jd})$ and $\sigma_{ig}(x_i)$ are basis function vectors, ε_{if} and ε_{ig} are approximation errors, and ε_{i1} and ε_{i2} are unknown positive constants.

Consider the ith subsystem of reconfigurable manipulator system (4), by employing local input/output data, the local NN identifier can be established as

$$\begin{cases} \dot{\hat{x}}_{i1} = \hat{x}_{i2} + k_{i1}\tilde{x}_{i1}, \\ \dot{\hat{x}}_{i2} = \hat{F}_i(\hat{x}_i,x_{jd}) + \hat{g}_i(\hat{x}_i)u_i(x_i) + k_{i2}\tilde{x}_{i2}, \end{cases} \tag{11}$$

where $\hat{x}_i = [\hat{x}_{i1}, \hat{x}_{i2}]^{\mathsf{T}} \in \mathbb{R}^2$ is the state vector of the identifier, $\tilde{x}_i = x_i - \hat{x}_i = [\tilde{x}_{i1}, \tilde{x}_{i2}]^{\mathsf{T}} \in \mathbb{R}^2$ is the local identification error vector, $\hat{F}_i(\hat{x}_i,x_{jd})$ and $\hat{g}_i(\hat{x}_i)$ are the estimation of nonlinear dynamics $F_i(x_i,x_{jd})$ and $g_i(x_i)$, respectively; k_{i1} and k_{i2} are positive constants.

Combining (4) with (11), the identification error dynamic can be described as

$$\begin{cases} \dot{\tilde{x}}_{i1} = \tilde{x}_{i2} - k_{i1}\tilde{x}_{i1}, \\ \dot{\tilde{x}}_{i2} = F_i(x_i, x_{jd}) - \hat{F}_i(\hat{x}_i, x_{jd}) + \Delta h_i(x, x_{jd}) - k_{i2}\tilde{x}_{i2} \\ \qquad + (g_i(x_i) - \hat{g}_i(\hat{x}_i))\, u_i(x_i). \end{cases}$$

Let $\hat{W}_{if}$ and $\hat{W}_{ig}$ be the estimations of W_{if} and W_{ig}, we have

$$\hat{F}_i(\hat{x}_i, x_{jd}) = \hat{W}_{if}^{\top} \sigma_{if}(\hat{x}_i, x_{jd}), \tag{12}$$

$$\hat{g}_i(\hat{x}_i) = \hat{W}_{ig}^{\top} \sigma_{ig}(\hat{x}_i), \tag{13}$$

where $\hat{W}_{if}$ and $\hat{W}_{ig}$ can be updated by the following adaptive laws as

$$\dot{\hat{W}}_{if} = \Gamma_{if}\tilde{x}_{i2}\sigma_{if}(\hat{x}_i, x_{jd}), \tag{14}$$

$$\dot{\hat{W}}_{ig} = \Gamma_{ig}\tilde{x}_{i2}\sigma_{ig}(\hat{x}_i)u_i, \tag{15}$$

where Γ_{if} and Γ_{ig} are positive constants.

Thus, from (9), (10), (12) and (13), we have

$$\begin{aligned} F_i(x_i, x_{jd}) - \hat{F}_i(\hat{x}_i, x_{jd}) &= W_{if}^{\top} \tilde{\sigma}_{if}(x_i, \hat{x}_i, x_{jd}) \\ &\quad + \tilde{W}_{if}^{\top} \sigma_{if}(\hat{x}_i, x_{jd}) + \varepsilon_{if}, \\ g_i(x_i) - \hat{g}_i(\hat{x}_i) &= W_{ig}^{\top} \tilde{\sigma}_{ig}(x_i, \hat{x}_i) + \tilde{W}_{ig}^{\top} \sigma_{ig}(\hat{x}_i) + \varepsilon_{ig}, \end{aligned}$$

where $\tilde{W}_{if} = W_{if} - \hat{W}_{if}$ and $\tilde{W}_{ig} = W_{ig} - \hat{W}_{ig}$ are the weight estimation errors, $\tilde{\sigma}_{if}(x_i, \hat{x}_i, x_{jd}) = \sigma_{if}(x_i, x_{jd}) - \hat{\sigma}_{if}(\hat{x}_i, x_{jd})$ and $\tilde{\sigma}_{ig}(x_i, \hat{x}_i) = \sigma_{ig}(x_i) - \sigma_{ig}(\hat{x}_i)$ are the estimation errors of activation functions, respectively.

Theorem 1: For ith subsystem of the reconfigurable manipulator (4), by using the local input/output data, the developed local identifier (11) can guarantee the identification error $\tilde{x}_i$ to be UUB with the help of the updated laws (14)-(15). In other words, the dynamic model of reconfigurable manipulator can be identified in a local manner.

Proof: Select a Lyapunov function candidate as

$$L_{i1} = \frac{1}{2}\tilde{x}_{i1}^2 + \frac{1}{2}\tilde{x}_{i2}^2 + \frac{1}{2}\tilde{W}_{if}^{\top}\Gamma_{if}^{-1}\tilde{W}_{if} + \frac{1}{2}\tilde{W}_{ig}^{\top}\Gamma_{ig}^{-1}\tilde{W}_{ig}. \tag{16}$$

The time derivative of (16) is

$$\begin{aligned} \dot{L}_{i1} &= \tilde{x}_{i1}\dot{\tilde{x}}_{i1} + \tilde{x}_{i2}\dot{\tilde{x}}_{i2} - \tilde{W}_{if}^{\top}\Gamma_{if}^{-1}\dot{\hat{W}}_{if} - \tilde{W}_{ig}^{\top}\Gamma_{ig}^{-1}\dot{\hat{W}}_{ig} \\ &= -k_{i1}\tilde{x}_{i1}^2 + \tilde{x}_{i1}\tilde{x}_{i2} - k_{i2}\tilde{x}_{i2}^2 \\ &\quad + \tilde{x}_{i2}\left(\tilde{W}_{if}^{\top}\sigma_{if}(\hat{x}_i, x_{jd}) + \tilde{W}_{ig}^{\top}\sigma_{ig}(\hat{x}_i)u_i(x_i) + w_{i1}\right) \\ &\quad - \tilde{W}_{if}^{\top}\Gamma_{if}^{-1}\dot{\hat{W}}_{if} - \tilde{W}_{ig}^{\top}\Gamma_{ig}^{-1}\dot{\hat{W}}_{ig}, \end{aligned} \tag{17}$$

where $w_{i1} = W_{if}^{\top}\tilde{\sigma}_{if}(x_i, \hat{x}_i, x_{jd}) + \left(W_{ig}^{\top}\tilde{\sigma}_{ig}(x_i, \hat{x}_i) + \varepsilon_{ig}\right)u_i + \varepsilon_{if} + \Delta h_i(x, x_{jd})$.

Assumption 3: The defined term w_{i1} is bounded, i.e., $|w_{i1}| \le \delta_{i1}$, where δ_{i1} is an unknown positive constant.

Substituting (14) and (15) into (17), we have

$$\begin{aligned} \dot{L}_{i1} &= -k_{i1}\tilde{x}_{i1}^2 + \tilde{x}_{i1}\tilde{x}_{i2} - k_{i2}\tilde{x}_{i2}^2 + \tilde{x}_{i2}w_{i1} \\ &\le -\left(k_{i1} - \frac{1}{2}\right)\tilde{x}_{i1}^2 - \frac{1}{2}\left(\tilde{x}_{i1} - \tilde{x}_{i2}\right)^2 \\ &\quad + |\tilde{x}_{i2}|\left(\left(k_{i2} - \frac{1}{2}\right)|\tilde{x}_{i2}| - \delta_{i1}\right). \end{aligned}$$

We can observe that $\dot{L}_{i1} \le 0$ when $\tilde{x}_{i2}$ lies outside of the compact set

$$\Omega_{\tilde{x}_{i2}} = \left\{\tilde{x}_{i2} : |\tilde{x}_{i2}| \le \frac{\delta_{i1}}{k_{i2} - \frac{1}{2}}\right\}$$

as long as $k_{i1} \ge \frac{1}{2}$ and $k_{i2} \ge \frac{1}{2}$. Therefore, according to Lyapunov's direct method, the identification error $\tilde{x}_{i2}$ can be guaranteed to be UUB. This completes the proof.

3.2. The decentralized tracking controller design

The optimal tracking control is generally composed of the feedforward control and the feedback control [21]. Thus, the local desired control as feedforward control should be designed as follows.

According to the local identifier (11), the identified dynamics of subsystem can be described as

$$\begin{cases} \dot{\hat{x}}_{i1} = \hat{x}_{i2}, \\ \dot{\hat{x}}_{i2} = \hat{F}_i(\hat{x}_i, x_{jd}) + \hat{g}_i(\hat{x}_i)u_i(\hat{x}_i). \end{cases}$$

By using the desired states x_{id} of the ith subsystem, the local desired control can be obtained as

$$u_{id}(x_{id}) = \hat{g}_i^{-1}(x_{id})\left(\dot{x}_{i2d} - \hat{F}_i(x_d)\right). \tag{18}$$

Remark 3: Unlike it must be assumed to be inversible [47–49], the assumption can be removed since $\hat{g}_i^{-1}(x_{id})$ in (18) denotes the estimated inertia matrix of ith subsystem of practical reconfigurable manipulator, which is a positive definite matrix.

In the following part, we turn to design the local tracking error control as a feedback control. To obtain a near-optimal control performance, the designed local tracking error control policy should be admissible. Therefore, before the algorithm is presented, the definition of admissible control is introduced [50].

Definition 1: For local tracking error dynamics (6), a local tracking error control policy $\mu_{ie}(e_i)$ is defined to be admissible if $\mu_{ie}(e_i)$ is continuous on a set Ω_i with $\mu_{ie}(0) = 0$, $\mu_{ie}(e_i)$ ensures the convergence of the ith subsystem (4) on Ω_i, and $J_i(e_i(t))$ is finite for all $e_i \in \Omega_i$.

For any admissible control policy $\mu_i(e_i) \in \psi_i(\Omega_i)$ of subsystem (4), where $\psi_i(\Omega_i)$ is the set of admissible control, if the improved cost function

$$V_i(e_i(t)) = \int_t^{\infty} \left(\hat{\delta}_i\|e_i(\tau)\| + U_i\left(e_i(\tau), \mu_{ie}(\tau)\right)\right) \mathrm{d}\tau \tag{19}$$

is continuously differentiable, then the infinitesimal version of (19) is the so-called local Lyapunov equation

$$0 = \hat{\delta}_i \|e_i(t)\| + U_i(e_i, \mu_{ie}) + (\nabla V_i(e_i))^{\mathsf{T}} \dot{e}_i \tag{20}$$

with $V_i(0) = 0$, and the term $\nabla V_i(e_i)$ denotes the partial derivative of $V_i(e_i)$ with respect to the local tracking error e_i, i.e., $\nabla V_i(e_i) = \partial V_i(e_i)/\partial e_i$.

Remark 4: Noticing that the right hand side of (19) is positive when $t > 0$, since $\hat{\delta}_i$ can be updated by (8). $\hat{\delta}_i$ will be guaranteed to be positive all the time as long as its initial value $\hat{\delta}_{i0} \geq 0$ for updating. Thus, (19) can be guaranteed to be a local Lyapunov equation with a proper initial value.

The Hamiltonian of the optimal problem and the optimal improved cost function can be formulated as

$$\begin{aligned} H_i\left(e_i, \mu_{ie}, \nabla V_i(e_i), \hat{\delta}_i\right) = \; & \hat{\delta}_i \|e_i(t)\| + U_i(e_i, \mu_{ie}) \\ & + (\nabla V_i(e_i))^{\mathsf{T}} \dot{e}_i, \end{aligned}$$

and

$$\begin{aligned} & J_i^*(e_i) \\ & = \min_{\mu_{ie} \in \psi_i(e_i)} \int_t^{\infty} \left(\hat{\delta}_i \|e_i(\tau)\| + U_i(e_i(\tau), \mu_{ie}(\tau))\right) \mathrm{d}\tau. \end{aligned}$$

Thus,

$$0 = \min_{\mu_{ie} \in \psi_i(e_i)} H_i\left(e_i, \mu_{ie}, \nabla J_i^*(e_i), \hat{\delta}_i\right),$$

where $\nabla J_i^*(e_i) = \partial J_i^*(e_i)/\partial e_i$. If the solution $J_i^*(e_i)$ exists and is continuously differentiable, the ideal local optimal tracking error control can be described as

$$u_{ie}^*(e_i) = -\frac{1}{2} R_i^{-1} \hat{g}_i^{\mathsf{T}}(x_i) \nabla J_i^*(e_i). \tag{21}$$

Remark 5: We can see that the error dynamics (6) depend on the desired trajectories x_{id}, it turns to be a time-varying system since the desired trajectories are time-varying. Different from [51, 52] tackled tracking control problems for finite horizon time-varying systems, the error system (6) in this paper is actually infinite horizon time-varying. Thus in this case, similar to and motivated by [53–57], the local optimal tracking error control (21) can be derived by solving the so-called local Lyapunov equation (20) which minimizes the improved local cost function (19).

In the process of ADP design, the cost function is often built by NN, which is a powerful tool for approximating nonlinear functions. Here, a local critic NN is employed to approximate the improved cost function on the compact set Ω_i as

$$V_i(e_i) = W_{ic}^{\mathsf{T}} \sigma_{ic}(e_i) + \varepsilon_{ic}(e_i),$$

where $W_{ic} \in \mathbb{R}^{l_i}$ is the ideal weight vector, $\sigma_{ic}(e_i) \in \mathbb{R}^{l_i}$ is the activation function, l_i is the number of neurons in the hidden layer, and $\varepsilon_{ic}(e_i)$ is the approximation error. Then, the gradient of $V_i(e_i)$ with respect to e_i is

$$\nabla V_i(e_i) = (\nabla \sigma_{ic}(e_i))^{\mathsf{T}} W_{ic} + \nabla \varepsilon_{ic}(e_i), \tag{22}$$

where $\nabla \sigma_{ic}(e_i) = \partial \sigma_{ic}(e_i)/\partial e_i \in \mathbb{R}^{l_i \times 2}$ and $\nabla \varepsilon_{ic}(e_i)$ are the gradients of the activation function and the approximation error, respectively.

Thus, according to (21), the ideal local optimal tracking error control can be derived as

$$\mu_{ie}(e_i) = -\frac{1}{2} R_i^{-1} \hat{g}_i^{\mathsf{T}}(x_i) \left((\nabla \sigma_{ic}(e_i))^{\mathsf{T}} W_{ic} + \nabla \varepsilon_{ic}(e_i)\right). \tag{23}$$

According to the framework of ADP-based approximate optimal control design, an approximate critic NN is established to estimate the infinite horizon cost function as

$$\hat{V}_i(e_i) = \hat{W}_{ic}^{\mathsf{T}} \sigma_{ic}(e_i), \tag{24}$$

where $\hat{W}_{ic} \in \mathbb{R}^{l_i}$ is the weight estimation. Similarly, the gradient of (24) with respect to e_i is

$$\nabla \hat{V}_i(e_i) = (\nabla \sigma_{ic}(e_i))^{\mathsf{T}} \hat{W}_{ic}. \tag{25}$$

Thus, using (23) and (25), the local optimal tracking error control can be obtained as

$$\hat{\mu}_{ie}(e_i) = -\frac{1}{2} R_i^{-1} \hat{g}_i^{\mathsf{T}}(x_i) (\nabla \sigma_{ic}(e_i))^{\mathsf{T}} \hat{W}_{ic}. \tag{26}$$

Considering (20) and (22), one can obtain

$$\begin{aligned} 0 = & \hat{\delta}_i \|e_i\| + U_i(e_i, \mu_{ie}) \\ & + \left((\nabla \sigma_{ic}(e_i))^{\mathsf{T}} W_{ic} + \nabla \varepsilon_{ic}(e_i)\right) \dot{e}_i. \end{aligned}$$

Therefore, the Hamiltonian can be expressed as

$$\begin{aligned} & H_i(e_i, \mu_{ie}, W_{ic}, \hat{\delta}_i) \\ & = \hat{\delta}_i \|e_i\| + U_i(e_i, \mu_{ie}) + W_{ic}^{\mathsf{T}} \nabla \sigma_i(e_i) \dot{e}_i \\ & = -\nabla \varepsilon_{ic}(x_i) \dot{e}_i = e_{icH}, \end{aligned} \tag{27}$$

where e_{icH} is the residual error caused by NN approximation.

In the same manner, the approximate Hamiltonian can be formulated by

$$\begin{aligned} & \hat{H}_i(e_i,\ \mu_{ie}, \hat{W}_{ic}, \hat{\delta}_i) \\ & = \hat{\delta}_i \|e_i\| + U_i(e_i, \mu_{ie}) + \hat{W}_{ic}^{\mathsf{T}} \nabla \sigma_i(e_i) \dot{e}_i = e_{ic}. \end{aligned} \tag{28}$$

Let $\theta_i = \nabla \sigma_i(e_i) \dot{e}_i$. From (27) and (28), we have

$$e_{ic} = e_{icH} - \tilde{W}_{ic}^{\mathsf{T}} \theta_i,$$

where $\tilde{W}_{ic} = W_{ic} - \hat{W}_{ic}$, and it can be updated as

$$\dot{\tilde{W}}_{ic} = -\dot{\hat{W}}_{ic} = \eta_{i1}(e_{icH} - \tilde{W}_{icH}^{\mathsf{T}} \theta_i) \theta_i, \tag{29}$$

where $\eta_{i1} > 0$ is the learning rate of the critic NN.

To obtain the update rule of the critic NN weight vector $\hat{W}_{ic}$, the objective function $E_{ic} = \frac{1}{2}e_{ic}^{\mathsf{T}}e_{ic}$ should be minimized with the steepest decent algorithm as

$$\dot{\hat{W}}_{ic} = -\dot{\tilde{W}}_{ic} = -\eta_{i1}e_{ic}\theta_i. \tag{30}$$

Theorem 2: For ith subsystem of reconfigurable manipulator, the local critic NN weight approximation error $\tilde{W}_{ic}$ can be guaranteed to be UUB as long as the weights of the local critic NN are updated by (30).

Proof: Select the Lyapunov function candidate as

$$L_{i2} = \frac{1}{2\eta_{i1}}\tilde{W}_{ic}^{\mathsf{T}}\tilde{W}_{ic}. \tag{31}$$

With the solution of (29), the time derivative of (31) is

$$\begin{aligned}\dot{L}_{i2} &= \frac{1}{\eta_{i1}}\tilde{W}_{ic}^{\mathsf{T}}\dot{\tilde{W}}_{ic}\\ &= \tilde{W}_{ic}^{\mathsf{T}}e_{icH}\theta_i - \left\|\tilde{W}_{ic}\theta_i\right\|^2\\ &\leq \frac{1}{2}e_{icH}^2 - \frac{1}{2}\left\|\tilde{W}_{ic}\theta_i\right\|^2.\end{aligned}$$

Assume $\|\theta_i\| \leq \theta_{iM}$, hence $\dot{L}_{i2} < 0$ whenever the approximation error of the local critic NN $\tilde{W}_{ic}$ lies outside of the compact set

$$\Omega_{\tilde{W}_{ic}} = \left\{\tilde{W}_{ic} : \left\|\tilde{W}_{ic}\right\| \leq \left\|\frac{e_{icH}}{\theta_{iM}}\right\|\right\}.$$

According to Lyapunov stability theorem, the weight approximate error of local critic NN is UUB. This completes the proof.

Now, we derive the near-optimal DTC by combining the local desired control (18) and local tracking error control (26) as

$$u_i = u_{id} + \hat{\mu}_{ie}. \tag{32}$$

3.3. Stability analysis

Theorem 3: Consider the n-DOF reconfigurable manipulator (1) and the improved local cost function (7). The proposed near-optimal DTC scheme (32) can guarantee the closed-loop system of reconfigurable manipulator to be asymptotically stable.

Proof: Select the Lyapunov function candidate as

$$L_{i3} = \frac{1}{2}e_i^{\mathsf{T}}e_i + V_i(e_i) + \Gamma_{i\delta}^{-1}\tilde{\delta}_i^2, \tag{33}$$

where $\tilde{\delta}_i = \delta_i - \hat{\delta}_i$ denotes the estimation error of δ_i.

According to (6), by adding and subtracting $F_i(x_d)$ and $\hat{g}_i(x_{id})$, the time derivative of (33) is

$$\begin{aligned}\dot{L}_{i3} &= e_i^{\mathsf{T}}\dot{e}_i + \nabla V_i^{\mathsf{T}}(e_i)\dot{e}_i - \Gamma_{i\delta}^{-1}\tilde{\delta}_i\dot{\hat{\delta}}_i\\ &= e_i^{\mathsf{T}}(\Theta_1 + \Theta_2 + \Theta_3) - \hat{\delta}_i\|e_i\|\\ &\quad - U_i(e_i,\mu_{ie}) - \Gamma_{i\delta}^{-1}\tilde{\delta}_i\dot{\hat{\delta}}_i,\end{aligned} \tag{34}$$

where $\Theta_1 = [e_{i1}, F_i(x_i,x_{jd}) - F_i(x_d)]^{\mathsf{T}}$, $\Theta_2 = [0, \hat{g}_i(x_{id})]^{\mathsf{T}}$, $\Theta_3 = [0, w_{i2}]^{\mathsf{T}}$, $w_{i2} = \tilde{F}_i(x_d) + \tilde{g}_i(x_{id})u_i(x_i) - \tilde{\mu}_{ie}$ with $\tilde{g}_i(x_i,x_{id}) = g_i(x_i) - \hat{g}_i(x_{id})$. And Θ_3 can be assumed to be boundedness, i.e., $\|\Theta_3\| \leq \delta_{i2}$.

As $F_i(\cdot)$ is locally Lipschitz, Θ_1 is locally Lipschitz, there exists a positive constant η_{if} such that $\|\Theta_1\| \leq \eta_{if}\|e_i\|$. Assuming that $\|\Theta_2\| \leq \eta_{ig}$, and introducing (32), (34) becomes

$$\begin{aligned}\dot{L}_{i3} &\leq \eta_{if}\|e_i\|^2 + e_i^{\mathsf{T}}\Theta_3 + e_i^{\mathsf{T}}\Delta h_i(x,x_{jd})\\ &\qquad + \eta_{ig}\|e_i\|\|\mu_{ie}\| - \hat{\delta}_i\|e_i\| - U_i(e_i,\mu_{ie}) - \Gamma_{i\delta}^{-1}\tilde{\delta}_i\dot{\hat{\delta}}_i\\ &\leq \eta_{if}\|e_i\|^2 + e_i^{\mathsf{T}}w_{i2} + e_i^{\mathsf{T}}\Delta h_i(x,x_{jd}) + \frac{1}{2}\|e_i\|^2\\ &\quad - \hat{\delta}_i\|e_i\| - \lambda_{\min}(Q_i)\|e_i\|^2 - \Gamma_{i\delta}^{-1}\tilde{\delta}_i\dot{\hat{\delta}}_i\\ &\quad - \left(\lambda_{\min}(R_i) - \eta_{ig}^2\right)\|\mu_{ie}\|^2.\end{aligned}$$

Thus, according to (5), we have

$$\begin{aligned}\dot{L}_3 &= \sum_{i=1}^{n}\dot{L}_{i3}\\ &\leq \sum_{i=1}^{n}\Big(\eta_{if}\|e_i\|^2 + e_i^{\mathsf{T}}\delta_{i2} + \|e_i\|\sum_{j=1,j\neq i}^{n}d_{ij}E_j\\ &\quad + \frac{1}{2}\|e_i\|^2 - \hat{\delta}_i\|e_i\| - \lambda_{\min}(Q_i)\|e_i\|^2\\ &\quad - \left(\lambda_{\min}(R_i) - \eta_{ig}^2\right)\|\mu_{ie}\|^2 - \Gamma_{i\delta}^{-1}\tilde{\delta}_i\dot{\hat{\delta}}_i\Big)\\ &\leq \sum_{i=1}^{n}\Big(\eta_{if}\|e_i\|^2 + e_i^{\mathsf{T}}\delta_{i2} + \frac{1}{2}\|e_i\|^2 - \hat{\delta}_i\|e_i\|\\ &\quad - \lambda_{\min}(Q_i)\|e_i\|^2 - \left(\lambda_{\min}(R_i) - \eta_{ig}^2\right)\|\mu_{ie}\|^2\\ &\quad - \Gamma_{i\delta}^{-1}\tilde{\delta}_i\dot{\hat{\delta}}_i\Big) + \sum_{i=1}^{n}\|e_i\|\sum_{j=1}^{n}d_{ij}E_j\\ &\leq \sum_{i=1}^{n}\Big(\eta_{if}\|e_i\|^2 + e_i^{\mathsf{T}}\delta_{i2} + \frac{1}{2}\|e_i\|^2 - \hat{\delta}_i\|e_i\|\\ &\quad - \lambda_{\min}(Q_i)\|e_i\|^2 - \left(\lambda_{\min}(R_i) - \eta_{ig}^2\right)\|\mu_{ie}\|^2\\ &\quad - \Gamma_{i\delta}^{-1}\tilde{\delta}_i\dot{\hat{\delta}}_i\Big) + \max_{ij}\{d_{ij}\}\sum_{i=1}^{n}\|e_i\|\sum_{j=1}^{n}E_j.\end{aligned} \tag{35}$$

Noticing that $\|e_i\| \leq \|e_j\| \Leftrightarrow E_i \leq E_j$. Using Chebyshev inequality, we have

$$\sum_{i=1}^{n}\|e_i\|\sum_{j=1}^{n}E_j \leq n\sum_{i=1}^{n}\|e_i\|E_j. \tag{36}$$

Combining (35) and (36), we have

$$\begin{aligned}\dot{L}_3 &\leq \sum_{i=1}^{n}\Big(\eta_{if}\|e_i\|^2 + \|e_i\|\delta_{i2} + \frac{1}{2}\|e_i\|^2 - \hat{\delta}_i\|e_i\|\\ &\quad - \lambda_{\min}(Q_i)\|e_i\|^2 - \left(\lambda_{\min}(R_i) - \eta_{ig}^2\right)\|\mu_{ie}\|^2\\ &\quad - \Gamma_{i\delta}^{-1}\tilde{\delta}_i\dot{\hat{\delta}}_i\Big) + n\max_{ij}\{d_{ij}\}\sum_{i=1}^{n}\|e_i\|E_i.\end{aligned} \tag{37}$$

Defining $\delta_i = \delta_{i2} + n\max_{ij}\{d_{ij}\}$, and we have

$$\dot{L}_3 \leq \sum_{i=1}^{n}\Big(-\Big(\lambda_{\min}(Q_i) - \eta_{if} - \frac{1}{2}\Big)\|e_i\|^2 + \|e_i\|\,\tilde{\delta}_i - \big(\lambda_{\min}(R_i) - \eta_{ig}^2\big)\|\mu_{ie}\|^2 - \Gamma_{i\delta}^{-1}\tilde{\delta}_i\dot{\hat{\delta}}_i\Big). \tag{38}$$

Substituting (8) into (38), we have

$$\dot{L}_3 \leq \sum_{i=1}^{n}\Big(-\Big(\lambda_{\min}(Q_i) - \eta_{if} - \frac{1}{2}\Big)\|e_i\|^2 - \big(\lambda_{\min}(R_i) - \eta_{ig}^2\big)\|\mu_{ie}\|^2\Big).$$

We can observe that $\dot{L}_{i3} \leq 0$ whenever the following conditions hold

$$\begin{cases} \lambda_{\min}(Q_i) \geq \eta_{if} + \frac{1}{2}, \\ \lambda_{\min}(R_i) \geq \eta_{ig}^2. \end{cases} \tag{39}$$

It implies that the developed model-free ADP based near-optimal DTC (32) ensures the closed-loop system of reconfigurable manipulator to be asymptotically stable based on Lyapunov stability theorem. This completes the proof.

Remark 6: We obviously know that the closed-loop system of reconfigurable manipulator is guaranteed to be asymptotically stable when conditions in (39) hold. However, it is difficult to choose appropriate weight matrices Q_i and R_i since the dynamics related parameters η_{if} and η_{ig} are unknown. Thus, these parameters are decided in simulations, which can provide suggestions before implementing the algorithm to reconfigurable manipulators.

Remark 7: Unlike [34] and [58], the improved local cost function is constructed by only tracking error and tracking error control in a local manner, rather than in an overall manner. Therefore, $\mu_{ie}^*(e_i)$ and $J_i^*(e_i)$ denote the near-optimal values for the overall reconfigurable manipulator. It means that the proposed scheme can only guarantee the local subsystem of reconfigurable manipulators to be optimal. That is to say, the proposed DTC can ensure tracking control performance of the overall reconfigurable manipulator in a near-optimal manner.

4. SIMULATION STUDIES

To show the effectiveness of the developed model-free ADP based near-optimal DTC scheme, simulations of two 2-DOF reconfigurable manipulators with different configurations [18] are given in this section.

Configuration A

The dynamic model of Configuration A is given as

$$M(q) = \begin{bmatrix} 0.36\cos(q_2)+0.6066 & 0.18\cos(q_2)+0.1233 \\ 0.18\cos(q_2)+0.1233 & 0.1233 \end{bmatrix},$$

$$C(q,\dot{q}) = \begin{bmatrix} -0.36\sin(q_2)\dot{q}_2 & -0.18\sin(q_2)\dot{q}_2 \\ 0.18\sin(q_2)(\dot{q}_1-\dot{q}_2) & 0.18\sin(q_2)\dot{q}_1 \end{bmatrix},$$

$$G(q) = \begin{bmatrix} -5.88\sin(q_1+q_2) - 17.64\sin(q_1) \\ -5.88\sin(q_1+q_2) \end{bmatrix}.$$

The desired trajectories of two subsystems are

$$q_d = \begin{bmatrix} q_{1d} \\ q_{2d} \end{bmatrix} = \begin{bmatrix} 0.2\sin(3t)+0.1\cos(4t) \\ 0.3\sin(2t)+0.2\cos(t) \end{bmatrix}.$$

In local identifier (11), we define $X_i = [x_i, x_{jd}]^\mathsf{T}$ as the state vector of the RBFNN, whose basis functions are chosen as Gaussian type as

$$\sigma_{if}(X_i) = \exp\left(\frac{-(X_i - c_{if})^\mathsf{T}(X_i - c_{if})}{b_{if}^2}\right),$$

$$\sigma_{ig}(x_i) = \exp\left(\frac{-(x_i - c_{ig})^\mathsf{T}(x_i - c_{ig})}{b_{ig}^2}\right),$$

where the centers of the basis functions are

$$c_{if} = \begin{bmatrix} -3 & -2 & -1 & 0 & 1 & 2 & 3 \\ -3 & -2 & -1 & 0 & 1 & 2 & 3 \\ -3 & -2 & -1 & 0 & 1 & 2 & 3 \\ -3 & -2 & -1 & 0 & 1 & 2 & 3 \end{bmatrix},$$

$$c_{ig} = \begin{bmatrix} -3 & -2 & -1 & 0 & 1 & 2 & 3 \\ -3 & -2 & -1 & 0 & 1 & 2 & 3 \end{bmatrix},$$

the widths of the basis functions are $b_{if} = b_{ig} = 0.5$.

Let the initial states of the subsystems be $x_{10} = x_{20} = [1,0]^\mathsf{T}$, the initial states of the observers be $\hat{x}_{10} = [2,-1]^\mathsf{T}$ and $\hat{x}_{20} = [1.5,-0.5]^\mathsf{T}$, the identification gains be $k_{i1} = 50$ and $k_{i2} = 250$, the RBFNN weights learning rates of the local identifier (11) be $\Gamma_{if} = 500$ and $\Gamma_{ig} = 1$, respectively. The improved local cost function (7) is approximated by a local critic NN, whose structure is chosen as 2-3-1 with 2 input neurons, 3 hidden neurons and 1 output neuron, and the weight vector as $\hat{W}_{ic} = \big[\hat{W}_{ic1}, \hat{W}_{ic2}, \hat{W}_{ic3}\big]^\mathsf{T}$ with initial values $\hat{W}_{1c} = [0.2, 1.5, 1.1]^\mathsf{T}$ and $\hat{W}_{2c} = [1.2, 0.8, 0.9]^\mathsf{T}$, as well as the learning rate be $\eta_{i1} = 0.0001$. The activation function of the critic NN is chosen as $\sigma_{ic}(e_i) = \big[e_{i1}^2, e_{i1}e_{i2}, e_{i2}^2\big]$. Let the weight learning rates of the critic NN be $\Gamma_{i\delta} = 7$, $Q_i = 2I_2$, $R_i = 0.0002I$, where I_n denotes the identity matrix with n dimensions, respectively.

The simulation results are shown as Figs. 1-4. The identification error curves in Fig. 1 show that subsystems of reconfigurable manipulator (2) can be identified to be UUB. The trajectories tracking curves are displayed in Fig. 2,

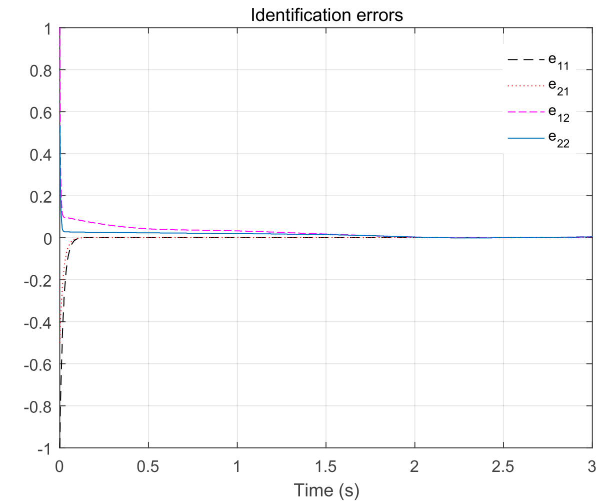

Fig. 1. The identification errors by using the local identifier of Configuration A.

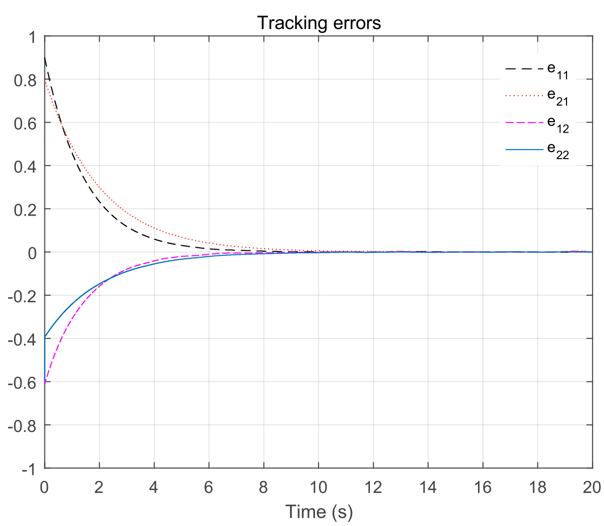

Fig. 2. The tracking errors of Configuration A.

from where we can see that the actual trajectories of subsystems can follow their desired ones after the reconfigurable manipulator is operated for a short time by using the developed near-optimal DTC scheme (32). Then, the tracking errors of subsystems illustrated as Fig. 3 show the excellent tracking performance intuitively. Fig. 4 described control input curves of the two subsystems. From these figures, the closed-loop reconfigurable manipulator system can be guaranteed to be asymptotically stable.

Configuration B

To further test the effectiveness of the proposed near-optimal DTC scheme, Configuration B which is also 2-DOF, but different shape from Configuration A, is employed in our simulation. The dynamic model of the Con-

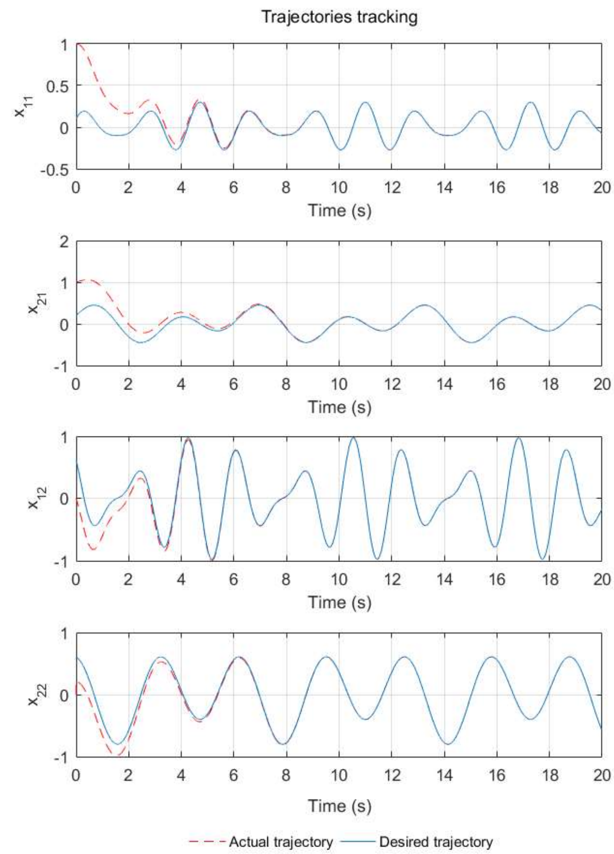

Fig. 3. The trajectories tracking performance of Configuration A.

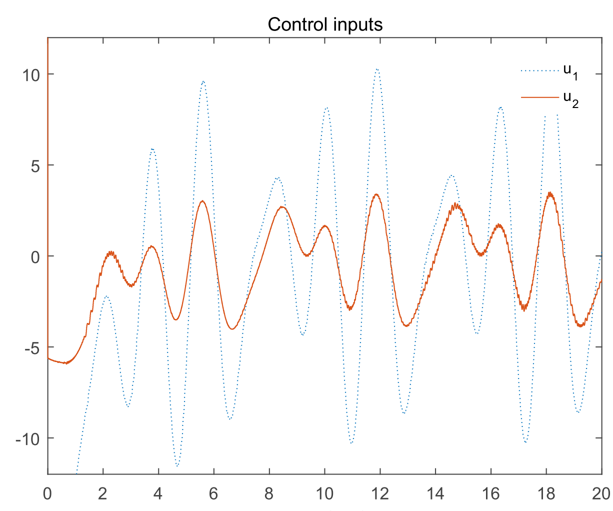

Fig. 4. The control inputs of Configuration A.

figuration B is described by

$$M(q) = \begin{bmatrix} 0.17 - 0.1166\cos^2(q_2) & -0.06\cos(q_2) \\ -0.06\cos(q_2) & 0.1233 \end{bmatrix},$$

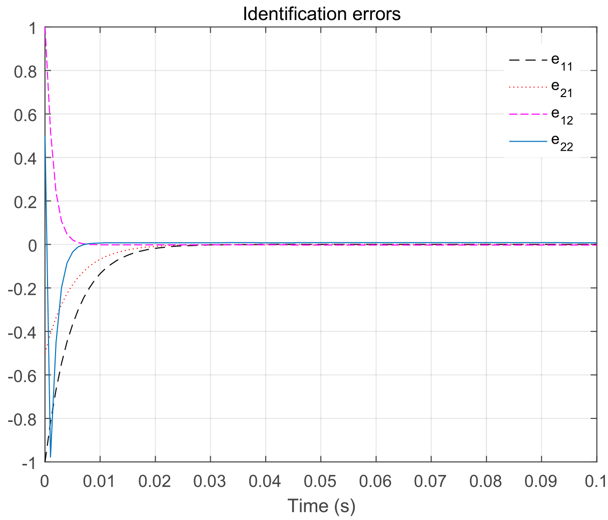

Fig. 5. The identification errors by using the local identifier of Configuration B.

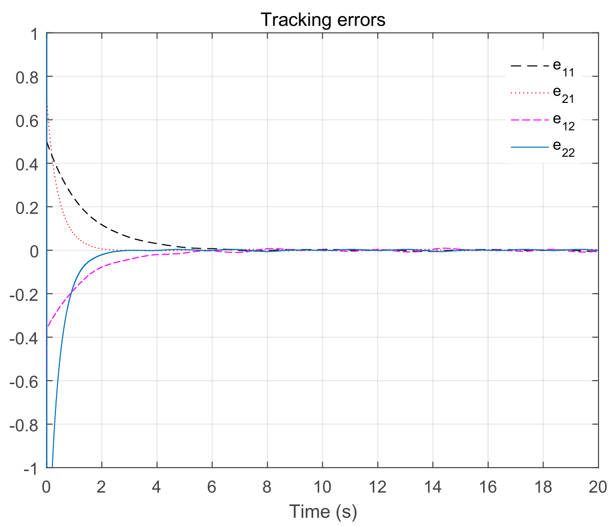

Fig. 6. The tracking errors of Configuration B.

$$C(q,\dot{q}) =$$
$$\begin{bmatrix} 0.1166\sin(2q_2)\dot{q}_2 & 0.06\sin(q_2)\dot{q}_2 \\ 0.06\sin(q_2)\dot{q}_2 - 0.0583\sin(q_2)\dot{q}_1 & -0.06\sin(q_2)\dot{q}_1 \end{bmatrix},$$
$$G(q) = \begin{bmatrix} 0 \\ -5.88\cos(q_2) \end{bmatrix}.$$

The desired trajectories are given as

$$q_d = \begin{bmatrix} q_{1d} \\ q_{2d} \end{bmatrix} = \begin{bmatrix} 0.5\cos(t) + 0.2\sin(3t) \\ 0.3\cos(3t) - 0.5\sin(2t) \end{bmatrix}.$$

Let the initial values of the subsystem trajectory states, initial values, structures and weight learning rates of local identifiers and local critic NNs to be the same as Configuration A. Let the identification gains be $k_{i1} = 200$ and $k_{i2} = 800$, the learning rate of the improved cost function be $\Gamma_{i\delta} = 60$, $Q_i = I_2$, $R_1 = 0.002I$ and $R_2 = 0.0001I$.

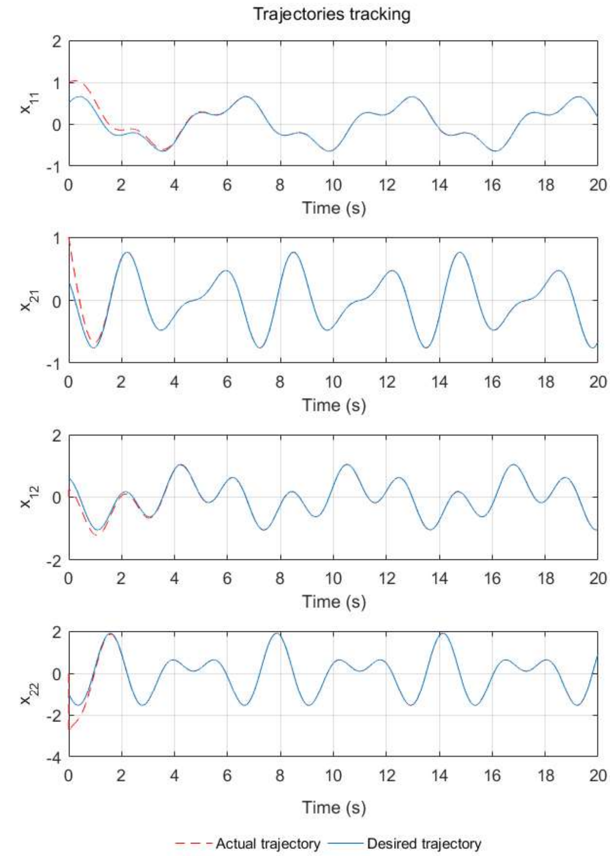

Fig. 7. The trajectories tracking performance of Configuration B.

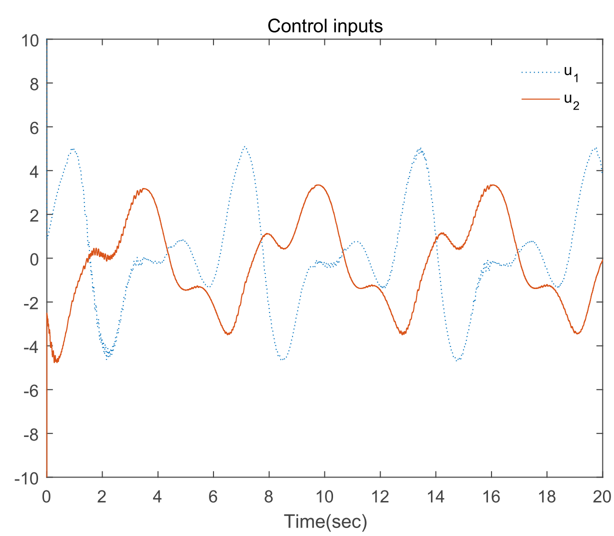

Fig. 8. The control inputs of Configuration B.

Figs. 5-8 display the simulation results. From Fig. 5, the identification performance of subsystem dynamic models can be shown by using the established local identifier (11). The trajectory tracking and their tracking errors are illustrated respectively in Fig. 6 and Fig. 7. The control inputs shown as Fig. 8 by using the proposed model-free ADP based near-optimal DTC ensure the trajectories of entire reconfigurable manipulator asymptotically follow the desired trajectories. Therefore, we can declare that the proposed near-optimal DTC can be applied to reconfigurable manipulators with different configurations successfully.

Remark 8: Many previous literature solved the control problems of reconfigurable manipulators. Different from [8–10], on one hand, the proposed DTC is designed model-free, rather than the available dynamics of reconfigurable manipulators. It implies that the proposed scheme can be utilized to reconfigurable manipulators with uncertain configurations and different DOF without changing control architecture. On the other hand, in contrast to our previous works [18, 20], the tracking delay is longer by using the proposed DTC in this paper. Fortunately, this DTC scheme presents a simple control structure in a near-optimal manner, which is different from the previous ones. In the future, composite learning algorithm [59, 60] with wavelet NNs, neuro-fuzzy networks, etc. are planed to be utilized to improve the tracking performance and avoid local optimal.

5. CONCLUSION

This paper addresses the DTC problems with ADP algorithm for reconfigurable manipulators in a near-optimal manner. The common boundedness assumption on interconnections is removed by substituting the actual states of the coupled subsystems with their desired states. Then, the unknown subsystem dynamics of reconfigurable manipulator can be identified by establishing local identifiers. With the help of identifiers, the local desired control can be derived with the desired trajectories of corresponding subsystems. By constructing improved local cost functions which reflect the overall error, regulation and control simultaneously, the local tracking error control can be obtained. Thus, the DTC can be constructed with the combination of the local desired control and local tracking error control. Two 2-DOF reconfigurable manipulators with different configurations are employed to demonstrate the effectiveness of proposed model-free ADP based near-optimal DTC scheme.

REFERENCES

[1] C. J. J. Paredis, H. B. Brown, and P. K.Khosla, "A rapidly deployable manipulator system," *Robotics and Autonomous Systems*, vol. 21, no. 3, pp. 289-304, 1997. [click]

[2] R. Naldi, F. Forte, A. Serrani, and L. Marconi, "Modeling and control of a class of modular aerial robots combining under actuated and fully actuated behavior," *IEEE Transactions on Control Systems Technology*, vol. 23, no. 5, pp. 1869-1885, 2015. [click]

[3] E. Meister, A. Gutenkunst, and P. Levi, "Dynamics and control of modular and self-reconfigurable robotic systems," *International Journal on Advances in Intelligent Systems*, vol. 6, no. 1 & 2, Number 1, 2013.

[4] Y. Liu, and Y. Li, "Dynamics and model-based control for mobile modular manipulators," *Robotica*, vol. 23, no. 6, pp. 795-797, 2005. [click]

[5] Y. Liu, and Y. Li, "Sliding mode adaptive neural-network control for nonholonomic mobile modular manipulators," *Journal of Intelligent and Robotic Systems*, vol. 44, no. 3, pp. 203-224, 2005.

[6] S. Kirchoff, and W. W.Melek, "A saturation-type robust controller for modular manipulators arms," *Mechatronics*, vol. 17, no. 4, pp. 175-190, 2007. [click]

[7] D. J. Christensen, U. P. Schultz, and K. Stoy, "A distributed and morphology-independent strategy for adaptive locomotion in self-reconfigurable modular robots," *Robotics and Autonomous Systems*, vol. 61, no. 9, pp. 1021-1035, 2013. [click]

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

[9] S. Ahmad, H. Zhang, and G. Liu, "Distributed fault detection for modular and reconfigurable robots with joint torque sensing: A prediction error based approach," *Mechatronics*, vol. 23, no. 6, pp. 607-616, 2013. [click]

[10] 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, 2008. [click]

[11] B. Zhao, Y. Li, and D. Liu, "Self-tuned local feedback gain based decentralized fault tolerant control for a class of large-scale nonlinear systems," *Neurocomputing*, vol. 235, pp. 147-156, 2017. [click]

[12] B. Zhao, C. Li, T. Ma, and Y. Li, "Multiple faults detection and isolation via decentralized sliding mode observer for reconfigurable manipulator," *Journal of Electrical Engineering & Technology*, vol. 10, no. 6, pp. 2393-2405, 2015. [click]

[13] T. Ababsa, N. Djedi, Y. Duthen, and S. C. Blanc, "Decentralized approach to evolve the structure of metamorphic robots," *2013 IEEE Symposium on Artificial Life (ALife)*, pp. 74-81, 2013. [click]

[14] Z. Butler, K. Kotay, D. Rus, and K. Tomita, "Generic decentralized control for lattice-based self-reconfigurable robots," *The International Journal of Robotics Research*, vol. 23, no. 9, pp. 919-937, 2004.

[15] J. Yuan, G. Liu, and B. Wu, "Power efficiency estimation-based health monitoring and fault detection of modular and reconfigurable robot," *IEEE Transactions on Industrial Electronics*, vol. 58, no. 10, pp. 4880-4887, 2011. [click]

[16] W. H. Zhu, T. Lamarche, E. Dupuis, D. Jameux, P. Barnard, and G. Liu, "Precision control of modular robot manipulators: the VDC approach with embedded FPGA," *IEEE Transactions on Robotics*, vol. 29, no. 5, pp. 1162-1179, 2013. [click]

[17] Z. Li, W. W. Melek, and C. Clark, "Decentralized robust control of robot manipulators with harmonic drive transmission and application to modular and reconfigurable serial arms," *Robotica*, vol. 27, no. 2, pp. 291-302, 2009. [click]

[18] M. Zhu, and Y. Li, "Decentralized adaptive fuzzy sliding mode control for reconfigurable modular manipulators," *International Journal of Robust and Nonlinear Control*, vol. 20, no. 4, pp. 472-488, 2010. [click]

[19] Y. Li, X. Liu, Z. Peng, and Y. Liu, "The identification of joint parameters for modular robots using fuzzy theory and a genetic algorithm," *Robotica*, vol. 20, no. 5, pp. 509-517, 2002.

[20] B. Zhao, and Y. Li, "Local joint information based active fault tolerant control for reconfigurable manipulator," *Nonlinear dynamics*, vol. 77, no. 3, pp. 859-876, 2014.

[21] X. Yang, D. Liu, Q. Wei, and D. Wang, "Guaranteed cost neural tracking control for a class of uncertain nonlinear systems using adaptive dynamic programming," *Neurocomputing*, vol. 198, pp. 80-90, 2016. [click]

[22] P. J. Werbos, "Approximate dynamic programming for real-time control and neural modeling," *Handbook of Intelligent Control: Neural, Fuzzy, and Adaptive Approaches*, vol. 15, pp. 493-525, 1992.

[23] Y. Pan, and H. Yu, "Biomimetic hybrid feedback feedforward neural-network learning control," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 28, no. 6, pp. 1481-1487, 2017. [click]

[24] D. V. Prokhorov, and D. C. Wunsch, "Adaptive critic designs," *IEEE Transactions on Neural Networks*, vol. 8, no. 5, pp. 997-1007, 1997. [click]

[25] D. P. Bertsekas, and J. N. Tsitsiklis, "Neuro-dynamic programming (optimization and neural computation series 3)," *Athena Scientific*, vol. 7, pp. 15-23, 1996.

[26] F. L. Lewis, and K. G. Vamvoudakis, "Reinforcement learning for partially observable dynamic processes: Adaptive dynamic programming using measured output data," *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, vol. 41, no. 1, pp. 14-25, 2011.

[27] 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 Transactions on Neural Networks and Learning Systems*, vol. 25, no. 2, pp. 418-428, 2014. [click]

[28] 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 Transactions on Systems, Man, and Cybernetics: Systems*, DOI: 10.1109/TSMC.2017.2690665, 2017.

[29] H. Lin, Q. Wei, and D. Liu, "Online identifier-actor-critic algorithm for optimal control of nonlinear systems," *Optimal Control Applications and Methods*, vol. 38, no. 3, pp. 317-335, 2017.

[30] Y. Jiang, and Z. P. Jiang, "Robust adaptive dynamic programming for large-scale systems with an application to multimachine power systems," *IEEE Transactions on Circuits and Systems II: Express Briefs*, vol. 59, no. 10, pp. 693-697, 2012. [click]

[31] W. C. Wong, and J. H. Lee, "A reinforcement learning-based scheme for direct adaptive optimal control of linear stochastic systems," *Optimal Control Applications and Methods*, vol. 31, no. 4, pp. 365-374, 2010.

[32] M. Sharma, and A.Verma, "Wavelet reduced order observer based adaptive tracking control for a class of uncertain nonlinear systems using reinforcement learning," *International Journal of Control, Automation and Systems*, vol. 11, no. 3, pp. 496-502, 2013. [click]

[33] B. Zhao, D. Liu, X. Yang, and D. Liu, " Observer-critic structure based adaptive dynamic programming for decentralized tracking control of unknown large-scale nonlinear systems," *International Journal of Systems Science*, 2017, vol. 48, no. 9, pp. 1978-1989, 2017. [click]

[34] X. Yang, D. Liu, and D. Wang, "Reinforcement learning for adaptive optimal control of unknown continuous-time nonlinear systems with input constraints," *International Journal of Control*, vol. 87, no. 3, pp. 553-566, 2014. [click]

[35] 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 & Applications*, vol. 10, no. 16, pp. 1816-1823, 2016. [click]

[36] B. Zhao, D. Liu, and Y. Li, "Observer based adaptive dynamic programming for fault tolerant control of a class of nonlinear systems," *Information Sciences*, vol. 384, pp. 21-33, 2017. [click]

[37] S. J. Chang, J. Y. Lee, J. B. Park, and Y. H. Choi, "An online fault tolerant actor-critic neuro-control for a class of nonlinear systems using neural network HJB approach," *International Journal of Control, Automation and Systems*, vol. 13, no. 2, pp. 311-318, 2015. [click]

[38] C. L. Chen, D. Y. Sun, and C. Y. Chang, "Numerical solution of time-delayed optimal control problems by iterative dynamic programming," *Optimal Control Applications and Methods*, vol. 21, no. 3, pp. 91-105, 2000. [click]

[39] H. Li, D. Liu, and D. Wang, "Integral reinforcement learning for linear continuous-time zero-sum games with completely unknown dynamics," *IEEE Transactions on Automation Science and Engineering*, vol. 11, no. 3, pp. 706-714, 2014. [click]

[40] S. Yasini, M. B. N. Sistani, and A. Karimpour, "Approximate dynamic programming for two-player zero-sum game related to H_∞ control of unknown nonlinear continuous-time systems," *International Journal of Control, Automation and Systems*, vol. 13, no. 1, pp. 99-109, 2015. [click]

[41] T. Bian, Y. Jiang, and Z. P. Jiang, "Decentralized adaptive optimal control of large-scale systems with application to power systems," *IEEE Transactions on Industrial Electronics*, vol. 62, no.4, pp. 2439-2447, 2015. [click]

[42] W. Gao, Y. Jiang, Z. P. Jiang, and T. Chai, "Output-feedback adaptive optimal control of interconnected systems based on robust adaptive dynamic programming," *Automatica*, vol. 72, pp. 37-45, 2016. [click]

[43] D. Hioe, N. Hudon, and J. Bao, "Decentralized nonlinear control of process networks based on dissipativity-A Hamilton-Jacobi equation approach," *Journal of Process Control*, vol. 24, no. 3, pp. 172-187, 2014. [click]

[44] D. Wang, D. Liu, C. Mu, and H. Ma, "Decentralized guaranteed cost control of interconnected systems with uncertainties: a learning-based optimal control strategy," *Neurocomputing*, vol. 214, pp. 297-306, 2016. [click]

[45] D. Liu, C. Li, H. Li, D. Wang, and H. Ma, "Neural-network-based decentralized control of continuous-time nonlinear interconnected systems with unknown dynamics," *Neurocomputing*, vol. 165, pp. 90-98, 2015. [click]

[46] S. Mehraeen, and S. Jagannathan, "Decentralized optimal control of a class of interconnected nonlinear discrete-time systems by using online Hamilton-Jacobi-Bellman formulation," *IEEE Transactions on Neural Networks*, vol. 22, no. 11, pp. 1757-1769, 2011. [click]

[47] D. Wang, D. Liu, and Q. Wei, "Finite-horizon neuro-optimal tracking control for a class of discrete-time nonlinear systems using adaptive dynamic programming approach," *Neurocomputing*, vol. 78, no. 1, pp. 14-22, 2012.

[48] Y. Huang, and D. Liu, "Neural-network-based optimal tracking control scheme for a class of unknown discrete-time nonlinear systems using iterative ADP algorithm," *Neurocomputing*, vol. 125, pp. 46-56, 2014. [click]

[49] Q. Wei, D. Liu, and Y. Xu, "Neuro-optimal tracking control for a class of discrete-time nonlinear systems via generalized value iteration adaptive dynamic programming approach," *Soft Computing*, vol. 20, no. 2, pp. 697-706, 2016. [click]

[50] Y. M. Park, M. S. Choi, and K. Y. Lee, "An optimal tracking neuro-controller for nonlinear dynamic systems," *IEEE Transactions on Neural Networks*, vol. 7, no. 5, pp. 1099-1110, 1996. [click]

[51] T. Cheng, F. L. Lewis, and M. Abu-Khalaf, "Fixed-final-time-constrained optimal control of nonlinear systems using neural network HJB approach," *IEEE Transactions on Neural Networks*, vol. 18, no. 6, pp. 1725-1737, 2007. [click]

[52] H. Wu, M. Li, and L. Guo, "Finite-Horizon Approximate Optimal Guaranteed Cost Control of Uncertain Nonlinear Systems With Application to Mars Entry Guidance," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 26. no. 7, pp. 1456-1467, 2015. [click]

[53] B. Kiumarsi, F. L. Lewis, H. Modares, A. Karimpour, and M. B. Naghibi-Sistani, "Reinforcement Q-learning for optimal tracking control of linear discrete-time systems with unknown dynamics," *Automatica*, vol. 50, no. 4, pp. 1167-1175, 2014. [click]

[54] H. Modares, and F. L.Lewis, "Linear quadratic tracking control of partially-unknown continuous-time systems using reinforcement learning," *IEEE Transactions on Automatic control*, vol. 59, no. 11, pp. 3051-3056, 2014. [click]

[55] H. Zhang, L. Cui, X. Zhang, and Y. Luo, "Data-driven robust approximate optimal tracking control for unknown general nonlinear systems using adaptive dynamic programming method," *IEEE Transactions on Neural Networks*, vol. 22, no. 12, pp. 2226-2236, 2011. [click]

[56] Y. Zhu, D. Zhao, and X. Li, "Using reinforcement learning techniques to solve continuous-time non-linear optimal tracking problem without system dynamics," *IET Control Theory & Applications*, vol. 10, no. 12, pp. 1339-1347, 2016. [click]

[57] R. Kamalapurkar, L. Andrews, P. Walters, and W. E. Dixon, "Model-based reinforcement learning for infinite-horizon approximate optimal tracking," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 28, no. 3, pp. 753-758, 2017. [click]

[58] A. Al-Tamimi, F. L. Lewis, and M. Abu-Khalaf, "Discrete-time nonlinear HJB solution using approximate dynamic programming: convergence proof," *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, vol. 38, no. 4, pp. 943-949, 2008. [click]

[59] Y. Pan, and H. Yu, "Composite learning from adaptive dynamic surface control," *IEEE Transactions on Automatic Control*, vol. 61, no. 9, pp. 2603-2609, 2016. [click]

[60] Y. Pan, J. Zhang, and H. Yu, "Model reference composite learning control without persistency of excitation," *IET Control Theory & Applications*, vol. 10, no. 16, pp. 1963-1971, 2016. [click]

Bo Zhao received his B.S. degree in Automation, and Ph.D. degree in Control Science and Engineering, all from Jilin University, Changchun, China, in 2009 and 2014, respectively. He is currently a Post-Doctoral Fellow with the State Key Laboratory of Management and Control for Complex Systems, Institute of Automation, Chinese Academy of Sciences, Beijing, China, from 2014. His research interests include fault diagnosis and tolerant control, neural-network based control, robot control, optimal control, and adaptive dynamic programming.

Yuanchun Li received his Ph.D. degree in General Mechanics from Harbin Institute of Technology, Harbin, China, in 1990. He was an Lecturer and Associate Professor at Jilin University of Technology, Changchun, China in 1990 and 1992, respectively. He was a Full Professor at Jilin University, Changchun, China from 1996 to 2012. He is currently a Full Professor in Control Science and Engineering, Changchun University of Technology, Changchun, China. His research interests include complex system modeling and robot control, intelligent control.