Journal of Control, Automation and Electrical Systems (2021) 32:281–290
https://doi.org/10.1007/s40313-020-00674-w

Sliding Variable-based Online Adaptive Reinforcement Learning of Uncertain/Disturbed Nonlinear Mechanical Systems

Van Tu Vu[1] · Phuong Nam Dao[2] · Pham Thanh Loc[2] · Tran Quang Huy[2]

Received: 12 May 2020 / Revised: 27 August 2020 / Accepted: 21 November 2020 / Published online: 2 January 2021

Abstract

In this work, the trajectory tracking control scheme is the framework of optimal control and robust integral of the sign of the error (RISE); sliding mode control technique for an uncertain/disturbed nonlinear robot manipulator without holonomic constraint force is presented. The sliding variable combining with RISE enables to deal with external disturbance and reduced the order of closed systems. The adaptive reinforcement learning technique is proposed by tuning simultaneously the actor–critic network to approximate the control policy and the cost function, respectively. The convergence of weight as well as tracking control problem was determined by theoretical analysis. Finally, the numerical example is investigated to validate the effectiveness of proposed control scheme.

Keywords Adaptive dynamic programming (ADP) · Robotic systems · Robust integral of the sign of the Error (RISE) · Sliding mode control (SMC)

1 Introduction

The motion of a physical systems group such as robotic manipulators, ship, surface vessels and quad-rotor can be considered as mechanical systems with dynamic uncertainties and external disturbances (Dupree et al. 2011). Furthermore, the actuator saturation and full-state constraint and finite time control have been mentioned in Hu et al. (2019), Yang and Yang (2011), Guo et al. (2019), He et al. (2015), He and Dong (2017), He et al. (2015). Dealing with unknown parameters and disturbances, the terminal sliding mode control (SMC) is one of the remarkable solutions with the consideration of finite time convergence. In Mondal and Mahanta (2014), the non-singular terminal sliding surface was employed to obtain the adaptive terminal SMC for a manipulator system. The work in Galicki (2015) was also based on the non-singular terminal sliding manifold to investigate the finite time control, which seems to be effective in counteracting not only uncertain dynamics but also unbounded disturbances. Authors in Madani et al. (2016) have extended terminal sliding mode technique to establish control design for exoskeleton systems to ensure the trajectory of the closed-loop system can be driven onto the sliding surface in finite time. In order to tackle the challenges of external disturbance, the classical robust control design was investigated the input-state stability (ISS) with equivalent attraction region. However, in the situation that the external disturbance can be a combination of finite number of step signals and sinusoidal signals, the closed-loop system in Lu et al. (2019) is asymptotic stability. In Huang et al. (2018), the optimal gain matrices-based disturbance observer, combining with SMC, was presented for under-actuated manipulators. Authors in Wang et al. (2018) considered the frame of generalized proportional integral observer(GPIO) technique and continuous SMC to overcome the matched/mismatched time-varying disturbances guaranteeing a high tracking performance in compliantly actuated robots. SMC technique is not only employed for classical manipulators but also for different types including bilateral teleoperators (BTs) and mobile robotic systems (wheeled mobile robotics, tractor–trailer systems) (Liu et al. 2020; Nguyena et al. 2019; Binh et al. 2019). Several control schemes have been considered for manipulators to handle the input saturation disadvantage by integrating the additional terms into the control structure (Hu et al. 2019; Yang and Yang 2011; Guo et al. 2019; He et al. 2015). In Hu et al. (2019), a new desired trajectory has been proposed due to the actuator saturation. The additional

✉ Phuong Nam Dao
nam.daophuong@hust.edu.vn

[1] Hai Phong University, Haiphong, Vietnam

[2] Hanoi University of Science and Technology, Hanoi, Vietnam

term would be obtained after taking the derivative of initial Lyapunov candidate function along the state trajectory in the presence of actuator saturation (Hu et al. 2019). Furthermore, a new approach was given in Hu et al. (2019) to tackle not only the actuator constraints but also handling external disturbances. The given sliding manifold was realized the Sat function of joint variables. The equivalent SMC scheme was computed, and then, the boundedness of input signal was mentioned. This approach leads to adjust absolutely input bound by choosing appropriate several parameters. The work in He et al. (2015) gives a technique to tackle the actuator saturation using a modified Lyapunov candidate function. Due to the actuator saturation, the Lyapunov function would be integrated more in the quadratic form from the relation between the input signal from controller and the real signal applied to object. The control design was obtained after considering the Lyapunov function derivative along the system trajectory. In order to tackle the drawback of state constraints in manipulator, the framework of barrier Lyapunov function and Moore–Penrose inverse matrix, fuzzy-neural network technique was proposed in Guo et al. (2019), He et al. (2015), Hu et al. (2019). *Furthermore, these techniques are also developed for the situation of output feedback control with the appropriate virtual control input* (He et al. 2020; Yu et al. 2020). *On the other hand, the uncertainties/disturbance terms in control design are approximated by neural network and fuzzy method* (He et al. 2020; Yu et al. 2020). However, these aforementioned classical nonlinear control techniques have several challenges, such as appropriate Lyapunov function and additional terms dynamic (He et al. 2015; He and Dong 2017; He et al. 2015). Optimal control solution has the remarkable way that can solve above constraint problems by considering the constraint-based optimization (Yang et al. 2020; Sun et al. 2017; Vamvoudakis et al. 2014; Yu et al. 2018; Zhu et al. 2016; Lv et al. 2016; Sun et al. 2017; Li et al. 2020) and model predictive control (MPC) is one of the most effective solutions to tackle the these constraint problems for manipulators (Yu et al. 2018). The terminal controller as well as equivalent terminal region has been established for a nominal system of disturb manipulators with finite horizon cost function (Yu et al. 2018). This technique of robust MPC was also considered for wheeled mobile robotics (WMRs) with the consideration of kinematic model after adding more disturbance observer (DO) (Sun et al. 2017). This work has been extended for the inner loop model by backstepping technique (Yang et al. 2020). Thanks to the advantages of the event-triggering mechanism, the computation load of robust MPC has been reduced in control systems for uni-cycle (Sun et al. 2017). The optimal control algorithm has been mentioned in the work of Dupree et al. (2011) after using classical nonlinear control law. However, the online computation technique has not considered yet in Dupree et al. (2011). Furthermore, it is difficult to find the explicit solution of Riccati equation and partial differential HJB (Hamilton–Jacobi–Bellman) equation in general nonlinear systems (Vamvoudakis et al. 2014). The reinforcement learning strategy was established to obtain the controller by Q learning and temporal difference learning and then was developed to a novel stage by the approximate/adaptive dynamic programming (ADP), which has been the appropriate solution in recent years. Thanks to the neural network approximation technique, authors in Vamvoudakis et al. (2014) proposed the novel online ADP algorithm which enables to tune simultaneously both actor and critic terms. The training problem of critic neural network (NN) was determined by modified Levenberg–Marquardt technique to minimize the square residual error. Furthermore, the weights convergence and convergence problem were shown by the weights in actor and critic NN tuning the need of persistence of excitation (PE) condition (Vamvoudakis et al. 2014). Considering the approximate Bellman error, the proposed algorithm in Vamvoudakis et al. (2014) enables to online simultaneously adjusted with unknown drift term. Extending this work, by using the special cost function, a model-free adaptive reinforcement learning has been presented without any information of the system dynamics (Zhu et al. 2016). Furthermore, by integrating the additional identifier, the nonlinear systems were controlled by online adaptive reinforcement learning with completely unknown dynamics (Lv et al. 2016; Bhasin et al. 2013). However, these three above works have not mentioned for robotic systems as well as non-autonomous systems yet (Zhu et al. 2016; Lv et al. 2016; Bhasin et al. 2013). In the work of Li et al. (2020), under the consideration of approximation and discrete time systems, online ADP tracking control was proposed for the dynamic of mobile robots. Inspired by the above works and analysis from traditional nonlinear control technique to optimal control strategy, the work focuses on the frame of online adaptive reinforcement learning for manipulators and nonlinear control with main contribution which are described in the following:

1) In comparison with the previous papers (Dupree et al. 2011; Hu et al. 2019; Yang and Yang 2011; Guo et al. 2019; He et al. 2015; He and Dong 2017; He et al. 2015; Mondal and Mahanta 2014; Lu et al. 2019; Galicki 2015; Madani et al. 2016; Huang et al. 2018; Wang et al. 2018), which were presented classical nonlinear controller in manipulator control systems, an adaptive reinforcement learning (ARL)-based optimal control design is proposed for a uncertain manipulator system in this paper. *Compared with the proposed optimal control in* Dupree et al. (2011) *using Kim–Lewis formula in special case of cost function, ARL-based optimal control design has the advantage in that it is able to deal with general performance index for non-autonomous system with appropriate transform.*

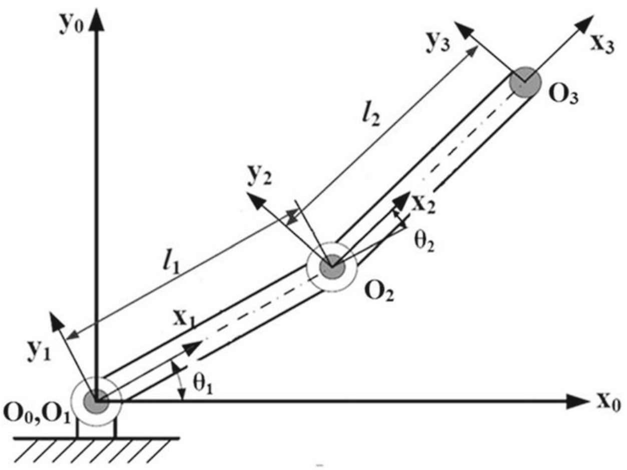

Fig. 1 2-DOF Planar Robot Manipulator

2) Unlike the reinforcement learning scheme-based optimal control in Vamvoudakis et al. (2014), Zhu et al. (2016), Lv et al. (2016), Li et al. (2020), Bhasin et al. (2013) is considered for mathematical systems of a first-order continuous-time nonlinear autonomous system without any external disturbance, the contribution is described that the adaptive dynamic programming combining with the sliding variable and the robust integral of the sign of error (RISE) was employed for second-order uncertain/disturbed manipulators in the situation of trajectory tracking control and non-autonomous systems.

The remainder of this paper is organized as follows. The dynamic model of robotic manipulators and control objective are given in Sect. 2. The proposed adaptive reinforcement learning algorithm and theoretical analysis are presented in Sect. 3. The offline simulation is shown in Sect. 4. Finally, the conclusions are pointed out in Sect. 5.

2 Dynamic Model of Robot Manipulator

Consider the planar robot manipulator systems described by the following dynamic equation:

$$M(\eta)\ddot{\eta} + C(\eta,\dot{\eta})\dot{\eta} + G(\eta) + F(\dot{\eta}) + d(t) = \tau(t) \tag{1}$$

where $M(\eta) \in \mathbb{R}^{n\times n}$ is a generalized inertia matrix, $C(\eta,\dot{\eta}) \in \mathbb{R}^{n\times n}$ is a generalized centripetal-Coriolis matrix, $G(\eta) \in \mathbb{R}^{n}$ is a gravity vector, $F(\dot{\eta}) \in \mathbb{R}^{n}$ is a generalized friction, $d(t)$ is a vector of disturbances, and $\tau(t)$ is the vector of control inputs. *It is worth emphasizing that the above manipulator belongs to the class of Euler–Lagrange systems, which has the following special property* (Guo et al. 2019)*:*

Property 01: The inertia symmetric matrix $M(\eta)$ is positive definite and satisfies $\forall\xi \in \mathbb{R}^{n}$:

$$a\|\xi\|^2 \le \xi^{TM}(\eta)\xi \le b(\eta)\|\xi\|^2 \tag{2}$$

$$\xi^{T}(\dot{M}(\eta) - 2C(\eta,\dot{\eta})\xi = 0 \tag{3}$$

where $a \in \mathbb{R}$ is a positive constant, and $b(\eta) \in \mathbb{R}$ is a positive function with respect to η. Several following assumptions will be employed in considering the stability later (Fig. 1).

Assumption 1 If $\eta(t), \dot{\eta}(t) \in \mathcal{L}_\infty$, then all these functions $C(\eta,\dot{\eta})$, $F(\dot{\eta})$, $G(\eta)$ and the first and second partial derivatives of all functions of $M(\eta)$, $C(\eta,\dot{\eta})$, $G(\eta)$ with respect to $\eta(t)$ as well as of the elements of $C(\eta,\dot{\eta})$, $F(\dot{\eta})$ with respect to $\dot{\eta}(t)$ exist and are bounded.

Assumption 2 The desired trajectory $\eta_d(t)$ as well as the first, second, third and fourth time derivatives of it exists and is bounded.

Assumption 3 The vector of external disturbance term $d(t)$ and the derivatives with respect to time of $d(t)$ are bounded by known constants.

The control objective is to ensure the system tracks a desired time-varying trajectory $\eta_d(t)$ in the presence of dynamic uncertainties by using the frame of online adaptive reinforcement learning-based optimal control design and disturbance attenuation technique. Considering the sliding variable $s(t) = \dot{e}_1 + \lambda_1 e_1$, $\left(\lambda_1 \in \mathbb{R}^{n\times n} > 0, e_1(t) = \eta^{ref} - \eta\right)$, and the corresponding sliding surface is as follows:

$$\mathbf{M} = \left\{e_1(t) \in \mathbb{R}^{n} : s(t) = 0\right\} \tag{4}$$

According to (1), the dynamic equation of the sliding variable $s(t)$ can be given as:

$$M\dot{s} = -Cs - \tau + f + d \tag{5}$$

where $f(\eta, \dot{\eta}, \eta_{ref}, \dot{\eta}_{ref}, \ddot{\eta}_{ref})$ is nonlinear function defined:

$$f = M(\ddot{\eta}^{ref} + \alpha_1\dot{e}_1) + C(\dot{\eta}^{ref} + \alpha_1 e_1) + G + F \tag{6}$$

Remark 1 The role of the above sliding variable is considered to reduce the order of second-order uncertain/disturbed manipulator systems. It enables us to employ the adaptive reinforcement learning for a first-order continuous-time nonlinear autonomous system. Additionally, the external disturbance $d(t)$ and nonlinear function f are handled by RISE in the next section.

3 Adaptive Reinforcement Learning-Based Optimal Control Design

Assume that the dynamic model of robot manipulator is known, the control input can be designed as

$$\boldsymbol{\tau} = \boldsymbol{f} + \boldsymbol{d} - \boldsymbol{u} \tag{7}$$

where the term $\boldsymbol{u}$ is designed by using optimal control algorithm and the remaining term $\boldsymbol{f} + \boldsymbol{d}$ will be estimated later. Therefore, it can be seen that

$$\boldsymbol{M}\dot{\boldsymbol{s}} = -\boldsymbol{C}\boldsymbol{s} + \boldsymbol{u} \tag{8}$$

According to (4) and (8), we obtain the following time-varying model

$$\dot{\boldsymbol{x}} = \begin{bmatrix} -\boldsymbol{\lambda}_1 \boldsymbol{e}_1 + \boldsymbol{s} \\ -\boldsymbol{M}\left(\boldsymbol{\eta}^{ref} - \boldsymbol{e}_1, \dot{\boldsymbol{\eta}}^{ref} + \boldsymbol{\lambda}_1 \boldsymbol{e}_1 - \boldsymbol{s}\right)\boldsymbol{s} \end{bmatrix} + \begin{bmatrix} \boldsymbol{0}_{n\times n} \\ \boldsymbol{M}^{-1} \end{bmatrix} \boldsymbol{u} \tag{9}$$

where $\boldsymbol{x} = [\boldsymbol{e}_1^T, \boldsymbol{s}^T]^T$ and the infinite horizon cost function to be minimized is

$$\boldsymbol{J}(\boldsymbol{x}, \boldsymbol{u}) = \int_0^\infty \left(\frac{1}{2}\boldsymbol{x}^T \boldsymbol{Q}\boldsymbol{x} + \frac{1}{2}\boldsymbol{u}^T \boldsymbol{R}\boldsymbol{u}\right) \boldsymbol{dt} \tag{10}$$

where $\boldsymbol{Q} \in \mathbb{R}^{2n\times 2n}$ and $\boldsymbol{R} \in \mathbb{R}^{n\times n}$ are positive definite symmetric matrices. However, in order to deal with the problem of tracking control, some additional states are given. This work leads us to avoid the non-autonomous systems. Subsequently, the adaptive reinforcement learning is considered to find optimal control solution for autonomous affine state-space model with the assumption that the desired trajectory $\boldsymbol{\eta}^{ref(t)}$ satisfies $\dot{\boldsymbol{\eta}}^{ref}(\boldsymbol{t}) = \boldsymbol{f}^{ref}(\boldsymbol{\eta}^{ref})$:

$$\dot{\boldsymbol{X}} = \boldsymbol{A}(\boldsymbol{X}) + \boldsymbol{B}(\boldsymbol{X})\boldsymbol{u} \tag{11}$$

where $\boldsymbol{X} = [\boldsymbol{x}^T, \boldsymbol{\eta}^{refT}, \dot{\boldsymbol{\eta}}^{refT}]^T$

$$\boldsymbol{A}(\boldsymbol{X}) = \begin{bmatrix} -\boldsymbol{\lambda}_1 \boldsymbol{e}_1 + \boldsymbol{s} \\ \boldsymbol{\Upsilon} \\ \boldsymbol{f}^{ref}(\boldsymbol{\eta}^{ref}) \\ \dot{\boldsymbol{f}}^{ref}(\boldsymbol{\eta}^{ref}) \end{bmatrix},$$

$$\boldsymbol{\Upsilon} = -\boldsymbol{M}(\boldsymbol{\eta}^{ref} - \boldsymbol{e}_1)^{-1}\boldsymbol{C}(\boldsymbol{\eta}^{ref} - \boldsymbol{e}_1, \dot{\boldsymbol{\eta}}^{ref} + \boldsymbol{\lambda}_1 \boldsymbol{e}_1 - \boldsymbol{s})\boldsymbol{s},$$

$$\boldsymbol{B}(\boldsymbol{X}) = \begin{bmatrix} \boldsymbol{0}_{n\times n} \\ \boldsymbol{M}^{-1} \\ \boldsymbol{0}_{2n\times n} \end{bmatrix},$$

Define the new infinite horizon integral cost function to be minimized is

$$\boldsymbol{J}(\boldsymbol{X}, \boldsymbol{u}) = \int_t^\infty \left(\frac{1}{2}\boldsymbol{X}^T \boldsymbol{Q}_T \boldsymbol{X} + \frac{1}{2}\boldsymbol{u}^T \boldsymbol{R}\boldsymbol{u}\right) \boldsymbol{d\tau} \tag{12}$$

where

$$\boldsymbol{Q}_T = \begin{bmatrix} \boldsymbol{Q} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} \end{bmatrix}. \tag{13}$$

In order to guarantee the stability in optimal control design, we can consider the class of "Admissible Policy" described in Vamvoudakis et al. (2014), Zhu et al. (2016):

Definition 1 Vamvoudakis et al. (2014), Zhu et al. (2016), (Admissible Policy): A control input $\boldsymbol{\mu}(\boldsymbol{X})$ is called as admissible in terms of (12) on $\boldsymbol{U}$, if $\boldsymbol{\mu}(\boldsymbol{X})$ is continuous on $\boldsymbol{U}$ and the affine system (11) was stabilized by this control signal $\boldsymbol{\mu}(\boldsymbol{X})$ on $\boldsymbol{U}$ and $\boldsymbol{J}(\boldsymbol{X})$ is finite for any $\boldsymbol{X} \in \boldsymbol{U}$.

The optimal control objective can now be considered finding an admissible control signal $\boldsymbol{\mu}^*(\boldsymbol{X})$ such that the cost function (12) associated with affine system (11) is minimized. According to the classical Hamilton–Jacobi–Bellman (HJB) equation theory (Bhasin et al. 2013), the optimal controller $\boldsymbol{u}^*(\boldsymbol{X})$ and equivalent optimal cost function $\boldsymbol{V}^*(\boldsymbol{X})$ are derived as:

$$\boldsymbol{u}^*(\boldsymbol{X}) = -\frac{1}{2}\boldsymbol{R}^{-1}\boldsymbol{B}^T(\boldsymbol{X})\frac{\partial \boldsymbol{V}^*(\boldsymbol{X})}{\partial \boldsymbol{X}}^T \tag{14}$$

$$\boldsymbol{H}^*\left(\boldsymbol{X}, \boldsymbol{u}^*, \frac{\partial \boldsymbol{V}^*}{\partial \boldsymbol{X}}\right) = \frac{\partial \boldsymbol{V}^*}{\partial \boldsymbol{X}}(\boldsymbol{A}\boldsymbol{X} + \boldsymbol{B}\boldsymbol{u}^*) + \frac{1}{2}\boldsymbol{X}^T \boldsymbol{Q}_T \boldsymbol{X} + \frac{1}{2}\boldsymbol{u}^{*T}\boldsymbol{R}\boldsymbol{u}^* \tag{15}$$

However, it is hard to directly solve the HJB equation as well as offline solution which requires complete knowledge of the mathematical model. Thus, the simultaneous learning-based online solution is considered by using neural networks to represent the optimal cost function and the equivalent optimal controller (Bhasin et al. 2013):

$$\boldsymbol{V}(\boldsymbol{X}) = \boldsymbol{W}^T \boldsymbol{\psi}(\boldsymbol{X}) + \boldsymbol{\epsilon}_v(\boldsymbol{X}), \tag{16}$$

$$\boldsymbol{u}^*(\boldsymbol{X}) = -\frac{1}{2}\boldsymbol{R}^{-1}\boldsymbol{B}^T(\boldsymbol{X})\left(\left(\frac{\partial \boldsymbol{\psi}}{\partial \boldsymbol{x}}\right)^T \boldsymbol{W} + \left(\frac{\partial \boldsymbol{\varepsilon}_v(\boldsymbol{x})}{\partial \boldsymbol{x}}\right)^T\right) \tag{17}$$

where $\boldsymbol{W} \in \mathbb{R}^N$ is vector of unknown ideal NN weights, N is the number of neurons, $\boldsymbol{\psi}(\boldsymbol{X}) \in \mathbb{R}^N$ is a smooth NN activation function, and $\boldsymbol{\epsilon}_v(\boldsymbol{X}) \in \mathbb{R}$ is the function reconstruction error. The objective of establishing the NN (16) is to find the actor–critic NN updating laws $\widehat{\boldsymbol{W}_a}$, $\widehat{\boldsymbol{W}_c}$ to approximate the actor and critic parts obtaining the optimal control law without solving the HJB equation. (For more details, see (Bhasin et al. 2013).) Moreover, the smooth NN activation function $\boldsymbol{\psi}(\boldsymbol{X}) \in \mathbb{R}^N$ is chosen depending on the description of manipulators (see chapter 4). In Bhasin et al. (2013), the Weierstrass approximation theorem enables us to uniformly approximate not only $\boldsymbol{V}^*(\boldsymbol{X})$ but also $\frac{\partial \boldsymbol{V}^*(\boldsymbol{X})}{\partial \boldsymbol{X}}$ with $\boldsymbol{\varepsilon}_{\boldsymbol{v}}(\boldsymbol{x}), \left(\frac{\partial \boldsymbol{\varepsilon}_{\boldsymbol{v}}(\boldsymbol{x})}{\partial \boldsymbol{x}}\right) \rightarrow \boldsymbol{0}$ as $\boldsymbol{N} \rightarrow \infty$. Consider to fix the number $\boldsymbol{N}$, the critic $\hat{\boldsymbol{V}}(\boldsymbol{X})$ and the actor $\hat{\boldsymbol{u}}(\boldsymbol{X})$ are employed to approximate the optimal cost function and the optimal controller as:

$$\hat{\boldsymbol{V}}(\boldsymbol{X}) = \hat{\boldsymbol{W}}_{\boldsymbol{c}}^{\boldsymbol{T}} \psi(\boldsymbol{X}) \tag{18}$$

$$\widehat{\boldsymbol{u}}(\boldsymbol{X}) = -\frac{\boldsymbol{1}}{\boldsymbol{2}} \boldsymbol{R}^{-\boldsymbol{1}} \boldsymbol{B}^{\boldsymbol{T}}(\boldsymbol{X}) \left(\frac{\partial \boldsymbol{\psi}}{\partial \boldsymbol{x}}\right)^{\boldsymbol{T}} \widehat{\boldsymbol{W}}_{\boldsymbol{a}} \tag{19}$$

The adaptation laws of critic $\hat{\boldsymbol{W}}_{\boldsymbol{c}}$ and actor $\hat{\boldsymbol{W}}_{\boldsymbol{a}}$ weights are simultaneously implemented to minimize the integral squared Bellman error and the squared Bellman error δ_{hjb}, respectively.

$$\begin{aligned}\boldsymbol{\delta}_{\boldsymbol{hjb}} &= \hat{\boldsymbol{H}}\left(\boldsymbol{X}, \hat{\boldsymbol{u}}, \frac{\partial \hat{\boldsymbol{V}}}{\partial \boldsymbol{X}}\right) - \boldsymbol{H}^*\left(\boldsymbol{X}, \boldsymbol{u}^*, \frac{\partial \boldsymbol{V}^*}{\partial \boldsymbol{X}}\right) \\ &= \hat{\boldsymbol{W}}_{\boldsymbol{c}}^{\boldsymbol{T}} \boldsymbol{\sigma} + \frac{\boldsymbol{1}}{\boldsymbol{2}} \boldsymbol{X}^{\boldsymbol{T}} \boldsymbol{Q}_{\boldsymbol{T}} \boldsymbol{X} + \frac{\boldsymbol{1}}{\boldsymbol{2}} \hat{\boldsymbol{u}}^{\boldsymbol{T}} \boldsymbol{R} \hat{\boldsymbol{u}}\end{aligned} \tag{20}$$

where $\boldsymbol{\sigma}(\boldsymbol{X}, \hat{\boldsymbol{u}}) = \frac{\partial \boldsymbol{\psi}}{\partial \boldsymbol{x}}(\boldsymbol{A} + \boldsymbol{B}\hat{\boldsymbol{u}})$ is the critic regression vector. Similar to the work in Bhasin et al. (2013), the adaptation law of critic weights is given:

$$\frac{d}{dt} \hat{W}_c = -k_c \lambda \frac{\boldsymbol{\sigma}}{1 + \nu \boldsymbol{\sigma}^T \lambda \boldsymbol{\sigma}} \delta_{hjb} \tag{21}$$

where $\nu, k_c \in \mathbb{R}$ are constant positive gains, and $\boldsymbol{\lambda} \in \mathbb{R}^{\boldsymbol{N} \times \boldsymbol{N}}$ is a symmetric estimated gain matrix computed as follows

$$\frac{\boldsymbol{d}}{\boldsymbol{dt}} \boldsymbol{\lambda} = -\boldsymbol{k}_{\boldsymbol{c}} \boldsymbol{\lambda} \frac{\boldsymbol{\lambda} \boldsymbol{\sigma}^{\boldsymbol{T}}}{\boldsymbol{1} + \boldsymbol{\nu} \boldsymbol{\sigma}^{\boldsymbol{T}} \boldsymbol{\Psi} \boldsymbol{\sigma}} \boldsymbol{\lambda}; \quad \boldsymbol{\lambda}(\boldsymbol{t}_s^+) = \boldsymbol{\lambda}(\boldsymbol{0}) = \boldsymbol{\varphi}_{\boldsymbol{0}} \boldsymbol{I} \tag{22}$$

where t_s^+ is resetting time satisfying $\boldsymbol{\alpha}_{\mathbf{min}}\{\boldsymbol{\lambda}(\boldsymbol{t})\} \leq \boldsymbol{\varphi}_{\boldsymbol{1}}, \boldsymbol{\varphi}_{\boldsymbol{0}} > \boldsymbol{\varphi}_{\boldsymbol{1}}$. It can be seen that ensure $\boldsymbol{\lambda}(\boldsymbol{t})$ is positive definite and prevent the covariance wind-up problem (Bhasin et al. 2013).

$$\boldsymbol{\varphi}_{\boldsymbol{1}} \boldsymbol{I} \leq \boldsymbol{\lambda}(\boldsymbol{t}) \leq \boldsymbol{\varphi}_{\boldsymbol{0}} \boldsymbol{I} \tag{23}$$

Moreover, the actor adaptation law can be described as:

$$\begin{aligned}\frac{\boldsymbol{d}}{\boldsymbol{dt}} \widehat{\boldsymbol{W}_a} = &-\frac{\boldsymbol{k}_{\boldsymbol{a1}}}{\sqrt{\boldsymbol{1} + \boldsymbol{\sigma}^{\boldsymbol{T}} \boldsymbol{\sigma}}} \frac{\partial \boldsymbol{\psi}}{\partial \boldsymbol{x}} \\ &\boldsymbol{B} \boldsymbol{R}^{-1} \boldsymbol{B}^{\boldsymbol{T}} \frac{\partial \boldsymbol{\psi}^{\boldsymbol{T}}}{\partial \boldsymbol{x}} \left(\widehat{\boldsymbol{W}_a} - \widehat{\boldsymbol{W}_c}\right) \boldsymbol{\delta}_{\boldsymbol{hjb}} \\ &- \boldsymbol{k}_{\boldsymbol{a2}} \left(\widehat{\boldsymbol{W}_a} - \widehat{\boldsymbol{W}_c}\right)\end{aligned} \tag{24}$$

Remark 2 The approximate/adaptive reinforcement learning (ARL) control law (actor) and approximately optimal cost function (critic) are obtained in (19), (18), respectively. Based on the optimization principle, the updated law of actor and critic are carried out as in (24), (22). Compared with the optimal control law in Dupree et al. (2011), the ARL control algorithm has the advantage in that it is able to handle for general performance index. The convergences of estimated actor/critic weights $\hat{\boldsymbol{W}}_{\boldsymbol{c}}$ and $\hat{\boldsymbol{W}}_{\boldsymbol{a}}$ depend on the PE condition of $\frac{\boldsymbol{\sigma}}{\sqrt{\boldsymbol{1} + \boldsymbol{\nu} \boldsymbol{\sigma}^{\boldsymbol{T}} \boldsymbol{\lambda} \boldsymbol{\sigma}}} \in \mathbb{R}^{\boldsymbol{N}}$ in Bhasin et al. (2013). Unlike the work in Bhasin et al. (2013), this algorithm does not mention the identifier design and focuses on the manipulator control design. Moreover, the learning technique in adaptation law (22), (24) is different from data-driven online integral reinforcement learning in Vamvoudakis et al. (2014), Zhu et al. (2016). In order to develop this adaptive reinforcement learning for manipulator systems in the trajectory tracking control problem, it is necessary to consider the manipulator dynamic as affine systems (11).

Consequently, the control design (7) is completed by implementing the estimation of $\boldsymbol{\epsilon} = \boldsymbol{f} + \boldsymbol{d}$, which is designed based on the robust integral of the sign of the error (RISE) framework (Dupree et al. 2011) as follows:

$$\epsilon_j(t) = (k_{sj} + 1)s_j(t) - (k_{sj} + 1)s_j(0) + \rho_j(t) \tag{25}$$

where $\rho(t) \in \mathbb{R}^n$ is computed by the following equation:

$$\frac{d}{dt} \rho_j = (k_{sj} + 1)\lambda_{2j} s_j + \gamma_{1j} sgn(s_j) \tag{26}$$

and $\boldsymbol{k}_{\boldsymbol{s}} \in \mathbb{R}^{\boldsymbol{n} \times \boldsymbol{n}}, \boldsymbol{\gamma}_{\boldsymbol{1}} \in \mathbb{R}^{\boldsymbol{n} \times \boldsymbol{n}}, \boldsymbol{\lambda}_{\boldsymbol{2}} \in \mathbb{R}^{n \times n}$ are the positive diagonal matrices and $\zeta_1 \in \mathbb{R}$, $\zeta_2 \in \mathbb{R}$ are the positive control gains selected satisfying the sufficient condition as::

$$\gamma_{1j} > \zeta_1 + \frac{1}{\lambda_{2j}} \zeta_2. \tag{27}$$

Remark 3 In early works (Dupree et al. 2011), the optimal control design was considered for uncertain/disturbed mechanical systems by the RISE framework. The tracking control objective of this optimal control law is satisfied by appropriate assumptions 1-3 (Dupree et al. 2011). However, it should be noted that the work in Dupree et al. (2011) is extended by integrating adaptive reinforcement learning in the trajectory tracking problem with the consideration of non-autonomous systems, which are not directly applied on the adaptive reinforcement learning. It can be seen that the proposed control scheme in Dupree et al. (2011) only considered the optimal control in the special case of cost function. It leads to the optimal control problem which was easily implemented by using the formula of Dupree et al. (2011) for this special case. However, it is worth emphasizing that the method of Kim and Lewis in Dupree et al. (2011) is not able to carry out for general case. Compared with the proposed controller in Dupree et al. (2011), RISE-based uncertainties/disturbance estimation has the advantage in that it is able to combine with adaptive reinforcement learning algorithm for HJB equation to deal with general performance index. Moreover, this work deals with optimal control problem (9) for the general performance index (10) required the appropriate algorithm being adaptive reinforcement learning (ARL) for HJB equation. Additionally, due to the non-autonomous property of model (9), it is not able to directly carry out the model (9) by ARL strategy. Therefore, we proposed the transform method to obtain the modified autonomous system (11) developed by ARL algorithm. On the other hand, it should be noted that authors in Bhasin et al. (2013) considered an online ARL-based method for a first-order continuous-time nonlinear autonomous system without any external disturbance. However, unlike the work in Bhasin et al. (2013), a disturbed manipulator is described by a second-order continuous-time nonlinear systems (1). Therefore, in order to employ ARL strategy, the sliding variable is proposed in this work to reduce the order of manipulator model.

4 Simulation Results

In this section, to verify the effectiveness of the proposed tracking control algorithm, the simulation is carried out by a 2-DOF planar robot manipulator system, which is modeled by Euler–Lagrange formulation (1). In the case of 2-DOF planar robot manipulator systems ($n = 2$), the above matrices in (1) can be represented as follows:

$$\begin{aligned}
\boldsymbol{M}(\boldsymbol{\eta}) &= \begin{bmatrix} \varrho_1 + 2\varrho_2 \cos\eta_2 & \varrho_3 + \varrho_2 \cos\eta_2 \\ \varrho_3 + \varrho_2 \cos\eta_2 & \varrho_3 \end{bmatrix}, \\
\boldsymbol{G}(\boldsymbol{\eta}) &= \begin{bmatrix} \varrho_4 \cos\eta_1 + \varrho_5 \cos(\eta_1 + \eta_2) \\ \varrho_5 \cos(\eta_1 + \eta_2) \end{bmatrix} \\
\boldsymbol{C}(\boldsymbol{\eta}, \dot{\boldsymbol{\eta}}) &= \begin{bmatrix} -\varrho_2 \sin\eta_2 \dot{\eta}_2 & -\varrho_2 \sin\eta_2(\dot{\eta}_1 + \dot{\eta}_2) \\ \varrho_2 \sin\eta_2 \dot{\eta}_1 & 0 \end{bmatrix} \qquad (28)
\end{aligned}$$

where $\varrho_i, i = 1...5$ are constant parameters depending on mechanical parameters and gravitational acceleration. In this simulation, these constant parameters are chosen as $\varrho_1 = 5, \varrho_2 = 1, \varrho_3 = 1, \varrho_4 = 1.2g, \varrho_5 = g$. The two simulation scenarios are considered to validate the performance of proposed controller as follows:

Case 1: The time-varying desired reference signal is defined as $\boldsymbol{\eta}_d = \begin{bmatrix} 3sin(t) & 3cos(t) \end{bmatrix}^T$ where the vector of disturbances is given as $\boldsymbol{d}(t) = \begin{bmatrix} 50sin(t) & 50cos(t) \end{bmatrix}^T$. *For the control objective of general cost function, the optimal control problem is implemented with the arbitrary positive definite symmetric matrices in cost function (10) as:*

$$\boldsymbol{Q} = \begin{bmatrix} 40 & 2 & -4 & 4 \\ 2 & 40 & 4 & -6 \\ -4 & 4 & 4 & 0 \\ 4 & -6 & 0 & 4 \end{bmatrix}, \quad \boldsymbol{R} = \begin{bmatrix} 0.25 & 0 \\ 0 & 0.25 \end{bmatrix}$$

Moreover, due to the stability description of sliding surface, the design parameters in sliding variable $\boldsymbol{s(t)} = \dot{\boldsymbol{e}}_1 + \boldsymbol{\lambda_1 e_1}$ are chosen to satisfy that $\boldsymbol{\lambda_1} \in \mathbb{R}^{\boldsymbol{n \times n}}$ is a constant positive definite matrix:

$$\boldsymbol{\lambda}_1 = \begin{bmatrix} 15.6 & 10.6 \\ 10.6 & 10.4 \end{bmatrix}$$

For the purpose of stability of the closed system as well as uncertainties/disturbance estimation, the remaining control gains in RISE framework are chosen satisfying (25), (26), (27) *as:*

$$\boldsymbol{\lambda}_2 = \begin{bmatrix} 60 & 0 \\ 0 & 35 \end{bmatrix}, \quad \boldsymbol{k}_s = \begin{bmatrix} 140 & 0 \\ 0 & 20 \end{bmatrix}, \quad \gamma_{1j} = 5$$

and the gains in actor–critic learning laws are selected guaranteeing (21)-(24) as:

$$k_c = 800, \quad \nu = 1, \quad k_{a1} = 0.01, \quad k_{a2} = 1,$$

On the other hand, according to Dupree et al. (2011), the consideration of V in (16) can be calculated precisely as

$$\begin{aligned} V &= 2x_1^2 - 4x_1x_2 + 3x_2^2 + 2.5x_3^2 + x_3^2 \cos(\eta_2) + x_3x_4 \\ &\quad + x_3x_4 \cos(\eta_2) + 0.5x_4^2 \qquad (29) \end{aligned}$$

Although we can choose the arbitrary $\boldsymbol{\psi}(\boldsymbol{X})$ in (16), however, for the comparison between result from experiences and result in (29), it leads to that the $\boldsymbol{\psi}(\boldsymbol{X})$ was chosen as

$$\begin{aligned} &\boldsymbol{\psi}(\boldsymbol{X}) \\ &= [x_1^2, x_1x_2, x_2^2, x_3^2, x_3^2 \cos(\eta_2), x_3x_4, x_3x_4 \cos(\eta_2), x_4^2]^T \\ &\qquad (30) \end{aligned}$$

and according to (29), exact values of $\hat{\boldsymbol{W}}_{\boldsymbol{c}}$ in (18) and $\hat{\boldsymbol{W}}_{\boldsymbol{a}}$ in (19) are

$$\begin{aligned} \hat{\boldsymbol{W}}_{\boldsymbol{c}} &= \begin{bmatrix} 2, & -4, & 3, & 2.5, & 1, & 1, & 1, & 0.5 \end{bmatrix} \\ \hat{\boldsymbol{W}}_{\boldsymbol{a}} &= \begin{bmatrix} 2, & -4, & 3, & 2.5, & 1, & 1, & 1, & 0.5 \end{bmatrix} \qquad (31) \end{aligned}$$

In the simulation, the covariance matrix is initialized as

$$\boldsymbol{\Psi}(\boldsymbol{0}) = diag\begin{bmatrix} 100 \; 300 \; 300 \; 1 \; 1 \; 1 \; 1 \; 1 \end{bmatrix}$$

, while all the NN weights $\boldsymbol{W_c}$, $\boldsymbol{W_a}$ are randomly initialized in $[-1, 1]$, and the states and the its first time derivative are initialized to random matrices $\boldsymbol{q}(\boldsymbol{0})$, $\dot{\boldsymbol{q}}(\boldsymbol{0}) \in \mathbb{R}^{\boldsymbol{2}}$. It is necessary to guarantee PE conditions of the critic regression vector (in Remark 1) in using this developed algorithm. Unlike linear systems, where PE conditions of the regression translate to sufficient richness of the external input, there is no verifiable method exists to ensure PE regression translates in nonlinear regulation problems. To deal with this situation, a small exploratory signal consisting of sinusoids of varying frequencies is added to the control signal for first 100 times. Each experiment was performed 150 times, and data from experiments are displayed in Figs. 2, 4, 5 depicting the tracking states and the updating of NN weights $\boldsymbol{W_c}$, $\boldsymbol{W_a}$, respectively. It is clear that the problem of tracking was satisfied after only about 2.5 times through Fig. 2. Meanwhile, the weights of NNs are compared to (31) as Table 1.

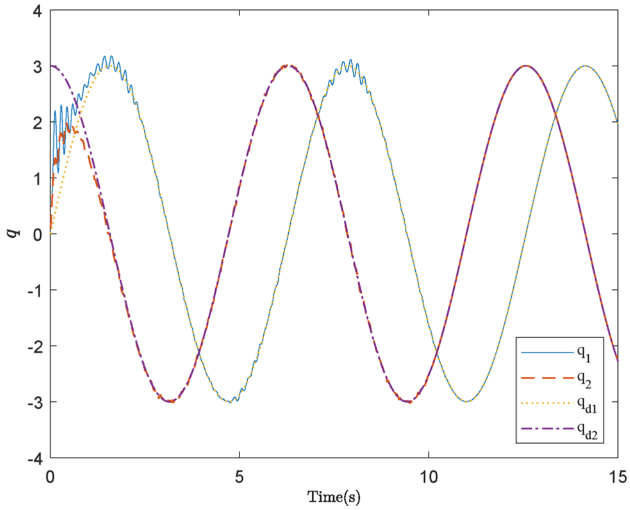

Fig. 2 System states $q(t)$ and its references $q_d(t) = \eta_d$ with persistently excited input for the first 100 times

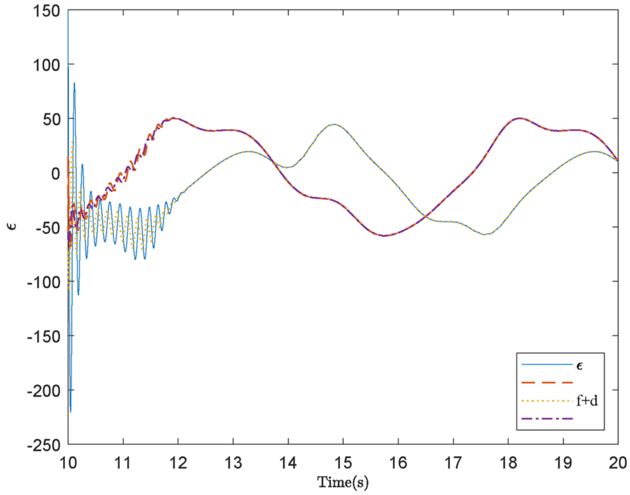

Fig. 3 Estimation of the total of external disturbance and nonlinear function by RISE control input

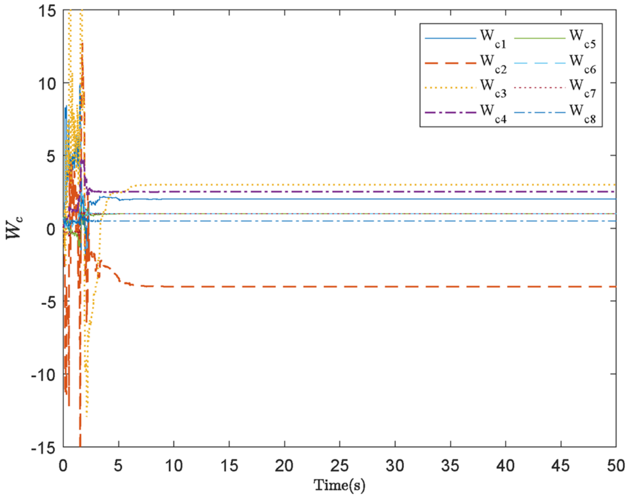

Fig. 4 The weight of NN for critic

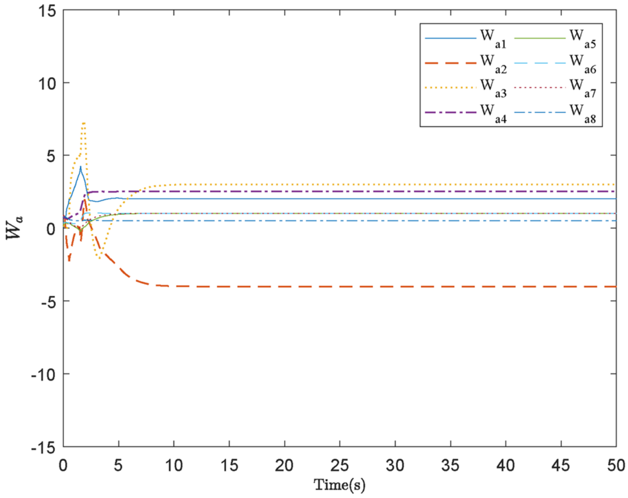

Fig. 5 The weight of NN for actor

Table 1 Comparison between the proposed algorithm and exact values

W	proposed algorithm	exact value
$\boldsymbol{W}_1$	2.02	2.00
$\boldsymbol{W}_2$	– 3.95	– 4.00
$\boldsymbol{W}_3$	2.98	3.00
$\boldsymbol{W}_4$	2.50	2.50
$\boldsymbol{W}_5$	1.00	1.00
$\boldsymbol{W}_6$	1.00	1.00
$\boldsymbol{W}_7$	1.00	1.00
$\boldsymbol{W}_8$	0.50	0.50

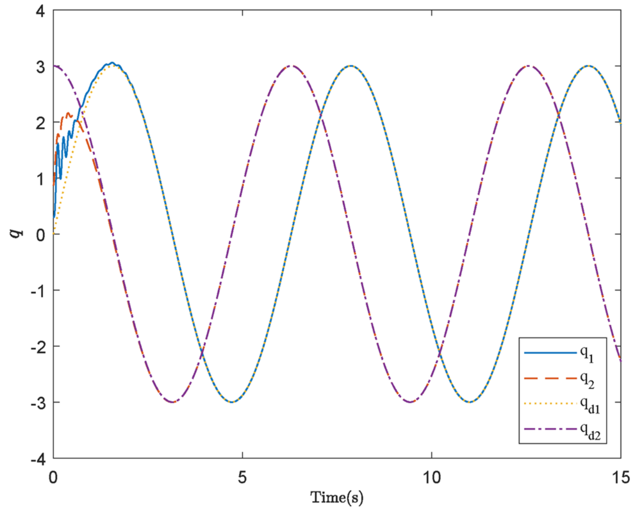

Fig. 6 System states $q(t)$ and its references $q_d(t) = \eta_d$ with persistently excited input for the first 100 times

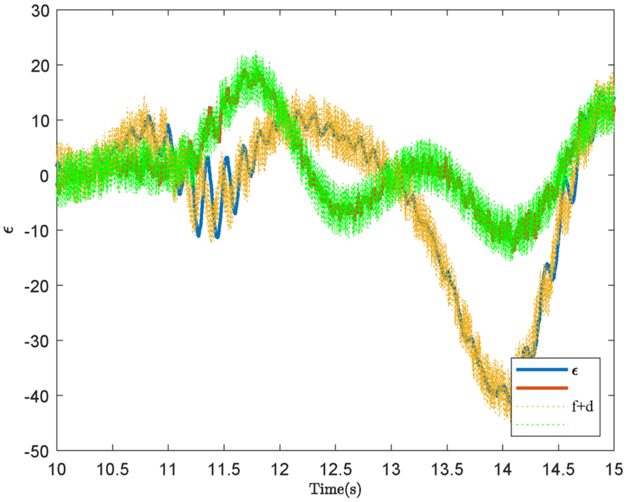

Fig. 7 Estimation of the total of external disturbance and nonlinear function by RISE control input

The highest error which is approximately 0.05 is a acceptable result although the time of convergence is still high. Furthermore, we obtain the tracking performance of the total of external disturbance $\boldsymbol{d}(t)$ and nonlinear function $\boldsymbol{f}(t)$, enabling the disturbance attenuation property of proposed control scheme in Fig. 3. These results proved the correctness of the algorithm.

Case 2: *In this case, we consider for the different case of disturbance. The time-varying desired reference signal is defined as* $\boldsymbol{\eta}_d = \begin{bmatrix} 3sin(t) & 3cos(t) \end{bmatrix}^T$ *where the vector of random disturbances is given as* $\boldsymbol{d}(t) = \begin{bmatrix} 10rand(1) & 10rand(1) \end{bmatrix}^T$. *Based on the simulation method, the parameters are chosen as described in case 1, and our algorithm is effectively verified in tracking problem of the desired reference, weight convergence and disturbance attenuation as shown in Figs.* 6, 7, 8, 9.

Case 3: *In this case, we consider the different case of desired trajectory. The step function desired reference signal is defined as* $\boldsymbol{\eta}_d = \begin{bmatrix} 2 * 1(t) & 3 * 1(t) \end{bmatrix}^T$ *where the disturbance is given as* $\boldsymbol{d}(t) = \begin{bmatrix} 50sin(t) & 50cos(t) \end{bmatrix}^T$. *The parameters in simulation are chosen as mentioned in case 1. It should be noted that our algorithm is effective in tracking the desired reference, weight convergence and disturbance attenuation as described in Figs.* 10, 11, 12, 13.

Remark 4 It is worth noting that the simulation results in Figs. 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13 illustrate the good behavior of trajectory tracking problem, the convergence in actor–critic neural network weights in the presence of dynamic uncertainties, external disturbances. This work is the remarkable extension of the work in Bhasin et al. (2013), which only mentions the first-order mathematical model without any disturbances. Additionally, the optimal control algorithm for manipulators was not considered the

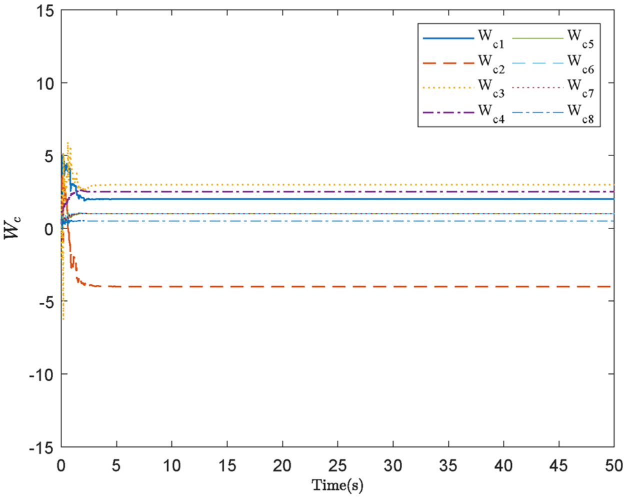

Fig. 8 The weight of NN for critic

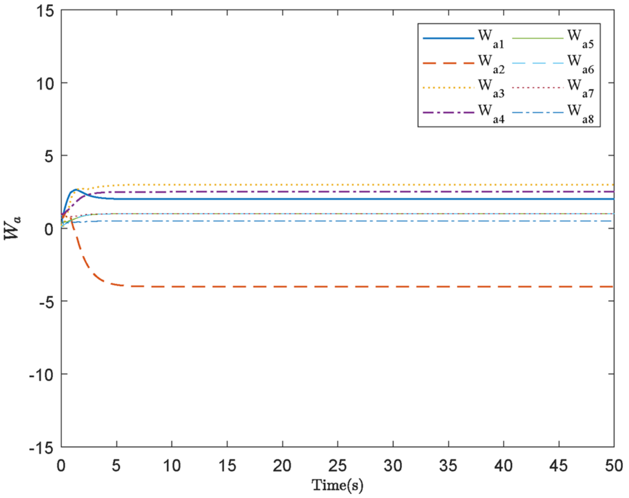

Fig. 9 The weight of NN for actor

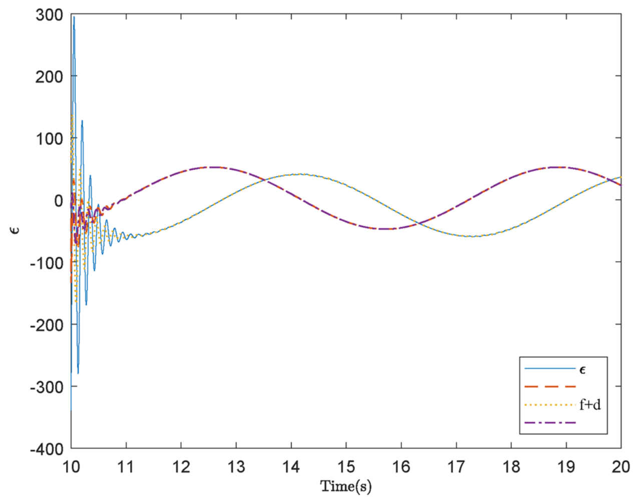

Fig. 10 Estimation of the total of external disturbance and nonlinear function by RISE control input

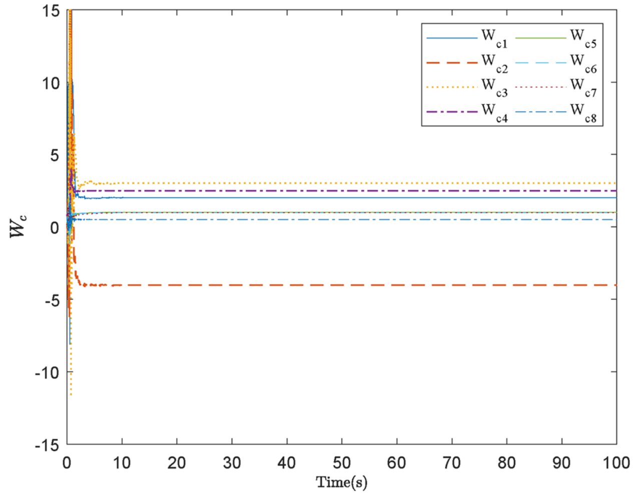

Fig. 11 The weight of NN for critic

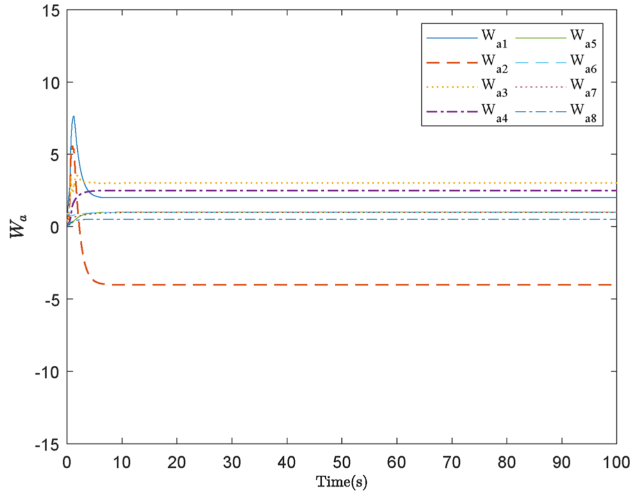

Fig. 12 The weight of NN for actor

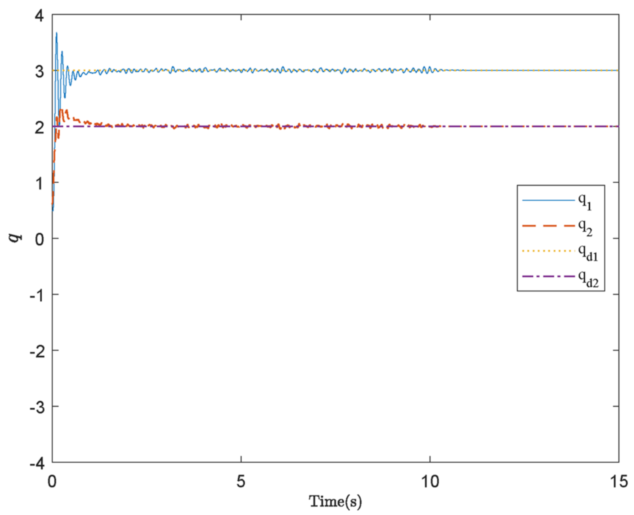

Fig. 13 System states $q(t)$ and its references $q_d(t) = \eta_d$ with persistently excited input for the first 100 times

adaptive dynamic programming technique (Dupree et al. 2011).

5 Conclusions

This paper addresses the problem of adaptive reinforcement learning design for a second-order uncertain/disturbed manipulators in connection with sliding variable and RISE technique. Thanks to the online ADP algorithm based on the neural network, the solution of HJB equation was achieved by iteration algorithm to obtain the controller satisfying not only the weight convergence but also the trajectory tracking problem in the situation of non-autonomous closed systems. Offline simulations were developed to demonstrate the performance and effectiveness of the optimal control for manipulators.

Acknowledgements This research is supported by the Ministry of Education and Training, Vietnam, under Grant B2020-BKA-05.

References

Bhasin, S., Kamalapurkar, R., Johnson, M., Vamvoudakis, K. G., Lewis, F. L., & Dixon, W. E. (2013). A novel actor-critic-identifier architecture for approximate optimal control of uncertain nonlinear systems. *Automatica*, *49*(1), 82–92.

Binh, N. T., Tung, N. A., Nam, D. P., & Quang, N. H. (2019). An adaptive backstepping trajectory tracking control of a tractor trailer wheeled mobile robot. *International Journal of Control, Automation and Systems*, *17*(2), 465–473.

Dupree, K., Patre, P. M., Wilcox, Z. D., & Dixon, W. E. (2011). Asymptotic optimal control of uncertain nonlinear Euler–Lagrange systems. *Automatica*, *1*, 99–107.

Galicki, M. (2015). Finite-time control of robotic manipulators. *Automatica*, *51*, 49–54.

Guo, Y., Huang, B., Li, A., & Wang, C. (2019). Integral sliding mode control for Euler-Lagrange systems with input saturation. *International Journal of Robust and Nonlinear Control*, *29*(4), 1088–1100.

He, W., Xue, C., Yu, X., Li, Z., & Yang, C. (2020). Admittance-Based Controller Design for Physical Human-Robot Interaction in the Constrained Task Space. In *IEEE Transactions on Automation Science and Engineering,Early Access*, (pp 1–13).

He, W., & Dong, Y. (2017). Adaptive fuzzy neural network control for a constrained robot using impedance learning. *IEEE Transactions on Neural Networks and Learning Systems*, *29*(4), 1174–1186.

He, W., Chen, Y., & Yin, Z. (2015). Adaptive neural network control of an uncertain robot with full-state constraints. *IEEE Transactions on Cybernetics*, *46*(3), 620–629.

He, W., Dong, Y., & Sun, C. (2015). Adaptive neural impedance control of a robotic manipulator with input saturation. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, *46*(3), 334–344.

Hu, X., Wei, X., Zhang, H., Han, J., & Liu, X. (2019). Robust adaptive tracking control for a class of mechanical systems with unknown disturbances under actuator saturation. *International Journal of Robust and Nonlinear Control*, *6*(29), 1893–1908.

Huang, J., Ri, S., Fukuda, T., & Wang, Y. (2018). A disturbance observer based sliding mode control for a class of underactuated robotic system with mismatched uncertainties. *IEEE Transactions on Automatic Control*, *64*(6), 2480–2487.

Li, S., Ding, L., Gao, H., Liu, Y.-J., Huang, L., & Deng, Z. (2020). ADP-based online tracking control of partially uncertain time-delayed nonlinear system and application to wheeled mobile robots. *IEEE Transactions on Cybernetics*, *50*(7), 3182–3194.

Liu, Y., Dao, N., & Zhao, K. Y. (2020). On robust control of nonlinearteleoperators under dynamic uncertainties with variable time delays and without relative velocity. *IEEE Transactions on Industrial Informatics*, *16*(2), 1272–1280.

Lu, M., Liu, L., & Feng, G. (2019). Adaptive tracking control of uncertain Euler–Lagrange systems subject to external disturbances. *Automatica*, *104*, 207–219.

Lv, Y., Na, J., Yang, Q., Wu, X., & Guo, Y. (2016). Online adaptive optimal control for continuous-time nonlinear systems with completely unknown dynamics. *International Journal of Control*, *89*(1), 99–112.

Madani, T., Daachi, B., & Djouani, K. (2016). Modular controller design based fast terminal sliding mode for articulated exoskeleton systems. *IEEE Transactions on Control Systems Technology*, *25*(3), 1133–1140.

Mondal, S., & Mahanta, C. (2014). Adaptive second order terminal sliding mode controller for robotic manipulators. *Journal of the Franklin Institute*, *351*(4), 2356–2377.

Nguyena, T., Hoang, T., Pham, M., Dao, N. (2019). A Gaussian wavelet network-based robust adaptive tracking controller for a wheeled mobile robot with unknown wheel slips. *International Journal of Control*, *92*(11), 2681–2692.

Sun, Z., Xia, Y., Dai, L., Liu, K., & Ma, D. (2017). Disturbance rejection MPC for tracking of wheeled mobile robot. *IEEE/ASME Transactions on Mechatronics*, *22*(6), 2576–2587.

Sun, Z., Dai, L., Xia, Y., & Liu, K. (2017). Event-based model predictive tracking control of nonholonomic systems with coupled input constraint and bounded disturbances. *IEEE Transactions on Automatic Control*, *63*(2), 608–615.

Vamvoudakis, K. G., Vrabie, D., & Lewis, F. L. (2014). Online adaptive algorithm for optimal control with integral reinforcement learning. *International Journal of Robust and Nonlinear Control*, *24*(17), 2686–2710.

Wang, H., Pan, Y., Li, S., & Yu, H. (2018). Robust sliding mode control for robots driven by compliant actuators. *IEEE Transactions on Control Systems Technology*, *27*(3), 1259–1266.

Yang, L., & Yang, J. (2011). Nonsingular fast terminal sliding-mode control for nonlinear dynamical systems. *International Journal of Robust and Nonlinear Control*, *21*(16), 1865–1879.

Yang, H., Guo, M., Xia, Y., & Sun, Z. (2020). Dual closed-loop tracking control for wheeled mobile robots via active disturbance rejection control and model predictive control. *International Journal of Robust and Nonlinear Control*, *30*(1), 80–89.

Yu, X., He, W.I, Li, H., & Sun, J. (2020). Adaptive fuzzy full-state and output-feedback control for uncertain robots with output constraint. In *IEEE Transactions on Systems, Man, and Cybernetics: Systems,Early Acess,* (pp 1–14).

Yu, Y, Dai, L., Sun, Z., & Xia, Y. (2018). Robust Nonlinear Model Predictive Control for Robot Manipulators with Disturbances. The 37th Chinese Control Conference (CCC), 3629–3633.

Zhu, Y., Zhao, D., & Li, X. (2016). Using reinforcement learning techniques to solve continuous-time non-linear optimal tracking problem without system dynamics. *IET Control Theory and Applications*, *10*(12), 1339–1347.