Adaptive Control of a Marine Vessel Based on Reinforcement Learning

Zhao Yin[a], Wei He[a] *, Chenguang Yang[b], Changyin Sun[c]

[a] *School of Automation and Electrical Engineering and Key Laboratory of Knowledge Automation for Industrial Processes, Ministry of Education, University of Science and Technology Beijing, Beijing 100083, China.*

[b] *Zienkiewicz Centre for Computational Engineering, Swansea University, SA1 8EN, UK.*

[c] *School of Automation, Southeast University, Nanjing 210096, China.*

Abstract — In this paper, our main goal is to solve optimal control problem by using reinforcement learning (RL) algorithm for marine surface vessel system with known dynamic. And this algorithm is an optimal control algorithm based on policy iteration (PI), and it can obtain the suitable approximations of cost function and the optimized control policy. There are two neural networks (NNs), where critic NN aims to estimate the cost-to-go and actor NN is utilized to design suitable input controller and minimize the tracking error. A novel tuning method is given for critic NN and actor NN. The stability and convergence are proven by Lyapunov's direct method. Finally, the numerical simulations are conducted to demonstrate the feasibility and superiority of presented algorithm.

Index Terms — Reinforcement Learning, Critic Neural Networks, Actor Neural Networks, Lyapunov Method, Marine Vessel.

1 Introduction

Recently, marine vessels have been used in various fields, for example, ocean exploration, marine transportation, etc. [1, 2, 3, 4, 5, 6, 7]. With the continuous development of society, the traditional control methods are unable to satisfy the growth in the marine transportation and the needs for

*E-mail: weihe@ieee.org

This work was supported in part by the National Natural Science Foundation of China under Grants 61522302, 61761130080, U1713209, Grant NA160436 and International Exchanges Grant IE 170247 from the Royal Society, UK, and the Fundamental Research Funds for the China Central Universities of USTB under Grant FRF-BD-17-002A.

modern navigation safety. In order to increase tracking accuracy, there are a lot of studies have been proposed with different control methods of marine surface vessels [8, 9, 10, 11, 2, 12, 13].

For marine surface vessel system, it is a difficult problem to ensure the stability in the brutal environment. Therefore, there have been many researches presented in the last couple of years. For example, an adaptive robust tracking control law with finite-time for a fully actuated marine vessel with unknown interference is proposed in [8]. In [14], a control law for trajectory tracking is proposed for the marine vessels system with state constraints and dynamics uncertainties. The authors present a control method of tracking the desired trajectory for a fully actuated marine vessel in [11]. And a control problem of a variable length crane system is investigated in [15]. In [10, 16], the authors propose the sliding-mode control method for a surface vessels system.

In the mathematical view, the optimal control problem is equal to solve Hamilton-Jacobi-Bellman (HJB) equation. Because of the difficulty of nonlinear nature of the HJB equation, more and more researchers put effort into this field in order to solve this puzzle. More achievements have presented the reasonable methods to cope with the discrete-time HJB equation. In [17, 18], many useful points about this problem have been given.

Reinforcement learning is an approach to deal with the aforesaid problem [19, 18, 20, 21, 22, 23]. For a typical structure of reinforcement learning, there includes two neural networks, and the actor neural network updates its output value based on the value of critic neural network. These two neural networks must execute coordinately, and the ultimate target is to reach the global optimum of cost function. The authors provide an adaptive neural network control by using RL algorithm for a robot manipulator systems with unknown functions and input dead-zone in [24]. In this paper, we propose a surface marine vessel by using reinforcement learning and prove its availability.

In recent years, PI has been discussed in [25, 26, 27, 28, 29, 30, 31]. This method belongs to optimal learning for dealing with optimal control problems. For the linear time-invariant system, it can reduce the problem of Kleinman algorithm to solve the Riccati function problem. It is the same as other reinforcement learning algorithms, PI is applied on critic/actor neural networks which are used to approximate the unknown parameters. In this paper, a method about synchronous policy iteration is investigated and it is inspired by PI [32]. This method is one of the generalized PI proposed in [33].

For the past few years, adaptive neural network has been applied for the nonlinear systems broadly, and it can be learned to approximate solution of any nonlinear equations as long as the hidden layer with enough nodes [34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45]. In [46, 47], authors use NN to approximate the unknown system parameters. Two NNs are utilized to approximate the input deadzone and unknown system dynamics in [48]. In [49], a novel critic neural network controller is presented for nonlinear feedback systems, and the control design is based on the predictor model. An adaptive neural network controller is presented to cope with the problem of system uncertainties [50, 51, 52, 53, 54, 55, 56, 57, 41, 12, 58, 59]. An adaptive NN control method based on radial basis function for nonlinear multiagent systems is investigated in [60]. In [61], the authors employ an adaptive NN method for an underactuated wheeled inverted pendulum model. In [62], a trajectory tracking control for marine vessel with full-state constraints and system unknown is designed. In the controller, an adaptive neural networks are used to compensate the dynamics uncertainties. To sum up, the NN is a more and more important technique and can be applied to many fields.

In this paper, there are several main contributions. i) The critic NN is designed to approach the optimal cost function of the marine vessel system, and we tune the critic NN weights when an adoptable policy is specified. ii) And an extra NN actor neural is proposed, and in standard policy iteration we adjust both NN synchronous in real time. iii) RL is applied to control the position of a three degrees of freedom multiple-input-multiple-output (MIMO) marine vessel system, which has a good control effect.

In what follows, section 2 covers problem formulation that contains system modeling and some necessary lemmas, assumptions and properties. The two neural networks control design and stability analysis are shown in Section 3. Next, the simulation is given to show the feasibility and effectiveness of our controller. At last, Section 5 concludes this paper.

2 Problem Formulation

Some notations are proposed as follows, and we will use some symbols: $\mathbb{R}^+$ denotes a positive real number, $\mathbb{R}^n$ is the n-dimensional Euclidean space, $||\cdot||$ is the norm of Euclidean vector, $|\varpi|$ is the absolute value of a scalar ϖ, $\|\varpi\|$ is the norm of vector ϖ, that is $\|\varpi\| = \sqrt{\varpi^T\varpi}$, and $\|\cdot\|_2$ represents the matrix 2-norm.

2.1 System Modeling

In this paper, the dynamic of a marine surface vessel [1] is described as

$$\begin{aligned} \dot{\eta} &= J(\eta)\upsilon \\ M\dot{\upsilon} + C(\upsilon)\upsilon + D(\upsilon)\upsilon + g(\eta) &= u \end{aligned} \tag{1}$$

where $\eta = [\eta_x, \eta_y, \eta_\psi]^T \in \mathbb{R}^3$ denotes the earth-frame positions and heading, $u \in \mathbb{R}^3$ presents the control input of the systems, $\upsilon = [\upsilon_x, \upsilon_y.\upsilon_\psi]^T \in \mathbb{R}^3$ presents the velocities of vessel in the vessel-frame. $M \in \mathbb{R}^{3\times3}$ is a symmetric positive definite inertia matrix, $C(\upsilon) \in \mathbb{R}^{3\times3}$ denotes centripetal and Coriolis torques, $D(\upsilon) \in \mathbb{R}^{3\times3}$ is the damping matrix, and $g(\eta)$ presents the restoring force, and $J(\eta)$ is the transformation matrix which is defined as

$$J(\eta) = \begin{bmatrix} \cos\eta_\psi & -\sin\eta_\psi & 0 \\ \sin\eta_\psi & \cos\eta_\psi & 0 \\ 0 & 0 & 1 \end{bmatrix} \tag{2}$$

We can let $x_1 = \eta$, $x_2 = \upsilon$, then we are able to get following description of our system:

$$\begin{aligned} \dot{x}_1 &= J(x_1)x_2 \\ \dot{x}_2 &= M^{-1}[u - C(x_2)x_2 - D(x_2)x_2 - g(x_1)] \end{aligned} \tag{3}$$

Then the vessel dynamical system is given by

$$\dot{x}(t) = A(x(t)) + B(x(t))u(x(t)); \;\; x(0) = x_0 \tag{4}$$

where

$$\begin{aligned} \dot{x}(t) &= \begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \end{bmatrix}, \\ A(x(t)) &= \begin{bmatrix} J(x_1)x_2 \\ M^{-1}[-C(x_2)x_2 - D(x_2)x_2 - g(x_1)] \end{bmatrix}, \\ B(x(t)) &= \begin{bmatrix} \mathbf{0}_{3\times3} \\ M^{-1} \end{bmatrix} \end{aligned} \tag{5}$$

with $\mathbf{0}_{3\times3}$ denoting 3×3 zero matrices.

Assumption 1 *[63] According to* (4)*, we can assume* $B(x)$ *is bounded, and matrix* $B(x)$ *has full column rank for all* $x \in \mathbb{R}^n$*, and we need to define* $B^+ = (B^T B)^{-1} B^T$ *is bounded and locally Lipschitz.*

Assumption 2 *[64] Let* $x_d(t)$ *be the bounded desired trajectory, and we can assume that there exists a Lipschitz continuous equation* $f_d(\cdot) \in \mathbb{R}^n$ *with* $f_d(0) = 0$ *such that*

$$\dot{x}_d(t) = f_d(x_d(t)) \tag{6}$$

Then denoting the tracking error as,

$$e = x(t) - x_d(t) \tag{7}$$

From (3), (6) and (7), we can obtain the tracking error dynamics

$$\dot{e}(t) = A(x(t)) + B(x(t))u(x(t)) - f_d(x_d(t)) \tag{8}$$

The input controller u_d corresponding to the desired trajectory x_d is

$$u_d(x_d) = B^+(x_d)\dot{x}_d - A(x_d) \tag{9}$$

Therefore, we need to define a new state $\varpi \in \mathbb{R}^{12}$ as

$$\varpi = [e^T, x_d^T]^T \tag{10}$$

According to (8) and Assumption 1, we can obtain the derivative of (10)

$$\dot{\varpi} = E(\varpi) + F(\varpi)\nu \tag{11}$$

where the functions $E \in \mathbb{R}^{12}, G \in \mathbb{R}^{12\times 3}$, and controller $\nu \in \mathbb{R}^3$, we have

$$E(\varpi) = \begin{bmatrix} A(e+x_d)-f_d(x_d)+B(e+x_d)u_d \\ f_d(x_d) \end{bmatrix}, \tag{12}$$

$$F(\varpi) = \begin{bmatrix} B(e+x_d) \\ 0^{6\times 3} \end{bmatrix}, \ \ \nu = u - u_d \tag{13}$$

Assumption 3 *[18] We can assume that, $A(0) = 0$, $A(x) + B(x)u$ is Lipschitz continuous on a set $\Omega \subseteq \mathbb{R}^6$ which contains the origin, and the dynamics system achieves stability on Ω. That is, there exists a continuous control torque $\nu(t) \in \mathbb{U}$ so that the system is asymptotically stable on Ω. On the other hand, we assume that the system parameters M, $C(\upsilon)$, $D(\upsilon)$, and $g(\eta)$ are all known.*

2.2 The Optical Control and Problem Formulation

In this paper, we define the integral cost function as [63]

$$V(\varpi) = \int_0^\infty r(\varpi(\tau), \nu(\tau))\mathrm{d}\tau \tag{14}$$

and we can define $r(\varpi, u) = Q(\varpi) + \nu^T R\nu$, where $Q(\varpi)$ is positive and chosen as $Q(\varpi) = \varpi^T q\varpi$, and $R \in \mathbb{R}^{3\times 3}$ is a symmetric positive definite matrix, $q \in \mathbb{R}^{12\times 12}$ is a positive semi-definite matrices.

Definition 1 *[18]. We define a control policy $\mu(x)$ as admissible with respect to* (14) *on Ω, and $\mu \in \Psi(\Omega)$, if $\mu(\varpi)$ is continuous on Ω, $\mu(0) = 0$, $\nu(\varpi) = \mu(\varpi)$ stabilizes* (4) *on Ω, and $V(\varpi_0)$ is finite $\forall \varpi_0 \in \Omega$.*

If $V(\varpi(t))$ is smooth, through differentiation, the nonlinear Lyapunov equation is represented with the feedback control policy as

$$0 = r(\varpi, \mu(\varpi)) + (V_\varpi^\mu)^T(E(\varpi) + F(\varpi)\mu(\varpi)), V^\mu(0) = 0 \tag{15}$$

where V_ϖ^μ is the partial derivative of V^μ with respect to ϖ. And $\mu(\varpi)$ denotes the admissible control policy, if $V^\mu(\varpi)$ conforms to (15), then $V^\mu(\varpi)$ presents a Lyapunov function for the marine vessel dynamics system (4).

Based on the optimal policy and the CT Hamiltonian function, we can define

$$H(\varpi, \nu, V_{\varpi}) = r(\varpi, \nu) + (V_{\varpi})^T (E(\varpi) + F(\varpi)\mu) \tag{16}$$

then, the optimal cost-to-go $V^*(\varpi)$ is expressed as

$$V^*(\varpi(t)) = \min_{\mu(t)\in\Omega} \Big(\int_0^{\infty} r(\varpi(\tau), \mu(\tau)) \mathrm{d}\tau \Big) \tag{17}$$

where ϖ_0 is known. Then, we can obtain

$$0 = \min_{\mu(t)\in\Omega} H(\varpi, \mu, V_{\varpi}^*) \tag{18}$$

We can assume that the minimum on the term $\min_{\mu\in\Omega} H(\varpi, \mu, V_{\varpi}^*)$ of (18) exists and is unique, we can design the optimal control as

$$\mu^*(\varpi) = -\frac{1}{2} R^{-1} F^T(\varpi) V_{\varpi}^* \tag{19}$$

Substituting (19) into (15), we can obtain

$$0 = Q(\varpi) + V_{\varpi}^{*T}(\varpi) E(\varpi) - \frac{1}{4} V_{\varpi}^{*T}(\varpi) F(\varpi) R^{-1} F(\varpi)^T V_{\varpi}^*(\varpi) \tag{20}$$

$$V^*(0) = 0 \tag{21}$$

According to (19) and (20), we can solve this problem by the optimum control scheme. However, it is difficult and impossible to find the solution as a result of the nonlinear characteristic of HJB function (20).

Based on [65], the method of PI is adopted to cope with the optimal control problems. Therefore, the method of synchronous PI used in this paper, and the of PI algorithm is designed as follows.

1. choosing admissible initial control $\mu^{(0)}(\varpi)$.

2. given $\mu^{(i)}(\varpi)$, solving for cost function $V^{\mu^{(i)}}$ through

$$0 = r(\varpi, \mu^{(i)}(\varpi)) + (\nabla V^{\mu^{(i)}})^T (E(\varpi) + F(\varpi)\mu^{(i)}(\varpi))$$

$$V^{\mu^{(i)}}(0) = 0 \tag{22}$$

3. updating the control policy

$$\mu^{(i+1)} = \arg\min_{\nu} H(\varpi, \nu, \nabla V_{\varpi}^{(i)}) \tag{23}$$

which are also represented as

$$\mu^{(i+1)}(\varpi) = -\frac{1}{2} R^{-1} F^T(\varpi) \nabla V_{\varpi}^{(i)} \tag{24}$$

The convergence of PI algorithm have been proven in [66].

PI is a based on Newton iteration method. For the case of linear time-invariant, it can reduce the problem of Kleinman algorithm [67] to solve the Riccati equation. Then, (22) becomes a Lyapunov function.

2.3 Neural Network

As a reinforcement learning algorithm, the PI is able to be applied in a critic/actor structure, and this structure contains two NNs to approximate the solutions of (22) and (23).

For this structure, the cost function $V^{\mu^{(i)}}$ and the controller $\mu^{(i+1)}(\varpi)$ are approximated by NNs at every step of the PI process. These NNs are designed as the critic NN and the actor NN, respectively. The critic NN aims to solve (22) and the actor NN is tuned to deal with (24).

Assumption 4 *[18] Eq.* (15) *is positive definite. It is guaranteed by the condition that* $Q(\varpi) > 0$, $\varpi \in \Omega_0$; $Q(0) = 0$ *is positive definite.*

Assumption 5 *[18] Eq.* (15) *is smooth and* $V(\varpi) \in \mathcal{V}^1(\Omega)$.

From Assumption 5, we can obtain that there exists a basis set $\{\varphi_i(\varpi)\}$ so that the solution $V(\varpi)$ to (17) and its gradient are estimated. In other word, the coefficients v_i are defined as follows:

$$V(\varpi) = \sum_{i=1}^{\infty} v_i \varphi_i(\varpi) = \sum_{i=1}^{N} v_i \varphi_i(\varpi) + \sum_{i=N+1}^{\infty} v_i \varphi_i(\varpi) \equiv \mathcal{V}^T \phi(\varpi) + \sum_{N+1}^{\infty} v_i \varphi_i(\varpi) \tag{25}$$

$$\frac{\partial V(\varpi)}{\partial \varpi} = \sum_{i=1}^{\infty} v_i \frac{\partial \varphi_i(\varpi)}{\partial \varpi} = \sum_{i=1}^{N} v_i \frac{\partial \varphi_i(\varpi)}{\partial \varpi} + \sum_{i=N+1}^{\infty} v_i \frac{\partial \varphi_i(\varpi)}{\partial \varpi} \tag{26}$$

where $\phi(\varpi) = [\varphi_1(\varpi), \varphi_2(\varpi), \cdots, \varphi_N(\varpi)]^T$, and the last term in (25) and (26) approach to zero as $N \to \infty$.

Thus, there exists weights W and the value $V(\varpi)$ is approximated as

$$V(\varpi) = W_c^T \phi(\varpi) + \epsilon(\varpi) \tag{27}$$

where $\phi(\varpi) \subset \mathbb{R}^N$ is the NN input vector, N states the number of neurons in the hidden layer, $W_c \in \mathbb{R}^N$ is the weight vector of NN, and $\epsilon(\varpi)$ denotes the error of NN. For the NN input functions, $\{\varphi_i(\varpi) : i = 1, 2, ..., N\}$ are selected such that $\{\varphi_i(\varpi) : i = 1, 2, ..., \infty\}$ provides a complete independent set. Therefore, the derivative of $V(\varpi)$ is represented as

$$\frac{\partial V}{\partial \varpi} = \Big(\frac{\partial \phi(\varpi)}{\partial \varpi}\Big)^T W_c + \frac{\partial \epsilon}{\partial \varpi} = \nabla \phi^T W_c + \nabla \epsilon \tag{28}$$

Then, as $N \to \infty$, the approximation errors $\epsilon \to 0$, $\nabla\epsilon \to 0$. Additionally, for fixed N, ϵ and $\nabla\epsilon$ are bounded.

Taking the fixed controller $\nu(t)$ into consideration, the nonlinear Lyapunov function (16) can be expressed as

$$H(\varpi, \nu, W_c) \;\; = \;\; Q(\varpi) + \nu^T R \nu + (\nabla\epsilon + W_c^T \nabla\phi)(E(\varpi) + F(\varpi)u) \tag{29}$$

According to Assumption 5 and (15), the nonlinear Lyapunov equation can be represented as

$$W_c^T \nabla\phi(E(\varpi) + F(\varpi)\nu) + Q(\varpi) + \nu^T R \nu = \epsilon_H \tag{30}$$

where the residual error ϵ_H is

$$\begin{aligned} \epsilon_H &= -(\nabla\epsilon)^T (E(\varpi) + F(\varpi)\nu) \\ &= -(\mathcal{V} - W_c)^T \nabla\phi(E(\varpi) + F(\varpi)\nu) - \sum_{i=N+1}^{\infty} v_i \nabla\varphi_i(\varpi)(E(\varpi) + F(\varpi)\nu) \end{aligned} \tag{31}$$

In the basis of the Lipschitz assumption, we can assume that there exits a bounded for the residual error ϵ_H.

Lemma 1 *[18] There is an unique least-squares solution for* (29) *with the control policy* $\mu(t)$*. we can define if denoting this solution as* W_c

$$V_0(\varpi) = W_c^T \phi(\varpi) \tag{32}$$

Then, as $N \to \infty$ *:*

$$a. \sup_{\varpi \in \Omega} |\epsilon_H| \to 0$$

$$b. \|W_c - \mathcal{V}\|_2 \to 0$$

$$c. \sup_{\varpi \in \Omega} |V_0 - V| \to 0$$

$$d. \sup_{\varpi \in \Omega} \|\nabla V_0 - \nabla V\| \to 0$$

According to the results, $V_0(\varpi)$ achieves convergence to the solution $V(\varpi)$ as $N \to \infty$, and the weights converge to the first N of weights, $\mathcal{V}$ is solved by (15). For HJB approximation error, we can substitute (27) to (20)

$$Q(\varpi) + (W_c^T \nabla\phi + \nabla\epsilon^T)E - \frac{1}{4}(W_c^T \nabla\phi + \nabla\epsilon^T)FR^{-1}F^T(\nabla\phi^T W_c + \nabla\epsilon) = 0 \tag{33}$$

then, we can let

$$\epsilon_{HJB} = -\nabla\epsilon^T E + \frac{1}{2}W_c^T \phi F R^{-1} F^T \nabla\epsilon + \frac{1}{4}\nabla\epsilon^T F R^{-1} F^T \nabla\epsilon \tag{34}$$

therefore, we have

$$Q(\varpi) + W_c^T \nabla\phi E - \frac{1}{4}W_c^T \nabla\phi F R^{-1} F^T \nabla\phi^T W_c = \epsilon_{HJB} \tag{35}$$

where $\forall\varepsilon > 0, \exists N(\epsilon) : \sup_{\varpi\in\Omega} \|\epsilon_{HJB}\| < \epsilon$.

3 Control Design

In this part, we design the critic NN and actor NN to optimize the cost function and obtain the optimal controller. Fig. 1 provides the block diagram of learning control process.

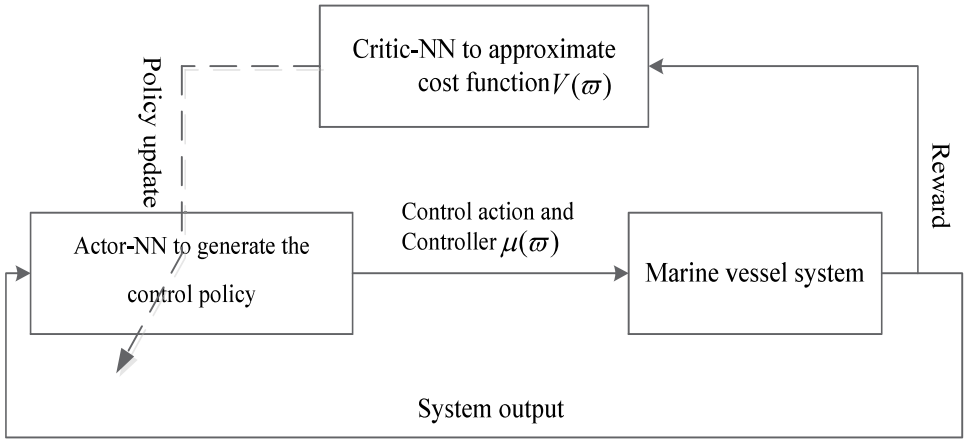

Figure 1: The block diagram of Reinforcement Learning structure.

3.1 Critic Neural Network and Adaptive Optimal Control

In this section, we focus on design of critic NN. The process investigates the adaptability and convergence of the critic NN weights.

The cost function is designed as an observer for the value function. Hence, we design the critic NN controller as

$$\hat{V}(\varpi) = \hat{W}_c^T \phi_c(\varpi) \tag{36}$$

where $\hat{W}_c$ expresses the estimated value of the desired critic NN weight vector W_c. Then, the nonlinear Lyapunov function is approximated as

$$H(\varpi, \hat{W}_c, \nu) = Q(\varpi) + \nu^T R \nu + \hat{W}_c^T \phi_c(E + F\nu(t)) \tag{37}$$

In this paper, the critic weight approximation error is defined as

$$\tilde{W}_c = W_c - \hat{W}_c \tag{38}$$

Then, we can let $e_c = H(\varpi, \hat{W}_c, \nu)$, and we have

$$e_c = -\tilde{W}_c^T \phi_c(E + F\nu(t)) + \epsilon_H \tag{39}$$

The system aims to minimize the squared error by choosing proper $\hat{W}_c$, and we can define Lyapunov function as

$$V_c = \frac{1}{2} e_c^T e_c \tag{40}$$

The critic NN updating law is designed as

$$\dot{\hat{W}}_c = -\Gamma_c \frac{\partial V_c}{\partial \hat{W}_c} = -\Gamma_c \frac{\alpha_c}{(\alpha_c^T \alpha_c + 1)^2} [\alpha_c^T \hat{W}_c + Q(\varpi) + \nu^T R \nu]$$

where $\alpha_c = \nabla \phi_c(E + F\nu(t))$

According to (31), we can obtain

$$Q(\varpi) + \nu^T R \nu = -W_c^T \phi_c(E + F\nu(t)) + \epsilon_H \tag{41}$$

Substituting (41) into updating law, we have

$$\dot{\tilde{W}}_c = -\Gamma_c \bar{\alpha}_c \bar{\alpha}_c^T \tilde{W}_c + \Gamma_c \frac{\bar{\alpha}_c}{\vartheta_c} \epsilon_H \tag{42}$$

where $\bar{\alpha}_c = \alpha_c / (\alpha_c^T \alpha_c + 1)$ and $\vartheta_c = \alpha_c^T \alpha_c + 1$.

The next assumption and lemmas are proposed to guarantee the optimization of weight of error $\tilde{W}_c$.

Assumption 6 *[68] Persistence of excitation (PE) assumption: with the persistently exciting $\bar{\alpha}_c$ belonging to $[t, t+T]$, there exist constants $\xi_1 > 0, \xi_2 > 0, T > 0$ satisfying that for t,*

$$\xi_1 I \leq S_0 = \int_t^{t+T} \bar{\alpha}_c(\tau) \bar{\alpha}_c^T(\tau) \mathrm{d}\tau \leq \xi_2 I \tag{43}$$

It is necessary to present the PE assumption in adaptive controller because one effectively desires to verify the critic arguments to approximate $V(x)$.

Lemma 2 *[69] We can define the error dynamics system as*

$$\dot{\tilde{W}}_c = -\Gamma_c \bar{\alpha}_c \bar{\alpha}_c^T \tilde{W}_c + \Gamma_c \bar{\alpha}_c \frac{\epsilon_H}{\vartheta_c} \tag{44}$$

$$y = \bar{\alpha}_c^T \tilde{W}_c \tag{45}$$

The PE condition (43) *is equal to the uniform complete observability (UCO) and there exist constants* $\xi_3 > 0, \xi_4 > 0, T > 0$ *satisfying that for all* t *[70],*

$$\xi_3 I \leq S_1 = \int_t^{t+T} \Phi^T(\tau, t) \bar{\alpha}_c(\tau) \bar{\alpha}_c^T(\tau) \Phi(\tau, t) \mathrm{d}\tau \leq \xi_4 I \tag{46}$$

where $\Phi(t_1, t_0), t_0 \leq t_1$ *is the state transition matrix of* (44)*, and* I *is an identity matrix.*

Proof: Consider the system $\dot{\tilde{W}}_c = \Gamma_c \bar{\alpha}_c u, y = \bar{\alpha}_c^T \tilde{W}_c$ based on the output feedback $u = -y + \epsilon_H / \vartheta_c$, it is equivalent to the error dynamics system (44). For the error system, (43) is the observability gramian. ■

UCO shows that if input and output are bounded, the state $\tilde{W}_c$ for error system is bounded.

Lemma 3 *[18] Take the error dynamics system* (44) *into consideration.* $\bar{\alpha}_c$ *is designed to be persistently exciting. Then, we have*

(a) The dynamic equation (44) *achieves exponentially stable. Actually, if* $\epsilon_H = 0$*, we can obtain*

$$\|\tilde{W}_c(kT)\| \leq e^{-\lambda k T} \|\tilde{W}_c(0)\| \tag{47}$$

where

$$\lambda = -\frac{1}{T} \ln(\sqrt{1 - 2\Gamma_c \xi_3}) \tag{48}$$

(b) We can let $\|\epsilon_H\| \leq \epsilon_{\max}$ *and* $\|y\| \leq y_{\max}$ *then* $\|\tilde{W}_c\|$ *achieves exponential convergence to the residual set*

$$\tilde{W}_c(t) \leq \frac{\sqrt{\xi_2 T}}{\xi_1} \{[y_{\max} + \delta \xi_2 \Gamma_c (\epsilon_{\max} + y_{\max})]\} \tag{49}$$

where δ is a positive constant. The proof of this lemma is shown in [18]

The following performance indicates that the adaptive law (41) is valid under the PE condition, the weights $\hat{W}_c$ can converge to the unknown weights W_c. Therefore, $\hat{V}(x)$ achieves the convergence approach to the value function under the controller $\nu(t)$.

Theorem 1 *With any admissible bounded control policy $\nu(t)$, given the adaptive law* (41) *for the critic NN and assuming that $\bar{\alpha}_1$ is persistently exciting, letting the residual error ϵ_H be bounded and satisfy $\|\epsilon_H\| < \|\epsilon_{\max}\|$, the critic parameter error achieves the exponential convergence with decay factor* (48) *to the residual set, which is defined as*

$$\tilde{W}_c(t) \leq \frac{\sqrt{\xi_2 T}}{\xi_1}\{[1 + 2\delta\xi_2\Gamma_c]\epsilon_{\max}\} \tag{50}$$

Proof: We design Lyapunov function candidate as

$$L(t) = \frac{1}{2}tr\{\tilde{W}_c^T\Gamma_c^{-1}\tilde{W}_c\} \tag{51}$$

Considering (42), the derivative of L becomes

$$\dot{L} = -tr\left\{\tilde{W}_c^T\frac{\alpha_c\alpha_c^T}{\vartheta_c^2}\tilde{W}_c\right\} + tr\left\{\tilde{W}_c^T\frac{\alpha_c}{\vartheta_c}\frac{\epsilon_H}{\vartheta_c}\right\} \tag{52}$$

Then, we obtain

$$\begin{aligned}\dot{L} &\leq -\left\|\frac{\alpha_c^T}{\vartheta_c}\tilde{W}_c\right\|^2 + \left\|\frac{\alpha_c^T}{\vartheta_c}\tilde{W}_c\right\|\left\|\frac{\epsilon_H}{\vartheta_c}\right\| \\ \dot{L} &\leq -\left\|\frac{\alpha_c^T}{\vartheta_c}\tilde{W}_c\right\|\left[\left\|\frac{\alpha_c^T}{\vartheta_c}\tilde{W}_c\right\| - \left\|\frac{\epsilon_H}{\vartheta_c}\right\|\right]\end{aligned} \tag{53}$$

If $\epsilon_{\max}$ satisfies

$$\left\|\frac{\epsilon_H}{\vartheta_c}\right\| < \epsilon_{\max} < \left\|\frac{\alpha_c^T}{\vartheta_c}\tilde{W}_c\right\| \tag{54}$$

therefore $\dot{L} \leq 0$ with $\|\vartheta_c\| \geq 1$.

If the condition (54) is satisfied, as $L(t)$ decreases, a proper bound for $\|\bar{\alpha}_c\tilde{W}_c\|$ is provided.

Take the error dynamics system (44) with the bounded output $\|y\| < \epsilon_{\max}$ into consideration. Therefore, $\|\tilde{W}_c\|$ achieves the exponential convergence to the residual set

$$\tilde{W}_c(t) \leq \frac{\sqrt{\xi_2 T}}{\xi_1}\{[1 + 2\delta\xi_2\Gamma_c]\epsilon_{\max}\} \tag{55}$$

■

Remark 1 *As $N \to \infty$, $\epsilon_H \to 0$ uniformly [66]. Thus, $\epsilon_{\max}$ decreases as the number of hidden layer neurons increases.*

Remark 2 *This theorem is obtained based on the assumption that the controller $\nu(t)$ is bounded, because ϵ_H is associated with $\nu(t)$.*

3.2 Action Neural Network and Adaptive Optimal Control

In this part, we propose an adaptive PI algorithm. Namely, we need to adjust the weights of actor NN and critic NN simultaneously. This algorithm is based on the Generalized Policy Iteration (GPI), and there is a particular introduction in [32].

For the actor NN, a rigorously justified form is required. Therefore, we take one step of the PI algorithm into account. According to (23) and (24), update controller is designed as

$$\nu = -\frac{1}{2}R^{-1}F^T(\varpi)\sum_{i=1}^{\infty} c_i \nabla\varphi_i(\varpi) \tag{56}$$

Lemma 4 *[18] Let W_c as the least-squares solution to* (29) *and we can design*

$$\nu_c(\varpi) = -\frac{1}{2}R^{-1}F^T(\varpi)\nabla V_0(\varpi) = -\frac{1}{2}R^{-1}F^T(\varpi)\nabla\phi_c^T(\varpi)W_c \tag{57}$$

where V_0 is defined in (32)*.*

Then, as $N \to \infty$:

(a) $\sup_{x\in\Omega}\|\nu_c - \nu\| \to 0$

(b) There exists an N_0 satisfying that for $N > N_0$, $\nu_c(x)$ is admissible.

The proof is presented in [18]. According to the above results, we can consider the desired control policy is as (57) with unknown weights. Consequently, based on the form of actor NN which can compute the control input. And then we design the control policy as follows

$$\nu_a(\varpi) = -\frac{1}{2}R^{-1}F^T(\varpi)\nabla\phi_c^T(\varpi)\hat{W}_a \tag{58}$$

where $\hat{W}_a$ represents the approximated values of the desired NN weights W_c for action NN. Then, we design the actor NN evaluated error as follows

$$\tilde{W}_a = W_c - \hat{W}_a \tag{59}$$

The following definition and assumption are necessary conditions for our studies.

Definition 2 *It is said to be uniformly ultimately bounded (UUB) for the equilibrium point* $\varpi_e = 0$ *of* (3) *with a compact set* $\Theta \subset \mathbb{R}^n$ *for all* $\varpi_0 \in \Theta$*, and there exits a boundary* B *and a time* T *with* $\|\varpi(t) - \varpi_e\| \leq B$ *for all* $t \geq t_0 + T$.

Assumption 7 *a. We can assume that* $E(\cdot)$ *is Lipschitz and* $F(\cdot)$ *is bounded*

$$\|E(\varpi)\| < d_e\|\varpi\|, \quad \|F(\varpi)\| < d_f.$$

b. We can assume that the evaluated error of actor NN and the gradient of it are bounded on a compact set Ω.

$$\|\epsilon\| < d_\epsilon, \quad \|\nabla\epsilon\| < d_{\epsilon_\varpi}$$

c. We can assume that the basis functions of NN and their gradients are bounded

$$\|\phi_c(\varpi)\| < d_\phi, \quad \|\nabla\phi_c(\varpi)\| < d_{\phi_\varpi}.$$

Then, we obtain the following main theorem. The adaptive laws for the actor NN and critic NN are designed to ensure the convergence of the synchronized PI algorithm to the optical control policy with the condition that the system is stable. Accordingly, we propose the theorem as follows.

Theorem 2 *With the dynamics described as* (3)*, the critic NN provided as* (36) *and the control input represented by actor NN* (58)*, then we can define the critic NN adaptive law as*

$$\dot{\hat{W}}_c = -\Gamma_c \frac{\alpha_a}{(\alpha_a^T \alpha_a + 1)^2}[\alpha_a^T \hat{W}_c + Q(\varpi) + \nu_a R \nu_a] \tag{60}$$

where $\alpha_a = \nabla\phi_c(E + F\nu_a(t))$*. With the assumption that* $\bar{\alpha}_a = \alpha_a/(\alpha_a^T\alpha_a + 1)$ *is persistently exciting, we have the adaptive law for the actor NN as*

$$\dot{\hat{W}}_a = -\Gamma_a\{(K_a\hat{W}_a - K_c\bar{\alpha}_a^T\hat{W}_c) - \frac{1}{4}\bar{D}_c(x)\hat{W}_a\vartheta_a^T\hat{W}_c\} \tag{61}$$

where

$$\bar{D}_c \equiv \nabla\phi_c(\varpi)F(\varpi)R^{-1}F^T(\varpi)\nabla\phi_c^T(\varpi) \tag{62}$$

$$\vartheta_a \equiv \frac{\alpha_a}{(\alpha_a^T\alpha_a + 1)^2} \tag{63}$$

$K_a > 0$ *and* $K_c > 0$ *are adaptive parameters, which are selected in the proof in detail. Then, there is a constant* N_0 *satisfying that for the nodes of hidden layer unites* $N > N_0$*, the state of closed-loop system, the weights error of critic NN* $\tilde{W}_c$*, and the weights error of actor NN* $\tilde{W}_a$ *are uniformly ultimately bounded. Furthermore, according to Theorem 1,* $\epsilon_{\max}$ *is given in the appendix part such that* $\tilde{W}_c$ *achieves the exponential convergence to the approximate optimal critic NN weight value* W_c*.*

Proof*: According to analysis of Lyapunov stability, the convergence of the system can be proved. We design the Lyapunov function candidate as*

$$L(t) = V(\varpi) + \frac{1}{2}tr\{\tilde{W}_c^T\Gamma_c^{-1}\tilde{W}_c\} + \frac{1}{2}tr\{\tilde{W}_a^T\Gamma_a^{-1}\tilde{W}_a\} \tag{64}$$

Through the proper selection of adaptive laws, the errors $\tilde{W}_c$ *and* $\tilde{W}_a$ *can be proved to be UUB, and convergence can be obtained. We can let*

$$L_v = V(\varpi), L_c = \frac{1}{2}tr\{\tilde{W}_c^T\Gamma_c^{-1}\tilde{W}_c\}, L_a = \frac{1}{2}tr\{\tilde{W}_a^T\Gamma_a^{-1}\tilde{W}_a\}$$

then,

$$\dot{L}(\varpi) = \dot{L}_v(\varpi) + \dot{L}_c(\varpi) + \dot{L}_a(\varpi)$$

For the first term L_v, its derivative is

$$\dot{L}_v = W_c^T(\nabla\phi_c e(x) - \frac{1}{2}\bar{D}_c(\varpi)\hat{W}_a) + \nabla\epsilon^T(\varpi)(E(\varpi) - \frac{1}{2}FR^{-1}F^T\nabla\phi_c^T\hat{W}_a) \tag{65}$$

Because of

$$\dot{\epsilon}(\varpi) = \nabla\epsilon^T(\varpi)(E(\varpi) - \frac{1}{2}FR^{-1}F^T\nabla\phi_c^T\hat{W}_a)$$

So $\dot{L}_v$ can be expressed as

$$\begin{aligned}
\dot{L}_v &= W_c^T(\nabla\phi_c E(\varpi) - \frac{1}{2}\bar{D}_c(x)\hat{W}_a) + \dot{\epsilon}(x) \\
&= W_c^T\nabla\phi_c E(\varpi) + \frac{1}{2}W_c^T\bar{D}_c(\varpi)(W_c - \hat{W}_a) - \frac{1}{2}W_c^T\bar{D}_c(\varpi)W_c + \dot{\epsilon}(\varpi) \\
&= W_c^T\nabla\phi_c E(\varpi) + \frac{1}{2}W_c^T\bar{D}_c(\varpi)\tilde{W}_a - \frac{1}{2}W_c^T\bar{D}_c(\varpi)W_c + \dot{\epsilon}(\varpi) \\
&= W_c^T\alpha_c + \frac{1}{2}W_c^T\bar{D}_c(z)\tilde{W}_a + \dot{\epsilon}(\varpi)
\end{aligned} \tag{66}$$

From the HJB equation, we can obtain

$$W_c^T\alpha_c = -Q(\varpi) - \frac{1}{4}W_c^T\bar{D}_c(\varpi)W_c + \epsilon_{HJB} \tag{67}$$

Then, we have

$$\dot{L}_v = -Q(\varpi) - \frac{1}{4}W_c^T\bar{D}_c(\varpi)W_c + \epsilon_{HJB} + \frac{1}{2}W_c^T\bar{D}_c(\varpi)\tilde{W}_a + \dot{\epsilon}(\varpi) \tag{68}$$

For L_c, its derivative is

$$\begin{aligned}
&\dot{L}_c = \tilde{W}_c^T \Gamma_c^{-1} \dot{\tilde{W}}_c \\
&= \tilde{W}_c^T \Gamma_c^{-1} \Gamma_c \frac{\alpha_a}{(\alpha_a^T \alpha_a + 1)^2} \Big(\alpha_a^T \hat{W}_c + Q(\varpi) + \frac{1}{4} \hat{W}_a^T \bar{D}_c \hat{W}_a \Big) \\
&= \tilde{W}_c^T \frac{\alpha_a}{(\alpha_a^T \alpha_a + 1)^2} \Big(\alpha_a^T \hat{W}_c - \alpha_c^T W_c + \frac{1}{4} \hat{W}_a^T \bar{D}_c \hat{W}_a - \frac{1}{4} W_c^T \bar{D}_c W_c + \epsilon_{HJB} \Big) \\
&= \tilde{W}_c^T \frac{\alpha_a}{(\alpha_a^T \alpha_a + 1)^2} \Big(- E^T \nabla \phi_c^T \tilde{W}_c + \frac{1}{2} \hat{W}_a^T \bar{D}_c \tilde{W}_c + \frac{1}{4} \tilde{W}_a^T \bar{D}_c \tilde{W}_a + \epsilon_{HJB} \Big)
\end{aligned} \tag{69}$$

Since we defined α_a as $\alpha_a = \nabla \phi_c (E + F \nu_a(t))$, we have

$$\begin{aligned}
\dot{L}_c &= \tilde{W}_c^T \frac{\alpha_a}{(\alpha_a^T \alpha_a + 1)^2} \Big(- \alpha_a^T \tilde{W}_c + \frac{1}{4} \tilde{W}_a^T \bar{D}_c \tilde{W}_a + \epsilon_{HJB} \Big) \\
&= \bar{L}_c + \frac{1}{4} \tilde{W}_c^T \frac{\alpha_a}{(\alpha_a^T \alpha_a + 1)^2} \tilde{W}_a^T \bar{D}_c \tilde{W}_a
\end{aligned} \tag{70}$$

where

$$\bar{L}_c = \tilde{W}_c^T \frac{\alpha_a}{(\alpha_a^T \alpha_a + 1)^2} (-\alpha_a^T \tilde{W}_c + \epsilon_{HJB}) = \tilde{W}_c^T \bar{\alpha}_a \Big(- \bar{\alpha}_a^T \tilde{W}_c + \frac{\epsilon_{HJB}}{\vartheta'_a} \Big) \tag{71}$$

Substituting (68) *and* (70) *into $\dot{L}$, we can obtain*

$$\begin{aligned}
\dot{L} = &-Q(\varpi) - \frac{1}{4} W_c^T \bar{D}_c(\varpi) W_c + \frac{1}{2} W_c^T \bar{D}_c(\varpi) \tilde{W}_a \\
&+ \epsilon_{HJB} + \dot{\epsilon} + \tilde{W}_a \Gamma_a^{-1} \dot{\tilde{W}}_a + \tilde{W}_c^T \frac{\alpha_a}{(\alpha_a^T \alpha_a + 1)^2} \Big(- \alpha_a^T \tilde{W}_c + \frac{1}{4} \tilde{W}_a^T \bar{D}_c \tilde{W}_a + \epsilon_{HJB} \Big)
\end{aligned} \tag{72}$$

Then we can obtain

$$\begin{aligned}
\dot{L} = &W_c^T \alpha_c + \bar{L}_c + \dot{\epsilon} + \frac{1}{2} W_a^T \bar{D}_c(\varpi) \tilde{W}_c - \tilde{W}_a \Gamma_a^{-1} \dot{\hat{W}}_a \\
&+ \frac{1}{4} \tilde{W}_a \bar{D}_c W_c \frac{\bar{\alpha}_a^T}{\vartheta'_a} \tilde{W}_c - \frac{1}{4} \tilde{W}_a \bar{D}_c W_c \frac{\bar{\alpha}_a^T}{\vartheta'_a} W_c + \frac{1}{4} \tilde{W}_a \bar{D}_c \tilde{W}_a \frac{\bar{\alpha}_a^T}{\vartheta'_a} W_c + \frac{1}{4} \tilde{W}_a \bar{D}_c \hat{W}_a \frac{\bar{\alpha}_a^T}{\vartheta'_a} \hat{W}_c
\end{aligned} \tag{73}$$

where $\bar{\alpha}_a = \frac{\alpha_a}{\alpha_a^T \alpha_a + 1}$ and $\vartheta'_a = \alpha_a^T \alpha_a + 1$.

For the selection of the update law for the action NN, we have

$$\begin{aligned}\dot{L} = & W_c^T\alpha_c + \bar{L}_c + \dot{\epsilon} - \tilde{W}_a^T\Big(\Gamma_a^{-1}\dot{\hat{W}}_a - \frac{1}{4}\tilde{W}_a\bar{D}_c\hat{W}_a\frac{\bar{\alpha}_a^T}{\vartheta_a'}\hat{W}_c\Big) \\ & + \frac{1}{2}W_a^T\bar{D}_c(\varpi)\tilde{W}_c + \frac{1}{4}\tilde{W}_a\bar{D}_cW_c\frac{\bar{\alpha}_a^T}{\vartheta_a'}\tilde{W}_c - \frac{1}{4}\tilde{W}_a\bar{D}_cW_c\frac{\bar{\alpha}_a^T}{\vartheta_a'}W_c + \frac{1}{4}\tilde{W}_a\bar{D}_c\tilde{W}_a\frac{\bar{\alpha}_a^T}{\vartheta_a'}W_c \end{aligned} \tag{74}$$

Therefore, we can define the actor adaptive law as

$$\dot{\hat{W}}_a = -\Gamma_a\{(K_a\hat{W}_a - K_c\bar{\alpha}_a^T\hat{W}_c) - \frac{1}{4}\bar{D}_c(\varpi)\hat{W}_a\vartheta_a^T\hat{W}_c\} \tag{75}$$

According to the redefined actor adaptive law (75)*, there exists*

$$\begin{aligned} & -\tilde{W}_a^T\Big(\Gamma_a^{-1}\dot{\hat{W}}_a - \frac{1}{4}\tilde{W}_a\bar{D}_c\hat{W}_a\frac{\bar{\alpha}_a^T}{\vartheta_a'}\hat{W}_c\Big) \\ = & \tilde{W}_a^TK_a\hat{W}_a - \tilde{W}_a^TK_c\bar{\alpha}_a^T\hat{W}_c \\ = & \tilde{W}_a^TK_a(W_c - \tilde{W}_a) - \tilde{W}_a^TK_c\bar{\alpha}_a^T(W_c - \tilde{W}_c) \\ = & \tilde{W}_a^TK_aW_c - \tilde{W}_a^TK_a\tilde{W}_a - \tilde{W}_a^TK_c\bar{\alpha}_a^TW_c + \tilde{W}_a^TK_c\bar{\alpha}_a{}^T\tilde{W}_c \end{aligned} \tag{76}$$

Finally, we have

$$\begin{aligned}\dot{L} = & -Q(\varpi) - \frac{1}{4}W_c^T\bar{D}_cW_c + \epsilon_{HJB} + \tilde{W}_c^T\bar{\alpha}_a\Big(-\bar{\alpha}_a^T\tilde{W}_c + \frac{\epsilon_{HJB}}{\vartheta_a'}\Big) + \tilde{W}_a^TK_aW_c \\ & -\tilde{W}_a^TK_c\bar{\alpha}_a^TW_c + \tilde{W}_a^TK_c\bar{\alpha}_a^T\tilde{W}_c - \tilde{W}_a^TK_a\tilde{W}_a + \dot{\epsilon} + \frac{1}{2}W_a^T\bar{D}_c\tilde{W}_c + \frac{1}{4}\tilde{W}_a\bar{D}_cW_c\frac{\bar{\alpha}_a^T}{\vartheta_a'}\tilde{W}_c \\ & -\frac{1}{4}\tilde{W}_a\bar{D}_cW_c\frac{\bar{\alpha}_a^T}{\vartheta_a'}W_c + \frac{1}{4}\tilde{W}_a\bar{D}_c\tilde{W}_a\frac{\bar{\alpha}_a^T}{\vartheta_a'}W_c \end{aligned} \tag{77}$$

It is necessary to utilize norm bounds at present. According to Assumption 7, for $\dot{\epsilon}$*, we have*

$$\|\dot{\epsilon}(\varpi)\| < d_{\epsilon_\varpi}d_a\|\varpi\| + \frac{1}{2}d_{\epsilon_\varpi}d_b^2d_{\phi_\varpi}\sigma_{\min}(R)\Big(\|W_c\| + \|\tilde{W}_a\|\Big) \tag{78}$$

Since $Q(\varpi) > 0$*, there exists* q' *satisfying* $\varpi^Tq'\varpi < Q(\varpi)$ *for* $\varpi \in \Omega$*. According to [18] that* ϵ_{HJB} *achieves uniform convergence as* N *increases.*

Choosing proper $\epsilon_{\max} > 0$ *and* $N_0(\epsilon_{\max})$*, there is* $\sup_{\varpi\in\Omega}\|\epsilon_{HJB}\| < \epsilon_{\max}$*. Next, let* $N > N_0$ *and*

$\tilde{X} = \begin{bmatrix} \varpi \\ \bar{\alpha}_a^T \tilde{W}_c \\ \tilde{W}_a \end{bmatrix}$, *we have*

$$\begin{aligned}
\dot{L} < & \frac{1}{4}\|W_c\|^2\|\bar{D}_c\| + \frac{1}{2}d_{\epsilon_\varpi} d_b^2 d_{\phi_\varpi}\sigma_{\min}(R)\|W_c\| + \epsilon_{\max} \\
& -\tilde{X}^T \begin{bmatrix} q'I & 0 & 0 \\ 0 & I & \left(-\frac{1}{2}K_c - \frac{1}{8\vartheta_a'}\bar{D}_c W_c\right)^T \\ 0 & -\frac{1}{2}K_c - \frac{1}{8\vartheta_a'}\bar{D}_c W_c & K_a - \frac{1}{8}(\bar{D}_c W_c \vartheta_a^T + \vartheta_a W_c^T \bar{D}_c) \end{bmatrix} \tilde{X} \\
& +\tilde{X}^T \begin{bmatrix} \frac{d_{\epsilon_x} d_a}{\frac{\epsilon_{\max}}{\vartheta_a'}} \\ \left(\frac{1}{2}\bar{D}_c + K_a - K_c\bar{\alpha}_a^T - \frac{1}{4}\bar{D}_c W_c \vartheta_a^T\right) W_c + \frac{1}{2}d_{\epsilon_\varpi} d_b^2 d_{\phi_\varpi}\sigma_{\min}(R) \end{bmatrix}
\end{aligned}$$

Then, we define

$$U = \begin{bmatrix} q'I & 0 & 0 \\ 0 & I & \left(-\frac{1}{2}K_c - \frac{1}{8\vartheta_a'}\bar{D}_c W_c\right)^T \\ 0 & -\frac{1}{2}K_c - \frac{1}{8\vartheta_a'}\bar{D}_c W_c & K_a - \frac{1}{8}(\bar{D}_c W_c \vartheta_a^T + \vartheta_a W_c^T \bar{D}_c) \end{bmatrix} \tilde{X}$$

and

$$\begin{aligned}
d = & \begin{bmatrix} \frac{d_{\epsilon_\varpi} d_a}{\frac{\epsilon_{\max}}{\vartheta_a'}} \\ \left(\frac{1}{2}\bar{D}_c + K_a - K_c\bar{\alpha_a}^T - \frac{1}{4}\bar{D}_c W_c \vartheta_a^T\right) W_c + \frac{1}{2}d_{\epsilon_\varpi} d_b^2 d_{\phi_\varpi}\sigma_{\min}(R) \end{bmatrix} \\
c = & \frac{1}{4}\|W_c\|^2\|\bar{D}_c\| + \frac{1}{2}d_{\epsilon_\varpi} d_b^2 d_{\phi_\varpi}\sigma_{\min}(R)\|W_c\|
\end{aligned}$$

By choosing proper parameters such that $U > 0$, we can obtain

$$\dot{L} < -\sigma_{\min}(U)\|\tilde{X}\|^2 + \|d\|\|\tilde{X}\| + c + \epsilon_{\max} \tag{79}$$

In order to make the Lyapunov derivative negative, the following condition should be satisfied

$$\|\tilde{X}\| > \frac{\|d\|}{2\sigma_{\min}(U)} + \sqrt{\frac{\|d\|^2}{4\sigma_{\min}^2(U)} + \frac{c + \epsilon_{\max}}{\sigma_{\min}(U)}} \tag{80}$$

■

4 Numerical simulations

In this part, we use the model of Cybership II, and it is built in a marine control laboratory in Norwegian University of Science and Technology with a 1:70 scale [1, 71]. We set the ideal trajectories of x_1 as

$$\begin{cases} x_{1xd}(t) = \sin(0.5t) \\ x_{1yd}(t) = 0.14\cos(2t) \\ x_{1\psi d}(t) = \tan^{-1}(\frac{\dot{x}_{1xd}}{\dot{x}_{1yd}}) \end{cases} \tag{81}$$

$$J^{-1}x_{1d} = x_{2d} \tag{82}$$

Force of gravity is defined as $g(x_1) = [0.4\cos(x_{1\psi})-0.72\cos(x_{1\psi}), 0.4\sin(x_{1\psi})+0.72\sin(x_{1\psi}), 0.36]$.

The symmetric positive definite inertia matrix M, the Centripetal and Coriolis torques C and the damping matrix D are expressed as

$$M = \begin{bmatrix} m-X_{du} & 0 & 0 \\ 0 & m-Y_{dv} & mx_g-Y_{dr} \\ 0 & mx_g-Y_{dr} & I_z-N_{dr} \end{bmatrix}$$

$$C(\upsilon) = \begin{bmatrix} C_{11} & C_{12} & C_{13} \\ C_{21} & C_{22} & C_{23} \\ C_{31} & C_{32} & C_{33} \end{bmatrix}$$

$$C_{11} = C_{12} = C_{21} = C_{22} = 0, \;\; C_{13} = (-m - Y_{dv})\upsilon_y - (mx_g - Y_{dr}\upsilon_\psi)$$
$$C_{23} = C_{32} = (m - X_{du})\upsilon_x, \;\; C_{31} = (m - Y_{dv})\upsilon_y + (mx_g - Y_{dr})\upsilon_\psi$$

$$D(\upsilon) = \begin{bmatrix} D_{11} & D_{12} & D_{13} \\ D_{21} & D_{22} & D_{23} \\ D_{31} & D_{32} & D_{33} \end{bmatrix}$$

$$D_{11} = -X_u - X_{uu}|\upsilon_x| - x_{uuu}\upsilon_x^2, \;\; D_{12} = D_{13} = D_{21} = D_{31} = 0$$
$$D_{22} = -Y_\upsilon - Y\upsilon\upsilon|\upsilon_y| - Y_{r\upsilon}|\upsilon_\psi|, \;\; D_{23} = -Y_r - Y_{\upsilon r}|\upsilon_y| - Y_{rr}|\upsilon_\psi|$$
$$D_{32} = -N_\upsilon - N_{\upsilon\upsilon}|\upsilon_y| - N_{r\upsilon}|\upsilon_\psi|, \;\; D_{33} = -N_r - N_{\upsilon r}|\upsilon_y| - N_{rr}|\upsilon_\psi|$$

In this paper, the parameters of system are chosen as follows

Table 1: Parameters of a marine vessel system

Parameter	Value	Parameter	Value	Parameter	Value
m	$23.8kg$	N_{d_r}	-1.0	Y_{d_ν}	-10.0
I_z	1.76	N_{d_ν}	0	$N_{\nu\nu}$	5.0437
x_g	0.046	Y_{d_r}	0	X_{d_ν}	-2.0
X_u	-0.7225	Y_{r_ν}	2	N_ν	0.1052
X_{uu}	-1.3274	Y_{ν_r}	1	N_{rr}	0.8
X_{uuu}	-5.8664	Y_{rr}	3	Y_r	0.1079
Y_ν	-0.8612	N_{r_ν}	5	N_{ν_r}	0.5
$Y_{\nu\nu}$	36.2823	N_r	4		

In our paper, we chose the initial parameters as: $\eta(0) = [0, 0.14, 0]^T$, $\upsilon(0) = [2, 0, 0]^T$, $q = 100I_{12}$, $R = 0.1I_3$. For the neural network, the parameters are selected as: $\Gamma_c = 100I_{150\times150}$, $\Gamma_a = 500I_{150\times150}$, $F_c = 0.02I_{150\times150}$, $F_a = 0.5I_{150\times150}$.

The performances of simulation are presented in Figs. 2-7. From Fig. 2, it is obvious that the real trajectory of position state x_1 can precisely track the desired trajectory of the closed-loop vessel system. The tracking error e_1 is shown in Fig. 3, from which we can obtain that the system achieves the stability with a very small tracking errors. The control input u is stated in Fig. 4. Fig. 5, and Fig. 6 present the values of the weight vectors $\hat{W}_c$ and $\hat{W}_a$, respectively. Approximations of integral cost function (14) with critic NN, actor NN and the real value are proposed in Fig. 7 respectively.

According to the above simulation results, we can draw the following conclusions. From Fig. 2 and Fig. 3, we can obtain the performance of trajectory tracking is very well and the tracking error is very small that close to zero almost. From Fig. 7 we can get a good approximation of the actual value function is being evolved. From these all figures, we can obtain the availability of our algorithm applied in marine vessel system, and our study direction will now be directed towards integrating the neural network with the actor/critic structure with the purpose of approximating the system dynamics.

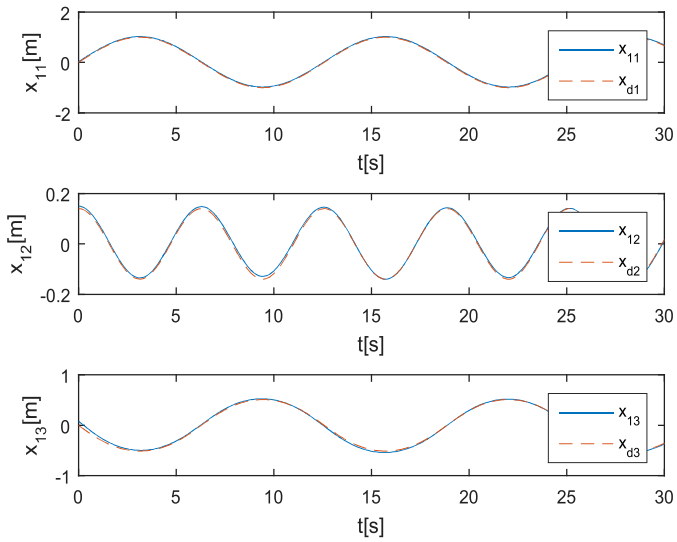

Figure 2: Tracking error e_1 of RL control.

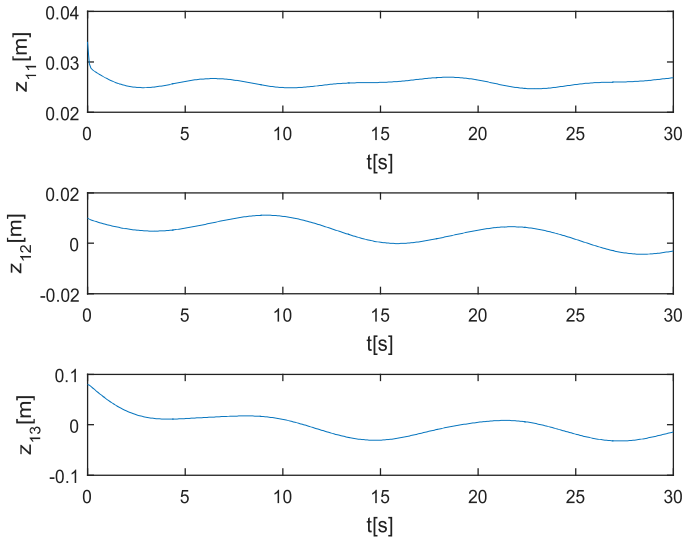

Figure 3: Tracking error e_1 of RL control.

5 Conclusion

A controller for the fully actuated vessel with known dynamic by using reinforcement learning algorithm is investigated in this paper. And then we have proven that the signals of closed-loop system are uniformly ultimately bounded, the performance of trajectory tracking is very well. Simulation part also has shown the effective and predominant results of the presented controller. In the future,

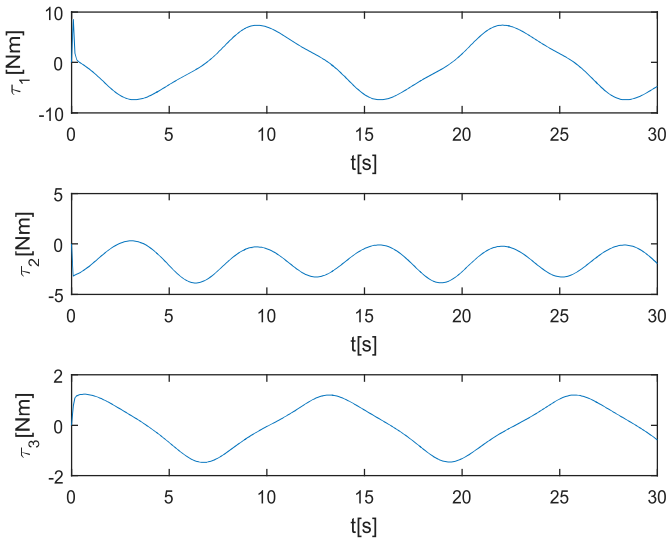

Figure 4: Control inputs τ under the RL control.

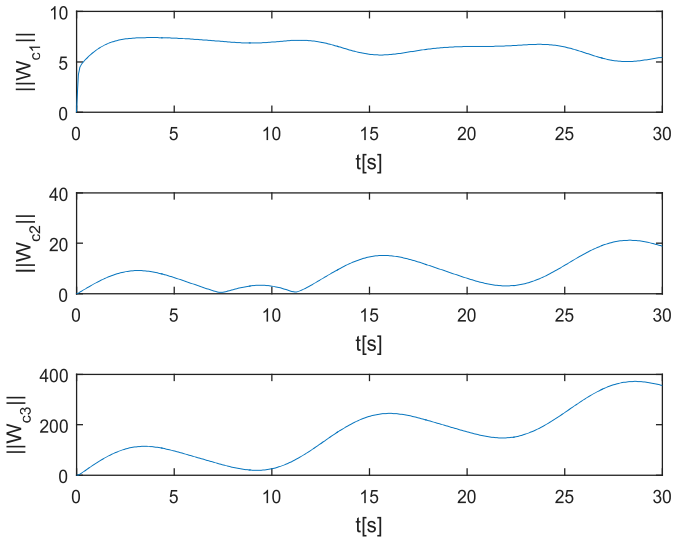

Figure 5: Weights of critic NN.

we will concentrate our attention on practical experiment, and we will also extend the proposed control method to deal with more complex issues such as constraints and deadzone. Furthermore, we will try our best to do more research about the novel control methods, and apply them to more domains, for example, robot arm, aircraft and so on. Due to the limitations in existing facilities and lack of resources on the vessel systems, we were not able to conduct the practical validation with good scaling for the proposed controls. Therefore, the next we will conduct some experiments to

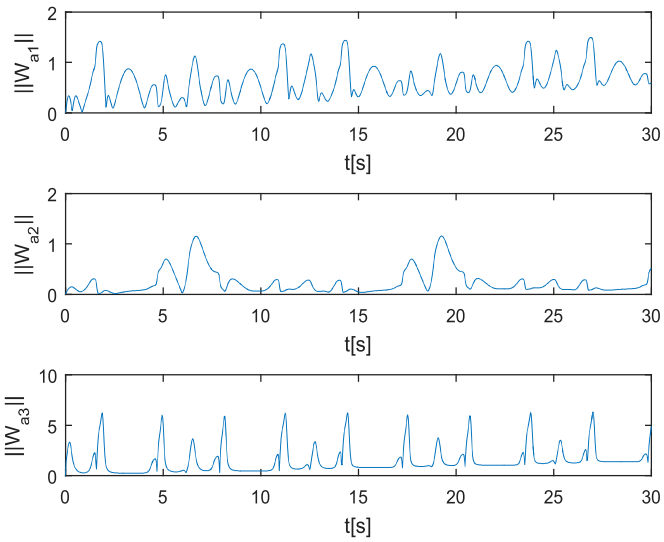

Figure 6: Weights of actor NN.

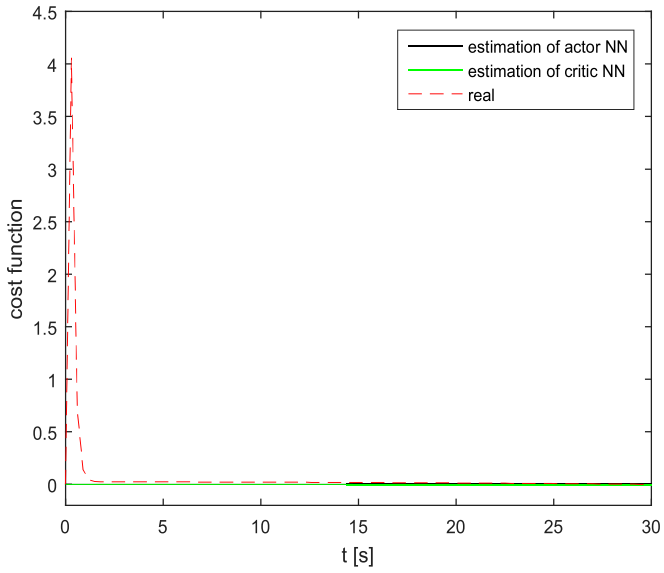

Figure 7: The cost function and its estimate.

verify the feasibility of these methods.

References

[1] K. P. Tee and S. S. Ge, "Control of fully actuated ocean surface vessels using a class of feed-forward approximators," *IEEE Transactions on Control Systems Technology*, vol. 14, no. 4, pp.

750–756, 2006.

[2] B. K. Wilhelm, R. B. Ivan, D. V. E. Karl, and M. R. Dhanak, "Control of an Unmanned Surface Vehicle With Uncertain Displacement and Drag," *IEEE Journal of Oceanic Engineering*, vol. 42, no. 2, pp. 458–476, 2016.

[3] B. V. E. How, S. S. Ge, and Y. S. Choo, "Dynamic Load Positioning for Subsea Installation via Adaptive Neural Control," *IEEE Journal of Oceanic Engineering*, vol. 35, no. 2, pp. 366–375, 2010.

[4] W. He and S. S. Ge, "Vibration control of a flexible string with both boundary input and output constraints," *IEEE Transactions on Control Systems Technology*, vol. 23, no. 4, pp. 1245–1254, 2015.

[5] Y. Yang, J. Du, H. Liu, C. Guo, and A. Ajith, "A trajectory tracking robust controller of surface vessels with disturbance uncertainties," *IEEE Transactions on Control Systems Technology*, vol. 22, no. 4, pp. 1511–1518, 2014.

[6] M. Chen, "Robust tracking control for self-balancing mobile robots using disturbance observer," *IEEE/CAA Journal of Automatica Sinica*, vol. 4, no. 3, pp. 458–465, 2017.

[7] Y. Yang, J. Du, H. Liu, C. Guo, and A. Abraham, "A trajectory tracking robust controller of surface vessels with disturbance uncertainties," *IEEE Transactions on Control Systems Technology*, vol. 22, no. 4, pp. 1511–1518, 2014.

[8] N. Wang, C. Qian, J. Sun, and Y. Liu, "Adaptive Robust Finite-Time Trajectory Tracking Control of Fully Actuated Marine Surface Vehicles," *IEEE Transactions on Control Systems Technology*, vol. 24, no. 4, pp. 1454–1462, 2016.

[9] N. Wang and M. J. Er, "Direct Adaptive Fuzzy Tracking Control of Marine Vehicles With Fully Unknown Parametric Dynamics and Uncertainties," *IEEE Transactions on Control Systems Technology*, vol. 24, no. 5, pp. 1845–1852, 2016.

[10] R. Yu, Q. Zhu, G. Xia, and Z. Liu, "Sliding mode tracking control of an underactuated surface vessel," *IET Control Theory and Applications*, vol. 6, no. 3, pp. 416–466, 2012.

[11] Z. Zhao and S. S. Ge, "Adaptive Neural Network Control of a Fully Actuated Marine Surface Vessel with Multiple Output Constraints," *IEEE Transactions on Control Systems Technology*, vol. 22, no. 4, pp. 1536–1543, 2014.

[12] Z. Zhao, X. Wang, C. Zhang, Z. Liu, and J. Yang, "Neural network based boundary control of a vibrating string system with input deadzone," *Neurocomputing*, vol. 275, pp. 1021–1027, 2018.

[13] M. Chen, "Disturbance attenuation tracking control for wheeled mobile robots with skidding and slipping," *IEEE Transactions on Industrial Electronics*, vol. 64, no. 4, pp. 3359–3368, 2017.

[14] W. He, Z. Yin, and C. Sun, "Adaptive Neural Network Control of a Marine Vessel With Constraints Using the Asymmetric Barrier Lyapunov Function," *IEEE Transactions on Cybernetics*, vol. 47, no. 7, pp. 1641–1651, 2017.

[15] X. He, W. He, J. Shi, and C. Sun, "Boundary vibration control of variable length crane systems in two-dimensional space with output constraints," *IEEE/ASME Transactions on Mechatronics*, vol. 22, no. 5, pp. 1952–1962, 2017.

[16] A. Hashem and L. C. M. R. A. S. Kenneth, R. Muske, "Sliding-Mode Tracking Control of Surface Vessels," *IEEE Transactions on Industrial Electronics*, vol. 55, no. 11, pp. 4004–4012, 2008.

[17] J. Si, A. G. Barto, W. B. Powell, and D. Wunsch, *Handbook of learning and approximate dynamic programming*. New Jersey: John Wiley, 2004.

[18] M. Abu-Khalaf and F. L. Lewis, "Nearly optimal control laws for nonlinear systems with saturating actuators using a neural network HJB approach," *Automatica*, vol. 41, no. 5, pp. 779–791, 2005.

[19] X. Xu, Z. Huang, L. Zuo, and H. He, "Manifold-based reinforcement learning via locally linear reconstruction," *IEEE transactions on Neural Networks and Learning Systems*, vol. 28, no. 4, pp. 934–947, 2017.

[20] X. Wu and D. Gao, "Fault tolerance control of SOFC systems based on nonlinear model predictive control," *International Journal of Hydrogen Energy*, vol. 42, no. 4, pp. 2288–2308, 2017.

[21] I. Grondman, L. Busoniu, G. A. Lopes, and R. Babuska, "A survey of actor-critic reinforcement learning: Standard and natural policy gradients," *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, vol. 42, no. 6, pp. 1291–1307, 2012.

[22] R. S. Sutton, "Generalization in reinforcement learning: Successful examples using sparse coarse coding," in *Advances in neural information processing systems*, 1996, pp. 1038–1044.

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

[24] L. Tang, Y. J. Liu, and S. Tong, "Adaptive neural control using reinforcement learning for a class of robot manipulator," *Neural Computing and Applications*, vol. 25, no. 1, pp. 135–141, 2014.

[25] D. V. Prokhorov and D. C. Wunsch, "Adaptive critic designs," *IEEE transactions on Neural Networks*, vol. 8, no. 5, pp. 997–1007, 1997.

[26] H. Xiao, Z. Li, C. Yang, L. Zhang, P. Yuan, L. Ding, and T. Wang, "Robust Stabilization of A Wheeled Mobile Robot Using Model Predictive Control Based on Neuro-dynamics Optimization," *IEEE Transactions on Industrial Electronics*, vol. 64, pp. 505–516, 2016.

[27] D. Wang, H. He, and D. Liu, "Adaptive critic nonlinear robust control: a survey," *IEEE transactions on cybernetics*, vol. 47, no. 10, pp. 3429–3451, 2017.

[28] W. He, Z. Li, 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.

[29] M. Riedmiller, "Neural fitted q iteration–first experiences with a data efficient neural reinforcement learning method," in *European Conference on Machine Learning*. Springer, 2005, pp. 317–328.

[30] Y. Zhang, P. Huang, Z. Meng, and Z. Liu, "Precise angles-only navigation for noncooperative proximity operation with application to tethered space robot," *IEEE Transactions on Control Systems Technology*, 2018, In Press, DOI: 10.1109/TCST.2018.2790400.

[31] T. Meng and W. He, "Iterative learning control of a robotic arm experiment platform with input constraint," *IEEE Transactions on Industrial Electronics*, vol. 65, no. 1, pp. 664–672, 2018.

[32] R. Sutton and A. Barto, *Reinforcement Learning:An Introduction*. MIT Press, 1998.

[33] R. W. Beard, G. N. Saridis, and J. T. Wen, "Galerkin approximations of the generalized Hamilton-Jacobi-Bellman equation," *Automatica*, vol. 33, no. 12, pp. 2159–2177, 1997.

[34] H. Qiao, J. Peng, and Z.-B. Xu, "Nonlinear measures: a new approach to exponential stability analysis for hopfield-type neural networks," *IEEE Transactions on Neural Networks*, vol. 12, no. 2, pp. 360–370, 2001.

[35] Z. Li, Z. Huang, W. He, and C.-Y. Su, "Adaptive impedance control for an upper limb robotic exoskeleton using biological signals," *IEEE Transactions on Industrial Electronics*, vol. 64, no. 2, pp. 1664–1674, 2017.

[36] B. Xu, F. Sun, Y. Pan, and B. Chen, "Disturbance Observer Based Composite Learning Fuzzy Control of Nonlinear Systems with Unknown Dead Zone," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 47, no. 8, pp. 1854–1862, 2017.

[37] L. Bai, Q. Zhou, L. Wang, Z. Yu, and H. Li, "Observer-based adaptive control for stochastic nonstrict-feedback systems with unknown backlash-like hysteresis," *International Journal of Adaptive Control and Signal Processing*, vol. 31, no. 10, pp. 1481–1490, 2017.

[38] Q. Zhou, H. Li, L. Wang, and R. Lu, "Prescribed performance observer-based adaptive fuzzy control for nonstrict-feedback stochastic nonlinear systems," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 2017, In Press, DOI: 10.1109/TSMC.2017.2738155.

[39] M. Hamdy, S. Abd-Elhaleem, and M. Fkirin, "Time-varying delay compensation for a class of nonlinear control systems over network via h adaptive fuzzy controller," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 47, no. 8, pp. 2114–2124, 2017.

[40] D. Wang and C. Mu, "A novel neural optimal control framework with nonlinear dynamics: Closed-loop stability and simulation verification," *Neurocomputing*, vol. 266, pp. 353–360, 2017.

[41] Z. Zhao, J. Shi, X. Lan, X. Wang, and J. Yang, "Adaptive neural network control of a flexible string system with non-symmetric dead-zone and output constraint," *Neurocomputing*, 2017, In Press.

[42] W. He, S. S. Ge, Y. Li, E. Chew, and Y. S. Ng, "Neural Network Control of a Rehabilitation Robot by State and Output Feedback," *Journal of Intelligent & Robotic Systems*, vol. 80, no. 1, pp. 15–31, 2015.

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

[44] Y. Song, X. Huang, and C. Wen, "Tracking control for a class of unknown nonsquare mimo nonaffine systems: A deep-rooted information based robust adaptive approach," *IEEE Transactions on Automatic Control*, vol. 61, no. 10, pp. 3227–3233, 2016.

[45] X. Cao, L. Liu, W. Shen, and Y. Cheng, "Distributed scheduling and delay-aware routing in multihop mr-mc wireless networks," *IEEE Transactions on Vehicular Technology*, vol. 65, no. 8, pp. 6330–6342, 2016.

[46] S.-L. Dai, C. Wang, and M. Wang, "Dynamic learning from adaptive neural network control of a class of nonaffine nonlinear systems," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 25, no. 1, pp. 111–123, 2014.

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

[48] W. He, B. Huang, Y. Dong, Z. Li, and C.-Y. Su, "Adaptive neural network control for robotic manipulators with unknown deadzone," *IEEE Transactions on Cybernetics*, 2017, In Press, DOI: 10.1109/TCYB.2017.2748418.

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

[50] H. Zhang, Y. Luo, and D. Liu, "Neural-network-based near-optimal control for a class of discrete-time affine nonlinear systems with control constraints." *IEEE Transactions on Neural Networks*, vol. 20, no. 9, pp. 1490–1503, 2009.

[51] Y. J. Liu, C. L. Chen, G. X. Wen, and S. Tong, "Adaptive neural output feedback tracking control for a class of uncertain discrete-time nonlinear systems." *IEEE Transactions on Neural Networks*, vol. 22, no. 7, pp. 1162–7, 2011.

[52] Z. Liu, G. Lai, Y. Zhang, X. Chen, and C. L. Chen, "Adaptive neural control for a class of nonlinear time-varying delay systems with unknown hysteresis." *IEEE Transactions on Neural Networks & Learning Systems*, vol. 25, no. 12, pp. 2129–40, 2014.

[53] L. Wang, M. Basin, H. Li, and R. Lu, "Observer-based composite adaptive fuzzy control for nonstrict-feedback systems with actuator failures," *IEEE Transactions on Fuzzy Systems*, 2017, In Press, DOI: 10.1109/TFUZZ.2017.2774185.

[54] C. Yang, Y. Jiang, Z. Li, W. He, and C.-Y. Su, "Neural control of bimanual robots with guaranteed global stability and motion precision," *IEEE Transactions on Industrial Informatics*, vol. 13, no. 3, pp. 1162–1171, 2017.

[55] D. Wang and C. Mu, "Adaptive-critic-based robust trajectory tracking of uncertain dynamics and its application to a spring–mass–damper system," *IEEE Transactions on Industrial Electronics*, vol. 65, no. 1, pp. 654–663, 2018.

[56] M. Chen and G. Tao, "Adaptive fault-tolerant control of uncertain nonlinear large-scale systems with unknown dead zone," *IEEE Transactions on Cybernetics*, vol. 46, no. 8, pp. 1851–1862, 2016.

[57] M. Chen, S.-Y. Shao, and B. Jiang, "Adaptive neural control of uncertain nonlinear systems using disturbance observer," *IEEE transactions on cybernetics*, vol. 47, no. 10, pp. 3110–3123, 2017.

[58] P. Huang, D. Wang, Z. Meng, F. Zhang, and Z. Liu, "Impact dynamic modeling and adaptive target capturing control for tethered space robots with uncertainties," *IEEE/ASME Transactions on Mechatronics*, vol. 21, no. 5, pp. 2260–2271, 2016.

[59] S. Zhang, Y. Dong, Y. Ouyang, Z. Yin, and K. Peng, "Adaptive neural control for robotic manipulators with output constraints and uncertainties," *IEEE Transactions on Neural Networks and Learning Systems*, 2018, In Press, DOI: 10.1109/TNNLS.2018.2803827.

[60] C. L. P. Chen, G.-X. Wen, Y.-J. Liu, and F.-Y. Wang, "Adaptive Consensus Control for a Class of Nonlinear Multiagent Time-Delay Systems Using Neural Networks," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 25, pp. 1217–1226, 2014.

[61] C. Yang, Z. Li, R. Cui, and B. Xu, "Neural network-based motion control of an underactuated wheeled inverted pendulum model," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 25, no. 11, pp. 2004–2016, 2014.

[62] Z. Yin, W. He, and C. Yang, "Tracking control of a marine surface vessel with full-state constraints," *International Journal of Systems Science*, vol. 48, no. 3, pp. 535–546, 2017.

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

[64] H. Modares, F. L. Lewis, 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.

[65] K. Doya, "Reinforcement learning in continuous time and space," *Neural Computation*, vol. 12, no. 1, pp. 219–45, 2000.

[66] M. Abu-Khalaf and F. L. Lewis, "Nearly optimal control laws for nonlinear systems with saturating actuators using a neural network HJB approach," *Automatica*, vol. 41, no. 5, pp. 779–791, 2005.

[67] K. David, "On an iterative technique for Riccati equation computations," *IEEE Transactions on Automatic Control*, vol. 13, no. 1, pp. 114–115, 1968.

[68] Y. Li, K. P. Tee, R. Yan, W. L. Chan, and Y. Wu, "A Framework of Human–Robot Coordination Based on Game Theory and Policy Iteration," *IEEE Transactions on Robotics*, vol. 32, no. 6, pp. 1408–1418, 2016.

[69] 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.

[70] P. Ioannou, "Adaptive Control Tutorial (Advances in Design and Control)," *Siggraph Acm Siggraph Courses*, 2007.

[71] R. Skjetne, T. I. Fossen, and P. V. Kokotović, "Adaptive maneuvering, with experiments, for a model ship in a marine control laboratory," *Automatica*, vol. 41, no. 2, pp. 289–298, 2005.