Journal of Computer Science and Cybernetics, V.39, N.1 (2023), 61–77
DOI no 10.15625/1813-9663/17641

OPTIMAL TRACKING CONTROL FOR ROBOT MANIPULATORS WITH ASYMMETRIC SATURATION TORQUES BASED ON REINFORCEMENT LEARNING

NGUYEN DUC DIEN[1], NGUYEN TAN LUY[2,*], LAI KHAC LAI[3]

[1] *University of Economics Technology for Industry, 456 Minh Khai Street, Hai Ba Trung District, Ha Noi, Viet Nam*

[2] *Ho Chi Minh City University of Technology–VNU–HCM, 268 Ly Thuong Kiet Street, District 10, Ho Chi Minh City, Viet Nam*

[3] *Thai Nguyen University of Technology, 666, 3/2 Street, Tich Luong Ward, Thai Nguyen City, Thai Nguyen Province Viet Nam*

Abstract. This paper introduces an optimal tracking controller for robot manipulators with asymmetric saturation torques and partially unknown dynamics based on a reinforcement learning method using a neural network. Firstly, the feedforward control inputs are designed based on the backstepping technique to convert the tracking control problem into the optimal tracking control problem. Secondly, a cost function of the system with asymmetric saturation inputs is defined, and the constrained Hamilton-Jacobi-Bellman equation is built, which is solved by the online reinforcement learning algorithm using only a single neural network. Then, the asymmetric saturation optimal control rule is determined. Additionally, the concurrent learning technique is used to relax the demand for the persistence of excitation conditions. The built algorithm ensures that the tracking error and the approximation error are uniformly ultimately bounded (UUB), and the cost function converges to the near-optimal value. Finally, the effectiveness of the proposed algorithm is shown through comparative simulations.

Keywords. Robot manipulators, reinforcement learning, optimal tracking control, asymmetric saturation inputs.

1. INTRODUCTION

Robot manipulators have been widely used and have played an essential role in many fields, such as production, assembly, transportation, and mass production lines [1, 2]. Improving control performance for robot manipulators has received much attention from researchers [3–6], where reference position tracking control is a vital control problem for robot manipulators [3, 6, 7]. In [5], an adaptive controller using a radial basic function (RBF) neural network (NN) was proposed to control the robot with an uncertain model parameter,

*Corresponding author.
E-mail addresses: nddien@uneti.edu.vn (N.D. Dien); ntanluy@hcmut.edu.vn (N.T. Luy); laikhaclai@tnu.edu.vn (l.K. Lai)

and a sliding controller was integrated to suppress the disturbance. In [4], an adaptive sliding fuzzy controller was used, where fuzzy logic for approximating the unknown nonlinear component and a disturbance observer was also proposed. In [8], an adaptive neural PID controller combined with an angular velocity acceleration observer was presented for robot manipulators with uncertain model parameters. Yilmaz et al. [9] proposed a self-tuning fuzzy controller for robot manipulators. In [10], a robust model predictive controller was used to control the trajectory tracking for the robot manipulators. Yang et al. [11] designed an adaptive controller that ensures precise gravity compensation for the robot manipulators without velocity measurement.

In addition to the problem that the robot manipulators are affected by disturbance, the model parameter changes, and the robot manipulators are also affected by the problem that the inputs are saturated. This problem will reduce the control performance and even make the system unstable [12–14]. Ling et al. [15] proposed an adaptive fuzzy controller for manipulators with symmetrically saturated inputs, where the tanh(.) function was used to solve the saturated input problem. Yang et al. [16] proposed a control scheme for the robot manipulators with symmetrically saturated inputs and output error constraints, in which a backend system is designed to solve the saturation input problem, and the RBF NN is used to approximate the uncertain components. Hwang and Chen [17] developed an adaptive finite-time saturation tracking controller for robots with partially known dynamics. The problem of asymmetric saturation of the robot manipulators can occur when the actuator is partially reduced in efficiency, changes in the mechanical structure, and the faulty power drive circuit [18]. Ma et al. [19] proposed an adaptive NN controller for a nonlinear system with model uncertainty, asymmetric saturation inputs, and external noise. Zhou et al. [20] designed a fast terminal sliding mode controller for robots with asymmetric saturation inputs. In [18], an adaptive neural controller was developed based on NN and the backstepping technique to control robot manipulators with asymmetric saturation inputs, where NN is used to approximate unknown dynamics. In [15–18, 20], the proposed algorithms have effectively solved the trajectory tracking problem for robot manipulators with saturated inputs, but they have not considered the optimal control problem.

The adaptive dynamic programming (ADP) algorithm is an effective tool for solving the optimal control problem, where solving the Hamilton-Jacobi-Bellman (HJB) equation is based on the function approximators [21]. Recently, reinforcement learning (RL) has become an effective tool for online approximation of solutions of HJB equations [22–24] and has been applied to the optimal tracking control problem for robot manipulators [25–28]. Kamalapurkar et al. [25] used RL to solve the optimal tracking control problem for robot manipulators, but the algorithm uses an approximator with 2 NNs. Long et al. [27] proposed an optimal tracking control algorithm using only a single NN. Kong et al. [26] proposed a robust optimal control algorithm, but the algorithm needs to know the dynamics. Zhao et al. [28] used RL for the problem of saturated optimal tracking control for robot manipulators in discrete time. In our previous works [29, 30], we designed optimal tracking controllers using RL for nonlinear systems with disturbances. However, the algorithms therein are only applied to large-scale robot systems with symmetrically saturated inputs. In contrast, although Bu [31] proposed an optimal algorithm with asymmetric saturation input and Xia et al. [32] proposed a data-driven off-policy RL algorithm to approximate the solution to the constrained HJB equation without requiring complete knowledge of the dynamics, the algorithms are only applied to a class of affine nonlinear systems. To our best knowledge, traditional RL control rarely considers the optimal tracking problem for robot manipulators with asymmetric saturation torques.

Motivated by the above challenges, in this paper, we design an optimal tracking controller for robot manipulators with asymmetric saturation torques and partially unknown dynamics based on an RL method using a single NN. The main contributions of this paper are summarized as follows: 1) Different from [28–30], for the first time, we propose a feedforward control design to transform a position-tracking control problem into an optimal control problem for robot manipulators with asymmetric saturation torques and partially unknown dynamics.
2) Different from [25], the cost function of the asymmetrical saturation system is proposed, which both satisfies the asymmetric torques and the solution to the HJB equation is positive. The constrained solution is solved by the online RL algorithm using only a single NN instead of two to reduce the computational cost.
3) The algorithm for updating the weights of the NN is determined with partially unknown dynamics. Therefore, the proposed algorithm is more efficient than the algorithm in [31] because it does not require known dynamics. Furthermore, the concurrent learning (CL) technique is used to relax the demand for the persistence of excitation (PE) condition and ensures convergence of NN weights.

The structure of the paper is organized as follows. Section 2 describes the robot manipulator's dynamics and feedforward control inputs' design problem. Section 3 offers a design of the optimal control rule with asymmetric saturation torques and partially unknown dynamics. Section 4 presents the simulation results, and Section 5 gives the paper's conclusion.

2. ROBOT MODEL AND FEEDFORWARD CONTROL

This section presents two problems: A dynamic model of the robot manipulators as a strict-feedback nonlinear system and the feedforward control inputs to convert the tracking problem into the optimal control problem.

2.1. Robot model

Consider a model of a robot manipulator in the form of an Euler – Lagrange system with full actuators. The method based on the Lagrange equation is used to build the dynamic equation for the robot manipulators, which has the following general form [3, 6, 33, 34]

$$M(q)\ddot{q} + C(q,\dot{q})\dot{q} + G(q) + F(\dot{q}) = \tau - \tau_0, \tag{1}$$

where $q \in \Re^{n\times 1}$, $\dot{q} \in \Re^{n\times 1}$, and $\ddot{q} \in \Re^{n\times 1}$ are the joint variables' position, angular velocity, and angular acceleration vectors. The inertia matrix $M(q) \in \Re^{n\times n}$ is symmetric positive definite, $C(q,\dot{q}) \in \Re^{n\times n}$ is the Coriolis-centripetal matrices, $G(q) \in \Re^{n\times 1}$ is the dynamic friction, $F(\dot{q}) \in \Re^{n\times 1}$ is the static friction, $\tau_0 \in \Re^{n\times 1}$ is the vector of external disturbances, and $\tau \in \Re^{n\times 1}$ is the control input vector of joints torques and satisfies

$$\lambda_1 \le \tau_i \le \lambda_2, i = 1, 2, \cdots, n, \tag{2}$$

where τ_i the element of τ, λ_1 and λ_2 are the lower and upper bounds of τ_i, respectively.

The dynamic equation 1 is characterized by the following properties [3, 33].

Property 1. *The matrix $M(q)$ is bounded. It means $m_{min} \leq \|M(q)\| \leq m_{max}$, where m_{min} and m_{max} is positive constant.*

Property 2. *The matrix $C(q,\dot{q})$ is bounded, i.e. $\|C(q,\dot{q})\| \leq c_{max}$, where c_{max} are positive constants.*

Property 3. $\dot{M(q)} - 2C(q,\dot{q})$ *is skew-symmetric.*

Property 4. *The matrix $G(q)$ is bounded, which means $\|G(q)\| \leq g_{max}$, where g_{max} is the upper bound of $\|G(q)\|$.*

Remark 1. Properties 1, 2, 3, and 4 are satisfied in the practical robot manipulators and are mentioned when designing the controller for them [3,33,35]. In addition, these properties are used when proving the system's stability; therefore, one does not necessarily define the upper and lower bounds of the $M(q)$, $C(q,\dot{q})$, and $G(q)$ matrices.

To facilitate the design of the controller, we transform dynamics 1 into a strict-feedback nonlinear system with asymmetry saturation torques as follows

$$\begin{cases} \dot{q} = f_q(q) + g_q(q)v \\ \dot{v} = f_v(q,v) + g_v(q,v)\tau, \end{cases} \tag{3}$$

where v is the angular velocity vector of the joint variables, $f_q(q) = 0_{n\times 1}$, $g_q(q) = I_n$, $f_v(q,v) = -M^{-1}(q)\left(C(q,\dot{q})v + G(q) + F(\dot{q})\right) \in \Re^{n\times 1}$ and is considered unknown, $g_v(q,v) = M^{-1}(q) \in \Re^{n\times n}$.

Remark 2. By Properties 1, 2, and 4, $f_v(q,v)$ is bounded, i.e., $\|f_v(q,v)\| \leq -m_{min}^{-1}\,(c_{max} + g_{max})\|v\|$. By Properties 1, $g_v(q,v)$ is bounded such that $\|g_v(q,v)\| \leq m_{min}^{-1}$. In addition, $g_v(q,v)$ is known because the matrix $M(q)$ is well-defined.

Assumptions 1. The reference position trajectory $q_d(t)$ is smooth and bounded.

Problem formulation: The main objective for the optimal tracking control problem for robot manipulators with asymmetric saturation torques and partially unknown dynamics is to design feedback control rules for the system 3, such that $\lim\limits_{t\to\infty}(q(t) - q_d(t)) \to 0$ with $q_d(t)$ as the reference position trajectory. Furthermore, a defined tracking cost function related to tracking errors and control inputs is minimized.

2.2. Feedforward control

In this section, we design the feedforward control inputs by applying the backstepping method and then transform 3 into an optimal control problem.

The virtual control inputs and the control inputs are determined as follows

$$\begin{cases} v_d = v_d^* + v_{da} \\ \tau = \tau^* + \tau_a, \end{cases} \tag{4}$$

where v_d^* is the virtual optimal control input vector, v_{da} is the feedforward virtual control input vector, τ^* is the optimal control input vector, and τ_a is the feedforward control input vector.

First, we define the tracking error as

$$\begin{cases} e_q = q - q_d \\ e_v = \dot{q} - v_d. \end{cases} \tag{5}$$

Then, we design the feedforward control inputs through the backstepping technique by following the steps.
Step 1: Calculate the derivative of the position tracking error and combine it with 4, we obtain

$$\dot{e}_q = -\dot{q}_d + g_q(q)v_d^* + g_q(q)v_{da} + g_q(q)e_v. \tag{6}$$

We design v_{da} as follows

$$v_{da} = g_q^{-1}(q)(\Lambda_1 e_v - \Lambda_2 e_q + \dot{q}_d). \tag{7}$$

By substituting 7 into 6, one has

$$\dot{e}_q = g_g(q)v_d^* + \Lambda_1 e_v - \Lambda_2 e_q + g_q(q)e_v = \bar{f}_q(e_q, e_v) + g_q(q)v_d^* + g_q(q)e_v, \tag{8}$$

where $\bar{f}_q(e_q, e_v) = \Lambda_1 e_v - \Lambda_2 e_q$.

Step 2: Taking the derivative e_v and noting 4, we get

$$\dot{e}_v = -\dot{v}_d + f_v(q, v) + g_v(q, v)\tau^* + g_v(q, v)\tau_a. \tag{9}$$

We design τ_a as follows

$$\tau_a = g_v^{-1}(q, v)[\dot{v}_d - g_q^T(q)e_q - \Lambda_3 e_v - \Lambda_4 e_q], \tag{10}$$

where τ_a is constrained by

$$\lambda_1 - \rho_1 \tanh(1) \leq \tau_{ai} \leq \lambda_2 - \rho_2 \tanh(1), \; i = 1, 2, \cdots, n, \tag{11}$$

where τ_{ai} is the element of τ_a, $|\rho_1| < |\lambda_1|$, $|\rho_2| < |\lambda_2|$. Noting 4 and 11, we see that if the element τ_i^* of τ^* is designed to satisfy the constraint $\rho_1 \leq \tau_i^* \leq \rho_2$, then τ_i will be constrained $\lambda_1 \leq \tau_i \leq \lambda_2$. By substituting 10 into 9, dynamics 9 becomes

$$\dot{e}_v = \bar{f}_v(e_q, e_v) + g_v(q, v)\tau^* - g_q^T(q)e_q, \tag{12}$$

where $\bar{f}_v(e_q, e_v) = f_v(q, v) - \Lambda_3 e_v - \Lambda_4 e_q$.

Lemma 1. *Consider the dynamics 3, the virtual control inputs and the control inputs are determined in 4, with the feedforward control inputs designed in 7 and 10. Then the control problem of the system 3 is transformed into the control problem of the following system*

$$\dot{z} = \bar{f}_{qv}(e_q, e_v) + g_{qv}(q, v)u^*, \tag{13}$$

with asymmetric saturation optimal control inputs vector u^*, *where* $z = [e_q^T, e_v^T]^T \in \Re^{2n\times 1}$, $\bar{f}_{qv}(e_q, e_v) = [\bar{f}_q^T(e_q, e_v), \bar{f}_v^T(e_q, e_v)]^T \in \Re^{2n\times 1}$, $g_{qv}(q, v) = diag[g_q(q), g_v(q, v)] \in \Re^{2n\times 2n}$, $u^* = [v_d^{*T}, \tau^{*T}]^{\mathrm{T}} \in \Re^{2n\times 1}$, $u_a = [v_{da}^T, \tau_a^T]^T \in \Re^{2n\times 1}$, *and* $u_\tau = [v_d^T, \tau^T]^T = u^* + u_a \in \Re^{2n\times 1}$.

Proof. Choose a Lyapunov function for system 3 as follows

$$J_1 = \frac{1}{2}e_q^T e_q + \frac{1}{2}e_v^T e_v. \tag{14}$$

Taking derivative 14 along trajectories of 8 and 12, one obtains

$$\begin{aligned}\dot{J}_1 &= e_q^T \bar{f}_q(e_q, e_v) + e_q^T g_q(q)v_d^* + e_q^T g_q(q)e_v + e_v^T \bar{f}_v(e_q, e_v) + e_v^T g_v(q, v)\tau^* - e_v^T g_q^T(q)e_q \\ &= e_q^T \left(\bar{f}_q(e_q, e_v) + g_q(q)v_d^*\right) + e_v^T \left(\bar{f}_v(e_q, e_v) + g_v(q, v)\tau^*\right) \\ &= z^T(\bar{f}_{qv}(e_q, e_v) + g_{qv}(q, v)u^*).\end{aligned} \tag{15}$$

On the other hand, choose a Lyapunov function for system 13 as follows

$$J_2 = \frac{1}{2}z^T z. \tag{16}$$

Taking derivative 16, one obtains

$$\dot{J}_2 = \frac{1}{2}z^T(\bar{f}_{qv}(e_q, e_v) + g_{qv}(q, v)u^*). \tag{17}$$

We see that if we design a control rule u^* that stabilizes the closed system 13, then $\dot{J}_2 < 0$. Comparing 17 with 15, $\dot{J}_2 < 0$ results in $\dot{J}_1 < 0$, therefore, the closed system 3 is also stable. In other words, the optimal tracking control problem of 3 is equivalent to the optimal tracking controller design for 13.

The proof is completed. ∎

3. OPTIMAL TRACKING CONTROL

In this section, we design the optimal tracking controller for system 13 with asymmetric saturation inputs and partially unknown dynamics.

To design the optimal control algorithm, we define the cost function for the system 13 as follows

$$V(z) = \int_t^\infty r(z(\tau), u(\tau))d\tau, \tag{18}$$

where $r(z, u) = z^T Qz + \Upsilon(u)$, Q is a positive definite matrix, $\Upsilon(u)$ is non-negative, and u is an approximation of u^* at each given time.

Inspired by [31, 36, 37], the energy cost function is defined as

$$\Upsilon(u) = (\rho_2 - \rho_1)\int_{\frac{\rho_2+\rho_1}{2}}^{u} \tanh^{-T}\left(\frac{2s - \rho_2 - \rho_1}{\rho_2 - \rho_1}\right) Rds, \tag{19}$$

where R is a positive definite diagonal matrix. Using integral by parts, 19 becomes

$$\Upsilon(u) = (\rho_2 - \rho_1)\tanh^{-T}\left(\frac{2u - \rho_2 - \rho_1}{\rho_2 - \rho_1}\right) Ru + \Upsilon_1(u), \tag{20}$$

where

$$\Upsilon_1(u) = -(\rho_2 - \rho_1)\bar{R}\int_{\frac{\rho_2+\rho_1}{2}}^{u} s\nabla\ell ds, \tag{21}$$

where $\bar{R}$ is a column vector of the diagonal elements of R, $\ell = \tanh^{-T}\left(\frac{2s-\rho_2-\rho_1}{\rho_2-\rho_1}\right)$, $\nabla\ell = \partial\ell/\partial s$.

We have $\ell = \tanh^{-T}\left(\frac{2s-\rho_2-\rho_1}{\rho_2-\rho_1}\right) \to s = \frac{1}{2}\left(\tanh(\ell)\left(\rho_2-\rho_1\right)+\left(\rho_2+\rho_1\right)\right)$, 21 becomes

$$\Upsilon_1(u) = -\frac{\rho_2-\rho_1}{2}\bar{R}\int_0^{\tanh^{-T}\left(\frac{2u-\rho_2-\rho_1}{\rho_2-\rho_1}\right)}\left(\tanh(\ell)\left(\rho_2-\rho_1\right)+\left(\rho_2+\rho_1\right)\right)d\ell. \tag{22}$$

Defining $j = \tanh(\ell)$, $\bar{1} = [1,1,\cdots,1]^T \in \Re^{2n}$, we have

$$d\ell = \frac{dj}{\bar{1}-\tanh^2(\ell)} = \frac{dj}{\bar{1}-j^2}. \tag{23}$$

Substituting 23 into 22, one obtains

$$\begin{aligned}\Upsilon_1(u) &= -\frac{\left(\rho_2-\rho_1\right)^2}{2}\bar{R}\int_0^{\frac{2u-\rho_2-\rho_1}{\rho_2-\rho_1}}\frac{j}{\bar{1}-j^2}dj - R\left(\rho_2+\rho_1\right)\frac{\rho_2-\rho_1}{2}\tanh^{-T}\left(\frac{2u-\rho_2-\rho_1}{\rho_2-\rho_1}\right)\\ &= \frac{\left(\rho_2-\rho_1\right)^2}{4}\bar{R}\ln\left(\bar{1}-\left(\frac{2u-\rho_2-\rho_1}{\rho_2-\rho_1}\right)^2\right)\\ &\quad - R(\rho_2+\rho_1)\frac{\rho_2-\rho_1}{2}\tanh^{-T}\left(\frac{2u-\rho_2-\rho_1}{\rho_2-\rho_1}\right).\end{aligned} \tag{24}$$

Combining 24 and 20, $\Upsilon(u)$ is written as

$$\begin{aligned}\Upsilon(u) &= (\rho_2-\rho_1)R\tanh^{-T}\left(\frac{2u-\rho_2-\rho_1}{\rho_2-\rho_1}\right)\left(u-\frac{\rho_2+\rho_1}{2}\right)\\ &\quad + \bar{R}\frac{(\rho_2-\rho_1)^2}{4}\ln\left(\bar{1}-\left(\frac{2u-\rho_2-\rho_1}{\rho_2-\rho_1}\right)^2\right).\end{aligned} \tag{25}$$

The Hamilton function for system 13 with a cost function 18 is defined as follows

$$H\left(z,u,V_z\right) = z^TQz + \Upsilon(u) + V_z^T\left(\bar{f}_{qv}+g_{qv}u\right). \tag{26}$$

where $V_z = \partial V(z)/\partial z$. The HJB equation is presented as

$$\min_u H(z,u,V_z^*) = 0, \tag{27}$$

where $V_z^* = \partial V^*(z)/\partial z$. Then, the optimal control rule is defined as

$$u^* = \arg\min_u H(z,u,V_z^*). \tag{28}$$

Using the stationary condition $\frac{\partial H(z,u,V_z^*)}{\partial u} = 0$, the optimal asymmetric constraint control rule for system 13 is determined as

$$u^* = -\frac{\rho_2-\rho_1}{2}\tanh\left(\frac{1}{\rho_2-\rho_1}R^{-1}g_{qv}^TV_z^*\right)+\frac{\rho_2+\rho_1}{2}. \tag{29}$$

Substituting 29 into 26, then equation HJB 27 becomes

$$H^*\left(z,u^*,V_z^*\right)=z^TQz+\Upsilon\left(u^*\right)+V_z^{*T}\left(\bar{f}_{qv}+g_{qv}u^*\right)=0. \tag{30}$$

If $V^*(z)$ is determined, one can determine the optimal control rule u^* as 29. However, equation 30 cannot be solved analytically. Thus, we use ADP combined with NN to approximate the solution of equation 30. $V^*(z)$ is approximated as follows

$$V^*(z)=W^T\theta(z)+\varepsilon(z), \tag{31}$$

$$V_z^*=W^T\frac{\partial\theta(z)}{\partial z}+\frac{\partial\varepsilon(z)}{\partial z}=W^T\theta_z+\varepsilon_z, \tag{32}$$

where $W\in\Re^h$ is the NN weight vector, $\theta(z):\Re^n\to\Re^h$ is the NN activation function vector with $\theta(0)=0$, and h is the number of neurons in the hidden layer, $\varepsilon(z)$ is the NN approximation error.

Assumptions 2. One can choose $\theta\left(z\right)$ to be a linearly independent basis set with quadratic, tanh, or sigmoid elements to satisfy $\left\|\theta\left(z\right)\right\|\le s_\theta$, $\left\|\theta_z\right\|=\left\|\partial\theta\left(z\right)/\partial z\right\|\le s_{\nabla\theta}$, $\left\|\varepsilon\left(z\right)\right\|\le s_\varepsilon$, $\left\|\varepsilon_z\right\|=\left\|\partial\varepsilon\left(z\right)/\partial z\right\|\le s_{\nabla\varepsilon}$, where s_θ, $s_{\nabla\theta}$, s_ε, and $s_{\nabla\varepsilon}$ are positive constants [38].

Using 32 for the control rule u^* 29, we have

$$u^*=-\frac{\rho_2-\rho_1}{2}\tanh\left(\frac{1}{\rho_2-\rho_1}R^{-1}g_{qv}^T\left(\theta_z^TW+\varepsilon_z\right)\right)+\frac{\rho_2+\rho_1}{2}. \tag{33}$$

The weight vector W 31 is unknown, thus $V^*(z)$ is approximated by

$$\hat{V}(z)=\hat{W}^T\theta(z), \tag{34}$$

where $\hat{W}\in\Re^h$ is the approximate NN weight vector. The control rule 33 is approximated as

$$\hat{u}=-\frac{\rho_2-\rho_1}{2}\tanh\left(\frac{1}{\rho_2-\rho_1}R^{-1}g_{qv}^T\theta_z^T\hat{W}\right)+\frac{\rho_2+\rho_1}{2}. \tag{35}$$

Substituting 35 into 30, the HJB equation 30 becomes

$$\hat{H}(z,\hat{u},\hat{W}^T\theta_z)=z^TQz+\Upsilon\left(\hat{u}\right)+\hat{W}^T\theta_z\left(\bar{f}_{qv}+g_{qv}\hat{u}\right)=e_{\hat{H}}. \tag{36}$$

It can be seen that $\hat{W}$ needs to be tuned to minimize the error of $\hat{H}(z,\hat{u},\hat{W}^T\theta_z)$. The error function is chosen as $E=\frac{1}{2}\varepsilon_E^T\varepsilon_E$, where

$$\begin{aligned}\varepsilon_E&=\int_{t-T}^{t}e_{\hat{H}}d\tau=\hat{W}^T\int_{t-T}^{t}\theta_z\left(\bar{f}_{qv}+g_{qv}\hat{u}\right)d\tau+\int_{t-T}^{t}\left(z^TQz+\Upsilon\left(\hat{u}\right)\right)d\tau\\&=\hat{W}^T\left(\theta\left(z\left(t\right)\right)-\theta\left(z\left(t-T\right)\right)\right)+\int_{t-T}^{t}\left(z^TQz+\Upsilon\left(\hat{u}\right)\right)d\tau\\&=\hat{W}^T\Delta\theta\left(z\right)+\int_{t-T}^{t}\left(z^TQz+\Upsilon\left(\hat{u}\right)\right)d\tau.\end{aligned} \tag{37}$$

where $T>0$ is the sampling period, $\Delta\theta\left(z\right)=\theta\left(z\left(t\right)\right)-\theta\left(z\left(t-T\right)\right)$.

Using the algorithm gradient descent $\dot{\hat{W}} = -\alpha\frac{\partial E}{\partial\hat{W}} = -\alpha\frac{\partial E}{\partial\varepsilon_E}\frac{\partial\varepsilon_E}{\partial\hat{W}}$ and the CL technique, the update rule $\hat{W}$ is proposed as

$$\begin{aligned}\dot{\hat{W}} &= -\alpha\Delta\theta\left(\Delta\theta^T\hat{W} + \int_{t-T}^{t}\left(\mathrm{z}^TQz + \Upsilon(\hat{u})\right)d\tau\right)\\ &- \alpha\sum\nolimits_{i=1}^{P}\Delta\theta\left(z\left(t_i\right)\right)\left(\Delta\theta^T\left(z\left(t_i\right)\right)\hat{W} + \int_{t_i-T}^{t_i}\left(\mathrm{z}^TQz + \Upsilon(\hat{u})\right)d\tau\right),\end{aligned} \tag{38}$$

where $\alpha > 0$ is the learning rate.

The update rule $\hat{W}$ 38 uses the CL technique to relax a persistent citation condition PE, with $\Delta\phi\left(z\left(t_i\right)\right)$, $r(t_i) = \int_{t_i-T}^{t_i}\left(\mathrm{z}^TQz + \Upsilon(\hat{u})\right)d\tau$. The past data is recorded and stored in $\{\Delta\theta\left(z\left(t_i\right)\right)\}_{i=1}^{P}$, $\{r(t_i)\}_{i=1}^{P}$, where $\{\Delta\theta\left(z\left(t_i\right)\right)\}_{i=1}^{P}$ must be linearly independent, i.e. $\mathrm{rank}(\Delta\theta\left(z\left(t_1\right)\right), \Delta\theta\left(z\left(t_2\right)\right), ..., \Delta\theta\left(z\left(t_P\right)\right)) = N$, the number of selected is $P \geq N$ [36].

Defining $\tilde{W} = W - \hat{W}$, we have

$$\dot{\tilde{W}} = -\alpha\Delta\theta\left(\Delta\theta^T\tilde{W} - \varepsilon_H\right) - \alpha\sum\nolimits_{i=1}^{P}\Delta\theta\left(z\left(t_i\right)\right)\left(\Delta\theta^T\left(z\left(t_i\right)\right)\tilde{W} - \varepsilon_H\left(t_i\right)\right), \tag{39}$$

where $\|\varepsilon_H(s)\| = \left\|\int_{s-T}^{t}\varepsilon_z^T\dot{z}d\tau\right\| \leq s_{\varepsilon H}$, with $s_{\varepsilon H}$ as a positive constant.

4. PROOF OF STABILITY AND CONVERGENCE

In this section, we analyze the stability and convergence of the proposed algorithm. To prove the stability of the algorithm, the following definition is given.

Definition 1. [39] The equilibrium point z_0 of system $\dot{z} = f(z, u), z \in \Re^n$ is said to be UUB if there exists a compact set $\Omega \in \Re^n$ so that for all $z_0 \in \Omega$, there exists a bound B and time $T(B, z_0)$ such that $\|z - z_0\| \leq B$ for all $t > t_0 + T$.

Based on the designed algorithm, we give the following theorem.

Theorem 1. *Considering the dynamics 13 with asymmetric saturation inputs, the HJB equation is given by equation 30, and NN is given by equation 34. Let the weight update rule of NN is defined in equation 38, and the optimal control rule is defined by equation 35. Then the optimal control algorithm ensures that the tracking error and the approximate error of NN are UUB stable, and the control rule 35 converges to the near-optimal value.*

Proof. Choose a Lyapunov function for system 13 as follows

$$J_3 = \underbrace{\int_{t-T}^{t}V^*(z)d\tau}_{J_{31}} + \underbrace{\frac{1}{2}\int_{t-T}^{t}trace\left(\tilde{W}^T\tilde{W}\right)d\tau}_{J_{32}}. \tag{40}$$

Taking derivative J_{31} along trajectories of $\dot{z} = \bar{f}_{qv} + g_{qv}\hat{u}$, one obtains

$$\begin{aligned}\dot{J}_{31} &= \int_{t-T}^{t}V_z^{*T}\dot{z}d\tau = \int_{t-T}^{t}V_z^{*T}\left(\bar{f}_{qv} + g_{qv}\hat{u}\right)d\tau\\ &= \int_{t-T}^{t}\underbrace{\left(V_z^{*T}\left(\bar{f}_{qv} + g_{qv}u^*\right) + V_z^{*T}g_{qv}\left(\hat{u} - u^*\right)\right)}_{\bar{J}_{31}}d\tau,\end{aligned} \tag{41}$$

According to 30 and 32, $\bar{J}_{31}$ of equation 41 can be written as

$$\bar{J}_{31} = -z^T Q z - \Upsilon\left(u^*\right) + g_{qv}\left(\theta_z^{*T} W + \varepsilon_z\right)\left(\hat{u} - u^*\right). \tag{42}$$

Applying Young's inequality to 42, we get

$$\bar{J}_{31} \leq -z^T Q z - \Upsilon\left(u^*\right) + \frac{1}{4}\left\|g_{qv}\left(\theta_z^{*T} W + \varepsilon_z\right)\right\|^2 + \left\|\hat{u} - u^*\right\|^2. \tag{43}$$

On the other hand, we have

$$z^T Q z \geq \lambda_{\min}\left(Q\right)\|z\|^2, \tag{44}$$

where $\lambda\left(.\right)$ is an eigenvalue of the matrix. Substituting 44 into 43 yields

$$\bar{J}_{31} \leq \lambda_{\min}\left(Q\right)\|z\|^2 + \frac{1}{4}\left\|g_{qv}\left(\theta_z^{*T} W + \varepsilon_z\right)\right\|^2 + \left\|\hat{u} - u^*\right\|^2 - \Upsilon\left(u^*\right). \tag{45}$$

Then, using inequality $(a+b)^2 \leq 2\left(a^2 + b^2\right)$ and noting 33-35, we have

$$\begin{aligned}\bar{J}_{31} \leq{}& \lambda_{\min}\left(Q\right)\|z\|^2 + \frac{1}{2}\left(\left\|g_{qv}\theta_z^{*T} W\right\|^2 + \left\|g_{qv}\theta_z^{*T}\varepsilon_z\right\|^2\right) \\ &+ \left\|\left(\rho_2 - \rho_1\right)\tanh(\hat{H})\right\|^2 + \left\|\left(\rho_2 - \rho_1\right)\tanh(H^*)\right\|^2 - \Upsilon\left(u^*\right),\end{aligned} \tag{46}$$

where $\hat{H} = \frac{1}{\rho_2 - \rho_1} R^{-1} g_{qv}^T(\theta_z^T \hat{W})$, $H^* = \frac{1}{\rho_2 - \rho_1} R^{-1} g_{qv}^T(\theta_z^T W + \varepsilon_z)$. Note $\Upsilon(u^*) \geq 0$, we have

$$\bar{J}_{31} \leq \lambda_{\min}\left(Q\right)\|z\|^2 + \gamma_1, \tag{47}$$

where $\gamma_1 = \dfrac{1}{2} g_{qv\max}^2 (s_{\nabla\theta}^2 s_w^2 + s_{\nabla\theta}^2 s_\varepsilon^2) + 2(\rho_2 - \rho_1)^2$, $\|W\| \leq s_w$, $s_w > 0$ is a constant. Taking derivative J_{32} along 39, one obtains

$$\dot{J}_{32} = \int_{t-T}^{t}\left(-\alpha\tilde{W}^T\psi\tilde{W} + \alpha\tilde{W}^T\left(\Delta\theta\varepsilon_H + \sum\nolimits_{i=1}^{P}\Delta\theta^T\left(z\left(t_i\right)\right)\varepsilon_H(t_i)\right)\right)d\tau, \tag{48}$$

where $\psi = \Delta\theta\Delta\theta^T + \sum_{i=1}^{P}\Delta\theta\left(z\left(t_i\right)\right)\Delta\theta^T\left(z\left(t_i\right)\right) > 0$. Using Young's inequality 48, we have

$$\begin{aligned}\dot{J}_{32} &\leq \int_{t-T}^{t}\left(-\alpha\lambda_{min}(\psi)\left\|\tilde{W}\right\|^2 + \alpha\tilde{W}^T\Delta\theta\varepsilon_H + \alpha\tilde{W}^T\sum\nolimits_{i=1}^{P}\Delta\theta^T\left(z\left(t_i\right)\right)\varepsilon_H(t_i)\right)d\tau \\ &\leq \int_{t-T}^{t}\left(-\alpha\lambda_{min}(\psi)\left\|\tilde{W}\right\|^2 + \left\|\tilde{W}\right\|^2\psi + \frac{{\alpha_1}^2}{4}(P+1)s_{\varepsilon H}^2\right)d\tau \\ &\leq -(\alpha-1)\lambda_{min}(\psi)\int_{t-T}^{t}\left\|\tilde{W}\right\|^2 d\tau + \frac{{\alpha_1}^2}{4}(P+1)\int_{t-T}^{t}s_{\varepsilon H}^2 d\tau.\end{aligned} \tag{49}$$

From 47 and 49, we have

$$\dot{J}_3 \leq \int_{t-T}^{t}\left(-\lambda_{\min}\left(Q\right)\|z\|^2 - \beta_1\left\|\tilde{W}\right\|^2 + \beta_2\right)d\tau, \tag{50}$$

where $\beta_1 = (\alpha-1)\lambda_{min}(\psi)$, $\beta_2 = \gamma_1 + \frac{\alpha^2}{4}(P+1)s_{\varepsilon H}^2$, $\alpha > 1$.

$\dot{J}_3 < 0$, if and only if

$$\begin{cases} \|z\| > \sqrt{\frac{\beta_2}{\lambda_{\min}(Q)}} = s_z \\ \left\|\tilde{W}\right\| > \sqrt{\frac{\beta_2}{\beta_1}} = s_{\tilde{W}}. \end{cases} \tag{51}$$

From 51, it can be seen that if $\|z\|$ or $\left\|\tilde{W}\right\|$ surpasses a certain bound s_z or $s_{\tilde{W}}$, then $\dot{J}_3 < 0$ which means that $\|z\|$ or $\left\|\tilde{W}\right\|$ will be pulled into the bound s_z or $s_{\tilde{W}}$. In other words, according to the extended Lyapunov technique [39], the tracking error and the approximate error of NN are UUB stable (see Definition 1).

From 31, 33, 34, and 35, we have $\left\|V^*(z) - \hat{V}(z)\right\| \leq s_{\tilde{\mathrm{W}}} s_\theta + s_\varepsilon = s_v$, and $\|u^* - \hat{u}\| \leq \lambda_{min}(R)\, g_{qv\max}\left(s_{\tilde{\mathrm{W}}} s_{\nabla\theta} + s_{\nabla\varepsilon}\right) = s_u$, with $s_v \geq 0$, $s_u \geq 0$. It can be seen that $\hat{V}(z)$, $\hat{u}$ converge to the near-optimal values. The desired convergence quality can be achieved by choosing the appropriate α factor.

The proof is completed. ■

5. SIMULATIONS

To verify the performance of the proposed algorithm, in this section, we perform simulations proposed algorithm for a typical robot manipulator with two degrees of freedom and compare it with the algorithm presented in [26].

Consider a two-link robot manipulator as [40], with the inertia, Coriolis-centripetal, and dynamic friction matrices are

$$\begin{aligned} &M = \begin{bmatrix} p_1 + 2p_3c_2 & p_2 + p_3c_2 \\ p_2 + p_3c_2 & p_2 \end{bmatrix}, \ C = \begin{bmatrix} -p_3s_2\dot{q}_2 & -p_3s_2(\dot{q}_1 + \dot{q}_2) \\ p_3s_2\dot{q}_1 & 0 \end{bmatrix}, \\ &G(q) = [8.45\tanh(\dot{q}_1), 2.35\tanh(\dot{q}_2)]^T, \end{aligned} \tag{52}$$

where $c_2 = \cos(q_2)$, $s_2 = \sin(q_2)$, $p_1 = 3.473kgm^2$, $p_2 = 0.196kgm^2$, and $p_3 = 0.242kgm^2$. The control inputs $\tau = [\tau_1, \tau_2]^T$ are limited asymmetry, defined as $\tau_{\min} \leq \tau_i \leq \tau_{\max}$, where $i = 1, 2$, $\tau_{\min} = -5N.m$, and $\tau_{\max} = 6N.m$. The desired trajectories are given as $q_d = [q_{1d}, q_{2d}]^T$, $q_{1d} = \sin(0.1t)$, and $q_{2d} = \sin(0.1t)$. The initial values are chosen $q(0) = [1.0, -1.5]^T$, $\dot{q}(0) = [0, 0]^T$.

We choose the proposed controller parameters as follows $\Lambda_1 = \Lambda_2 = \Lambda_3 = \Lambda_4 = \mathrm{diag}[5, 3]$, the activation function is determined as

$$\phi(e) = [e_{q1}^2, e_{q1}e_{q2}, e_{q1}e_{v1}, e_{q1}e_{v2}, e_{q2}^2, e_{q2}e_{v1}, e_{q2}e_{v2}, e_{v1}^2, e_{v1}e_{v2}, e_{v2}^2],$$

$Q = I \in R^{4\times4}$, $R = 1$, $\alpha = 100$, $\rho_2 = 0.6$, $\rho_1 = -0.5$, $\lambda_2 = 6$, $\lambda_1 = -5$. The initial values for the weights of NN are 0. The external disturbance τ_0 is selected as $\tau_0 = 0.5[\sin(t), \cos(2t)]^T$. The small noise is added to

$$\begin{aligned} \varsigma = 0.5(&\sin^2(t)\cos(t) + \sin^2(2t)\cos(0.1t) + \sin^2(1.2t)\cos(0.5t) + \sin^5(t) + \sin^2(1.12t) \\ &+ \cos(2.4t)\sin^3(2.4t)) \end{aligned}$$

for the first $0.04s$, and the data stack size for CL is $P = 20$.

We perform a simulation with time $t = 100s$, sampling period $T = 0.001s$. We change the load at $t = 50s$, i.e., $p_1 = 5.473kgm^2$, $p_2 = 2.196kgm^2$, and $p_3 = 2.242kgm^2$. The

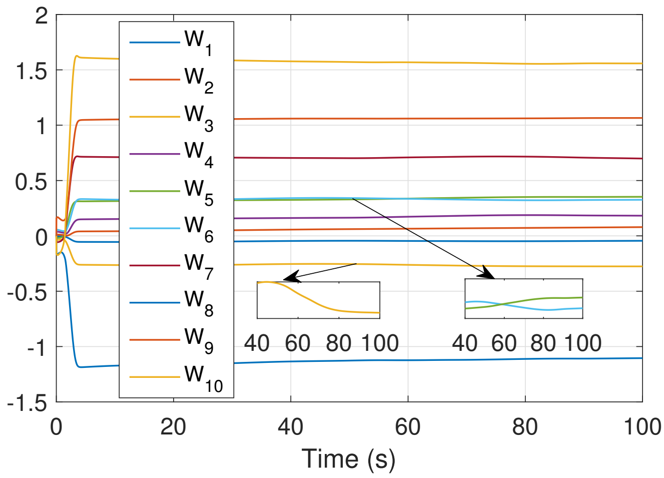

Figure 1: Convergence of NN weights: at the 50th second some weights change as the load changes

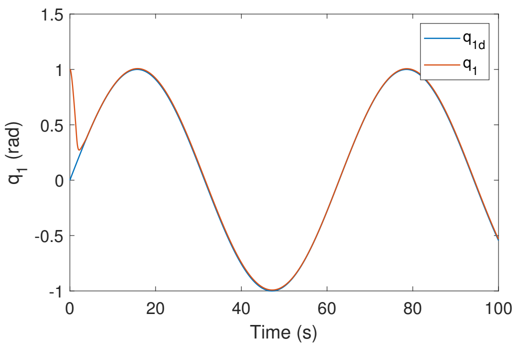

Figure 2: Comparison of the position of the first joint with the reference position

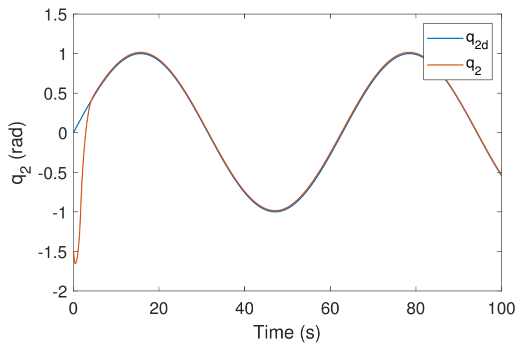

Figure 3: Comparison of the position of the second joint with the reference position

results are shown in Figures 1 to 7. The weights of NN converge to the optimal value $W = [-0.0438, 0.0785, -0.2745, 0.1801, 0.35275, 0.326, 0.6968, -1.103, 1.063, \ 1.554]^T$ after about $5s$, as shown in Figure 1. At the 50th second, when the load changes, the weights W_5, W_6, and W_{10} fluctuate slightly and then continue to converge. The joint positions' tracking quality is shown in Figures 2 and Figure 3. Figure 4 shows that the tracking error is approximately 10^{-3} after learning, and at the time of load change, the tracking error quickly reached approximately 10^{-3}. Figures 2, 3, and 4 show that the proposed algorithm has good tracking quality. The algorithm provides optimal control inputs, as shown in Figure 5, and feedforward input signals, as shown in Figure 6. The result of control torques is shown in Figure 7, and it shows that the values of the control torque are within the allowable asymmetry limit region. At the early stage, as the initial parameters of the controller have not achieved the optimal values, the control signals can exceed the saturation limits. The algorithm is, therefore, responsible for changing the control signals when they approach maximum and minimum limits. When the parameters of the controller converge to the near-optimal values, the algorithm makes the control energies as small as possible, followed by the minimization

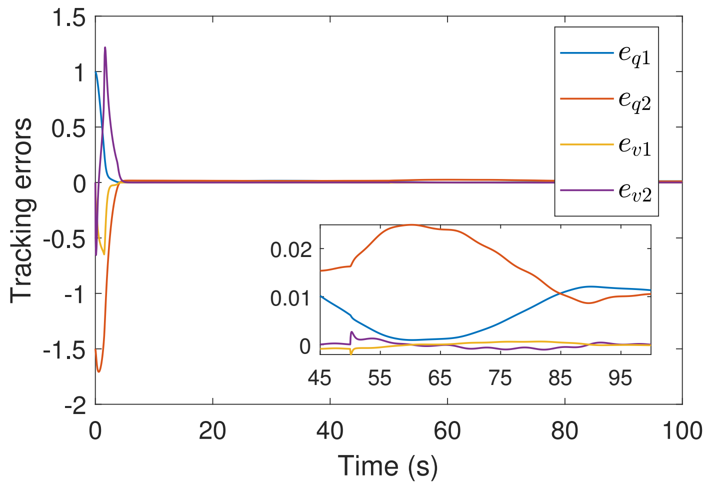

Figure 4: Tracking errors of positions and velocities

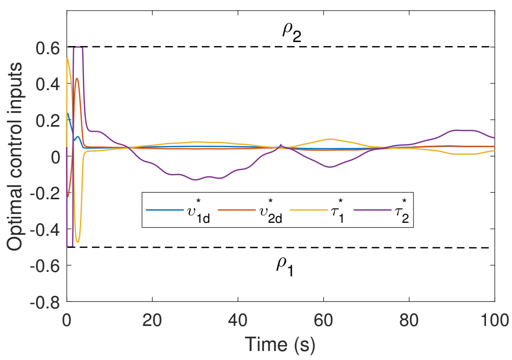

Figure 5: Optimal feedback control inputs

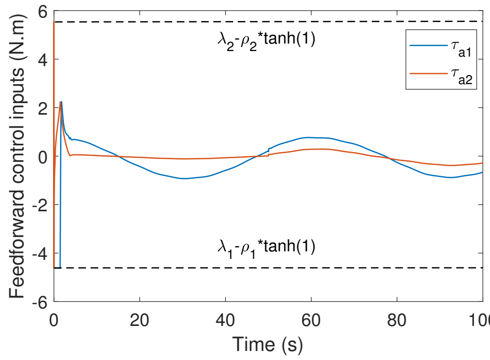

Figure 6: Feedforward control inputs

of the cost function. During this process, the control signals will not exceed the saturation limit. It is worth emphasizing that to balance and maintain the optimal performance between the tracking error and control energy in the cost function 18 at the time of load change, it is only necessary to change some of the weights slightly (Figure 1) such that the control signals in Figure 7 are just large enough to ensure the stability of the system and bring the tracking errors back to small values (Figure 4). This balance can be adjusted by the weight matrices Q and R.

To further illustrate the performance of the asymmetric saturation optimal tracking controller (ASOTC), we simulate a comparison with an optimal tracking controller (OTC) proposed in [26]. Select the parameters for the OTC as follows $Q = I \in \Re^{4\times 4}$, $R = \text{diag}[20, 100]$, $\Lambda = \text{diag}[200, 50]$.

The comparative simulation results are shown in Figures 8, 9, and 10. Figure 8 illustrates the tracking error of OTC (case one is OTC without torque limitation) and the ASOTC. The results show that the tracking error of the ASOTC approaches zero faster than that of the OTC. The OTC provides the control torques, as shown in Figure 9, which shows that

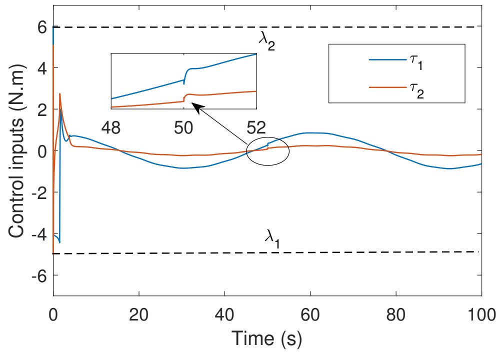

Figure 7: Control torques of ASOTC

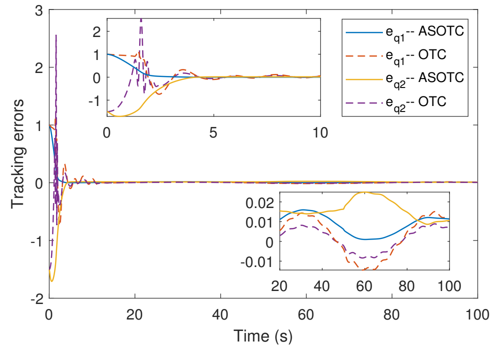

Figure 8: Tracking errors of ASOTC and OTC

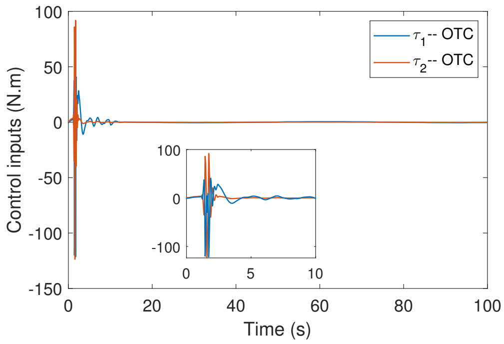

Figure 9: Input torques of OTC without asymmetric constraints

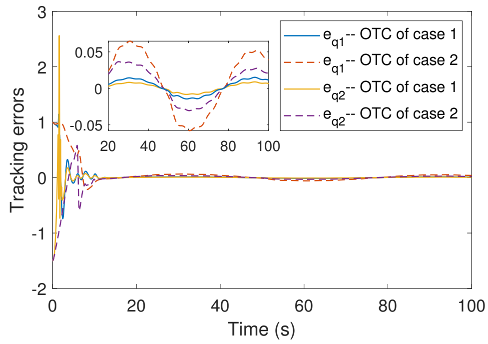

Figure 10: Tracking errors of OTC in two cases

the torque values are out of the allowable limits. Thus, the ASOTC ensures fault tracking with limited torque. The simulation results of the OTC with asymmetric saturation torques (case two), as shown in Figure 10, show that the error tracking performance is significantly reduced compared to case one. Therefore, through the simulation results, the effectiveness of the proposed algorithm is verified.

Remark 3. The restriction of the proposed control method is that the controller is only applied to the system with partially unknown dynamics. Therefore, the next direction will be to develop an algorithm with a completely unknown system.

6. CONCLUSION

This paper proposed an optimal tracking controller for robot manipulators with asymmetric saturation torques and partially unknown dynamics based on an RL method. The feedforward control inputs were designed to transform the position tracking control problem into the optimal tracking control problem. Then, the cost function with asymmetric saturation input was determined and approximated based on the online RL algorithm using a single NN, and the optimal control rule was built. Lyapunov analysis shows that the proposed controller ensures that the tracking error and the approximate error of NN are UUB stable according to Definition 1, and the cost function converges to a near-optimal value. Moreover, the proposed controller has eliminated using the long-term excitation condition and system identification procedures. In future work, we will focus on applying the proposed algorithm in an experiment with asymmetrically saturating actuators and develop a control algorithm for a completely unknown system.

ACKNOWLEDGMENT

We acknowledge Ho Chi Minh City University of Technology (HCMUT), VNU-HCM for supporting this study.

REFERENCES

[1] Z. L. W. He and C. P. Chen, "A survey of human-centered intelligent robots: issues and challenges," *IEEE/CAA Journal of Automatica Sinica*, vol. 4, no. 4, pp. 602–609, 2017.

[2] C. M. K. De Backer, T. DeStefano and J. R. Suh, "Industrial robotics and the global organisation of production," *OECD Science, Technology and Industry Working Papers*, 2018.

[3] F. L. Lewis, D. M. Dawson, and C. T. Abdallah, *Robot Manipulator Control: Theory and Practice*. CRC Press, 2003.

[4] M. Van and S. S. Ge, "Adaptive fuzzy integral sliding-mode control for robust fault-tolerant control of robot manipulators with disturbance observer," *IEEE Transactions on Fuzzy Systems*, vol. 29, no. 5, pp. 1284–1296, 2020.

[5] T.-N. M. R.-D. Xi, X. Xiao and Z.-X. Yang, "Adaptive sliding mode disturbance observer based robust control for robot manipulators towards assembly assistance," *IEEE Robotics and Automation Letters*, vol. 7, no. 3, pp. 6139–6146, 2022.

[6] M. A. L. S. Zeghloul and J.-P. Gazeau, "Robotics and mechatronics," in *Proceedings of the 4th IFToMM International Symposium on Robotics and Mechatronics*. Springer, 2015.

[7] L. C. C. Yang, X. Wang and H. Ma, "Neural-learning-based telerobot control with guaranteed performance," *IEEE Transactions on Cybernetics*, vol. 47, no. 10, pp. 3148–3159, 2017.

[8] A. K. K. Shojaei and A. Chatraei, "An observer-based neural adaptive pid^2 controller for robot manipulators including motor dynamics with a prescribed performance," *IEEE/ASME Transactions on Mechatronics*, vol. 26, no. 3, pp. 1689–1699, 2021.

[9] A. S. B. M. Yilmaz, E. Tatlicioglu and M. Alci, "Self-adjusting fuzzy logic based control of robot manipulators in task space," *IEEE Transactions on Industrial Electronics*, vol. 69, no. 2, pp. 1620–1629, 2022.

[10] D.-H. Z. T. H. L. Dai, Y. Yu and Y. Xia, "Robust model predictive tracking control for robot manipulators with disturbances," *IEEE Transactions on Industrial Electronics*, vol. 68, no. 5, pp. 4288–4297, 2021.

[11] Y. F. X. X. T. Yang, N. Sun and H. Chen, "New adaptive control methods for n-link robot manipulators with online gravity compensation: Design and experiments," *IEEE Transactions on Industrial Electronics*, vol. 69, no. 1, pp. 539–548, 2021.

[12] Y. D. W. He and C. Sun, “Adaptive neural impedance control of a robotic manipulator with input saturation,” *IEEE Transactions on Systems, Man and Cybernetics: Systems*, vol. 46, no. 3, pp. 334–344, 2015.

[13] P. Hippe, *Windup in Control: Its Effects and Their Prevention.* Springer Science & Business Media, 2006.

[14] S. G. M. Nakamura and N. Kyura, “Torque saturation of a mechatronic servo system,” in *Mechatronic Servo System Control.* Springer, 2004, pp. 97–119.

[15] S. Ling, H. Wang, and P. X. Liu, “Adaptive fuzzy dynamic surface control of flexible-joint robot systems with input saturation,” *IEEE/CAA Journal of Automatica Sinica*, vol. 6, no. 1, pp. 97–107, 2019.

[16] W. H. C. Yang, D. Huang and L. Cheng, “Neural control of robot manipulators with trajectory tracking constraints and input saturation,” *IEEE Transactions on Neural Networks and Learning Systems*, vol. 32, no. 9, pp. 4231–4242, 2021.

[17] C.-L. Hwang and B.-S. Chen, “Adaptive finite-time saturated tracking control for a class of partially known robots,” *IEEE Transactions on Systems, Man and Cybernetics: Systems*, vol. 51, no. 9, pp. 5674–5685, 2021.

[18] 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*, vol. 51, no. 3, pp. 1735–1746, 2021.

[19] Z. Z. J. Ma, S. S. Ge and D. Hu, “Adaptive nn control of a class of nonlinear systems with asymmetric saturation actuators,” *IEEE Transactions on Neural Networks and Learning Systems*, vol. 26, no. 7, pp. 1532–1538, 2015.

[20] L. H. Z. Zhou, G. Tang H. Huang and R. Xu, “Adaptive nonsingular fast terminal sliding mode control for underwater manipulator robotics with asymmetric saturation actuators,” *Control Theory and Technology*, vol. 18, no. 1, pp. 81–91, 2020.

[21] 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.

[22] D. Liu, S. Xue, B. Zhao, B. Luo, and Q. Wei, “Adaptive dynamic programming for control: A survey and recent advances,” *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 51, no. 1, pp. 142–160, 2021.

[23] 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.

[24] T. D. H. Zargarzadeh and S. Jagannathan, “Adaptive neural network-based optimal control of nonlinear continuous-time systems in strict-feedback form,” *International Journal of Adaptive Control and Signal Processing*, vol. 28, no. 3-5, pp. 305–324, 2014.

[25] S. B. R. Kamalapurkar, H. Dinh and W. E. Dixon, “Approximate optimal trajectory tracking for continuous-time nonlinear systems,” *Automatica*, vol. 51, pp. 40–48, 2015.

[26] L. Kong, W. He, C. Yang, and C. Sun, “Robust neurooptimal control for a robot via adaptive dynamic programming,” *IEEE Transactions on Neural Networks and Learning Systems*, vol. 32, no. 6, pp. 2584–2594, 2021.

[27] X. Long, Z. He, and Z. Wang, "Online optimal control of robotic systems with single critic nn-based reinforcement learning," *Complexity*, vol. 2021, 2021.

[28] L. Q. X. Zhao, B. Tao and H. Ding, "Model-based actor-critic learning for optimal tracking control of robots with input saturation," *IEEE Transactions on Industrial Electronics*, vol. 68, no. 6, pp. 5046–5056, 2021.

[29] L. N. Tan, "Distributed optimal control for nonholonomic systems with input constraints and uncertain interconnections," *Nonlinear Dynamics*, vol. 93, no. 2, pp. 801–817, 2018.

[30] ——, "Distributed h optimal tracking control for strict-feedback nonlinear large-scale systems with disturbances and saturating actuators," *IEEE Transactions on Systems, Man and Cybernetics: Systems*, vol. 50, no. 11, pp. 4719–4731, 2018.

[31] X. Bu, "An improvement of single-network adaptive critic design for nonlinear systems with asymmetry constraints," *Journal of the Franklin Institute*, vol. 356, no. 16, pp. 9646–9664, 2019.

[32] R. S. L. Xia, Q. Li and H. Modares, "Optimal synchronization control of heterogeneous asymmetric input-constrained unknown nonlinear mass via reinforcement learning," *IEEE/CAA Journal of Automatica Sinica*, vol. 9, no. 3, pp. 520–532, 2021.

[33] J. J. Craig, *Introduction to Robotics: Mechanics and Control.* Pearson Educacion, 2005.

[34] W. Khalil and E. Dombre, *Modeling Identification and Control of Robots.* CRC Press, 2002.

[35] J. Liu, *Intelligent Control Design and MATLAB Simulation.* Springer Singapore, 2018.

[36] M. F. M. K. G. Vamvoudakis and J. P. Hespanha, "Asymptotically stable adaptive–optimal control algorithm with saturating actuators and relaxed persistence of excitation," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 27, no. 11, pp. 2386–2398, 2016.

[37] D. L. S. Xue, B. Luo and Y. Gao, "Event-triggered integral reinforcement learning for nonzero-sum games with asymmetric input saturation," *Neural Networks*, vol. 152, pp. 212–223, 2022.

[38] L. N. Tan and T. C. Pham, "Optimal tracking control for pmsm with partially unknown dynamics, saturation voltages, torque and voltage disturbances," *IEEE Transactions on Industrial Electronics*, vol. 69, no. 4, pp. 3481–3491, 2021.

[39] F. L. Lewis, K. Liu, and A. Yesildirek, "Neural net robot controller with guaranteed tracking performance," *IEEE Transactions on Neural Networks*, vol. 6, no. 3, pp. 703–715, 1995.

[40] F. L. L. H. Modares and Z.-P. Jiang, "H tracking control of completely unknown continuous-time systems via off-policy reinforcement learning," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 26, no. 10, pp. 2550–2562, 2015.

Received on November 09, 2022
Accepted on March 25, 2023