Robust Neuro-Optimal Control for a Robot via Adaptive Dynamic Programming

Linghuan Kong, Wei He, *Senior Member, IEEE*, Chenguang Yang, *Senior Member, IEEE,* Changyin Sun

***Abstract*—We aim at optimization of the tracking control of a robot to improve the robustness, under the effect of unknown nonlinear perturbations. First, an auxiliary system is introduced and an optimal control of the auxiliary system can be seen as an approximate optimal control of the robot. Then, neural networks are employed to approximate the solution of the Hamilton-Jacobi-Isaacs (HJI) equation under the frame of adaptive dynamic programming (ADP). Next, based on the standard gradient attenuation algorithm and adaptive critic design, neural networks are trained depending on the designed updating law with relaxing the requirement of initial stabilizing control. In light of the Lyapunov stability theory, all the error signals can be proved to be uniformly ultimately bounded (UUB). A series of simulation studies are carried out to show the effectiveness of the proposed control.**

***Index Terms*—Neural Networks (NNs), Robots, Robust Optimal Control, Adaptive Dynamic Programming (ADP)**

I. Introduction

Robots are intelligent systems, which has the ability to perform some dangerous tasks alone for human beings, and play significant roles in search and rescue, military reconnaissance, industrial surveillance and medical endoscope [1]–[5]. It is therefore important to derive full information of the robotic system and control it better. However, robotic systems are highly nonlinear and strongly coupled, and these characteristics make accurate robotic models difficult to obtain [6]–[12]. Meanwhile, their motion control is exceedingly challenging due to the lack of accurate robotic models.

Many control policies including proportion integration differentiation (PID) control [13]–[15], decentralized control [16], neural network (NN) control [17]–[21], adaptive control [22]–[29], fuzzy control [30]–[32], etc., have been widely used in control theory and applications. In [33], [34], PID control is applied on a robotic manipulator such that the tracking error is convergent exponentially. Although PID control has simple structure and strong robustness, if the controlled systems are highly nonlinear and strongly coupled, selecting appropriate PID parameters will be exceedingly difficult. Furthermore, as requirements on speed and accuracy of motion increase, PID control often cannot satisfy the performance requirement due to the lack of the adaptive or learning capability of dealing with dynamic uncertainties in system parameters [35]. NNs are abstractions and simulations of the basic characteristics of the human brain [36]–[39], and also are a kind of imitating the behavior characteristics of animal NNs for distributed parallel algorithm which is for mathematical model of information processing [40]–[42], which therefore are widely used to achieve the identification of the unknown dynamics in control theory and applications [43]–[46]. NNs are a kind of learning models [47]. NN control has been proved to have an powerful ability to model the uncertain dynamics of the controlled system in real time, which leads to the improvement of system robustness. In [48], unknown robotic dynamics are online approximated by NNs such that the boundness of tracking errors is guaranteed. In [49], NN control is developed for an uncertain multi-input and multi-output nonlinear systems, where NNs are utilized to deal with unknown functions. In [50], based on the learning ability of NNs, an adaptive control scheme is proposed for a rigid robot such that finite-time convergence is achieved.

Tracking control is a hot topic in the robot community. In [51], the control method just cares about the stability and ignores the analysis on optimal performances. It is therefore essential to design an optimal tracking control that not only stabilizes a robotic system, but also minimizes its cost function, i.e., tracking control of the robot should be achieved under the optimal performance. ADP firstly proposed by [52] serves as an effective way to achieve the optimal performance for the tracking control of nonlinear systems. Under the frame of ADP, the solution of the optimization problem is obtained by solving the Hamilton-Jacobi-Bellman (HJB) equation. However, it is because of the nonlinear characteristics of HJB equation such that deriving an analytical solution for it is hardly possible. Traditionally, policy iteration is an efficient way to obtain an approximate solution of HJB equation [53], for which initial stabilizing control is necessary. However, in practical implementation, initial stabilizing control is often difficult to derive. Then, as the improvement, the adaptive critic-based learning algorithm is developed in [54] to derive an approximate solution of HJB equation with the universal approximate ability of NNs, and an additional stabilizing term is designed for removing the need of initial stabilizing control.

This work was supported in part by the National Natural Science Foundation of China under Grant 61933001 and 61921004, in part by the Joint Funds of Equipment Pre-Research and Ministry of Education of China under Grant 6141A02033339, in part by the Beijing Top Discipline for Artificial Intelligent Science and Engineering, University of Science and Technology Beijing, and in part by the Fundamental Research Funds for the China Central Universities of USTB under Grant FRF-TP-19-001C2.

Linghuan Kong and Wei He are with the Institute of Artificial Intelligence, University of Science and Technology Beijing, Beijing 100083, China, also with the School of Automation and Electrical Engineering, University of Science and Technology Beijing, Beijing 100083, China, and also with the Key Laboratory of Knowledge Automation for Industrial Processes, Ministry of Education, University of Science and Technology Beijing, Beijing 100083, China. (Corresponding author: Wei He, Email: weihe@ieee.org)

Chenguang Yang is with the Bristol Robotics Laboratory, University of the West of England, Bristol, BS16 1QY, UK.

Changyin Sun is with School of Automation, Southeast University, Nanjing 210096, China.

Based on the adaptive critic-based idea, [55]–[57] proposes an approximate optimal control policy for affine systems subject to the upper bounded disturbance, leading to the result that the controlled system is stable and achieves optimal performances.

Motivated by the mentioned-above discussions, our paper aim to solve optimal problems of tracking control for a robot. In [33]–[35], [48], PID control or NN control is developed for controlling the robotic systems, however these control algorithms mainly focus on stability without considering optimal control problems. In the field of robotics, how to optimize the path-tracking and minimize the design cost is an important topic. Few literatures of optimal control of robots using ADP can be found. In [55], [56], unknown nonlinear perturbation is assumed to be upper bounded and this boundary is asked to be known in practical implementation. However, the appropriate boundary is often difficult for a robot to obtain since the robotic system is highly nonlinear, strongly coupled and unknown. In this paper, this requirement is relaxed successfully.

Compared with the existing references, the main contributions of this paper are summarized as follows:

1) Under the frame of ADP, the adaptive critic-based learning algorithm is proposed to solve optimal control problems for a class of typical robotic systems.
2) An additional stabilizing term is incorporated into the updating algorithm of the critic network weight, and it releases the requirement of initial stabilizing control.
3) In [54]–[56], overall nonlinear perturbation is upper bounded and this boundary is assumed to be known, which limits the application scope of the designed algorithm. In this paper, this requirement is relaxed successfully.

Notations 1: Let $\|*\|$ denote the Euclidean norm of $*$. Let $\nabla * \triangleq \frac{\partial *}{\partial x}$ denote the gradient operator with x being a column vector and $\nabla *$ being a row vector. Let $\mathbf{0}$ denote a vector with appropriate dimensions, and its every element is zero. Let $\hat{*}$ denote the approximation of $*$. Let $I \in \mathbb{R}^{n\times n}$ denote the identity matrix with appropriate dimensions. Let $* \in \mathbb{R}^n$ be a vector whose every element is defined as $*_i$, $i = 1, \ldots, n$.

II. Preliminaries and Problem Formulation

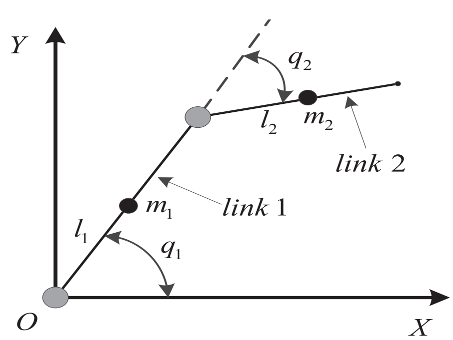


Fig. 1: Sketch of a two-link robot manipulator.

The parameters in Fig. 1 are defined as follows: $q = [q_1, q_2]^T$ denotes the position vector in joint space. $\mu = [\mu_1, \mu_2]^T$ denotes the control input. l_1, l_2 denote the length of link 1 and link 2, respectively. m_1, m_2 denote the mass of link 1 and link 2, respectively. I_1, I_2 denote the moment of inertial of link 1 and link 2 with respect to their own center of mass, respectively. l_{c1} and l_{c2} denote the half of length of link 1 and link 2, respectively.

By applying the Euler-Lagrange method, a two-link robot manipulator [44] shown in Fig. 1 is modeled as follows

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

with

$$M(q) = \begin{bmatrix} p_1 + p_2 + 2p_3\cos q_2 & p_2 + p_3\cos q_2 \\ p_2 + p_3\cos q_2 & p_2 \end{bmatrix} \tag{2}$$

$$C(q,\dot{q}) = \begin{bmatrix} -p_3\dot{q}_2\sin q_2 & -p_3(\dot{q}_1 + \dot{q}_2)\sin q_2 \\ p_3\dot{q}_1\sin q_2 & 0 \end{bmatrix} \tag{3}$$

$$G(q) = \begin{bmatrix} p_4 g\cos q_1 + p_5 g\cos(q_1+q_2) \\ p_5 g\cos(q_1+q_2) \end{bmatrix} \tag{4}$$

where $p_1 = m_1 l_{c1}^2 + m_2 l_1^2 + I_1$, $p_2 = m_2 l_{c2}^2 + I_2$, $p_3 = m_2 l_1 l_{c2}$, $p_4 = m_1 l_{c2} + m_2 l_1$ and $p_5 = m_2 l_{c2}$. For convenience, M, C, G are used to denote $M(q), C(q,\dot{q}), G(q)$, respectively.

Given the desired trajectory q_d, error z is defined as follows

$$z = q - q_d \tag{5}$$

and then the first derivative and the second derivative are calculated as $\dot{z} = \dot{q} - \dot{q}_d$ and $\ddot{z} = \ddot{q} - \ddot{q}_d$, respectively. Then, the sliding model surface e is defined by

$$e = \Lambda z + \dot{z} \tag{6}$$

where $\Lambda \in \mathbb{R}^{2\times 2}$ denotes the constant gain matrix, which implies

$$\dot{q} = e - \Lambda z + \dot{q}_d \tag{7}$$

Differentiating (6) leads to $\dot{e} = \Lambda\dot{z} + \ddot{z}$. Then, it follows that $\dot{e} = \Lambda\dot{z} + \ddot{q} - \ddot{q}_d$. Therefore, we have

$$\ddot{q} = \dot{e} - \Lambda\dot{z} + \ddot{q}_d \tag{8}$$

By substituting (7) and (8) into (1), error dynamics is rewritten as

$$\begin{aligned}\dot{e} = &- M^{-1}C(\dot{q}_d - \Lambda z + e) - M^{-1}G \\ &- \ddot{q}_d + \Lambda\dot{z} + M^{-1}\mu\end{aligned} \tag{9}$$

Define $x = [z, e]^T \in \mathbb{R}^4$, and the following augment system is obtained.

$$\dot{x} = \begin{bmatrix} e - \Lambda z, \\ -M^{-1}C(\dot{q}_d - \Lambda z + e) \\ -M^{-1}G - \ddot{q}_d + \Lambda\dot{z} \end{bmatrix} + \begin{bmatrix} \mathbf{0} \\ M^{-1} \end{bmatrix}\mu \tag{10}$$

It is assumed that mass m_i and length l_i are all known and called the nominal value such that system matrixes M, C, G are all known. Consequently, augment system (10) is called the nominal system. It is worth pointing out that nominal system (10) is based an assumption that mass of i-th link is uniformly distributed and m_i, l_i are accurate mass and length of i-th link, respectively. However, due to the complexity of the robotic structure and movement, this assumption is not always satisfied.

Define Δm_i as an unknown perturbation of mass of link i, and define Δl_i as an unknown perturbation of length of link

i. Then, $m_i + \Delta m_i$ and $l_i + \Delta l_i$ are accurate mass and length of link i. Define $\Delta M, \Delta C, \Delta G$ as the unknown nonlinear perturbation of M, C and G, respectively. It should be pointed out that $\Delta M, \Delta C, \Delta G$ are functions of $\Delta m_i, \Delta l_i$ and will be specified in simulation. By considering mass perturbation Δm_i and length perturbation Δl_i, augment system (10) is rewritten as

$$\dot{x} = \begin{bmatrix} e - \Lambda z, \\ -(M+\Delta M)^{-1}(C \\ +\Delta C)(\dot{q}_d - \Lambda z + e) \\ -(M+\Delta M)^{-1}(G \\ +\Delta G) - \ddot{q}_d + \Lambda \dot{z} \end{bmatrix} + \begin{bmatrix} \mathbf{0} \\ (M+\Delta M)^{-1} \end{bmatrix} \mu \tag{11}$$

Considering $(\Delta M + M)(\Delta M + M)^{-1} = I = MM^{-1}$ and then adding and subtracting ΔMM^{-1} to the right-hand, lead to

$$(\Delta M + M)(\Delta M + M)^{-1} = (\Delta M + M)M^{-1} - \Delta MM^{-1} \tag{12}$$

Thus, we obtain

$$(\Delta M + M)^{-1} = M^{-1} - (\Delta M + M)^{-1}\Delta MM^{-1} \tag{13}$$

Substituting (13) into (11) leads to

$$\dot{x} = f(x) + g(x)\mu + d(x) \tag{14}$$

with nonlinear functions $f : \mathbb{R}^4 \to \mathbb{R}^4, g : \mathbb{R}^4 \to \mathbb{R}^{4\times 2}$ and $d : \mathbb{R}^4 \to \mathbb{R}^4$ being specified by

$$f(x) = \begin{bmatrix} e - \Lambda z, \\ -M^{-1}C(\dot{q}_d - \Lambda z + e) \\ -M^{-1}G - \ddot{q}_d + \Lambda \dot{z} \end{bmatrix}, g(x) = \begin{bmatrix} \mathbf{0} \\ M^{-1} \end{bmatrix},$$

$$\begin{aligned} \bar{d}(x) &= -M^{-1}\Delta C(\dot{q}_d - \Lambda z + e) + (\Delta M + M)^{-1}\Delta MM^{-1} \\ &\times (C + \Delta C)(\dot{q}_d - \Lambda z + e) - M^{-1}\Delta G + (\Delta M + M)^{-1} \\ &\times \Delta MM^{-1}(G + \Delta G) - (\Delta M + M)^{-1}\Delta MM^{-1}\mu \end{aligned}$$

Let $d(x) = [0, \bar{d}(x)]^T$. It is assumed that there exists a certain state x_c satisfying $f(x_c) = \mathbf{0}$ and $d(x_c) = \mathbf{0}$. Assume that $\|g(x)\| \le g_c$ and $\|d(x)\| \le d_c$ with g_c and d_c being positive constants. In light of the definition of $f(x), g(x)$ and $d(x)$, it is found that $f(x)$ and $g(x)$ consist of nominal system dynamics and thus are known. $d(x)$ consists of both nominal system dynamics and the corresponding unknown nonlinear perturbation, and consequently can be considered as the equivalent unknown nonlinear perturbation.

III. Robust Optimal Control Using ADP

A. Robust Optimal Control

An auxiliary system is introduced as

$$\dot{x} = f(x) + g(x)\mu + (I - g(x)g^+(x))v \tag{15}$$

where $g^+(x)$ denotes the moore-penrose pseudoinverse of $g(x)$ and v denotes an auxiliary control input. It should be emphasized that the fist two terms of system (15) are the same as the first two term of system (14). For auxiliary system (15), for the sake of tackling the infinite horizon optimal control problem, optimal control μ and auxiliary optimal control v should be found to minimize the infinite horizon cost function given by

$$J(x(t)) = \int_t^\infty (\Theta(x(\tau)) + U(\mu(x(\tau)), v(x(\tau))))d\tau \tag{16}$$

where $\Theta(x(\tau)) = x(\tau)^T Q x(\tau)$, $U(\mu(x(\tau)), v(x(\tau))) = \mu(x(\tau))^T R \mu(x(\tau)) + \kappa^2 v(x(\tau))^T v(x(\tau))$, $Q = \tilde{Q} + \rho^2 I$, ρ and κ are positive numbers, $\tilde{Q} \in \mathbb{R}^{4\times 4}$ and $R = \tilde{R}^T\tilde{R} \in \mathbb{R}^{2\times 2}$ are symmetric positive definite matrixes. For simplicity, $J(x)$ denotes $J(x(t))$ for short. In terms of optimal theory, optimal control μ and auxiliary optimal control v stabilize auxiliary system (15), and make cost function (16) finite, i.e., the feedback control law should be admissible. Define χ_1, χ_2 as the sets of admissible controls μ and v, respectively. Then, for admissible controls $\forall \mu \in \chi_1$ and $\forall v \in \chi_2$, if cost function $J(x)$ given in (16) is continuously differentiable, then its infinitesimal version is the nonlinear Lyapunov equation given by

$$\begin{aligned} \Theta(x) + U(\mu(x), v(x)) + (\nabla J(x))^T(f(x) \\ + g(x)\mu + (I - g(x)g^+(x))v) = 0 \end{aligned} \tag{17}$$

Then, define the Hamiltonian as

$$\begin{aligned} H(x,\mu,v,\nabla J(x)) = \Theta(x) + U(\mu(x), v(x)) + (\nabla J(x))^T \\ \times (f(x) + g(x)\mu + (I - g(x)g^+(x))v) \end{aligned} \tag{18}$$

The optimal cost function of auxiliary system (15) is defined as follows

$$J^*(x) = \min_{\mu\in\chi_1, v\in\chi_2} \int_t^\infty (\Theta(x(\tau)) + U(\mu(x(\tau)), v(x(\tau))))d\tau \tag{19}$$

with $J^*(\mathbf{0}) = 0$. In order to obtain a controller that minimizes the cost function in the worst-case disturbance, it incorporates the requirement of finding the Nash equilibrium solution corresponding to HJI equation. In terms of optimal theory, the following HJI equation is obtained.

$$0 = \min_{\mu\in\chi_1, v\in\chi_2} H(x, u, v, \nabla J^*(x)) \tag{20}$$

It is assumed that the minimum on the right-hand side of (20) exists and is unique. In other words, the control inputs μ and v should minimize the Hamiltonian (18). Thus, applying $\frac{\partial H(x,\mu,v,\nabla J^*(x))}{\partial \mu} = 0$ and $\frac{\partial H(x,\mu,v,\nabla J^*(x))}{\partial v} = 0$ to (18), we derive the optimal control expressed as

$$\mu = -\frac{1}{2}R^{-1}g^T(x)\nabla J^*(x) \tag{21}$$

$$v = -\frac{1}{2\kappa^2}(I - g(x)g^+(x))^T\nabla J^*(x) \tag{22}$$

It should be noted that optimal control μ and auxiliary optimal control v lead to $H(x, \mu, v, \nabla J^*(x)) = 0$.

Theorem 1: Optimal control μ given by (21) ensures that state x in (14) is UUB and eventually converges to the set Ω defined by

$$\Omega := \{x | \|x\| \le \frac{\sqrt{c}}{\rho}\} \tag{23}$$

where c is defined by

$$\sup_{\substack{v\in\chi_1\\ \mu\in\chi_2}} (2\kappa^2\|v\|^2+\kappa^2\|d(x)\|^2+\|\tilde{R}g^{+}(x)d(x)\|^2)=c \quad (24)$$

Proof: By noticing optimal cost function (19), we know that $J^*(x)>0$ as $x\neq \mathbf{0}$ and $J^*(x)=0$ if and only if $x=\mathbf{0}$, which implies that $J^*(x)$ is a positive Lyapunov function candidate of nonlinear system (14) with dynamical uncertainty. Consider (14), the time derivative of $J^*(x)$ is given as

$$\dot{J}^*(x)=(\nabla J^*(x))^T(f(x)+g(x)\mu+d(x)) \quad (25)$$

We add and subtract $(\nabla J^*(x))^T(I-g(x)g^{+}(x))v$ and $(\nabla J^*(x))^T g(x)g^{+}(x)d(x)$ to the right-hand of (25), therefore the following equation holds

$$\begin{aligned}\dot{J}^*(x)=&(\nabla J^*(x))^T(f(x)+g(x)\mu+(I-g(x)g^{+}(x))v)\\ &-(\nabla J^*(x))^T(I-g(x)g^{+}(x))v\\ &+(\nabla J^*(x))^T g(x)g^{+}(x)d(x)\\ &+(\nabla J^*(x))^T(I-g(x)g^{+}(x))d(x)\end{aligned} \quad (26)$$

Substituting (17), (21) and (22) into (26) yields

$$\begin{aligned}\dot{J}^*(x)=&-x^TQx-\mu^TR\mu+\kappa^2\|v\|^2\\ &-2\mu^TR^Tg^{+}(x)d(x)-2\kappa^2v^Td(x)\end{aligned} \quad (27)$$

In terms of the Young's inequality, we have $-2\mu^TR^Tg^{+}(x)d(x)\leq \mu^T\tilde{R}^T\tilde{R}\mu+\|\tilde{R}g^{+}(x)d(x)\|^2$ and $-2\kappa^2v^Td(x)\leq\kappa^2\|v\|^2+\kappa^2\|d(x)\|^2$. Substituting the above inequalities into (27) yields

$$\begin{aligned}\dot{J}^*(x)\leq&-x^T\tilde{Q}x+2\kappa^2\|v\|^2\\ &+\|\tilde{R}g^{+}(x)d(x)\|^2+\kappa^2\|d(x)\|^2-\rho^2\|x\|^2\end{aligned} \quad (28)$$

Assume that the upper bound of $2\kappa^2\|v\|^2+\|\tilde{R}g^{+}(x)d(x)\|^2+\kappa^2\|d(x)\|^2$ is c for $v\in\chi_1,\mu\in\chi_2$ with c being a positive constant. If $c<\rho^2\|x\|^2$, i.e. $\|x\|>\frac{\sqrt{c}}{\rho}$, we have $\dot{J}^*(x)\leq -x^T\tilde{Q}x<0$ as $x\neq\mathbf{0}$, which illustrates that x decreases and eventually converges to the set $\Omega:=\{x|\|x\|\leq\frac{\sqrt{c}}{\rho}\}$. Consequently, we can conclude that nonlinear system (14) with dynamical uncertainty is UUB. This finishes the proof.

Remark 1: According to the mentioned-above proof, we know that x will go into the set Ω and that the size of the set Ω is affected by parameters κ and ρ. Hence, the satisfactory tracking performance can be obtained by adjusting κ and ρ. In this paper, parameter c is just for analysis and is not required to be known in practice, which implies that the requirement of the known upper bound of overall nonlinear perturbation is released.

Remark 2: According to (21) and (22), we know the optimal control of auxiliary system (15) is obtained by solving HJI equation. According to Theorem 1, we find that the optimal control of auxiliary system (15) with cost function (16) can be considered as a robust optimal control of nonlinear system (14) with dynamical uncertainty. However, the available solution of the HJI equation is always difficult or even impossible to derive [54]. The ADP approach is effective in solving optimal problems [58]. Then, based on ADP, we will develop an approximate robust optimal control and give the detailed proof of system stability.

B. Adaptive Critic Design Based on ADP

It is assumed that optimal cost function $J^*(x)$ is continuously differentiable. In light of the approximation ability of NNs, the following equality holds

$$J^*(x)=\omega^T\sigma(x)+\epsilon(x) \quad (29)$$

where $\omega\in\mathbb{R}^l$ denotes the optimal constant weight vector, $\sigma:\mathbb{R}^4\to\mathbb{R}^l$ denotes the activation function which is chosen according to engineering experience, $\epsilon(x)$ denotes the approximation error, l denotes the node number of the hidden layer. Then, the derivative of (29) with respect to x yields

$$\nabla J^*(x)=(\nabla\sigma(x))^T\omega+\nabla\epsilon(x) \quad (30)$$

Then, substituting (30) into (21) and (22) yields

$$\mu=-\frac{1}{2}R^{-1}g^T(x)((\nabla\sigma(x))^T\omega+\nabla\epsilon(x)) \quad (31)$$

$$v=-\frac{1}{2\kappa^2}(I-g(x)g^{+}(x))^T((\nabla\sigma(x))^T\omega+\nabla\epsilon(x)) \quad (32)$$

Define $B=g(x)R^{-1}g^T(x),\mathcal{B}=(I-g(x)g^{+}(x))(I-g(x)g^{+}(x))^T$. Based on (30), HJI equation (20) becomes

$$\begin{aligned}H(x,\mu,v,\omega)=&\Theta(x)+U(\mu(x),v(x))+\omega^T\nabla\sigma(x)\\ &\times(f(x)+g(x)\mu+(I-g(x)g^{+}(x))v)=e_c\end{aligned} \quad (33)$$

where

$$e_c=-(\nabla\epsilon(x))^T(f(x)+g(x)\mu+(I-g(x)g^{+}(x))v) \quad (34)$$

Assume that e_c is bounded, i.e., there exists a positive constant λ_{e_c} such that $|e_c|\leq\lambda_{e_c}$ holds [59]. Since the optimal weight ω is unknown, the optimal cost function $J^*(x)$ is approximated in the following form.

$$\hat{J}^*(x)=\hat{\omega}^T\sigma(x) \quad (35)$$

The derivative of (35) with respect to x yields

$$\nabla\hat{J}^*(x)=(\nabla\sigma(x))^T\hat{\omega} \quad (36)$$

Similarly, the following approximate optimal control is obtained.

$$\hat{\mu}=-\frac{1}{2}R^{-1}g^T(x)(\nabla\sigma(x))^T\hat{\omega} \quad (37)$$

$$\hat{v}=-\frac{1}{2\kappa^2}(I-g(x)g^{+}(x))^T(\nabla\sigma(x))^T\hat{\omega} \quad (38)$$

Similarly, the approximate Hamiltonian $\hat{H}(x,\hat{\mu},\hat{v},\hat{\omega})$ of Hamiltonian $H(x,\mu,v,\omega)$ is given by

$$\begin{aligned}\hat{H}(x,\hat{\mu},\hat{v},\hat{\omega})=&\Theta(x)+U(\mu,v)+\hat{\omega}^T\nabla\sigma(x)\\ &\times(f(x)+g(x)\hat{\mu}+(I-g(x)g^{+}(x))\hat{v})=e_H\end{aligned} \quad (39)$$

Define the approximation error of critic NN weights as $\tilde{\omega}=\omega-\hat{\omega}$. By defining $e_H=\hat{H}(x,\hat{\mu},\hat{v},\hat{\omega})-H(x,\mu,v,\omega)$, we further have

$$\begin{aligned}e_H=&e_c+\omega^T\nabla\sigma(x)(f(x)-g(x)\tilde{\mu}-(I-g(x)g^{+}(x))\tilde{v})\\ &-\tilde{\omega}^T\nabla\sigma(x)(f(x)+g(x)\hat{\mu}+(I-g(x)g^{+}(x))\hat{v})\end{aligned} \quad (40)$$

where $\tilde{\mu} = \mu - \hat{\mu}$ and $\tilde{v} = v - \hat{v}$. It should be emphasized that $H(x, \mu, v, \omega) = 0$ leads to $e_H = \hat{H}(x, \hat{\mu}, \hat{v}, \hat{\omega})$. Then, to ensure that e_H is upper bounded, we need to minimize the objective function $E = \frac{1}{2}e_H^2$. Thus, an appropriate updating law $\dot{\hat{\omega}}$ should be designed to make the approximate optimal weight $\hat{\omega}$ converge to the optimal weight ω. To release the need of initial stabilizing control, motivated by [60], an additional stabilizing term is incorporated into the standard steepest descent algorithm. Hence, the following weight updating law is obtained.

$$\dot{\hat{\omega}} = -\frac{\alpha_H}{(1+\delta^T\delta)^2}\left(\frac{\partial E}{\partial \hat{\omega}}\right) + \frac{1}{2}\alpha_c h \nabla\sigma(x)(B + \frac{1}{\kappa^2}\mathcal{B})\nabla J_s(x) \tag{41}$$

where α_H is the learning factor of the standard steepest descent algorithm, α_c is the stabilizing term learning rate, and $\delta \in \mathbb{R}^l$ is given as

$$\delta = \nabla\sigma(x)(f(x) + g(x)\hat{\mu} + (I - g(x)g^+(x))\hat{v}) \tag{42}$$

with $(1+\delta^T\delta)^2$ being utilized for normalization. h is defined as follows

$$h = \begin{cases} 1, & \text{if } (\nabla J_s(x))^T(f(x) + g(x)\hat{\mu} \\ & \quad + (I - g(x)g^+(x))\hat{v}) \geq 0 \\ 0, & \text{others} \end{cases} \tag{43}$$

where $J_s(x)$ is a continuous and differentiable Lyapunov function candidate such that $(\nabla J_s(x))^T(f(x) + g(x)\mu + (I - g(x)g^+(x))v) = -(\nabla J_s(x))^T \tilde{G} \nabla J_s(x) < 0$ with $\tilde{G}$ being a positive definite matrix [55]. Due to $\dot{\omega} = 0$, we have $\dot{\tilde{\omega}} = -\dot{\hat{\omega}}$, and error dynamics of critic networks is calculated as

$$\begin{aligned} \dot{\tilde{\omega}} = & \frac{\alpha_H \delta}{(1+\delta^T\delta)^2}(\Theta(x) + U(\mu, v) + \delta^T\hat{\omega}) \\ & - \frac{1}{2}\alpha_c h \nabla\sigma(x)(B + \frac{1}{\kappa^2}\mathcal{B})\nabla J_s(x) \end{aligned} \tag{44}$$

Then define $\delta_1 = \frac{\delta}{(1+\delta^T\delta)}$, $\delta_2 = 1 + \delta^T\delta \geq 1$. (44) becomes

$$\begin{aligned} \dot{\tilde{\omega}} = & -\alpha_H \delta_1 \delta_1^T \tilde{\omega} + \alpha_H \frac{\delta_1}{\delta_2} e_c \\ & - \frac{1}{2}\alpha_c h \nabla\sigma(x)(B + \frac{1}{\kappa^2}\mathcal{B})\nabla J_s(x) \end{aligned} \tag{45}$$

The persistence of excitation (PE) condition is essential to carry out system identification for adaptive control system. In the paper, adaptive technique is employed to identify the weight vector of NNs. Therefore, an assumption on PE condition is given.

Assumption 1: [61] The signal δ_1 is persistently exciting within the interval $[t, t+T]$ with $T > 0$, i.e.,

$$\sigma_1 I_l \leq \int_t^{t+T} \delta_1(y)\delta_1^T(y)dy \leq \sigma_2 I_l \tag{46}$$

holds for $t > 0$, where σ_1 and σ_2 denotes positive constants, and I_l is a l-dimensional identify matrix.

In light of Assumption 1, it is easy to obtain that PE condition guarantees $\lambda_{\min}(\delta_1\delta_1^T) > 0$, which is exceedingly important in stability analysis.

The structure diagram of the proposed robust optimal control is given in Fig. 2.

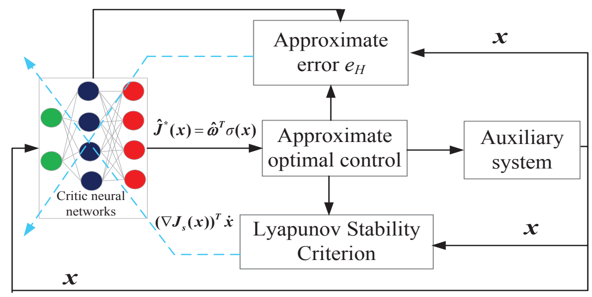


Fig. 2: Structure of the proposed robust optimal control.

Remark 3: Traditional adaptive critic design often relies on the selection of initial stability control. The appropriate initial stabilizing control is exceedingly difficult to derive in practical systems. In this paper, the great difference is that an additional stabilizing term is added for reinforcing the learning process, releasing the need of initial stabilizing control. The last term of (41) is an additional stabilizing term defined in order to reinforce the learning process. According to (43), it is known that when $(\nabla J_s(x))^T(f(x)+g(x)\hat{\mu}+(I-g(x)g^+(x))\hat{v}) \geq 0$, the system is unstable and auxiliary signal h will be activated to make the second term of (41) work such that the learning process can be improved. Hence, initial weight vectors can be set to zero in our paper.

C. Stability Analysis

In this section, we will provide the detailed proof that weight approximation error $\tilde{\omega}$ and the error state x of the auxiliary system given by (15) are UUB.

Theorem 2: Auxiliary system (15) with approximate optimal control (37) and (38), the critic network weight updating law (41), then it is concluded that weight approximation error $\tilde{\omega}$ and error state x of auxiliary system (15) are UUB.

Proof: See the Appendix.

Remark 4: In [62], optimal control was proposed for a class of affine systems where nonlinear functions are all known and no unknown disturbances exist. In our paper, with a series of transformation, the robotic system is transformed into an affine system with unknown disturbances. Although the design process is similar to that in [62], there are also some apparent differences, given as follows: 1) In [62], an affine system is considered without unknown disturbances. In our paper, an auxiliary system is introduced, and compared with [62], additional term $(I - g(x)g^+(x))\hat{v}$ is involved for satisfying the condition of Theorem 1 and further eliminating the effect of the unknown disturbances; 2) In [62], the only feedback optimal control is designed. However, optimal controls $\hat{\mu}$ and $\hat{v}$ in our paper are designed simultaneously for ensuring that the closed-loop system is stable and that disturbances are attenuate.

According to Theorem 2, it can be obtained that $\tilde{\omega}$ is uniformly ultimately bounded, i.e., $\|\tilde{\omega}\| \leq \lambda_\omega$ with λ_ω being a positive constant.

Corollary 1: Approximate optimal control given in (37) eventually converges to a small neighborhood of optimal control given in (31).

Proof: Recall (31) and (37), we know that

$$\mu - \hat{\mu} = -\frac{1}{2}R^{-1}g^T(x)((\nabla\sigma(x))^T\tilde{\omega} + \nabla\epsilon(x)) \tag{47}$$

By observing the proof of Theorem 1, it follows that $\|\nabla\epsilon(x)\| \leq \lambda_\epsilon$ and $\|\nabla\sigma(x)\| \leq \lambda_{2\sigma}$. Assume that $\|g^T(x)\| \leq \lambda_g$ with λ_g being a positive constant. Therefore, we further have

$$\begin{aligned}\|\mu - \hat{\mu}\| &= \|\frac{1}{2}R^{-1}g^T(x)((\sigma(x))^T\tilde{\omega} + \nabla\epsilon(x))\| \\ &\leq \frac{1}{2}\lambda_{\max}(R^{-1})\lambda_g(\lambda_{2\sigma}\lambda_\omega + \lambda_\epsilon) \triangleq \lambda_\mu\end{aligned} \tag{48}$$

with $\lambda_{\max}(R^{-1})$ denoting the maximum eigenvalue of matrix R^{-1}. This finishes the proof.

IV. Simulation Studies

In this section, we would use a typical robotic system with two degrees of freedom to verify the effectiveness of the proposed approximate robust optimal control. For the convenience of simulation implementation, nonlinear perturbations ΔM, ΔC and ΔG are given as follows

$$\Delta M = \begin{bmatrix} \Delta p_1 + \Delta p_2 + 2\Delta p_3 \cos q_2 & \Delta p_2 + \Delta p_3 \cos q_2 \\ \Delta p_2 + \Delta p_3 \cos q_2 & \Delta p_2 \end{bmatrix}$$

$$\Delta C = \begin{bmatrix} -\Delta p_3 \dot{q}_2 \sin q_2 & -\Delta p_3(\dot{q}_1 + \dot{q}_2)\sin q_2 \\ \Delta p_3 \dot{q}_1 \sin q_2 & 0 \end{bmatrix}$$

$$\Delta G = \begin{bmatrix} \Delta p_4 g \cos q_1 + \Delta p_5 g \cos(q_1 + q_2) \\ \Delta p_5 g \cos(q_1 + q_2) \end{bmatrix}$$

where $\Delta p_1 = (m_1 + \Delta m_1)(0.5\Delta l_1^2 + l_1\Delta l_1) + 2\Delta m_1 l_{c1}^2 + (m_2 + \Delta m_2)(\Delta l_1^2 + 2l_1\Delta l_1) + \Delta m_2 l_1^2$, $\Delta p_2 = (m_2 + \Delta m_2)(0.5\Delta l_2^2 + l_2\Delta l_2) + 2\Delta m_2 l_{c2}^2$, $\Delta p_3 = m_2\Delta l_1(l_{c2} + 0.5\Delta l_2) + \Delta m_2(l_1 + \Delta l_1)(l_{c2} + 0.5\Delta l_2)$, $\Delta p_4 = 0.5m_1\Delta l_2 + \Delta m_1(l_{c2} + 0.5\Delta l_2) + m_2\Delta l_1 + \Delta m_2(l_1 + \Delta l_1)$ and $\Delta p_5 = 0.5\Delta l_2(m_2 + \Delta m_2) + 0.5\Delta m_2 l_2$. It should be noted that ΔM, ΔC and ΔG are unknown in practical robotic systems. Therefore, the aim of the simulation is to verify the effectiveness of the proposed control when there exist unknown dynamical perturbations ΔM, ΔC and ΔG in the robotic system. The parameters of nominal system (10) are given in Table 1.

Table 1: Parameters of the robot

Parameter	Description	Value
m_1	Mass of link 1	2.0 kg
m_2	Mass of link 2	0.85 kg
l_1	Length of link 1	0.35 m
l_2	Length of link 2	0.31 m
I_1	Inertia of link 1	61.25×10^{-3} kgm^2
I_2	Inertia of link 2	20.42×10^{-3} kgm^2

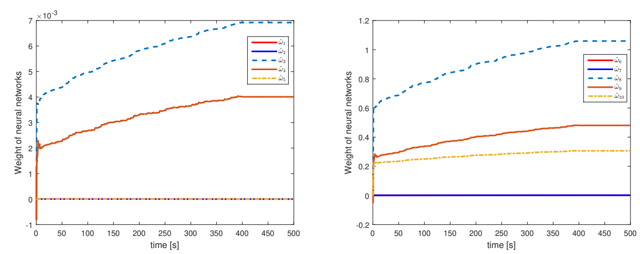
Initial values are $q(0) = [1.1, -0.1]^T$rad and $\dot{q}(0) = [0, 0]^T$rad/s. The reference trajectory q_d is set as $q_d = [\frac{\pi}{2} + \cos(t)e^{-0.2t}, \cos(t)e^{-0.3t}]^T$rad.

A. The Proposed Approximate Robust Optimal Control

We choose $\tilde{Q} = \text{diag}[1, 1, 1, 1]$, $R = \text{diag}[20, 100]$, $\Lambda = \text{diag}[9, 9]$, $\rho = 141.4$, $\kappa = 0.01$. A critic NN is constructed to approximate the optimal cost function as follows

$$\begin{aligned}\hat{J}^*(x) = &\hat{\omega}_1 x_1^2 + \hat{\omega}_2 x_1x_2 + \hat{\omega}_3 x_1x_3 + \hat{\omega}_4 x_1x_4 + \hat{\omega}_5 x_2^2 \\ &+ \hat{\omega}_6 x_2x_3 + \hat{\omega}_7 x_2x_4 + \hat{\omega}_8 x_3^2 + \hat{\omega}_9 x_3x_4 + \hat{\omega}_{10} x_4^2\end{aligned} \tag{49}$$

Note that $\sigma(x) = [x_1^2, x_1x_2, x_1x_3, x_1x_4, x_2^2, x_2x_3, x_2x_4, x_3^2, x_3x_4, x_4^2]^T$ and $\hat{\omega} = [\hat{\omega}_1, \hat{\omega}_2, \hat{\omega}_3, \hat{\omega}_4, \hat{\omega}_5, \hat{\omega}_6, \hat{\omega}_7, \hat{\omega}_8, \hat{\omega}_9, \hat{\omega}_{10}]^T$ are the activation function and the estimated weight vector of the NN, respectively. To obtain the optimal weight ω, the learning rate of the NN is chosen as $\alpha_H = 1.4$. To release the need of initial stabilizing control, the learning rate of the additional stabilizing term is chosen as $\alpha_c = 0.1 \times 10^{-3}$. Therefore, the initial value of the critic NN weight can be set as $\hat{\omega}_1(0) = \hat{\omega}_2(0) = \hat{\omega}_3(0) = \hat{\omega}_4(0) = \hat{\omega}_5(0) = \hat{\omega}_6(0) = \hat{\omega}_7(0) = \hat{\omega}_8(0) = \hat{\omega}_9(0) = \hat{\omega}_{10}(0) = 0$. Let $J_s(x) = \frac{1}{2}x^2$. During the NN learning process, we bring in an exploration noise $N(t) = \sin^2(t)\cos(t) + \sin^2(2t)\cos(0.1t) + \sin^2(1.2t)\cos(0.5t) + \sin^5(t) + \sin^2(1.12t) + \cos(2.4t)\sin^3(2.4t))$ to satisfy the PE condition. The exploration noise is introduced into the control input and thus affects the system state. After a learning stage, the weight of the critic NN converges to $[-1.65 \times 10^{-6}, -5.46 \times 10^{-7}, 0.0069, 0.0040, -6.342 \times 10^{-8}, 0.0018, 0.0014, 1.0592, 0.4804, 0.3061]^T$ as shown in Fig. 3, which presents the learning process of the critic NN during the first 500s. Fig. 4 gives the trajectory of tracking error z in the learning process of the critic NN.



(a) Convergence of $\hat{\omega}_1, \hat{\omega}_2, \hat{\omega}_3, \hat{\omega}_4, \hat{\omega}_5$ (b) Convergence of $\hat{\omega}_6, \hat{\omega}_7, \hat{\omega}_8, \hat{\omega}_9, \hat{\omega}_{10}$

Fig. 3: Convergence of the weight vector in the learning process of the critic NN.

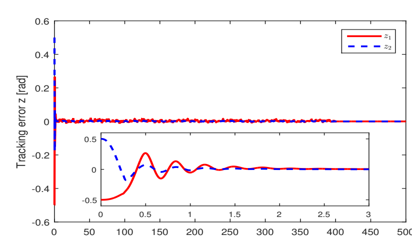
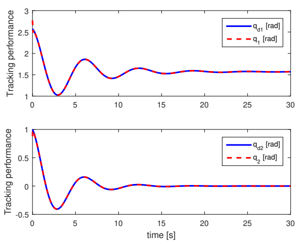


Fig. 4: Evolution of tracking error z in the learning process of the critic NN. (The inserted plot is a magnification of the evolution of tracking error z from 0s to 3s)

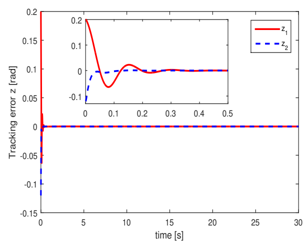
To fully illustrate the robustness of the propose control, three different mass perturbation Δm_i and length perturbation Δl_i are considered as follows: case one: $\Delta m_1 = 0.1$kg, $\Delta m_2 = 0.1$kg, $\Delta l_1 = 0.05$m, $\Delta l_2 = 0.05$m; case two:

$\Delta m_1 = 0.2$kg, $\Delta m_2 = 0.2$kg, $\Delta l_1 = 0.1$m, $\Delta l_2 = 0.1$m; case three: $\Delta m_1 = 0.45$kg, $\Delta m_2 = 0.3$kg, $\Delta l_1 = 0.15$m, $\Delta l_2 = 0.15$m. According to the converged critic NN weight vector, simulation results are given in Figs. 5-8.

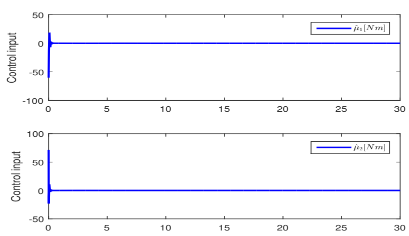
Simulation results of the three cases are given in Figs. 5-7. Figs. 5(a)-7(a) give the trajectories of q and q_d, respectively. From Figs. 5(a)-7(a), it is clear that although there are unknown perturbations ΔM, ΔC and ΔG, under the action of the proposed control (37), q can still converge to a small neighborhood of q_d. In Figs. 5(b)-7(b), the trajectories of error z are presented. From Figs. 5(b)-7(b), error z decreases rapidly as time approaches zero and converges to a small neighborhood of zero. In Figs. 5(c)-7(c), approximate optimal control $\hat{\mu}$ is plotted.



(a) q and q_d

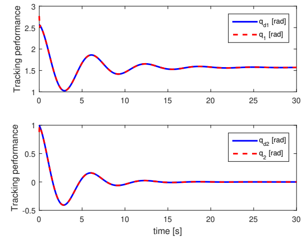


(b) Tracking error $z = q - q_d$

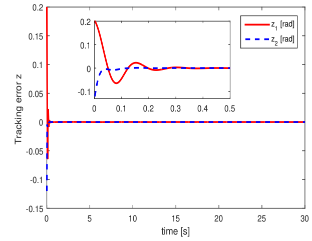


(c) Control input $\hat{\mu}$

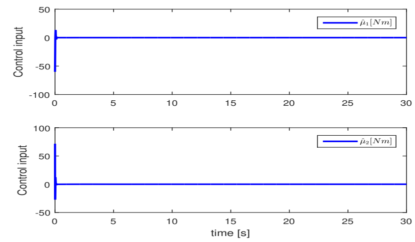
Fig. 5: Case one: simulation results when $\Delta m_1 = 0.1$kg, $\Delta m_2 = 0.1$kg, $\Delta l_1 = 0.05$m, $\Delta l_2 = 0.05$m. (In (b), the inserted plot is a magnification of tracking error z from 0s to 0.5s)



(a) q and q_d

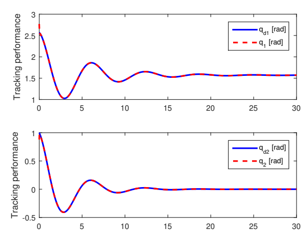


(b) Tracking error $z = q - q_d$

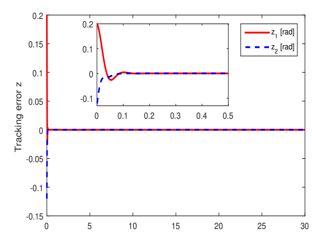


(c) Control input $\hat{\mu}$

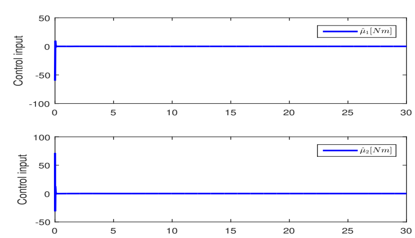
Fig. 6: Case two: simulation results when $\Delta m_1 = 0.2$kg, $\Delta m_2 = 0.2$kg, $\Delta l_1 = 0.1$m, $\Delta l_2 = 0.1$m. (In (b), the inserted plot is a magnification of tracking error z from 0s to 0.5s)



(a) q and q_d



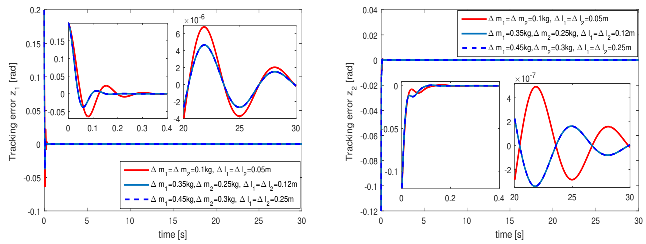
(b) Tracking error $z = q - q_d$



(c) Control input $\hat{\mu}$

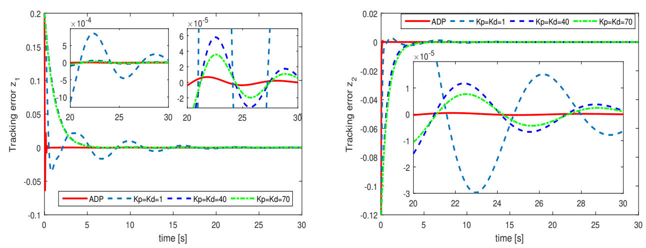
Fig. 7: Case three: simulation results when $\Delta m_1 = 0.45$kg, $\Delta m_2 = 0.3$kg, $\Delta l_1 = 0.15$m, $\Delta l_2 = 0.15$m. (In (b), the inserted plot is a magnification of tracking error z from 0s to 0.5s)

Performance comparison of the three cases is given in Fig. 8. According to Fig. 8, we know that with different mass perturbations and length perturbations, tracking error z still decreases rapidly and converges to a small neighborhood of zero, which shows that approximate optimal control (37) is strongly robust. It should be noted that when mass perturbation and length perturbation are very great, the proposed optimal control (37) will be invalid. However, mass perturbation and length perturbation corresponding to nominal system (10) in practice cannot be very great and must be upper bounded. Consequently, the proposed approximate optimal control (37) is valid and reasonable in dealing with optimal problems of robotic systems.



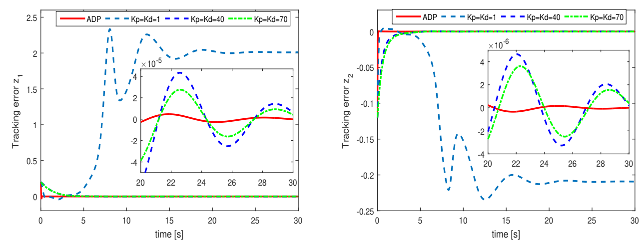
(a) (b)

Fig. 8: Performance comparison of the three different cases under the action of the proposed optimal control (37). (The inserted plots are a magnification of tracking error z from 0s to 0.4s and from 20s to 30s). (a) Tracking error z_1. (b) Tracking error z_2



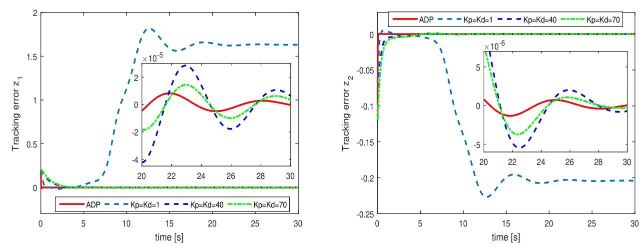
(a) (b)

Fig. 9: Performance comparison of the proposed optimal control and PD control of case one. (The inserted plots are a magnification of tracking error z from 20s to 30s) (a) Tracking error z_1. (b) Tracking error z_2



(a) (b)

Fig. 10: Performance comparison of the proposed optimal control and PD control of case two. (The inserted plots are a magnification of tracking error z from 20s to 30s) (a) Tracking error z_1. (b) Tracking error z_2



(a) (b)

Fig. 11: Performance comparison of the proposed optimal control and PD control of case three. (The inserted plots are a magnification of tracking error z from 20s to 30s) (a) Tracking error z_1. (b) Tracking error z_2

B. Performance Comparison between Approximate Robust Optimal Control and PD Control

To illustrate the superiority of the proposed control (37), PD control is performed. PD control is designed as follows

$$\mu = -K_p z - K_d \dot{z} \tag{50}$$

where $K_p \in \mathbb{R}^{2\times 2}$ denotes the proportion relation and $K_d \in \mathbb{R}^{2\times 2}$ denotes the differential gain.

For each case, three different gains are set. During PD control simulation implementation, many different gains are tried to control the robotic system, finally the three different gains which can stabilize the robotic system better are chosen as $K_p = K_d = 1$, $K_p = K_d = 40$ and $K_p = K_d = 70$. Simulation results are given in Figs. 9-11.

It is known from Figs. 9-11 that under the proposed optimal control (37), tracking error z decreases more rapidly and converges to a smaller neighborhood of zero than that under the action of PD control, when there exist different mass perturbation and length perturbation. That is because PD control possess a simple structure and does not have any built-in capability to handle changes in unknown nonlinear perturbations ΔM, ΔC and ΔG. However, based on an appropriate cost function, the proposed optimal control not only stabilizes the robotic system, but also minimizes the cost function such that the robotic system can possess a satisfactory performance and a strong robustness. Therefore, this type of control is suitable for accurate position tracking.

V. Conclusion

The paper proposed on the NN-based robust optimal control for a robot with dynamical uncertainty. With the help of an appropriate cost function, the optimal control of the introduced auxiliary system is regarded as an approximate robust optimal control of the robot with dynamical uncertainty. The stability of the closed-loop system including the auxiliary system and unknown plants has been proved in details.

It should be emphasized that the proposed optimal control is based on a two-link robotic manipulator. However, the proposed method is still effective on a multi-link robotic manipulator. In practice, nominal mass and nominal length of robotic links are known, and the weights of the critic NN are adjusted adaptively and eventually converge to optimal values. The optimal control with optimal weights is regarded as a robust optimal control of unknown robotic systems.

In practice, a robot is often subject to input saturation which can degrade the system performance and even incur instability. Under the frame of optimal algorithms, such as Q-learning algorithm [63], data-based control [64] and dynamic programming [65], how to optimize the system performance and eliminate the effect of input saturation is an open problem.

REFERENCES

[1] C. Yang, C. Zeng, P. Liang, Z. Li, R. Li, and C. Y. Su, “Interface design of a physical human robot interaction system for human impedance adaptive skill transfer,” *IEEE Transactions on Automation Science and Engineering*, vol. 15, pp. 329–340, Jan 2018.

[2] P. Huang, F. Zhang, J. Cai, D. Wang, Z. Meng, and J. Guo, “Dexterous tethered space robot: Design, measurement, control, and experiment,” *IEEE Transactions on Aerospace and Electronic Systems*, vol. 53, pp. 1452–1468, June 2017.

[3] H. Pu, N. Liu, J. Yu, Y. Yang, Y. Sun, Y. Peng, S. Xie, J. Luo, L. Dong, H. Chen, and Y. Sun, “Micropipette aspiration of single cells for both mechanical and electrical characterization,” *IEEE Transactions on Biomedical Engineering*, vol. 66, no. 11, pp. 3185–3191, 2019.

[4] 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, pp. 1–3, 2020.

[5] X. Yu, S. Zhang, L. Sun, Y. Wang, and B. Li, “Cooperative control of dual-arm robots in different human-robot collaborative tasks,” *Assembly Automation*, vol. 40, no. 1, pp. 95–104, 2019.

[6] S. Wu, Z. Wang, B. Shen, J.-H. Wang, and L. Dongdong, “Human-computer interaction based on machine vision of a smart assembly workbench,” *Assembly Automation*, vol. 40, no. 3, pp. 475–482, 2020.

[7] C. Yang, H. Wu, Z. Li, W. He, N. Wang, and C. Su, “Mind control of a robotic arm with visual fusion technology,” *IEEE Transactions on Industrial Informatics*, vol. 14, pp. 3822–3830, Sept 2018.

[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, 2015.

[9] 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, 2018.

[10] K. D. Kallu, W. Jie, and M. C. Lee, “Sensorless reaction force estimation of the end effector of a dual-arm robot manipulator using sliding mode control with a sliding perturbation observer,” *International Journal of Control, Automation and Systems*, vol. 16, pp. 1367–1378, Jun 2018.

[11] Z. Li, T. Zhao, F. Chen, Y. Hu, C. Su, and T. Fukuda, “Reinforcement learning of manipulation and grasping using dynamical movement primitives for a humanoidlike mobile manipulator,” *IEEE/ASME Transactions on Mechatronics*, vol. 23, pp. 121–131, Feb 2018.

[12] Z. Li, W. Yuan, Y. Chen, F. Ke, X. Chu, and C. L. P. Chen, “Neural-dynamic optimization-based model predictive control for tracking and formation of nonholonomic multirobot systems,” *IEEE Transactions on Neural Networks and Learning Systems*, vol. 29, pp. 6113–6122, Dec 2018.

[13] J. Zhang, “Design of a new pid controller using predictive functional control optimization for chamber pressure in a coke furnace.,” *Isa Transactions*, vol. 67, p. 208, 2017.

[14] J. Viola, L. Angel, and J. M. Sebastian, “Design and robust performance evaluation of a fractional order pid controller applied to a dc motor,” *IEEE/CAA Journal of Automatica Sinica*, vol. 4, no. 2, pp. 304–314, 2017.

[15] Y. Su and C. Zheng, “Pid control for global finite-time regulation of robotic manipulators,” *International Journal of Systems Science*, vol. 48, no. 3, pp. 547–558, 2017.

[16] C. Hua, Y. Li, and X. Guan, “Finite/fixed-time stabilization for nonlinear interconnected systems with dead-zone input,” *IEEE Transactions on Automatic Control*, vol. 62, pp. 2554–2560, May 2017.

[17] B. Xu, “Composite learning finite-time control with application to quadrotors,” *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 48, pp. 1806–1815, Oct 2018.

[18] C. Yang, K. Huang, H. Cheng, Y. Li, and C. Y. Su, “Haptic identification by elm-controlled uncertain manipulator,” *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 47, pp. 2398–2409, Aug 2017.

[19] S. Dai, S. He, H. Lin, and C. Wang, “Platoon formation control with prescribed performance guarantees for usvs,” *IEEE Transactions on Industrial Electronics*, vol. 65, pp. 4237–4246, May 2018.

[20] H. Li, L. Bai, L. Wang, Q. Zhou, and H. Wang, “Adaptive neural control of uncertain nonstrict-feedback stochastic nonlinear systems with output constraint and unknown dead zone,” *IEEE Transactions on Systems Man & Cybernetics Systems*, vol. 47, no. 8, pp. 2048–2059, 2017.

[21] B. Xu, D. Yang, Z. Shi, Y. Pan, B. Chen, and F. Sun, “Online recorded data-based composite neural control of strict-feedback systems with application to hypersonic flight dynamics,” *IEEE Transactions on Neural Networks and Learning Systems*, vol. 29, pp. 3839–3849, Aug 2018.

[22] D. Duan and G. Zong, “Exponential l_1 filtering of networked linear switched systems: An event-triggered approach,” *Journal of Systems ence and Complexity*, no. 7, pp. 383–400, 2020.

[23] 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. 1, pp. 119203–, 2020.

[24] 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, doi: 10.1109/TSMC.2019.2922305, 2020.

[25] J. Fu, R. Ma, and T. Chai, “Adaptive finite-time stabilization of a class of uncertain nonlinear systems via logic-based switchings,” *IEEE Transactions on Automatic Control*, vol. 62, no. 11, pp. 5998–6003, 2017.

[26] 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, 2020.

[27] Y. Song and X. Yuan, “Low-cost adaptive fault-tolerant approach for semiactive suspension control of high-speed trains,” *IEEE Transactions on Industrial Electronics*, vol. 63, no. 11, pp. 7084–7093, 2016.

[28] Z. Zhang and W. Chen, “Adaptive output feedback control of nonlinear systems with actuator failures,” *Information Sciences*, vol. 179, no. 24, pp. 4249–4260, 2009.

[29] C. L. P. Chen, G.-X. Wen, Y.-J. Liu, and Z. Liu, “Observer-based adaptive backstepping consensus tracking control for high-order nonlinear semi-strict-feedback multiagent systems,” *IEEE Transactions on Cybernetics*, vol. 46, no. 7, pp. 1591–1601, 2016.

[30] L. Kong, W. He, C. Yang, Z. Li, and C. Sun, “Adaptive fuzzy control for coordinated multiple robots with constraint using impedance learning,” *IEEE Transactions on Cybernetics*, vol. 49, no. 8, pp. 3052–3063, 2019.

[31] Q. Zhou, H. Li, L. Wang, and R. Lu, “Prescribed performance observer-based adaptive fuzzy control for nonstrict-feedback stochastic nonlinear systems,” *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 48, pp. 1747–1758, Oct 2018.

[32] L. Wang, M. V. Basin, H. Li, and R. Lu, “Observer-based composite adaptive fuzzy control for nonstrict-feedback systems with actuator failures,” *IEEE Transactions on Fuzzy Systems*, vol. 26, pp. 2336–2347, Aug 2018.

[33] V. Parra-Vega, S. Arimoto, Y. H. Liu, and G. Hirzinger, “Dynamic sliding pid control for tracking of robot manipulators: theory and experiments,” *IEEE Transactions on Robotics & Automation*, vol. 19, no. 6, pp. 967–976, 2003.

[34] J. T. Wen and S. H. Murphy, “Pid control for robot manipulator,” *Cirsse Document Rensselaer Polytechnic Institute*, vol. 54, 1990.

[35] F. L. Lewis, A. Yesildirak, and S. Jagannathan, *Neural Network Control of Robot Manipulators and Nonlinear Systems*. Taylor & Francis, Inc., 1998.

[36] L. Kong, W. He, Y. Dong, L. Cheng, C. Yang, and Z. Li, “Asymmetric bounded neural control for an uncertain robot by state feedback and output feedback,” *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, in press, 2019, doi: 10.1109/TSMC.2019.2901277.

[37] P. Liu, Z. Zeng, and J. Wang, “Multiple mittag–leffler stability of fractional-order recurrent neural networks,” *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 47, no. 8, pp. 2279–2288, 2017.

[38] G. Xie, A. Shangguan, R. Fei, W. Ji, W. Ma, and X. Hei, “Motion trajectory prediction based on cnn-lstm sequential model,” *SCIENCE CHINA Information Sciences*, in press, 2020, doi: 10.1007/s11432-019-2761-y.

[39] Y. Ouyang, L. Dong, L. Xue, and C. Sun, “Adaptive control based on neural networks for an uncertain 2-dof helicopter system with input deadzone and output constraints,” *IEEE/CAA Journal of Automatica Sinica*, vol. 6, no. 3, pp. 807–815, 2019.

[40] A. Wu and Z. Zeng, “Global mittag-leffler stabilization of fractional-order memristive neural networks,” *IEEE Transactions on Neural Networks and Learning Systems*, vol. 28, no. 1, pp. 206–217, 2017.

[41] J. Chen and H. Qiao, “Muscle-synergies-based neuromuscular control for motion learning and generalization of a musculoskeletal system,” *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, in press, 2020, doi: 10.1109/TSMC.2020.2966818.

[42] X. Yu, W. He, Y. Li, C. Xue, Y. Sun, and Y. Wang, “Adaptive nn impedance control for an sea-driven robot,” *SCIENCE CHINA Information Sciences*, vol. 63, no. 5, pp. 159207–, 2020.

[43] C. Mu, Z. Ni, C. Sun, and H. He, "Air-breathing hypersonic vehicle tracking control based on adaptive dynamic programming," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 28, pp. 584–598, March 2017.

[44] W. He, S. S. Ge, Y. Li, E. Chew, and Y. S. Ng, "Neural network control of a rehabilitation robot by state and output feedback," *Journal of Intelligent & Robotic Systems*, vol. 80, no. 1, pp. 15–31, 2015.

[45] C. Hua, L. Zhang, and X. Guan, "Distributed adaptive neural network output tracking of leader-following high-order stochastic nonlinear multiagent systems with unknown dead-zone input," *IEEE Transactions on Cybernetics*, vol. 47, pp. 177–185, Jan 2017.

[46] C. Yang, Y. Jiang, Z. Li, W. He, and C. Y. Su, "Neural control of bimanual robots with guaranteed global stability and motion precision," *IEEE Transactions on Industrial Informatics*, vol. 13, pp. 1162–1171, June 2017.

[47] L. Chen, Q. Wang, X. Lu, D. Cao, and F. Wang, "Learning driving models from parallel end-to-end driving data set," *Proceedings of the IEEE*, vol. 108, no. 2, pp. 262–273, 2020.

[48] S. Zhang, Y. Dong, Y. Ouyang, Z. Yin, and K. Peng, "Adaptive neural control for robotic manipulators with output constraints and uncertainties," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 29, pp. 5554–5564, Nov 2018.

[49] 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, 2017.

[50] C. Yang, T. Teng, B. Xu, Z. Li, J. Na, and C.-Y. Su, "Global adaptive tracking control of robot manipulators using neural networks with finite-time learning convergence," *International Journal of Control, Automation and Systems*, vol. 15, pp. 1916–1924, Aug 2017.

[51] W. He, W. Ge, Y. Li, Y. Liu, C. Yang, and C. Sun, "Model identification and control design for a humanoid robot," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 47, pp. 45–57, Jan 2017.

[52] P. J. Werbos, "Approximate dynamic programming for real-time control and neural modeling," *Handbook of Intelligent Control Neural Fuzzy & Adaptive Approaches*, 1992.

[53] D. Wang, D. Liu, and H. Li, "Policy iteration algorithm for online design of robust control for a class of continuous-time nonlinear systems," *IEEE Transactions on Automationence & Engineering*, vol. 11, no. 2, pp. 627–632, 2014.

[54] D. Liu, D. Wang, F. Y. Wang, H. Li, and X. Yang, "Neural-network-based online hjb solution for optimal robust guaranteed cost control of continuous-time uncertain nonlinear systems.," *IEEE Transactions on Cybernetics*, vol. 44, no. 12, pp. 2834–2847, 2014.

[55] D. Wang, D. Liu, C. Mu, and Y. Zhang, "Neural network learning and robust stabilization of nonlinear systems with dynamic uncertainties," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 29, pp. 1342–1351, April 2018.

[56] D. Wang, H. He, and D. Liu, "Adaptive critic nonlinear robust control: A survey," *IEEE Transactions on Cybernetics*, vol. 47, no. 10, pp. 3429–3451, 2017.

[57] Y. Yang, D. Wunsch, and Y. Yin, "Hamiltonian-driven adaptive dynamic programming for continuous nonlinear dynamical systems," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 28, pp. 1929–1940, Aug 2017.

[58] D. Wang, H. He, and D. Liu, "Improving the critic learning for event-based nonlinear h_∞ control design," *IEEE Transactions on Cybernetics*, vol. 47, no. 10, pp. 3417–3428, 2017.

[59] D. Wang, H. He, C. Mu, and D. Liu, "Intelligent critic control with disturbance attenuation for affine dynamics including an application to a microgrid system," *IEEE Transactions on Industrial Electronics*, vol. 64, no. 6, pp. 4935–4944, 2017.

[60] D. Nodland, H. Zargarzadeh, and S. Jagannathan, "Neural network-based optimal adaptive output feedback control of a helicopter uav.," *IEEE Transactions on Neural Networks & Learning Systems*, vol. 24, no. 7, pp. 1061–1073, 2013.

[61] 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, 2010.

[62] T. Dierks and S. Jagannathan, "Optimal control of affine nonlinear continuous-time systems," in *Proceedings of the 2010 American Control Conference*, pp. 1568–1573, June 2010.

[63] Y. Yang, K. G. Vamvoudakis, H. Ferraz, and H. Modares, "Dynamic intermittent q-learning–based model-free suboptimal co-design of-stabilization," *International Journal of Robust and Nonlinear Control*, vol. 29, no. 9, pp. 2673–2694, 2019.

[64] J. Fu, T.-F. Li, T. Chai, and C.-Y. Su, "Sampled-data-based stabilization of switched linear neutral systems," *Automatica*, vol. 72, pp. 92–99, 2016.

[65] 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, 2018.

Appendix

Proof: The Lyapunov function candidate is chosen as

$$V = \frac{1}{2}\tilde{\omega}^T\tilde{\omega} + \alpha_c J_s(x) \tag{51}$$

Substituting (45) into the time derivative of (51) yields

$$\begin{aligned}\dot{V} =& \tilde{\omega}^T(-\alpha_H\delta_1\delta_1^T\tilde{\omega} + \alpha_H\frac{\delta_1}{\delta_2}e_c - \frac{1}{2}\alpha_c h\nabla\sigma(x) \\ & \times (B + \frac{1}{\kappa^2}\mathcal{B})\nabla J_s(x)) + \alpha_c(\nabla J_s(x))^T\dot{x}\end{aligned} \tag{52}$$

Note that $\delta_1 = \frac{\delta}{(1+\delta^T\delta)}$ and $\delta_2 = 1 + \delta^T\delta \geq 1$, we have $\alpha_H\tilde{\omega}^T\frac{\delta_1}{\delta_2}e_c \leq \frac{1}{2}\lambda_{\max}(\delta_1\delta_1^T)\|\tilde{\omega}\|^2 + \frac{1}{2}\alpha_H^2\lambda_{e_c}^2$. Let $\lambda_{\max}(\delta_1\delta_1^T) = k\lambda_{\min}(\delta_1\delta_1^T)$ with k being a positive constant, and (52) is written as

$$\begin{aligned}\dot{V} =& -(\alpha_H - \frac{k}{2})\lambda_{\min}(\delta_1\delta_1^T)\|\tilde{\omega}\|^2 + \frac{1}{2}\alpha_H^2\lambda_{e_c}^2 - \frac{1}{2}\alpha_c\tilde{\omega}^T h\nabla\sigma(x) \\ & \times (B + \frac{1}{\kappa^2}\mathcal{B})\nabla J_s(x) + \alpha_c(\nabla J_s(x))^T\dot{x}\end{aligned} \tag{53}$$

with $\alpha_H > \frac{k}{2}$. Furthermore, it is required to satisfy $\lambda_{\min}(\delta_1\delta_1^T) > 0$, which has been guaranteed according to PE condition presented in Assumption 1.

Case one: $h = 0$. Since $(\nabla J_s(x))^T\dot{x} < 0$, it follows that $-(\nabla J_s(x))^T\dot{x} > 0$. According to the density property of real numbers, we know that there exist a positive constant λ_s satisfying $0 < \lambda_s\alpha_c\|\nabla J_s(x)\| < -\alpha_c(\nabla J_s(x))^T\dot{x}$. Hence, we further have $\dot{V} \leq -(\alpha_H - \frac{k}{2})\lambda_{\min}(\delta_1\delta_1^T)\|\tilde{\omega}\|^2 + \frac{1}{2}\alpha_H^2\lambda_{e_c}^2 - \lambda_s\alpha_c\|\nabla J_s(x)\|$. Then, if the following inequality hold: $\|\tilde{\omega}\| > \frac{\alpha_H\lambda_{e_c}}{\sqrt{(2\alpha_H - k)\lambda_{\min}(\delta_1\delta_1^T)}} \triangleq \mathscr{A}_1$ or $\|\nabla J_s(x)\| > \frac{\alpha_H^2\lambda_{e_c}^2}{2\lambda_s\alpha_c} \triangleq \mathscr{B}_1$, we would obtain $\dot{V} < 0$.

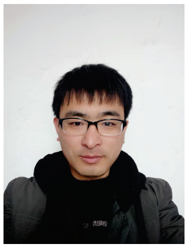
Case two: $h = 1$. Note that $\dot{x}^* = f(x) + g(x)\mu + (I - g(x)g^+(x))v$, we further have $\dot{x}^* = f(x) - \frac{1}{2}B(\nabla\sigma(x))^T\omega - \frac{1}{2}B\nabla\epsilon(x) - \frac{1}{2}\mathcal{B}(\nabla\sigma(x))^T\omega - \frac{1}{2}\mathcal{B}\nabla\epsilon(x)$. Therefore, we have $\dot{x} = \dot{x}^* + \frac{1}{2}B(\nabla\sigma(x))^T\tilde{\omega} + \frac{1}{2}B\nabla\epsilon(x) + \frac{1}{2\kappa^2}\mathcal{B}(\nabla\sigma(x))^T\tilde{\omega} + \frac{1}{2\kappa^2}\mathcal{B}\nabla\epsilon(x)$. Therefore, (53) is rewritten as

$$\begin{aligned}\dot{V} \leq& -(\alpha_H - \frac{k}{2})\lambda_{\min}(\delta_1\delta_1^T)\|\tilde{\omega}\|^2 + \frac{1}{2}\alpha_H^2\lambda_{e_c}^2 + \alpha_c(\nabla J_s(x))^T \\ & \times (f(x) + g(x)\mu + (I - g^+(x)g(x))v) \\ & + \frac{\alpha_c}{2}(\nabla J_s(x))^T(B + \frac{1}{\kappa^2}\mathcal{B})\nabla\epsilon(x)\end{aligned} \tag{54}$$

According to the definition of $J_s(x)$, we know that $(\nabla J_s(x))^T(f(x) + g(x)\mu + (I - g(x)g^+(x))v) = -(\nabla J_s(x))^T\tilde{G}\nabla J_s(x) < -\lambda_1\|\nabla J_s(x)\|^2$, where λ_1 is the minimum eigenvalue of matrix $\tilde{G}$. Assume that $\|B + \frac{1}{\kappa^2}\mathcal{B}\| \leq \lambda_{2M}$, where λ_{2M} is a positive constant. In terms of Young's inequality, it follows that $\dot{V} \leq -\lambda_2\|\tilde{\omega}\|^2 + \frac{1}{2}\alpha_H^2\lambda_{e_c}^2 - \lambda_1\|\nabla J_s(x)\|^2 + \frac{1}{2}\alpha_c\lambda_\epsilon\lambda_{2M}\|\nabla J_s(x)\|$, where $\lambda_2 = (\alpha_H - \frac{k}{2})\lambda_{\min}(\delta_1\delta_1^T)$. Then, if the following inequality holds: $\|\tilde{\omega}\| > \sqrt{\frac{\alpha_H^2\lambda_{e_c}^2}{2\lambda_2} + \frac{\alpha_c^2\lambda_\epsilon^2\lambda_{2M}^2}{16\lambda_2\lambda_1}} \triangleq \mathscr{A}_2$ or

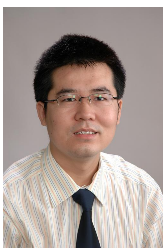
$\|\nabla J_s(x)\| > -\frac{\alpha_c\lambda_\epsilon\lambda_{2M}}{4\lambda_1} + \frac{\sqrt{\alpha_c^2\lambda_\epsilon^2\lambda_{2M}^2+8\lambda_1\lambda_H^2\lambda_{e_c}^2}}{4\lambda_1} \triangleq \mathscr{B}_2$, we would obtain $\dot{V} < 0$.

In the end, define $\mathscr{A} = \max\{\mathscr{A}_1, \mathscr{A}_2\}$ and $\mathscr{B} = \max\{\mathscr{B}_1, \mathscr{B}_2\}$. If $\|\omega\| \geq \mathscr{A}$ or $\|\nabla J_s(x)\| \geq \mathscr{B}$, we can conclude that $\dot{V} < 0$. Consequently, we conclude that $\tilde{\omega}$ and x are UUB. This finished the proof.



Linghuan Kong received the B.Eng. degree from the College of Engineering, Qufu Normal University, Rizhao, China, in 2016 and the M.Eng. degree from the School of Automation Engineering, University of Electronic Science and Technology of China, Chengdu, China, in 2019. He is currently pursuing the Ph.D. degree with the School of Automation and Electrical Engineering, University of Science and Technology Beijing, Beijing, China.

His current research interests include robotics, neural network control, and adaptive control.



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. He was a recipient of the IEEE SMC Society Andrew P. Sage Best Transactions Paper Award. 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*, *Assembly Automation*, and an Editor of *Journal of Intelligent & Robotic Systems*.

His current research interests include robotics, distributed parameter systems and intelligent control systems.

Chenguang Yang (M'10-SM'16) received the Ph.D. degree in control engineering from the National University of Singapore, Singapore, in 2010, and postdoctoral training in human robotics from the Imperial College London, London, U.K.

He has been awarded EU Marie Curie International Incoming Fellowship, UK EPSRC UKRI Innovation Fellowship, and the Best Paper Award of the IEEE Transactions on Robotics as well as over ten international conference Best Paper Awards. He is a Co-Chair of the Technical Committee on Bio-mechatronics and Bio-robotics Systems (B2S), IEEE Systems, Man, and Cybernetics Society, and a Co-Chair of the Technical Committee on Collaborative Automation for Flexible Manufacturing (CAFM), IEEE Robotics and Automation Society. He serves as Associate Editor of a number of IEEE Transactions and other international leading journals.

His research interest lies in human robot interaction and intelligent system design.

Changyin Sun received the B.S. degree in applied mathematics from the College of Mathematics, Sichuan University, Chengdu, China, in 1996, and the M.S. and Ph.D. degrees in electrical engineering from Southeast University, Nanjing, China, in 2001 and 2003, respectively.

He is currently a Professor with the School of Automation, Southeast University. His research interests include intelligent control, flight control, pattern recognition, optimal theory, etc. Dr. Sun is an Associate Editor for the *IEEE Transactions on Neural Networks and Learning Systems*, *Neural Processing Letters*, and *IEEE/CAA Journal of Automatica Sinica*.