Reinforcement Learning Control of a Flexible Two-link Manipulator: An Experimental Investigation

Wei He, Hejia Gao, Chen Zhou, Chenguang Yang, and Zhijun Li

***Abstract*—This paper discusses control design and experiment validation of a flexible two-link manipulator (FTLM) system represented by ordinary differential equations (ODEs). A reinforcement learning (RL) control strategy is developed that is based on actor-critic structure to enable vibration suppression while retaining trajectory tracking. Subsequently, the closed-loop system with the proposed RL control algorithm is proved to be semi-global uniform ultimate bounded (SGUUB) by Lyapunov's direct method. In the simulations, the control approach presented has been tested on the discretized ODE dynamic model and the analytical claims have been justified under the existence of uncertainty. Eventually, a series of experiments in a Quanser laboratory platform are investigated to demonstrate the effectiveness of the presented control and its application effect is compared with PD control.**

***Index Terms*—Reinforcement Learning, Vibration Control, Robots, Flexible Structure, Neural Networks.**

Nomenclature

XOY	The inertial frame of reference
x_iOy_i	The rotating coordinate system, $i = 1,\ 2$.
M_i	The mass of the ith link
m_i	The tip mass of the ith link
L_i	The length of the ith link
l_i	The same length of each element for the ith link
EI_i	The bending rigidity of the ith link
ρ_i	The mass per unit length of the ith link
I_{hi}	The ith hub moment of inertia
I_{ti}	The ith tip load moment of inertia
I_{oi}	The ith link moment of inertia
$\theta_i(t)$	The angular position of the ith hub
$\omega_i(x_i,t)$	The elastic deformation at x_i
$\tau_i(t)$	The control torque of the ith hub
$Y_i(x_i,t)$	The position at x_i of the ith link under XOY

Notations: $(\dot{*}) = \frac{\partial *}{\partial t}$, $(\ddot{*}) = \frac{\partial^2(*)}{\partial t^2}$, $(*)' = \frac{\partial *}{\partial x_i}$, $(*)'' = \frac{\partial^2(*)}{\partial x_i^2}$, $(*)''' = \frac{\partial^3(*)}{\partial x_i^3}$, $(*)'''' = \frac{\partial^4(*)}{\partial x_i^4}$.

I. Introduction

Recently, advanced adaptive control is a research frontier of robotics and distributed artificial intelligence [1], [2]. Vibration suppression of flexible robots has been an ongoing interest for many researchers due to its potential astronautical [3], biomedical [4] and industrial applications [5] such as extravehicular activities, surgery, and rescue operating. Due to their light weight, low energy consumption, flexible operation and fast response, flexible robots are ideal candidates for using in dynamically demanding applications [6], [7], [8]. The main control goal of the flexible system is to achieve precise positioning, therefore, the damping of oscillations, also referred to as vibration suppression, attributed to low stiffness is imperative [9], [10].

It is well known that flexible robots are distributed parameter systems, which are generally represented by partial differential equations (PDEs) [11], [12]. This also contributes to the complexity and difficulty of vibration control design based on PDEs model [13], especially when there exists model uncertainty [14]. Due to the intractability of distributed parameter systems, only particular solutions are available (e.g. for the uncertain infinite dimensional distributed systems case, dimensionality reduction and approximation can be employed to handle this). For this reason the establishment of discretized model, also referred to as transformation of PDEs model to ODEs model [15], [16], [17], is the primary motivating factor for studies leading to better understanding of flexible vibration mechanisms and identifications of appropriate vibration control strategy.

Owing to highly adverse effects of elastic vibration, considerable efforts have been made to build dynamical model of flexible robots, such as finite difference method (FDM) [18], lumped parameter method (LPM) [19], assumed mode method (AMM) [20], finite element method (FEM) [21], etc. [22]. The dynamic modeling problem of a planar cable-actuated system was addressed using lumped-mass method in [23]. An efficient modeling technique using Lagrange's equation was introduced in [24], from the computational point of view and/or also valid in the cases of varying cross sections, of large link deformations and of time-varying geometrical.

This work was supported by the National Natural Science Foundation of China under Grant 61873298, 61933001, Joint Funds of Equipment Pre-Research and Ministry of Education of China under Grant 6141A02033339 and the National Key Research and Development Program of China under Grant 2019YFB1703600, and in part by the Beijing Top Discipline for Artificial Intelligent Science and Engineering, University Science and Technology Beijing.

W. He, H. Gao and C. Zhou are with Institute of Artificial Intelligence, University of Science and Technology Beijing, Beijing 100083, China, and also with Key Laboratory of Knowledge Automation for Industrial Processes of Ministry of Education, School of Automation and Electrical Engineering, University of Science and Technology Beijing, Beijing 100083, China (Corresponding author: Wei He. E-mail: weihe@ieee.org).

C. Yang is with the Bristol Robotics Laboratory, University of the West of England, Bristol BS16 1QY, U.K.

Z. Li is with Department of Automation, University of Science and Technology of China, Hefei 230026, China.

The technical challenges associated with the vibration control problem have also attracted the attention of many scholars [25], [26], [27]. For conventional proportion-integral-derivative (PID) control, there is a contradiction between “speediness” and “overshoot” in the closed-loop system. Additionally, the closed-loop dynamic properties are sensitive to the change of PID gains [28]. Therefore, when the controlled object is in a constantly changing environment, PID gains need to be adjusted according to the change of the environment. Therefore, numerous researchers have used adaptive method to control the flexible structure [29], [30]. In [31] and [32], the authors proposed adaptive control scheme for a robot via sliding mode method. In addition, a number of authors have dealt with adaptive neural network (NN) control problem for a class of uncertain nonlinear systems [33], [34], [35]. Some other control methods have also been applied to nonlinear systems [36], [37], [38]. In [39], the authors proposed an adaptive dynamic programming (ADP) approach for a class of nonlinear, time-varying, indefinite and complex systems. In [40], an adaptive model-free robust control strategy was presented for a humanoid robot with flexible joints.

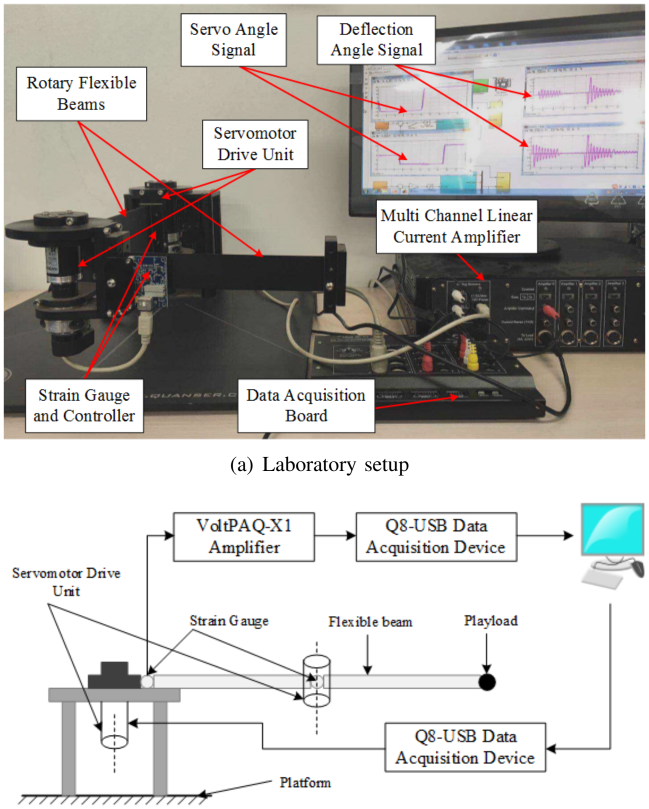
Some of the methods involve an efficient adaptive vibration control techniques with online learning ability known as RL control [41], [42]. RL is bridging the gap between traditional adaptive control and bio-inspired learning techniques [43]. It is shown how a system consisting of two neuron-like adaptive elements can solve a difficult learning control problem [44]. Actor-critic algorithms, however, had eluded satisfactory convergence analysis until a heuristic analysis was introduced in [45]. Later, in [46], actor-critic algorithms and their convergence analysis were discussed. Strictly speaking, training NNs to find approximate solutions is the key to solve the uncertain dynamical equations. An online adaptive synchronous policy iteration algorithm which involves both actor and critic NNs are used in results such as [47], [48] and [49] to solve optimal control problems for continuous-time nonlinear systems. From a practical point of view, extension of the RL strategy in [50], [51] and [52] to solve the optimal tracking control problem is of critical importance. In [53], a novel ADP reinforcement learning (RL) control, which combines RL and optimal control theory to develop an optimal policy on-line is implemented for a humanoid robot arm. The successful real-time learning results presented in [54] and [55] are also highly encouraging for the applicability of RL in practice.

As a matter of fact, the challenge of the control problem for FTLM is to track the desired position with minimum vibration. This paper concentrates on an efficient modeling technique to achieve the transformation of PDEs model to ODEs model and is concerned with developing an online adaptive RL algorithm built on actor-critic structure, to achieve the mentioned control objectives of FTLM systems with uncertain dynamics. Closed-loop stability while learning the parameters is guaranteed via Lyapunov design techniques [56], [57]. A primary contribution of this paper is to successfully present an experimental investigation of the proposed RL control on Quanser Flexible Link System.

The paper is organized as follows: subsequent to the presentation of the discretized dynamic model for the FTLM system a RL controller and discussions about convergence and stability analysis are investigated. Then, the performance of the approach is tested for the tracking control and vibration control of a flexible two-link robotic manipulator through simulations and experiments.

II. Problem Formulation

The Quanser Flexible Link System described in Fig. 1 which connects the FLEXIBLE module with the physical model and realizes real-time monitoring is a simplified model of a FTLM system. The purpose of the paper is to propose a control scheme to position the robot tip in a plane as rapidly as possible with minimum elastic deflection. Table I lists the main parameters associated with the Quanser laboratory platform.



(a) Laboratory setup

(b) Schematic showing details

Fig. 1. The flexible two-link manipulator

Fig. 2 shows the coordinate axes and symbols in dynamic modeling. Due to the detailed PDE model given in [58], the following governing and boundary equations are directly used:

$$\rho_i\ddot{Y}_i(x_i) = -EI_i\omega_i''''(x_i), \tag{1}$$

$$\tau_1 = I_{h1}\ddot{\theta}_1 - EI_1\omega_1''(0), \tag{2}$$

$$\begin{aligned}&(I_{t1}+I_{h2})\tau_2 = I_{t1}I_{h2}\ddot{\theta}_2 - I_{h2}EI_1\omega_1''(L_1)\\ &-I_{t1}EI_2\omega_2''(L_2).\end{aligned} \tag{3}$$

$$o(t)\ddot{Y}_1(L_1) + \dot{o}(t)\dot{Y}_1(L_1) + EI_2\dot{v}_1(t) = EI_1\omega_1'''(L_1), \tag{4}$$

$$\begin{aligned}&\rho_2L_2[\dot{Y}_1(L_1)]^2\sin\theta_2\cos\theta_2 - EI_1\omega_1''(L_1)\\ &+EI_2\dot{Y}_1(L_1)v_2(t)\sin\theta_2 + EI_2\omega_2''(0)\end{aligned}$$

TABLE I
PARAMETERS ASSOCIATED WITH THE QUANSER LABORATORY PLATFORM

Symbol	Definition	Value	Unit
M_1	Mass of link 1	0.49	kg
M_2	Mass of link 2	0.42	kg
m_{h2}	Mass of drive joint 2	1.00	kg
m_{t2}	Tip mass of link 2	0.157	kg
L_1	Length of link 1	0.35	m
L_2	Length of link 2	0.30	m
I_{h1}	Inertia of hub 1	0.0063	kgm^2
I_{h2}	Inertia of hub 2	0.0026	kgm^2
I_{o1}	Inertia of link 2	0.02	kgm^2
I_{o2}	Inertia of link 2	0.013	kgm^2
I_{t2}	Tip inertia of link 2	0.0064	kgm^2

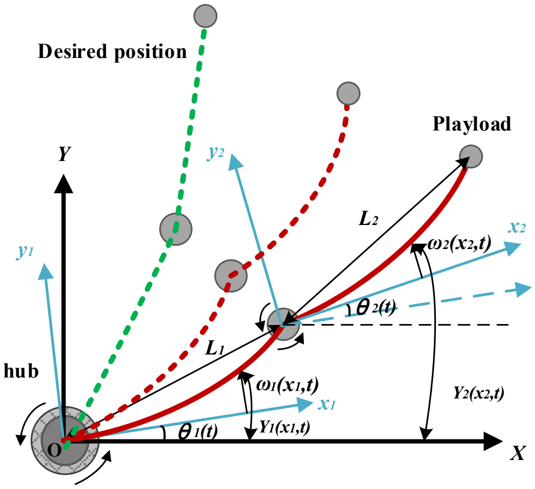


Fig. 2. Coordinate axes and symbols in dynamic modeling

$$
\begin{aligned}
&-I_{h2}[\ddot{\theta}_1+\ddot{\theta}_2]-I_{t1}[\ddot{\theta}_1+\ddot{\omega}_1'(L_1)]\\
&-m_2\dot{Y}_1(L_1)\sin\theta_2[2L_2\dot{\theta}_2+\dot{\omega}_2(L_2)]=0, \qquad (5)\\
&m_2\ddot{Y}_2(L_2)=EI_2\omega_2'''(L_2), \qquad (6)\\
&I_{t2}[\ddot{\theta}_1+\ddot{\theta}_2+\ddot{\omega}_2'(L_2)]+EI_2\omega_2''(L_2)=0, \qquad (7)\\
&\omega_i(0)=\omega_i'(0)=0, i=1,2. \qquad (8)
\end{aligned}
$$

where

$$
\begin{aligned}
&Y_1(x_1)=x_1\theta_1(t)+\omega_1(x_1),\\
&Y_2(x_2)=x_2\theta_2(t)+\omega_2(x_2)+\int_0^t[\dot{Y}_1(L_1,\xi)\cos\theta_2(\xi)]d\xi,\\
&o(t)=(M_2+m_2)\sin^2\theta_2(t)+m_1+M_2,\\
&v_1(t)=\cos\theta_2(t)\int_0^t\omega_2'''(0,\xi)d\xi,\\
&v_2(t)=\int_0^t[\omega_2'''(L_2,\xi)-\omega_2'''(0,\xi)]d\xi.
\end{aligned}
$$

ω_i which satisfies (1)-(8) can be described via AMM:

$$
\omega_i(x_i,t)=\sum_{j=1}^{\mu_i}F_{ij}(x_i)p_{ij}(t),\ \ i=1,2. \qquad (9)
$$

where $F_{ij}(x_i)$ denotes the assumed spatial mode shapes, $p_{ij}(t)$ is the time-varying variable, and μ_i is the order of finite-dimensional model. Considering $(\frac{\beta}{L})^4=\frac{\rho}{EI}\omega^2$, the general solution of $F_{ij}(x_i)$:

$$
\begin{aligned}
F_{ij}(x_i)\ =\ &B_{ij}[\cos(\beta_{ij}x_i)-\cosh(\beta_{ij}x_i)\\
&-\gamma_{ij}(\sin(\beta_{ij}x_i)-\sinh(\beta_{ij}x_i))]. \qquad (10)
\end{aligned}
$$

where β_{ij} is the solution of the following function:

$$
\begin{aligned}
&[1+\cos(\beta_{ij}l_i)\cosh(\beta_{ij}l_i)]\\
&-\frac{m_{hi}\beta_{ij}}{\rho_i}[\sin(\beta_{ij}l_i)\cosh(\beta_{ij}l_i)-\cos(\beta_{ij}l_i)\sinh(\beta_{ij}l_i)]\\
&-\frac{I_{ti}\beta_{ij}^3}{\rho_i}[\sin(\beta_{ij}l_i)\cosh(\beta_{ij}l_i)+\cos(\beta_{ij}l_i)\sinh(\beta_{ij}l_i)]\\
&+\frac{I_{ti}m_{hi}\beta_{ij}^4}{\rho_i^2}(1-\cos(\beta_{ij}l_i)\cosh(\beta_{ij}l_i))=0. \qquad (11)
\end{aligned}
$$

The time function $p_{ij}(t)$ is shown as follows:

$$
p_{ij}(t)=\exp(\mathrm{j}\nu_{\mathrm{ij}}\mathrm{t}), \qquad (12)
$$

Then, we can obtain γ_{ij}:

$$
\gamma_{ij}=\frac{m_{hi}\beta_{ij}k_1+\rho_ik_2}{m_{hi}\beta_{ij}k_3+\rho_ik_4} \qquad (13)
$$

where $k_1=\cos(\beta_{ij}l_i)-\cosh(\beta_{ij}l_i)$, $k_2=\sin(\beta_{ij}l_i)-\sinh(\beta_{ij}l_i)$, $k_3=\sin(\beta_{ij}l_i)-\sinh(\beta_{ij}l_i)$ and $k_4=\cos(\beta_{ij}l_i)+\cosh(\beta_{ij}l_i)$.

When $m_{ti}=0$, $B_{ij}=1/\sqrt{L}$; when $m_{ti}>0$,

$$
B_{ij}=\frac{1}{\sqrt{L+\frac{4\rho_im_{ti}(\sin(\beta_{ij}l_i)\sinh(\beta_{ij}l_i))^2}{[m_{hi}\beta_{ij}M_3-\rho_iM_4]^2}}} \qquad (14)
$$

Define the state as $q=[\theta,\ p]^T$, where $\theta=[\theta_1,\ \theta_2]^T$, the flexible generalized coordinate vector is represented by $p=[p_{11},\ ...,\ p_{1n_1},\ p_{21},\ ...,\ p_{2n_2}]^T$, where $N=n_1+n_2$. A linear dynamic model is achieved via Lagrange equations:

$$
A(q)\ddot{q}+O(q,\dot{q})\dot{q}+H(q)=\tau(t). \qquad (15)
$$

where $\tau\in\mathbb{R}^{(N+2)}=[\tau_1,\ \tau_2,\ 0,...,0]^T$ denotes the control force at each joint. $A(q)$, $O(q,\dot{q})\in\mathbb{R}^{(N+2)\times(N+2)}$ are the inertia matrix, the matrix of coriolis and centripetal forces, respectively. $H(q)\in\mathbb{R}^{(N+2)}$ represents the stiffness matrix.

III. CONTROL DESIGN

The actor-critic algorithm is an adaptive iterative method which consists of a strategy evaluation section and a strategy improvement section. The actor neural network uses radial basis function (RBF) NN to gradually accumulate the system experience to generate the appropriate control strategy, and the critic neural network is used to approximate the evaluation function for the current strategy. The actor-critic structure is shown in Fig. 3.

A. Design for Critic Neural Network

By comparing the difference between the output of the controlled object and the reference input, the critic neural network tests the performance of the current control strategy and generates rewards/punishments as the feedback value for

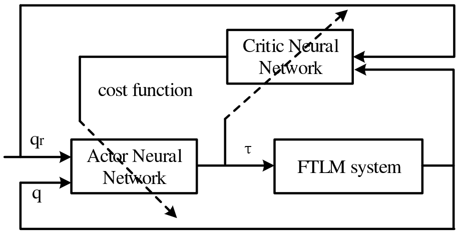


Fig. 3. Actor-critic structure

adaptive learning. Introduce a long-term cost function:

$$I(t) = \int_t^{\infty} e^{-\frac{m-t}{\psi}} \varphi(m) dm. \tag{16}$$

where ψ represents a constant to discount the future cost. $\varphi(t)$ is an instant cost function:

$$\varphi(t) = (q - q_r)^T D(q - q_r) + \tau^T R \tau. \tag{17}$$

where $D > 0$ and $R > 0$, and q_r is the desired state. To achieve optimal control, we need to minimize the cost-to-go.

A RBFNN is a function approximator, which has advantages of parallel computing, fault tolerance and self-learning. For a continuous function $f(Z) : R_k \to R$, the RBFNN is as

$$f(Z) = W^T S(Z). \tag{18}$$

where $Z \in \Omega \subset \mathbb{R}^k$ represents the input, $W \in \mathbb{R}^\ell$ is the weight with nodes number $\ell > 1$. $S(Z) = [s_1(Z), s_2(Z), ..., s_\ell(Z)]^T$ is the basic function, where

$$s_i(Z) = exp[\frac{-(Z - u_i)^T (Z - u_i)}{\eta_i^2}], \ i = 1, 2, ..., \ell. \tag{19}$$

where $u_i = [u_{i1}, \ u_{i2}, \ ..., \ u_{ik}]$ is the center of receptive field and η_i is the width of the Gaussian function. In addition, RBFNN can be an approximation of any continuous function over a compact set $\Omega_z \subset \mathbb{R}^k$ to any desired precision.

$$f(Z) = W^{*T} S(Z) + \epsilon, \ \forall Z \in \Omega_z. \tag{20}$$

where ϵ is a bounded approximation error.

Define $I = W_c^{*T} S_c(Z_c) + \epsilon_c$ and $\hat{I} = \hat{W}_c^T S_c(Z_c)$ with $Z_c = z_1 = q - q_r$, $S_c(Z_c)$ is the basic function of the critic NN. According to (17), the approximation error of the cost-to-go function can be represented as

$$\gamma(t) = \varphi(t) - \frac{1}{\psi} \hat{I}(t) + \dot{\hat{I}}(t). \tag{21}$$

As the constant $\psi \to \infty$, $\gamma(t)$ can be obtained:

$$\gamma(t) = \varphi(t) + \dot{\hat{I}}(t) = \varphi(t) + \bigtriangledown \hat{I}(t) \dot{Z}_c. \tag{22}$$

where $\bigtriangledown$ is defined as the gradient to Z_c. Design the updating law of the critic NN:

$$\dot{\hat{W}}_c = -\sigma_c \frac{\partial E_c}{\partial \hat{W}_c}. \tag{23}$$

where $E_c = \frac{1}{2} \gamma^T \gamma$. Substituting (22) into (23), we have

$$\begin{aligned} \dot{\hat{W}}_c &= -\sigma_c \gamma(t) \frac{\partial \gamma}{\partial \hat{W}_c} = -\sigma_c \gamma(t) \frac{\partial [\varphi(t) - \frac{1}{\psi} \hat{I}(t) + \dot{\hat{I}}(t)]}{\partial \hat{W}_c} \\ &= -\sigma_c \gamma(t) [-\frac{1}{\psi} \frac{\partial \hat{I}}{\partial \hat{W}_c} + \frac{\partial}{\partial \hat{W}_c} \frac{\partial \hat{I}}{\partial Z_c} \dot{Z}_c] \\ &= -\sigma_c (\varphi(t) + \hat{W}_c^T \Lambda) \Lambda. \end{aligned} \tag{24}$$

where $\sigma_c > 0$ represents the learning rate, $\Lambda = -\frac{S_c}{\psi} + \nabla S_c \dot{Z}_c$.

B. Design for Actor Neural Network

The actor neural network consists of a RBFNN and the appropriate control strategy is generated by gradually accumulating the system experience.

Given the tracking errors z_1 and z_2 as $z_1 = q - q_r$, $z_2 = \dot{q} - \alpha_1$, where define $\alpha_1 = \dot{q}_r - K_1 z_1$, $K_1 \in \mathbb{R}^{(N+2)\times(N+2)} = K_1^T > 0$, we have

$$\dot{z}_1(t) = z_2 + \alpha_1 - \dot{q}_r = z_2 - K_1 z_1, \tag{25}$$

$$\dot{z}_2(t) = A^{-1}[\tau - H(q) - O(q, \dot{q})\dot{q}] - \dot{\alpha}_1. \tag{26}$$

Consider a Lyapunov function $V_1 = \frac{1}{2} z_1^T z_1$.

$$\dot{V}_1 = z_1^T \dot{z}_1 = z_1^T z_2 - z_1^T K_1 z_1. \tag{27}$$

To deal with $z_1^T z_2$ in (27), define $V_2 = V_1 + \frac{1}{2} z_2^T A z_2$, so

$$\begin{aligned} \dot{V}_2 &= z_1^T z_2 - z_1^T K_1 z_1 \\ &\quad + z_2^T [\tau - O(q, \dot{q})\dot{q} - H(q) - A\dot{\alpha}_1]. \end{aligned} \tag{28}$$

Therefore, design the desired control torque as

$$\tau_1 = -z_1 - K_2 z_2 + H(q) + O(q, \dot{q})\dot{q} + A\dot{\alpha}_1. \tag{29}$$

where $K_2 \in \mathbb{R}^{(N+2)\times(N+2)}$, $\lambda_{\min}(K_2) > 0$. However, since the system dynamics information $H(q)$, $O(q, \dot{q})\dot{q}$ and $A\dot{\alpha}_1$ is unavailable, we propose an actor neural network to approximate the system uncertainties.

$$\tau_2 = -z_1 - K_2 z_2 + W_a^{*T} S_a(Z_a) + \epsilon_a. \tag{30}$$

where $\hat{W}_a$ and W_a^* are the estimated value and the optimal value of the weight and $\hat{W}_a = \tilde{W}_a + W_a^*$ with $Z_a = [q^T, \ \dot{q}^T, \ q_r^T, \ \dot{q}_r^T, \ \ddot{q}_r^T]$, $S_a(Z_a)$ is the basic function of the actor NN and ϵ_a is the estimation error of the actor NN. Therefore, the control input can be described:

$$\tau_3 = -z_1 - K_2 z_2 + \hat{W}_a^T S_a(Z_a). \tag{31}$$

Define the current estimation error as

$$\zeta_a = \tilde{W}_a^T S_a(Z_a). \tag{32}$$

Then, design the error with the actor neural network

$$e_a = \zeta_a + K_I(\hat{I}(t) - I_d(t)). \tag{33}$$

where $I_d(t) = 0$ represents the ideal value of cost-to-go and $K_I \subset \mathbb{R}^{N+2} > 0$. Subsequently, design:

$$E_a = \frac{1}{2} e_a^T e_a. \tag{34}$$

Design the updating law for the actor neural network as

$$\dot{\hat{W}}_a = -\sigma_a \frac{\partial E_a}{\partial \hat{W}_a}. \tag{35}$$

Substituting (34) to (35), we have

$$\dot{\hat{W}}_a = -\sigma_a \frac{\partial E_a}{\partial e_a} \frac{\partial e_a}{\partial \zeta_a} \frac{\partial \zeta_a}{\partial \hat{W}_a} = -\sigma_a(\zeta_a + K_I \hat{I}) S_a(Z_a). \tag{36}$$

where the learning rate $\sigma_a > 0$. When ζ_a is unknown, a new updating law is defined as

$$\dot{\hat{W}}_a = -\sigma_a(\hat{W}_a^T S_a(Z_a) + K_I \hat{I}) S_a(Z_a). \tag{37}$$

C. Stability Analysis

Design V_c as

$$V_c = \frac{1}{2}\tilde{W}_c^T \tilde{W}_c. \tag{38}$$

Substituting (24) to (38),

$$\dot{V}_c = \tilde{W}_c^T \dot{\tilde{W}}_c = \tilde{W}_c^T \dot{\hat{W}}_c = -\sigma_c \tilde{W}_c^T (\varphi(t) + \hat{W}_c^T \Lambda)\Lambda. \tag{39}$$

Lemma 1: [59] T is a finite moment at the end of the reinforcement learning process. When T is infinite, there is a feasible method (gradient descent) to make the cost function converge to a smaller range at the minimum.

That is, $\gamma(t) \le \hbar$, where $\hbar$ is a small constant. Thus, $\varphi(t) \le \frac{1}{\psi} I - \dot{I} + \hbar$. Then, we have

$$\begin{aligned} \varphi(t) \; &\le \; W_c^{*T}\frac{S_c}{\psi} + \frac{\epsilon_c}{\psi} - \nabla I \dot{Z}_c + \hbar \\ &\le \; W_c^{*T}\frac{S_c}{\psi} + \frac{\epsilon_c}{\psi} - \nabla (W_c^{*T} S_c(Z_c) + \epsilon_c) \dot{Z}_c + \hbar \\ &\le \; -W_c^{*T}\Lambda + \varepsilon_c. \end{aligned} \tag{40}$$

where $\varepsilon_c = \frac{\epsilon_c}{\psi} + \nabla \epsilon_c \dot{Z}_c + \hbar$, and ε_c is bounded, $\|\varepsilon_c\| \le \varepsilon_{c,\max}$. Combining (39) and (40), we have

$$\begin{aligned} \dot{V}_c \; &\le \; -\sigma_c \tilde{W}_c^T (\tilde{W}_c \Lambda + \varepsilon_c)\Lambda \\ &\le \; -\sigma_c \Lambda^T \Lambda \tilde{W}_c^T \tilde{W}_c - \sigma_c \tilde{W}_c^T \varepsilon_c \Lambda \\ &\le \; \frac{-\sigma_c \Lambda^T \Lambda}{2} \tilde{W}_c^T \tilde{W}_c + \frac{\sigma_c}{2} \varepsilon_c^T \varepsilon_c \end{aligned} \tag{41}$$

Introduce a Lyapunov function as

$$V = \frac{1}{2} z_1^T z_1 + \frac{1}{2} z_2^T B z_2 + \frac{1}{2} \tilde{W}_a^T \tilde{W}_a + \frac{1}{2} \tilde{W}_c^T \tilde{W}_c. \tag{42}$$

So its derivative is expressed as

$$\dot{V} = z_1^T \dot{z}_1 + z_2^T B \dot{z}_2 + \tilde{W}_a^T \dot{\tilde{W}}_a + \tilde{W}_c^T \dot{\tilde{W}}_c. \tag{43}$$

Substituting (24) and (37) to (43), we have

$$\begin{aligned} \dot{V} \; &\le \; -z_1^T K_1 z_1 - z_2^T K_2 z_2 + z_2^T (\tilde{W}_a^T S_a - \varepsilon_a) \\ &\quad -\sigma_a \tilde{W}_a^T S_a (\hat{W}_a^T S_a(Z_a) + K_I \hat{I}) \\ &\quad -\sigma_c \tilde{W}_c^T (\tilde{W}_c^T \Lambda + \varepsilon_c)\Lambda. \end{aligned} \tag{44}$$

As $\hat{I} = W_c^{*T} S_c(Z_c) + \tilde{W}_c^T S_c(Z_c)$, we can obtain

$$\hat{I}^T \hat{I} \le 2(W_c^{*T} S_c)^T W_c^{*T} S_c + 2(\tilde{W}_c^T S_c)^T \tilde{W}_c^T S_c. \tag{45}$$

Substituting (45) to (44), we have

$$\begin{aligned} \dot{V} \; &\le \; -z_1^T K_1 z_1 - z_2^T (K_2 - E) z_2 - \frac{\sigma_a - 1}{2} \parallel \tilde{W}_a \parallel^2 \parallel S_a \parallel^2 \\ &\quad - \frac{\sigma_c \Lambda^T \Lambda - 2\sigma_a K_I^2 \parallel S_c \parallel^2}{2} \parallel \tilde{W}_c \parallel^2 \\ &\quad + \frac{\sigma_a}{2} \parallel W_c^* \parallel^2 \parallel S_c \parallel^2 + \sigma_a K_I^2 \parallel W_c^* \parallel^2 \parallel S_c \parallel^2 \\ &\quad + \frac{1}{2} \parallel \epsilon_a \parallel^2 + \frac{1}{2} \parallel \varepsilon_{c,\max} \parallel^2 \\ &\le \; -aV + b \end{aligned} \tag{46}$$

where E presents a identity matrix and

$$\begin{aligned} a \; &= \; \min(K_1, K_2 - E, \frac{\sigma_a - 1}{2} b_s^2, \frac{\sigma_c b_\Lambda^2 - 2\sigma_a K_I^2 \parallel S_c \parallel^2}{2}) \\ b \; &= \; \frac{\sigma_a}{2} \parallel W_a^* \parallel^2 \parallel S_c \parallel^2 + \sigma_a K_I^2 \parallel W_c^* \parallel^2 \parallel S_c \parallel^2 \\ &\quad + \frac{1}{2} \parallel \epsilon_a \parallel^2 + \frac{1}{2} \parallel \varepsilon_{c,\max} \parallel^2 \end{aligned} \tag{47}$$

where $b_s \le \parallel S_a \parallel$ and $b_\Lambda \le \parallel \Lambda \parallel$. In order to ensure $a > 0$, the following conditions need to be considered.

$$\lambda_{\min}(K_1) > 0, \; \lambda_{\min}(K_2 - E) > 0, \; \lambda_{\min}(\sigma_a - 1) > 0, \\ \lambda_{\min}(\sigma_c b_\Lambda^2 - 2\sigma_a K_I^2 \parallel S_c \parallel^2) > 0.$$

Theorem 1: When there exists the bounded initial states, z_1, z_2, $\tilde{W}_a$ and $\tilde{W}_c$ are semi-global uniform ultimate bounded (SGUUB). Besides, z_1, z_2, $\tilde{W}_a$ and $\tilde{W}_c$, will ultimately remain within Ω_{z_1} , Ω_{z_2} , $\Omega_{\tilde{W}_a}$ and $\Omega_{\tilde{W}_c}$ respectively, defined as

$$\Omega_{z_1} \; = \; \{z_1 \in \mathbb{R}^{N+2} \mid \parallel z_1 \parallel \le \sqrt{P}\}, \tag{48}$$

$$\Omega_{z_2} \; = \; \{z_1 \in \mathbb{R}^{N+2} \mid \parallel z_2 \parallel \le \sqrt{\frac{P}{\lambda_{\min}(B)}}\}, \tag{49}$$

$$\Omega_{\tilde{W}_a} \; = \; \{\tilde{W}_a \in \mathbb{R}^{\ell \times (N+2)} \mid \parallel \tilde{W}_a \parallel \le \sqrt{P}\}, \tag{50}$$

$$\Omega_{\tilde{W}_c} \; = \; \{\tilde{W}_c \in \mathbb{R}^{\ell} \mid \parallel \tilde{W}_c \parallel \le \sqrt{P}\}. \tag{51}$$

where $P = 2(V(0) + \frac{b}{a})$, $a > 0$ and $b > 0$.

Proof: Multiplying (46) by e^{at} yields

$$\frac{d}{dt}(V e^{at}) \le b e^{at}. \tag{52}$$

Based on (52),

$$V \le (V(0) - b/a)e^{-at} + b/a \le V(0) + b/a. \tag{53}$$

It is obvious that $\frac{1}{2} z_2^T B z_2 \le V(0) + \frac{a}{b}$, $\frac{1}{2} z_1^T z_1 \le V(0) + \frac{b}{a}$, $\frac{1}{2} \tilde{W}_a^T \tilde{W}_a \le V(0) + \frac{b}{a}$ and $\frac{1}{2} \tilde{W}_c^T \tilde{W}_c \le V(0) + \frac{b}{a}$, then

$$\parallel z_1 \parallel^2 \le 2(V(0) + b/a), \tag{54}$$

$$\parallel z_2 \parallel^2 \le 2\frac{V(0) + b/a}{\lambda_{\min}(B)}, \tag{55}$$

$$\parallel \tilde{W}_a \parallel^2 \le 2(V(0) + b/a), \tag{56}$$

$$\parallel \tilde{W}_c \parallel^2 \le 2(V(0) + b/a). \tag{57}$$

By the above theoretical discussion, the closed-loop system is proved to be semi-global uniformly ultimately bounded (SGUUB), with output error converging to a residual set.

IV. Simulations

To observe the performance comparison of traditional control and the RL control presented in this paper for flexible two-link manipulator systems, simulation results without control, with PD control and RL control are offered. The detailed parameters are specified as Table I.

A. Simulation Results Without Control

Based on the analytical and numerical method, free vibration of the FTLM system with rotating motion is analyzed. When there is interference τ_i shown in Fig. 4, flexible links are affected by external force so that they are not steady with continuous vibration. The reference trajectories θ_{id} and the tracking trajectories θ_i for the open-loop system are shown in Fig. 4. Based on the discretized model, when we give the system a small disturbance, The angular positions θ_i increase gradually, even exceeding the desired angular position. The tracking trajectory will be far from the expected value without control.

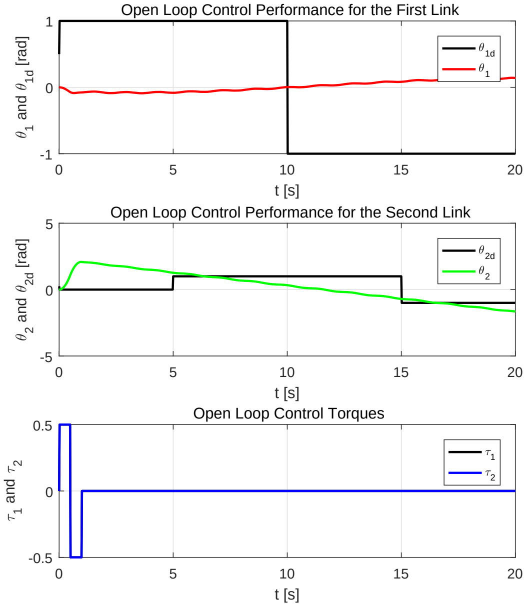


Fig. 4. Tracking trajectories and control torques of the open-loop system

B. Simulation Results using PD Control

Based on the established ODE model with the initial states $q = 0$, design a PD controller with two control gains $K_p = [5, 5]^T$ and $K_d = [1.5, 1.5]^T$:

$$\tau_i = -K_p(\theta_i - \theta_{id}) - K_d\dot{\theta}_i, \; i = 1, 2. \tag{58}$$

Based on θ_i, τ_i shown in Fig. 5 and tip position $Y(x, t)$ shown in Fig. 6, θ_i is steady in 2 s and θ_1 has a large overshoot within the range of -0.2 to 0.2 rad and $Y(L, t)$ has a large vibration within the range of -0.05 to 0.05 m when $t = 1, \; 5, \; 10, \; 15$ s.

Using the PD controller, the system can track the desired trajectory. However, there occurs a large vibration which is not allowed in practice engineering. Therefore, suppressing vibration is an urgent problem to be solved.

C. Simulation Results for Reinforcement Learning Control

Considering a RL control:

$$\tau_i = -z_1 - K_2 z_2 + \hat{W}_a^T S_a(Z_a). \tag{59}$$

The principle of actor-critic algorithm is that the actor neural network generates a control input according to the actor function, then the critic neural network evaluates the control performance according to the critic function. Then, adjust the action function to increase the selection probability of the control input with good performance. Subsequently, adjust the value function to make the value function more in line with expectations. At first, the actor neural network randomly generates control inputs, and the critic neural network scores randomly. After repeated iterations, the critic neural network is more and more accurate, and the control performance of the actor neural network is getting better and better.

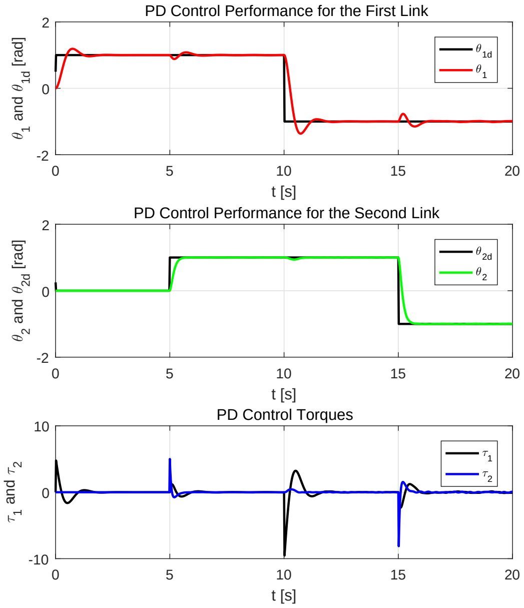
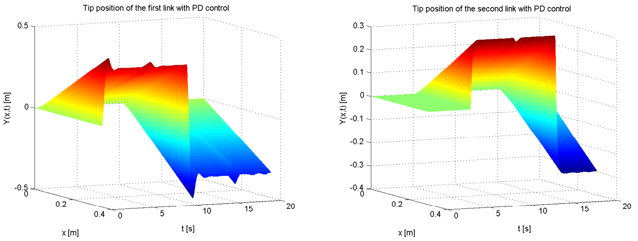


Fig. 5. Tracking trajectories and control torques using PD control



(a) Tip position for the first link (b) Tip position for the second link

Fig. 6. Tip positions using PD control

Model parameters are selected as $n_1 = n_2 = 2$. In addition, 256 nodes and 64 nodes are considered in actor and critic neural network, respectively. The center parameters are chosen as either -1 or 1. $\eta_a = 2$, $\eta_c = 0.5$, $\hat{W}_{ai} = 0 \; (i = 1, 2, ..., 256)$ and $\hat{W}_{ci} = 0 \; (i = 1, 2, ..., 64)$. Learning rates σ_a and σ_c are chosen as 100 and 0.1 respectively. Additionally, the control gains $K_1 = 3$, $K_2 = 8$, and $K_I = 50$. In the cost function, $Q = R = 0.1E_{7\times7}$.

Based on θ_i, τ_i shown in Fig. 7 and tip position $Y(x, t)$ shown in Fig. 8, θ_i is steady in 1 s and large overshoots of θ_1 when $t = 1, \; 5, \; 10, \; 15$ are reduced to a small neighborhood of zero via RL control.

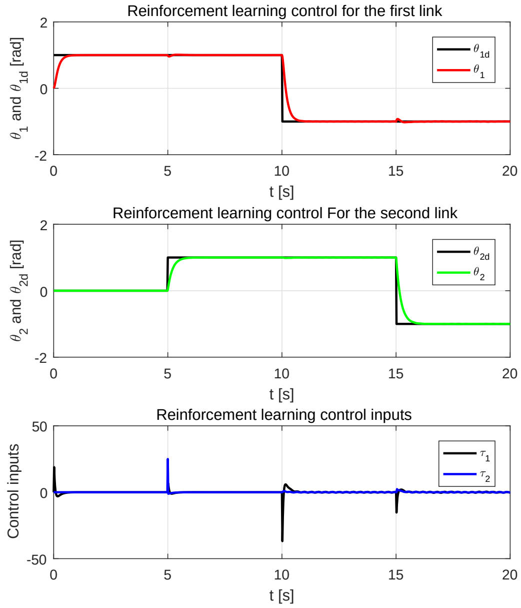
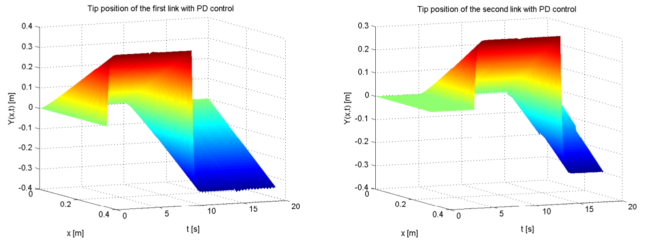


Fig. 7. Tracking trajectories and control torques using reinforcement learning control



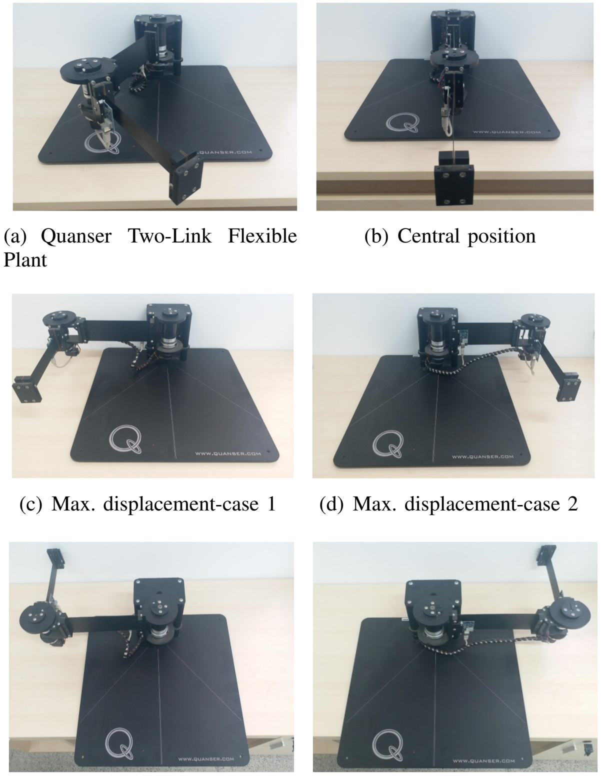
(a) Tip position for the first link (b) Tip position for the second link

Fig. 8. Tip positions using reinforcement learning control

V. Experiments

In the Quanser Two-Link Flexible Plant described in Fig. 9(a), the SRV02 device consists of DC motor loaded with solid aluminum frame and a harmonic gear gearbox. Quanser SRV02 incorporates a Faulhaber Coreless DC Motor model 2338S006 which is a high efficiency, low inductance motor and can obtain a much faster response than a conventional dc motor. Besides, the SRV02 have an optical encoder (The resolution is 4096 times per rotation in orthogonal mode) installed that measures the angular position of the load shaft θ_i. The strain gauge provides a tip vibration measurement α_i. The tachometer can be used to measure the velocity. In addition, the experimental platform also includes other important components, including multi channel linear current amplifier, filter device, data acquisition board and host. VoltPAQ-X1 amplifier processes sensor signals θ_i and α_i. Subsequently, A-D conversion is carried out in the data acquisition board which is connected with the host computer. Then, the control algorithm is implemented and control torques signals are generated.

As shown in Fig. 9 (b), the central position is displayed. Fig. 9 (b-f) the specification of the Quanser Two-Link Flexible Plant shows the max. displacement (+/- 90 degrees) of Axis 1, 2, respectively.



(a) Quanser Two-Link Flexible Plant
(b) Central position
(c) Max. displacement-case 1
(d) Max. displacement-case 2
(e) Max. displacement-case 3
(f) Max. displacement-case 4

Fig. 9. specification of the Quanser Two-Link Flexible Plant

The PD and RL controller are designed by MATLAB Simulink models shown in Fig. 10 and Fig. 11, respectively. In the PD Simulink shown in Fig. 10, the amplitude of reference trajectories θ_{id} $(i = 1, 2)$ are chosen as 15° and 10°, respectively. Gains are set as $K_p = [5, 8]^T$, $K_d = [0.01, 0.01]^T$. In the RL Simulink shown in Fig. 11, control gains are chosen as $K_1 = 3$, $K_2 = 8$, and $K_I = 50$. The other parameters are same as Section IV.

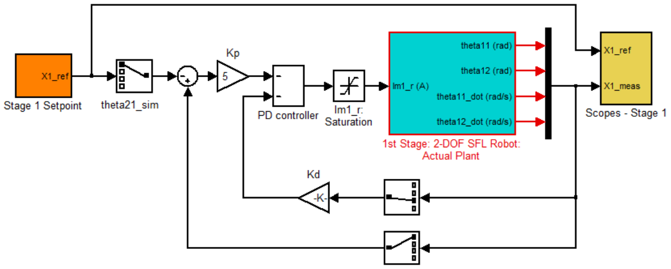


Fig. 10. Simulink diagram for PD control

The real time plots θ_i and α_i of PD control are revealed in Fig. 12 and Fig. 13, respectively. As for the tracking performance of PD control, θ_1 shown in Fig. 12 (a), has an overshoot of -1° to 1° during $t = 0$ s to $t = 2$ s and an overshoot of -2.5° to 2.5° deg during $t = 10$ s to $t = 12$ s. θ_2 shown in Fig. 12 (b), has an overshoot of -3° to 3° during

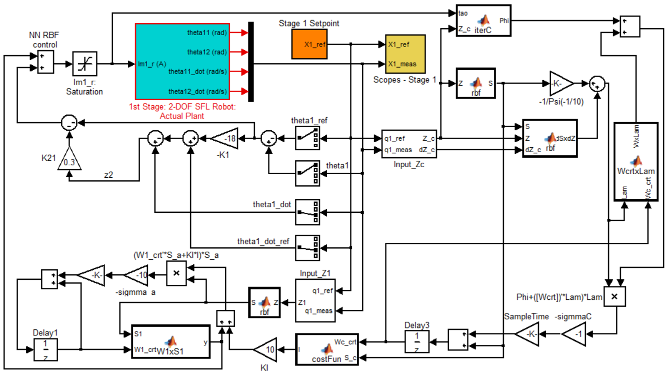
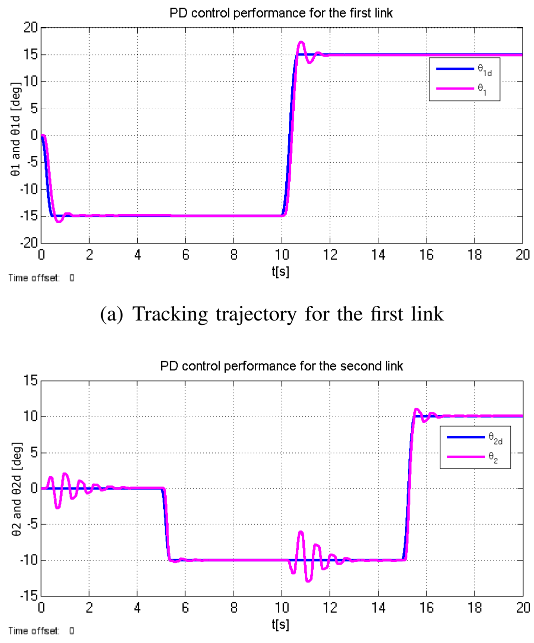


Fig. 11. Simulink diagram for RL control

$t = 0$ s to $t = 4$ s, an overshoot of -4.5° to 4.5° during $t = 10$ s to $t = 14$ s and an overshoot of -1° to 1° during $t = 15$ s to $t = 17$ s. As for the vibration suppression performance of PD control, α_1 shown in Fig. 13 (a) and α_2 shown in Fig. 13 (b), reach a steady-state finally with a vibration range from -0.6° to 0.6°.

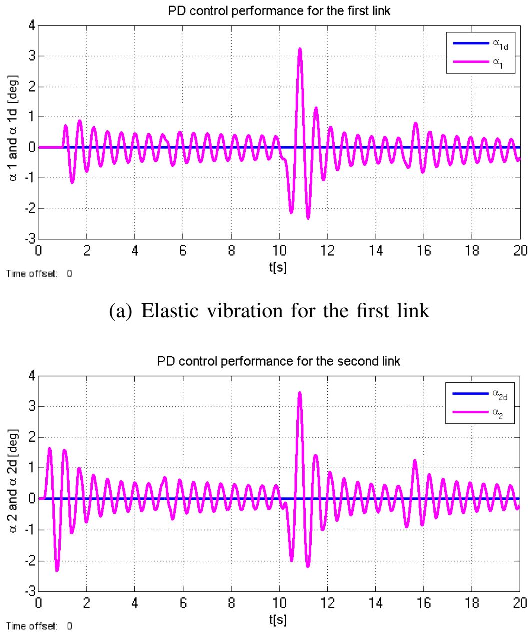


(a) Tracking trajectory for the first link

(b) Tracking trajectory for the second link

Fig. 12. Tracking trajectories for PD control

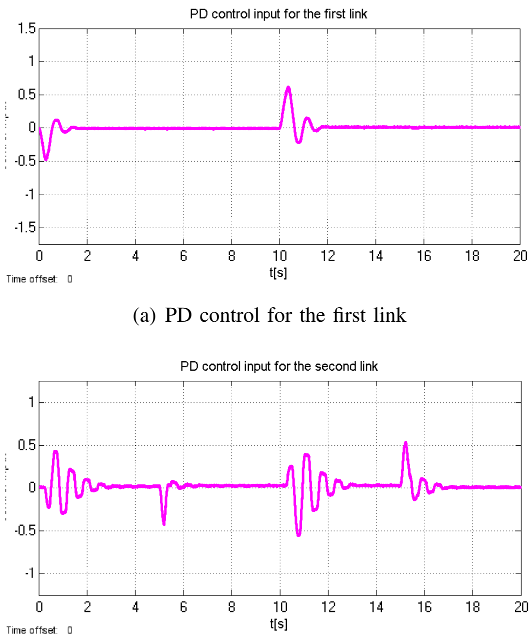
The real time plots θ_i and α_i of RL control are revealed in Fig. 15 and Fig. 16, respectively. As for the tracking performance of RL control, θ_1 shown in Fig. 15 (a), tracks the desired trajectory accurately and quickly. The rise time has been greatly reduced, and large overshoots have also been resolved. θ_2 shown in Fig. 15 (b), has an overshoot of -1° to 1° during $t = 0$ s to $t = 4$ s and an overshoot of -2° to 2° during $t = 10$ s to $t = 14$ s. And the overshoot during $t = 15$ s to $t = 17$ s shown in Fig. 12 (b) are removed. As for the vibration suppression performance of RL control, α_1 shown in Fig. 16 (a) and α_2 shown in Fig. 16 (b), reach a steady-state finally with a vibration range from -0.1° to 0.1°. Therefore, the vibration suppression performance is also multiplied.



(a) Elastic vibration for the first link

(b) Elastic vibration for the second link

Fig. 13. Tip deflections for PD control

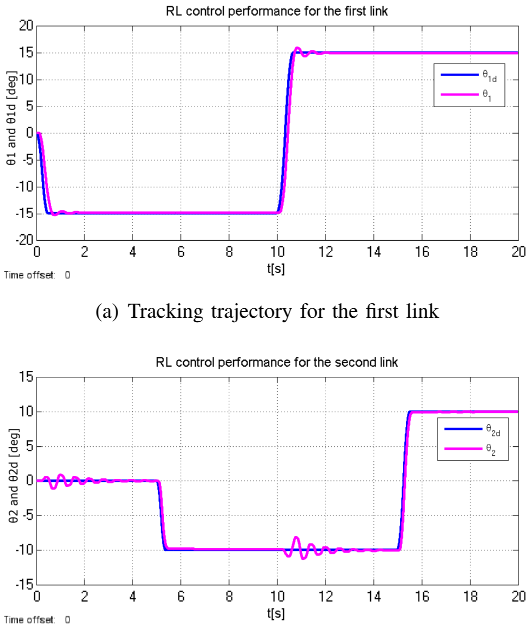


(a) PD control for the first link

(b) PD control for the second link

Fig. 14. PD control inputs

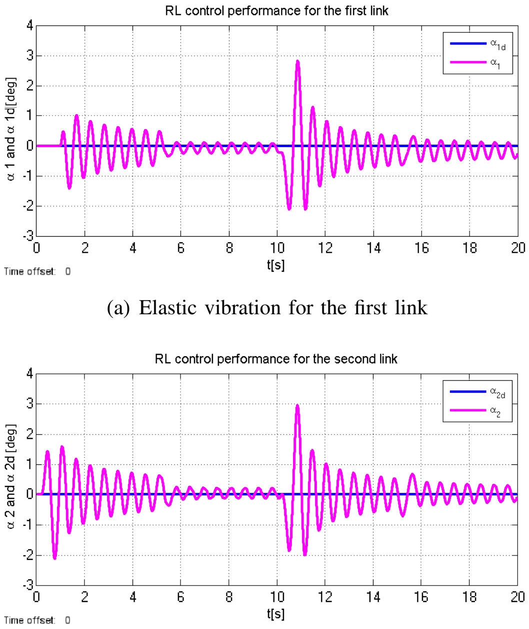
Comparing the experimental results of PD and RL control strategy, it can be observed that the vibration of the flexible two-link manipulator reaches ±4.5° during the trajectory tracking process, and the vibration attenuation time reaches 4 seconds under PD control. The RL approach will control the maximum vibration of the flexible two-link manipulator at ±2°. In addition, the end vibration is only suppressed within ±0.6° in half tracking period (10 seconds), while the



(a) Tracking trajectory for the first link

(b) Tracking trajectory for the second link

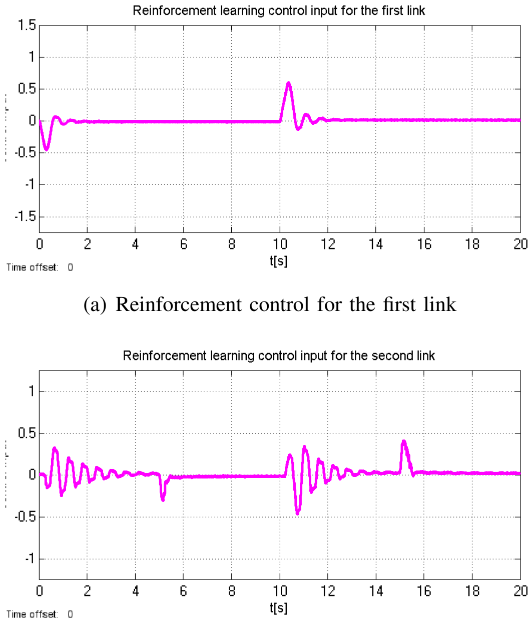
Fig. 15. Tracking trajectories for reinforcement learning control



(a) Elastic vibration for the first link

(b) Elastic vibration for the second link

Fig. 16. Tip deflections for reinforcement learning control



(a) Reinforcement control for the first link

(b) Reinforcement control for the second link

Fig. 17. Reinforcement control inputs

RL control can suppress the end vibration within ±0.1° in half tracking period (10 seconds). Therefore, the RL control strategy has feasibility and stability in suppressing vibration, especially in engineering with high precision requirements.

VI. Conclusions

In this paper, we present a RL controller for vibration suppressing of a FTLM system while trajectory tracking. The RL controller consists of an actor neural network, where the appropriate control strategy is generated by gradually accumulating the system experience, and a critic neural network, where the evaluation function for the current strategy is approximated. The persistent feasibility and stability of the RL controller are proved. We thereafter detail an implementation of the RL controller on a Quanser test platform, where its application effect is compared with PD control. In a comprehensive manner, the experimental results indicate the practical applicability of the RL controller.

In the future, the application of reinforcement learning technology to other complex flexible structures will be a worthwhile research direction. We also will further consider a integral term to the Quanser platform and introduce a PI or PID controller as the comparison. In addition to tracking control, reinforcement learning can also be used to complete fixed-point control in location space. Future works will focus on optimizing vibration control performance while achieving fixed-point control.

Acknowledgment

The authors would like to thank the Editor-in-Chief, the Associate Editor, and the anonymous reviewers for their constructive comments which helped to improve the quality and presentation of this paper.

References

[1] Y. Zhang, J. Sun, H. Liang, and H. Li, "Event-triggered adaptive tracking control for multiagent systems with unknown disturbances," *IEEE Transactions on Cybernetics*, vol. 50, no. 3, pp. 890–901, Mar 2020.

[2] R. Li and H. Qiao, "A survey of methods and strategies for high-precision robotic grasping and assembly tasksłsome new trends," *IEEE/ASME Transactions on Mechatronics*, vol. 24, no. 6, pp. 2718–2732, Dec 2019.

[3] R. J. Caverly and J. R. Forbes, "Dynamic modeling, trajectory optimization, and control of a flexible kiteplane," *IEEE Transactions on Control Systems Technology*, vol. 25, no. 4, pp. 1297–1306, Jul 2017.

[4] H. Lee, Y. Choi, and B. J. Yi, “Stackable 4-BAR manipulators for single port access surgery,” *IEEE/ASME Transactions on Mechatronics*, vol. 17, no. 1, pp. 157–166, Feb 2012.

[5] J. Yu, C. Wang, and G. Xie, “Coordination of multiple robotic fish with applications to underwater robot competition,” *IEEE Transactions on Industrial Electronics*, vol. 63, no. 2, pp. 1280–1288, Feb 2016.

[6] T. Xu, J. Yu, C.-I. Vong, B. Wang, X. Wu, and L. Zhang, “Dynamic morphology and swimming properties of rotating miniature swimmers with soft tails,” *IEEE/ASME Transactions on Mechatronics*, vol. 24, no. 3, pp. 924–934, Jun 2019.

[7] T. Xu, Y. Guan, J. Liu, and X. Wu, “Image-based visual servoing of helical microswimmers for planar path following,” *IEEE Transactions on Automation Science and Engineering*, in press, May 2019. DOI: 10.1109/TASE.2019.2911985.

[8] H. Qiao, M. Wang, J. Su, S. Jia, and R. Li, “The concept of “attractive region in environment” and its application in high-precision tasks with low-precision systems,” *IEEE/ASME Transactions on Mechatronics*, vol. 20, no. 5, pp. 2311–2327, Oct 2015.

[9] H. Yang and J. Liu, “An adaptive RBF neural network control method for a class of nonlinear systems,” *IEEE/CAA Journal of Automatica Sinica*, vol. 5, no. 2, pp. 457–462, Mar 2018.

[10] S. L. Dai, M. Wang, and C. Wang, “Neural learning control of marine surface vessels with guaranteed transient tracking performance,” *IEEE Transactions on Industrial Electronics*, vol. 63, no. 3, pp. 1717–1727, Mar 2016.

[11] Y. Ren, M. Chen, and J. Liu, “Bilateral coordinate boundary adaptive control for a helicopter lifting system with backlash-like hysteresis,” *Science China Information Sciences*, vol. 63, no. 119203, pp. 1–3, Jan 2020.

[12] Y. Liu, F. Guo, X. He, and Q. Hui, “Boundary control for an axially moving system with input restriction based on disturbance observers,” *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 49, no. 11, pp. 2242–2253, Nov 2019.

[13] Z. Zhao, X. He, Z. Ren, and G. Wen, “Boundary adaptive robust control of a flexible riser system with input nonlinearities,” *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 49, no. 10, pp. 1971–1980, Oct 2019.

[14] M. Chen, S.-Y. Shao, and B. Jiang, “Adaptive neural control of uncertain nonlinear systems using disturbance observer,” *IEEE Transactions on Cybernetics*, vol. 47, no. 10, pp. 3110–3123, Oct 2017.

[15] Z. Li, B. Huang, Z. Ye, M. Deng, and C. Yang, “Physical human-robot interaction of a robotic exoskeleton by admittance control,” *IEEE Transactions on Industrial Electronics*, vol. 65, no. 12, pp. 9614–9624, Dec 2018.

[16] Z. Li, B. Huang, A. Ajoudani, C. Yang, C.-Y. Su, and A. Bicchi, “Asymmetric bimanual control of dual-arm exoskeletons for human-cooperative manipulations,” *IEEE Transactions on Robotics*, vol. 34, no. 1, pp. 264–271, Feb 2018.

[17] X. Yu, S. Zhang, L. Sun, Y. Wang, C. Xue, and B. Li, “Cooperative control of dual-arm robots in different human-robot collaborative tasks,” *Assembly Automation*, in press, Jun 2019. DOI: 10.1108/AA-12-2018-0264.

[18] A. R. Tavakolpour, I. Z. M. Darus, and M. Mailah, “Modeling and simulation of an active vibration control system for a flexible structure using finite difference method,” in *Third Asia International Conference on Modelling and Simulation*, May 2009, pp. 448–453.

[19] C. Sun, W. He, and J. Hong, “Neural network control of a flexible robotic manipulator using the lumped spring-mass model,” *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 47, no. 8, pp. 1863–1874, Aug 2017.

[20] S. S. K. Tadikonda, T. G. Mordfin, and T. G. Hu, “Assumed modes method and articulated flexible multibody dynamics,” *Journal of Guidance Control and Dynamics*, vol. 18, no. 3, pp. 404–410, May 1995.

[21] J. Gerstmayr and J. Schoberl, “A 3D finite element method for flexible multibody systems,” *Multibody System Dynamics*, vol. 15, no. 4, pp. 305–320, May 2006.

[22] H. Wang, W. Chao, W. Chen, X. Liang, and Y. Liu, “Three dimensional dynamics for cable-driven soft manipulator,” *IEEE/ASME Transactions on Mechatronics*, vol. 22, no. 1, pp. 18–28, Feb 2017.

[23] R. J. Caverly and J. R. Forbes, “Dynamic modeling and noncollocated control of a flexible planar cable-driven manipulator,” *IEEE Transactions on Robotics*, vol. 30, no. 6, pp. 1386–1397, Dec 2014.

[24] A. Walsh and J. R. Forbes, “Modeling and control of flexible telescoping manipulators,” *IEEE Transactions on Robotics*, vol. 31, no. 4, pp. 936–947, Aug 2015.

[25] W. He, X. He, and C. Sun, “Vibration control of an industrial moving strip in the presence of input deadzone,” *IEEE Transactions on Industrial Electronics*, vol. 64, no. 6, pp. 4680–4689, Jun 2017.

[26] H. Yang, X. Fan, P. Shi, and C. Hua, “Nonlinear control for tracking and obstacle avoidance of a wheeled mobile robot with nonholonomic constraint,” *IEEE Transactions on Control Systems Technology*, vol. 24, no. 2, pp. 741–746, Mar 2016.

[27] A. Mohanty and B. Yao, “Integrated direct/indirect adaptive robust control of hydraulic manipulators with valve deadband,” *IEEE/ASME Transactions on Mechatronics*, vol. 16, no. 4, pp. 707–715, Aug 2011.

[28] X. Chen, C.-Y. Su, Z. Li, and F. Yang, “Design of implementable adaptive control for micro/nano positioning system driven by piezoelectric actuator,” *IEEE Transactions on Industrial Electronics*, vol. 63, no. 10, pp. 6471–6481, Oct 2016.

[29] S. E. Talole, J. P. Kolhe, and S. B. Phadke, “Extended-state-observer-based control of flexible-joint system with experimental validation,” *IEEE Transactions on Industrial Electronics*, vol. 57, no. 4, pp. 1411–1419, Apr 2010.

[30] J. Kim and E. A. Croft, “Full-state tracking control for flexible joint robots with singular perturbation techniques,” *IEEE Transactions on Control Systems Technology*, vol. 27, no. 1, pp. 63–73, Oct 2017.

[31] D. Li and W. Wang, “An intelligent sliding mode controller for vibration suppression in flexible structures,” *Journal of Vibration and Control*, vol. 17, no. 17, pp. 2187–2198, Jun 2011.

[32] S. Zhang, P. Yang, L. Kong, W. Chen, Q. Fu, and K. Peng, “Neural networks-based fault tolerant control of a robot via fast terminal sliding mode,” *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, in press, Aug 2019. DOI: 10.1109/TSMC.2019.2933050.

[33] B. Niu, H. Li, Z. Zhang, J. Li, T. Hayat, and E. A. Fuad, “Adaptive neural-network-based dynamic surface control for stochastic interconnected nonlinear nonstrict-feedback systems with dead zone,” *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 49, no. 7, pp. 1386–1398, Jul 2019.

[34] X. Wu, J. Liu, C. Huang, M. Su, and T. Xu, “3D path following of helical micro-swimmers with an adaptive orientation compensation model,” *IEEE Transactions on Automation Science and Engineering*, in press, Nov 2019. DOI: 10.1109/TASE.2019.2947071.

[35] S.-L. Dai, S. He, H. Lin, and C. Wang, “Platoon formation control with prescribed performance guarantees for USVs,” *IEEE Transactions on Industrial Electronics*, vol. 65, no. 5, pp. 4237–4246, May 2018.

[36] Y. Wu, B. Jiang, and Y. Wang, “Incipient winding fault detection and diagnosis for squirrel-cage induction motors equipped on CRH trains,” *ISA Transactions*, in press, Sep 2019. DOI: 10.1016/j.isatra.2019.09.020.

[37] Y. Wu, B. Jiang, and N. Lu, “A descriptor system approach for estimation of incipient faults with application to high-speed railway traction devices,” *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 49, no. 10, pp. 2108–2118, Oct 2019.

[38] G. Xie, L. Sun, T. Wen, X. Hei, and F. Qian, “Adaptive transition probability matrix-based parallel IMM algorithm,” *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, in press, Jun 2019. DOI: 10.1109/TSMC.2019.2922305.

[39] D. Liu, Y. Xu, Q. Wei, and X. Liu, “Residential energy scheduling for variable weather solar energy based on adaptive dynamic programming,” *IEEE/CAA Journal of Automatica Sinica*, vol. 5, no. 1, pp. 36–46, Jan 2018.

[40] M. Jin, J. Lee, and N. G. Tsagarakis, “Model-free robust adaptive control of humanoid robots with flexible joints,” *IEEE Transactions on Industrial Electronics*, vol. 64, no. 2, pp. 1706–1715, Feb 2017.

[41] B. Xu, C. Yang, and Z. Shi, “Reinforcement learning output feedback NN control using deterministic learning technique,” *IEEE Transactions on Neural Networks and Learning Systems*, vol. 25, no. 3, pp. 635–641, Mar 2014.

[42] B. Xu, “Composite learning control of flexible-link manipulator using NN and DOB,” *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 48, no. 11, pp. 1979–1985, Nov 2018.

[43] Z. Chao, Y. Chenguang, and C. Zhaopeng, “Bio-inspired robotic impedance adaptation for human-robot collaborative tasks,” *Science China Information Sciences*, in press, Feb 2020. DOI: 10.1007/s11432-019-2748-x.

[44] A. G. Barto, R. S. Sutton, and C. W. Anderson, “Neuronlike adaptive elements that can solve difficult learning control problems,” *IEEE Transactions on Systems, Man, and Cybernetics*, vol. SMC-13, no. 5, pp. 834–846, Sep 1983.

[45] H. Kimura and S. Kobayashi, “An analysis of actor-critic algorithms using eligibility traces: reinforcement learning with imperfect value functions,” *15th International Conference on Machine Learning*, pp. 278–286, Mar 1998.

[46] V. Konda, "Actor-critic algorithms," *Siam Journal on Control and Optimization*, vol. 42, no. 4, pp. 1143–1166, Jul 2003.

[47] K. G. Vamvoudakis and F. L. Lewis, "Online actor-critic algorithm to solve the continuous-time infinite horizon optimal control problem," *Automatica*, vol. 46, no. 5, pp. 878–888, Feb 2010.

[48] L. M. Zhu, H. Modares, O. P. Gan, F. L. Lewis, and B. Yue, "Adaptive suboptimal output-feedback control for linear systems using integral reinforcement learning," *IEEE Transactions on Control Systems Technology*, vol. 23, no. 1, pp. 264–273, Jun 2014.

[49] 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, Mar 2017.

[50] G. Wu, J. Sun, and J. Chen, "Optimal linear quadratic regulator of switched systems," *IEEE Transactions on Automatic Control*, vol. 64, no. 7, pp. 2898–2904, Jul 2019.

[51] L.-T. Luc and A.-S. Alin, "Robust adaptive tracking control based on state feedback controller with integrator terms for elastic joint robots with uncertain parameters," *IEEE Transactions on Control Systems Technology*, vol. 26, no. 6, pp. 2259–2267, Oct 2017.

[52] G. Schultz and K. Mombaur, "Modeling and optimal control of human-like running," *IEEE/ASME Transactions on Mechatronics*, vol. 15, no. 5, pp. 783–792, Oct 2010.

[53] S. G. Khan, G. Herrmann, F. L. Lewis, T. Pipe, and C. Melhuish, "Reinforcement learning and optimal adaptive control: An overview and implementation examples," *Annual Reviews in Control*, vol. 36, no. 1, pp. 42–59, Apr 2012.

[54] S. Adam, L. Busoniu, and R. Babuska, "Experience replay for real-time reinforcement learning control," *IEEE Transactions on Systems, Man, and Cybernetics, Part C*, vol. 42, no. 2, pp. 201–212, Feb 2012.

[55] Y. Ouyang, W. He, and X. Li, "Reinforcement learning control of a single-link flexible robotic manipulator," *IET Control Theory and Applications*, vol. 11, no. 9, pp. 1426–1433, May 2017.

[56] H. Li, L. Wang, H. Du, and A. Boulkroune, "Adaptive fuzzy backstepping tracking control for strict-feedback systems with input delay," *IEEE Transactions on Fuzzy Systems*, vol. 25, no. 3, pp. 642–652, Jun 2017.

[57] W. He and Y. Dong, "Adaptive fuzzy neural network control for a constrained robot using impedance learning," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 29, no. 4, pp. 1174–1186, Apr 2018.

[58] W. He and S. S. Ge, "Vibration control of a flexible beam with output constraint," *IEEE Transactions on Industrial Electronics*, vol. 62, no. 8, pp. 5023–5030, Aug 2015.

[59] L. Busoniu, R. Babuska, B. De Schutter, and D. Ernst, *Reinforcement learning and dynamic programming using function approximators*. CR-C press, 2017.

Wei He (S'09-M'12-SM'16) received his B.Eng. and his M.Eng. degrees from College of Automation Science and Engineering, South China University of Technology (SCUT), China, in 2006 and 2008, respectively, and his PhD degree from Department of Electrical & Computer Engineering, the National University of Singapore (NUS), Singapore, in 2011. He is currently working as a full professor in School of Automation and Electrical Engineering, University of Science and Technology Beijing, Beijing, China. He has co-authored 3 books published in Springer and published over 100 international journal and conference papers. He was awarded a Newton Advanced Fellowship from the Royal Society, UK in 2017. He was a recipient of the IEEE SMC Society Andrew P. Sage Best Transactions Paper Award in 2017. He is serving the Chair of IEEE SMC Society Beijing Capital Region Chapter. From 2018, he has been the chair of Technical Committee on Autonomous Bionic Robotic Aircraft (TC-ABRA), IEEE Systems, Man and Cybernetics Society. He is serving as an Associate Editor of *IEEE Transactions on Robotics*, *IEEE Transactions on Neural Networks and Learning Systems*, *IEEE Transactions on Control Systems Technology*, *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, *SCIENCE CHINA Information Sciences*, *IEEE/CAA Journal of Automatica Sinica*, *Neurocomputing*, and an Editor of *Journal of Intelligent & Robotic Systems*. His current research interests include robotics, distributed parameter systems and intelligent control systems.

Hejia Gao (S'16) received the B.Eng. degree in intelligence science and technology from the School of Automation and Electrical Engineering, University of Science and Technology Beijing, Beijing, China, in 2016. She is currently working toward the Ph.D. degree with the School of Automation and Electrical Engineering, University of Science and Technology Beijing.

Her research interests include neural network control, vibration control, flexible robots, reinforcement learning, etc.

Chen Zhou received the B.Eng. degree in control theory and engineering from the School of Automation and Electrical Engineering, University of Science and Technology Beijing, Beijing, China, in 2017. He is currently working toward the M.S. degree with the School of Electrical and Computer Engineering, Georgia Institute of Technology, Atlanta, GA, USA.

His research interests include neural network control, vibration control, machine learning, computer systems, and software.

Chenguang Yang (M'10-SM'16) received the B.Eng. degree in measurement and control from Northwestern Polytechnical University, Xi'an, China, in 2005, and the Ph.D. degree in control engineering from the National University of Singapore, Singapore, in 2010.

He is a Professor of Robotics with the University of the West of England, Bristol, U.K. He was a Post-Doctoral Fellow with Imperial College London, London, U.K., from 2009 to 2010. His current research interests include human-robot interaction and intelligent system design. Dr. Yang was a recipient of the Marie Curie International Incoming Fellowship Award, the EPSRC Innovation Fellowship, and the Best Paper Award of the *IEEE Transactions on Robotics* as well as over ten conference best paper awards.

Zhijun Li (M'07-SM'09) received the Ph.D. degree in mechatronics from Shanghai Jiao Tong University, Shanghai, China, in 2002.

From 2003 to 2005, he was a Postdoctoral Fellow with the Department of Mechanical Engineering and Intelligent systems, University of Electro-Communications, Tokyo, Japan. From 2005 to 2006, he was a Research Fellow with the Department of Electrical and Computer Engineering, National University of Singapore, Singapore, and Nanyang Technological University, Singapore. Since 2012, he has been a Professor with the College of Automation Science and Engineering, South China University of Technology, Guangzhou, China. In 2017, he is a Professor with the Department of Automation, University of Science and Technology of China, Hefei, China.

From 2016, he has been the Co-Chairs of IEEE SMC Technical Committee on Bio-Mechatronics and Bio-Robotics Systems, and IEEE-RAS Technical Committee on Neuro-Robotics Systems. He is serving as an Editor-at-large of *journal of Intelligent and Robotic Systems*, and Associate Editors of several IEEE Transactions. Dr. Li's current research interests include wearable robotics, tele-operation systems, nonlinear control, neural network optimization, etc.