Optimal Control for fully-actuated Surface Vessel Systems

Van Tu Vu, Thanh Loc Pham, Quang Huy Tran, Phuong Nam Dao

***Abstract*— This paper deals with the design of an optimal tracking control for fully-actuated Surface Vessel Systems with completely unknown dynamics. A feed-forward term in proposed controller is introduced for obtaining the corresponding autonomous tracking error model. An integral reinforcement learning (IRL) and an Actor/Critic technique are then developed to solve Hamilton-Jacobi-Bellman (HJB) equation in optimal control term. The convergence of the proposed technique to the analytical solution of HJB equation is guaranteed. Additionally, the trajectory tracking effectiveness is also mentioned. Simulation studies are given to evaluate the quality of the proposed method.**

Keywords- Surface Vessels (SVs), Integral Reinforcement Learning (IRL), Adaptive Dynamic Programming (ADP), Lyapunov Stability Theory.

I. INTRODUCTION

THE optimal control has been extensively considered in the development to improve control system performance of Surface Vessels (SVs). The application of optimal control theory has mostly concentrated on solving Hamilton-Jacobi-Bellman (HJB) equation using reinforcement learning (RL) technique because it is impossible to analytically solve this equation. In practice, it is usually utilized to develop traditional nonlinear controller in SVs control system. Existing control schemes to SVs are often considered by cascade control system [1-4]. In [1], full-state regulation control problem is addressed by dividing into 2 sub-tasks and then translation control scheme and rotation controller are presented with exact observers. Furthermore, finite time sliding mode control design is developed for SVs with the sliding variable to be obtained from observer of not only uncertainties/disturbances but also velocity error [2]. In [3], although the cascade control system is handled, but it is obviously different from the existing method in [1, 2], the tan-Barrier Lyapunov technique addressed the error constraint and finite time control problem. Additionally, authors in [3] solve the actuator saturation by inserting additional term into given control scheme and then obtaining its updating law. In [4], although this model is also considered as Under-Actuated system of Wheeled Mobile Robots (WMRs) but in contrast to designing WMRs control system, the outer sub-system is fully-actuated and the inner-subsystem is under-actuated. Additionally, actuator saturation and observer are also similarly implemented as in [2, 3]. The application of sliding mode control (SMC) for SVs is extend to integral SMC technique by the work in [6]. For the purpose of developing the control design for SVs with input/output and state constraint, the optimal control algorithm is considered as remarkable solution with the difficulties in solving HJB equation. In recent years, Reinforcement Learning (RL) algorithm has been remarkably mentioned with many approaches, such as Actor/Critic, On/Off policy Integral Reinforcement Learning (IRL) technique, Q learning, etc. [5-9]. In [5], the updating law in actor and critic are implemented simultaneously based on the consideration of Hamiltonian function. This method is also used in [7] for all control loops of cascade controller with modified critic NNs. However, because of time-varying desired trajectory, the tracking error model needs to be known as autonomous system. Therefore, it leads to the effect of control performance. Moreover, authors in [9] proposed the additional term of Nussabaum function to handle unknown control direction with modified performance index.

In this paper, two ARL learning based optimal control schemes are presented for SVs to find the solution to optimal tracking problem, including online on-policy IRL algorithm and Actor/Critic Learning structure. A transformation method to autonomous tracking error model is also introduced for developing reinforcement learning algorithm. This is in contrast to the existing method [7] that implement Actor/Critic RL for non-autonomous systems.

II. PROBLEM FORMULATION AND PRELIMINARIES

A class of fully-actuated Surface Vessel (SV) systems are considered as:

$$\begin{cases} \dot{q} = J(q)\upsilon \\ M'(q)\dot{\upsilon} + C'(\upsilon)\upsilon + D'(\upsilon)\upsilon + g(q) = \tau + d(t) \end{cases} \tag{1}$$

where the joints vector $q = [x, y, \theta]^T \in \Re^3$ describes the potition and heading angle under the earth-fixed coorfinate. $\tau \in \Re^3$ is the vector of control input. The transform matrix $J(q)$ is known as:

$$J(q) = \begin{bmatrix} \cos\theta & -\sin\theta & 0 \\ \sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{bmatrix} \tag{2}$$

The knowledge of classical matrices $M'(q), C'(\upsilon), D'(\upsilon)$ is shown in [1,2,3]. Furthermore, in this work, these matrices are considered as unknown matrices. They can be decoupled by known estimated terms $M(q), C(\upsilon), D(\upsilon)$ and unknown terms $\Delta M(q), \Delta C(\upsilon), \Delta D(\upsilon)$ as:

This paper is submitted on March 11th, 2021.

Phuong Nam Dao and Thanh Loc Pham are with School of Electrical Engineering, Hanoi University of Science and Technology (HUST), Hanoi, Vietnam. (e-mail: nam.daophuong@ hust.edu.vn).

Van Tu Vu is with Haiphong University, Haiphong, Vietnam.

Quang Huy Tran is with National Cheng Kung University, Tainan, Taiwan.

$$M'(q)=M(q)+\Delta M(q), C'(\upsilon)=C(\upsilon)+\Delta C(\upsilon)$$
$$D'(\upsilon)=D(\upsilon)+\Delta D(\upsilon) \tag{3}$$

According to (1) and (3), it leads to the model:

$$\begin{cases}\dot{q}=J(q)\upsilon \\ M(q)\dot{\upsilon}+C(\upsilon)\upsilon+D(\upsilon)\upsilon+g(q)=\tau+\Delta(q,\upsilon,\dot{\upsilon},t)\end{cases} \tag{4}$$

It is assummed that the lumped uncertainty is bounded $\|\Delta(q,\upsilon,\dot{\upsilon},t)\|\le\rho$. In this paper, the control purpose is to design an IRL based optimal control design in presence of Uncertainties/Disturbances.

Assumption 1: The desired trajectory $\eta_d(t)$ satisfies that there exists a function $k_1(.):\Re^n\mapsto\Re^n$ such that $\dot{\eta}_d=k_1(\eta_d)$

III. OPTIMAL CONTROL ALGORITHM FOR SVS

A. INTEGRAL REINFORCEMENT LEARNING ALGORITHM FOR SVS

This section introduces the application of optimal control in surface vessels by solving optimal control problem with Integral RL technique. This method enables us to develop optimal control scheme for uncertain SVs because of computing control policy from data collection. The Online On-Policy Integral Reinforcement Learning (IRL) is introduced to find control design for SVs with the advantage of eliminating the knowledge of internal dynamics.

For the purpose of implementing the trajectory tracking control scheme, the control input τ in (2) is utilized the feed-forward term τ_d to develop the remaning optimal control u for autonomous systems as follows:

$$\tau=u+\tau_d;$$
$$\tau_d=M(\nu)\dot{\nu}_d+C(\nu_d)\nu_d+D(\nu_d)\nu_d+g(\eta) \tag{5}$$

According to (1), (2) and assumption 1, we achieve the corresponding autonomous tracking error model as:

$$\dot{x}(t)=F(x)+G(x)u+\Delta \tag{6}$$

$$x=\begin{bmatrix}z_v^T & z_\eta^T & \eta_d^T\end{bmatrix}^T,$$

where $F(x)=\begin{bmatrix}-M^{-1}l_1+M^{-1}l\left(v_d\left(z_\eta,\eta_d\right)\right)\\ J\left(z_\eta+\eta_d\right)z_v-\beta_\eta z_\eta\\ h_1(\eta_d)\end{bmatrix};G(x)=\begin{bmatrix}M^{-1}\\0\\0\end{bmatrix}$

The term $u(t)$ in control input (2) plays the role of optimal control to minimize the following cost function:

$$J(x,u)=\int_0^\infty \alpha(x,u)d\theta \tag{7}$$

As we all known the optimal control is solve from HJB (Hamilton-Jacobi-Bellman) equation. However, it is hard to analytically solve this equation in theory and practical problem. To overcome this challenge, this work presents the application of On Policy IRL algorithm to find the term $u(t)$ as follows:

Step 1: Initializing admissible policy $u_0(t)$, $V_o(x(t))=0$ and using $u_0(t)$ for system.

- Collecting the data $x(t)$ and $u_0(t)$ in the time interval $[0,T]$ at N sampling time;
- $i\leftarrow 0$.

Step 2: Finding the function $V_{i+1}(x(t))$ approximating Bellman function $V^*(x(t))$ based on the data collection with the control input being $u_i(x(t))$:

$$\begin{cases}V_{i+1}(x(t))=\int_t^{t+\Delta}\alpha(x,u_i)d\tau+V_{i+1}(x(t+\Delta))\\ V_{i+1}(0)=0\end{cases} \tag{8}$$

Step 3: Updating the control policy $u_{i+1}(x(t))$:

$$u_{i+1}(x(t))=-\frac{1}{2}R^{-1}g^T(x(t))\frac{\partial V_{i+1}(x(t))^T}{\partial x} \tag{9}$$

- If $\|V_{i+1}(x(t))-V_i(x(t))\|\le\varepsilon$ then $u^*(x(t))=u_{i+1}(x(t));V^*(x(t))=V_{i+1}(x(t))$ and Stop Algorithm
- If $\|V_{i+1}(x(t))-V_i(x(t))\|>\varepsilon$ then
  - $i\leftarrow i+1$;
  - Using $u_i(t)$ for system and collecting the data $x(t)$ and $u_i(t)$ in the time interval $[0,T]$ at N sampling time;
  - Comming back Step 2;

In order to solve the equation (8) in the algorithm based on data collection, we will approximate $V_i(x(t))$ by Neural Network $V_i(x)=\hat{W}\Phi(x)$ and finding the updating law of $\hat{W}$ by optimality principle. According to (8), it follows that:

$$e(t)=\hat{W}^T\left\{\Phi(x(t+\Delta))-\Phi(x(t))\right\}-\int_t^{t+\Delta}\alpha(x,u)d\tau=\hat{W}^Th(t)-y(t) \tag{10}$$

Based on the data collection in N sampling times $H=[h(t_1),...,h(t_N)];Y=[y(t_1),...,y(t_N)]^T$ and Least-Squares solution, we imply the training weight of Critic NN as:

$$\hat{W}=(HH^T)^{-1}HY$$

B. ACTOR/CRITIC LEARNING ALGORITHM FOR SVS

This section introduces the second approach of optimal control in surface vessels by using Actor/Critic learning structure. This method is able to address optimal control scheme for uncertain SVs by obtaining the adjusting mechanism of Weights in Actor/Critic parts from arbitrary values.

As described in [9], we can utilize a Neural Networks (NN) based approximation method to develop the ARL algorithm in SVs controller. Because the Bellman function $V^*\left(x(t)\right)$ and optimal control input $u^*\left(x(t)\right)$ can be known as smooth functions with respect to the state $x(t)$, they are represented over any compact domain.

$$V^*(x) = W^T\psi(x) + \epsilon(x) \tag{11}$$

$$u^*\left(x\right) = -\frac{1}{2}R^{-1}G^T\left(x\right)\left(\left(\frac{\partial\psi}{\partial x}\right)^T W + \left(\frac{\partial\epsilon\left(x\right)}{\partial x}\right)^T\right) \tag{12}$$

where $W \in \Re^N$ is a vector of unknown ideal NN weights, N is the number of neurons of the proposed Neural Network, $\psi(X) \in \Re^N$ is a smooth NN activation function vector with $\psi_j(0) = 0$ and $\frac{\partial\psi_j}{\partial x}|_{x=0} = 0 \quad \forall j = 1,...,N$, $\epsilon(X) \in \Re$ is the reconstruction error of the Bellman function $V^*(x)$.

It is because of uncertain ideal NN weights, one need to find appropriate updating laws $\hat{W}_a, \hat{W}_c$ with the purpose of approximating the actor/critic parts and obtaining the optimal controller without solving analytically the HJB equation. In addition, the smooth NN activation function vector $\psi(x) \in \Re^N$ is chosen based on the description of SVs (see section 4). In [9], the Weierstrass approximation theorem is able to uniformly approximate not only $V^*\left(x\right)$ but also $\frac{\partial V^*\left(x\right)}{\partial x}$ with $\epsilon\left(x\right), \left(\frac{\partial\epsilon\left(x\right)}{\partial x}\right) \to 0 \text{ as } N \to \infty$

The estimated Bellman function of critic part $\hat{V}(X)$ and the estimated optimal control policy of actor part $\hat{u}(X)$ are employed to approximate the Bellman function and the optimal control input as:

$$\hat{V}(X) = \hat{W}_c^T\psi(X) \tag{13}$$

$$\hat{u}\left(x\right) = -\frac{1}{2}R^{-1}G^T\left(x\right)\left(\frac{\partial\psi}{\partial x}\right)^T\hat{W}_a \tag{14}$$

Based on the property of Hamiltonian $H(X,u,\frac{\partial V}{\partial x}) = r\left(x(\tau),u(\tau)\right) + \frac{\partial V}{\partial x}(F\left(x\right) + G\left(x\right)u)$ under the optimal control input $u^*\left(x(t)\right)$ and associated value function $V^*\left(x(t)\right)$, the adaptation laws of critic $\hat{W}_a, \hat{W}_c$ weights are simultaneously trained to minimize the squared Bellman error δ_{hjb} and the corresponding integral, respectively.

Due to the error between estimated functions $\hat{V}(X), \hat{u}(X)$ and optimal results $V^*\left(X\right), u^*\left(X\right)$, the Bellman error δ_{hjb} can be computed as:

$$\begin{aligned}\delta_{hjb} &= \hat{H}\left(X,\hat{u},\frac{\partial\hat{V}}{\partial X}\right) - H^*\left(X,u^*,\frac{\partial V^*}{\partial X}\right)\\ &= \hat{W}_c^T\sigma(X,\hat{u}) + \frac{1}{2}X^TQ_TX + \frac{1}{2}\hat{u}^TR\hat{u}\end{aligned} \tag{15}$$

where $\sigma(X,\hat{u}) = \frac{\partial\psi}{\partial X}(F(X) + G(X)\hat{u})$ is the regression vector of critic part.

The adaptation law of Critic weights is given:

$$\frac{d}{dt}\hat{W}_c = -k_c\lambda\frac{\sigma}{1+\nu\sigma^T\lambda\sigma}\delta_{hjb} \tag{16}$$

where $\nu, k_c \in \Re$ are constant positive gains, and $\lambda(t) \in \Re^{N\times N}$ is a estimated symmetric gain matrix obtained from the differential equation as:

$$\frac{d}{dt}\lambda = -k_c\lambda\frac{\sigma\sigma^T}{1+\nu\sigma^T\lambda\sigma}\lambda; \quad \lambda(t_s^+) = \lambda(0) = \varphi_0 I \tag{17}$$

where t_s^+ is resetting time satisfying the property of eigenvalue $\lambda_{\min}\left\{\lambda(t)\right\} \le \varphi_1, \varphi_0 > \varphi_1 > 0$. To ensure $\lambda(t) \in \Re^{N\times N}$ is positive definite and prevent the covariance wind-up problem, the covariance matrix $\lambda(t) \in \Re^{N\times N}$ can be satisfied as:

$$\varphi_1 I \le \lambda(t) \le \varphi_0 I$$

In addition, the adaptation law of actor NN part is proposed using the minimization of squared Bellman error.

$$\begin{aligned}\frac{d}{dt}\hat{W}_a &= -\frac{k_{a1}}{\sqrt{1+\sigma^T\sigma}}\frac{\partial\psi}{\partial x}GR^{-1}G^T\frac{\partial\psi^T}{\partial x}\left(\hat{W}_a - \hat{W}_c\right)\delta_{hjb}\\ &-k_{a2}\left(\hat{W}_a - \hat{W}_c\right)\end{aligned} \tag{18}$$

Remark 1: The proofs of stability in these two proposed solutions can be implemented by Lyapunov stability theory with the Lyapunov function being established from the Bellman function. Moreover, it is obviously different from [1,2,3,6] studying traditional nonlinear control approaches, this work considers the optimal control solution for SVs with two methodologies for finding the HJB solution. Moreover, this work extends the ARL based control for the case of non-autonomous system with time-varying reference.

IV. Simulation Studies

In this section, the proposed on-policy IL method is applied to a SV to show the convergence of actual trajectory and training weights. The parameters of a SV are considered with the following inertia, Coriolis matrices:

$$M = \begin{bmatrix} 20 & 0 & 0 \\ 0 & 19 & 0.72 \\ 0 & 0.72 & 2.7 \end{bmatrix};$$

$$C = \begin{bmatrix} 0 & 0 & -19v_y - 0.72v_z \\ 0 & 0 & 20v_x \\ 19v_y + 0.72v_z & -20v_x & 0 \end{bmatrix};$$

In this work, we only discuss the motion of SVs on the surface. It implies that $g(\eta)=0$ because there is no change of potential energy. The desired trajectory is given as $\eta_d(t)=\left[12\sin(0.2t),\ -12\sin(0.2t),\ 0.2t\right]^T$. The Critic Neural Network of $V_i(x(t))$ is choose by RBF with 12 nodes and the smooth activation function $\Phi(x)$ to be appropriately chosen. Fig.1, 2, 3 show the join variables converge to their desired values and confirm that the convergence of training weights. Furthermore, Off-Policy IRL is also considered for this SV to achieved the responses in Fig. 4, 5. On the other hand, Actor/Critic learning structure is also implemented to obtain the corresponding results in Fig. 6-17, in which, Fig 6-9 show the results in case of no Disturbance Observer and Fig10-17 show the results in case of using Disturbance Observer.

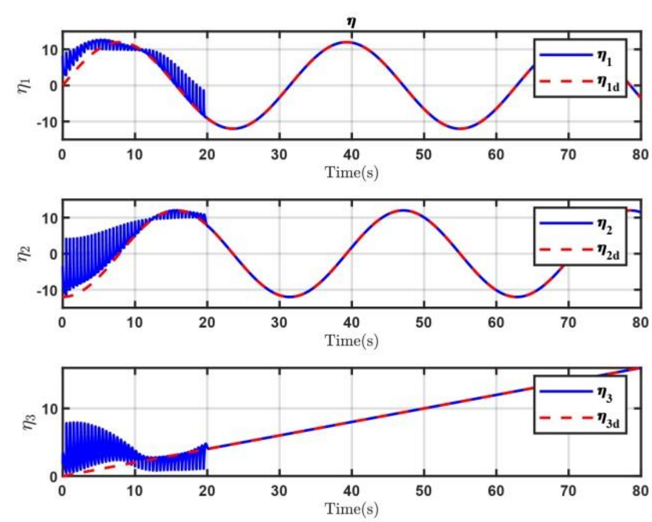

Fig. 1. The response of trajectory using on-Policy IRL algorithm.

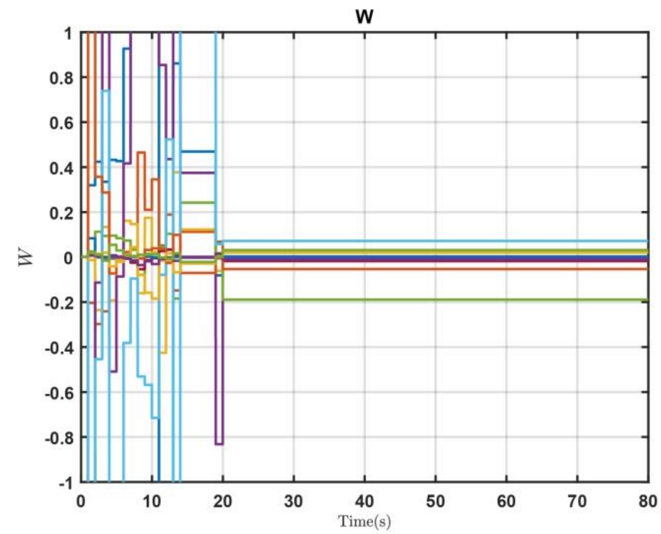

Fig. 2. The convergence of Critic Weights using on-Policy IRL algorithm.

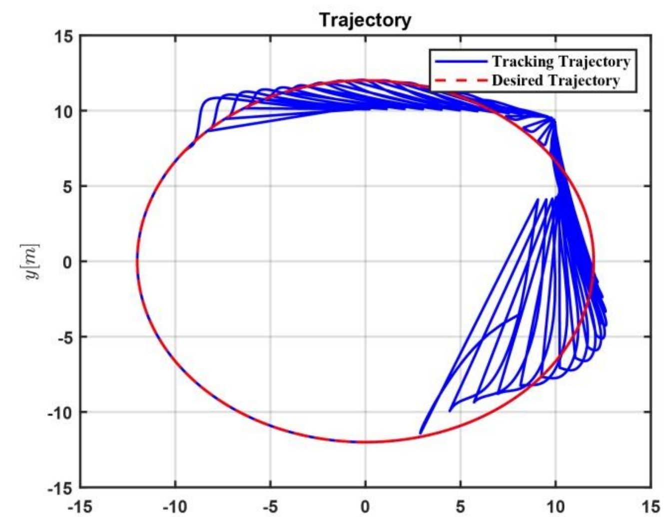

Fig. 3. The response of trajectory in Descartes coordinates.

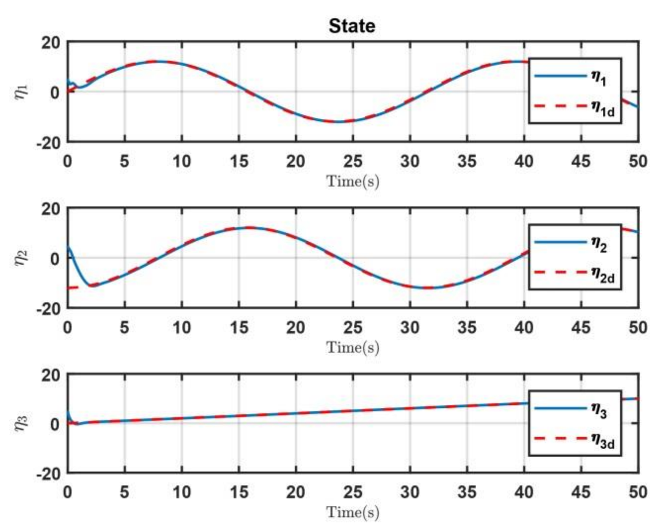

Fig. 4. The response of trajectory using off-Policy IRL algorithm.

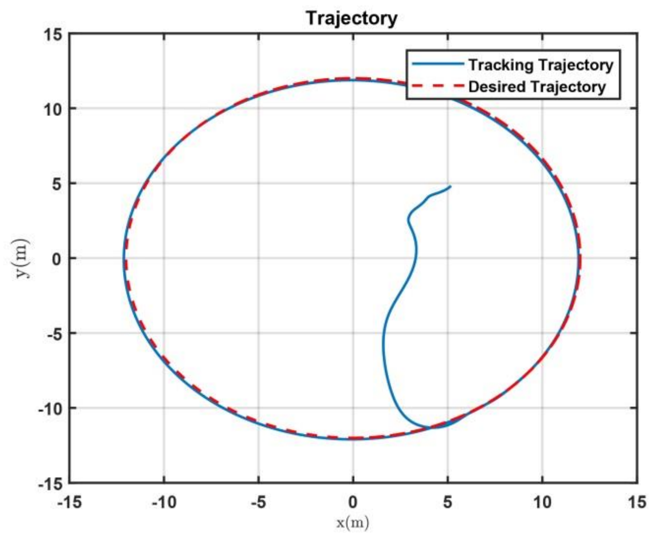

Fig. 5 The response of trajectory in Descartes coordinates using off-Policy IRL algorithm.

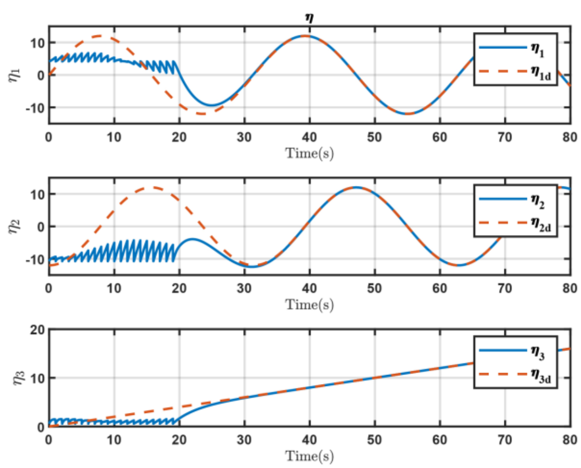

Fig. 6. The response of trajectory using Actor/Critic algorithm (Case 2).

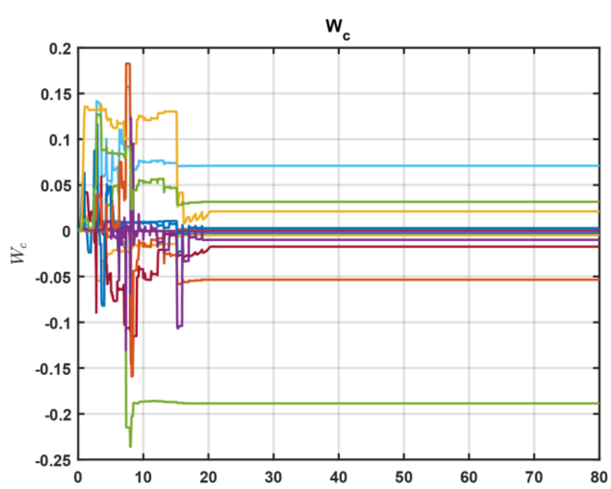

Fig. 7. The convergence of Critic Weights using Actor/Critic algorithm (Case 2).

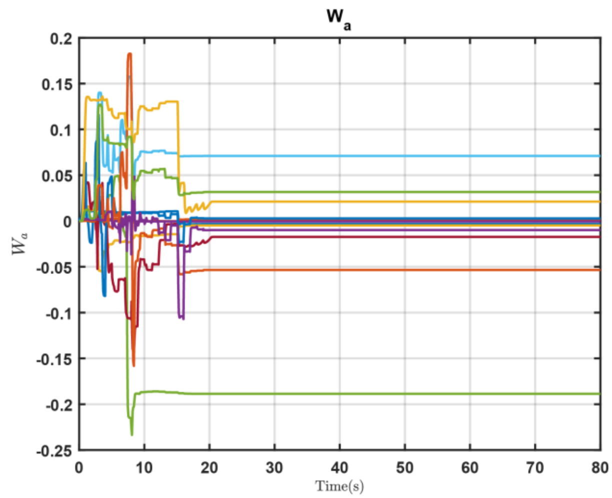

Fig. 8. The convergence of Actor Weights using Actor/Critic algorithm (Case 2).

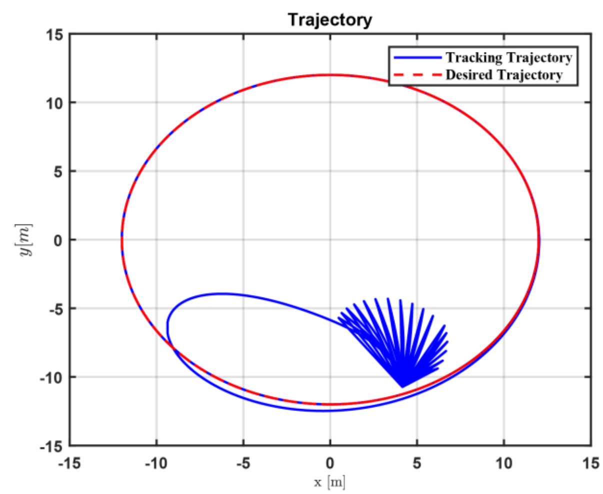

Fig. 9 The response of trajectory in Descartes coordinates using Actor/Critic algorithm (Case 2).

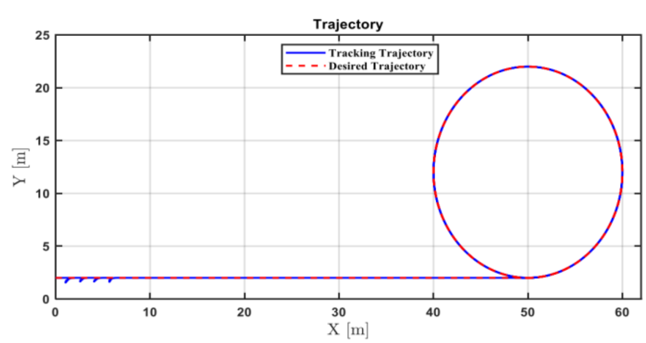

Fig. 10 The response of trajectory in Descartes coordinates using Actor/Critic algorithm (Case 2).

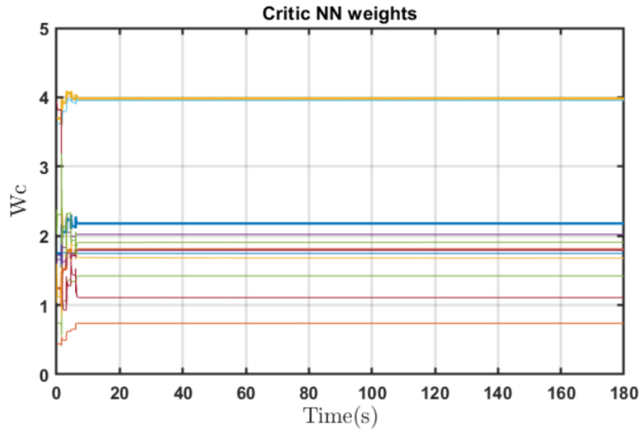

Fig. 11. The convergence of Critic Weights using Actor/Critic algorithm (Case 2).

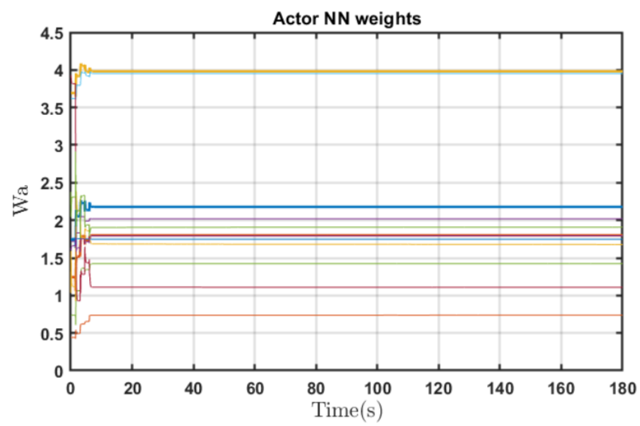

Fig. 12. The convergence of Actor Weights using Actor/Critic algorithm (Case 2).

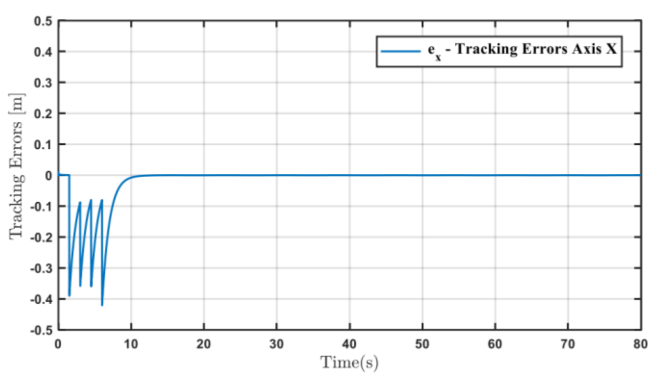

Fig. 13. The Tracking error in Axis X using Actor/Critic algorithm (Case 2).

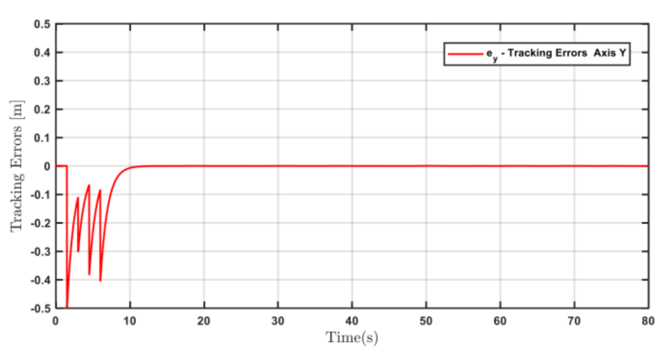

Fig. 14. The Tracking error in Axis Y using Actor/Critic algorithm (Case 2).

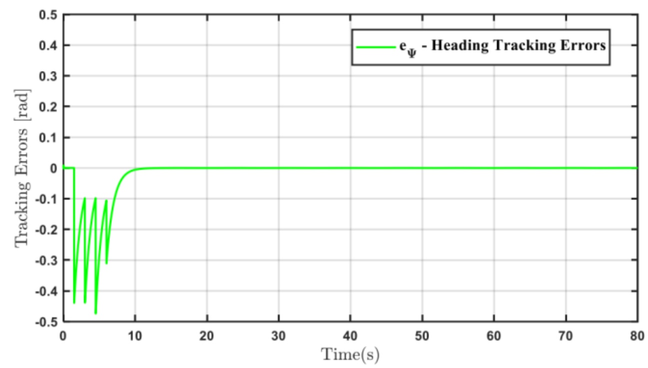

Fig. 15. The Tracking error in Angle using Actor/Critic algorithm (Case 2).

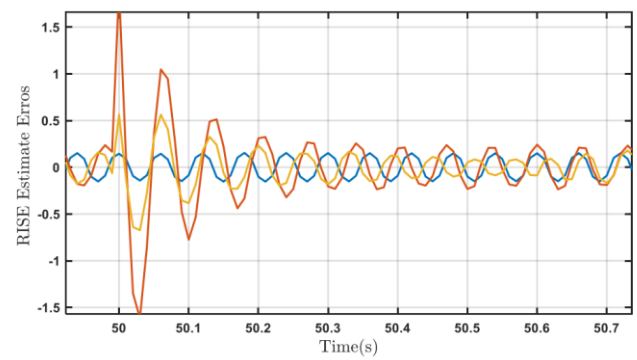

Fig. 16. The RISE Estimation using Actor/Critic algorithm (Case 2).

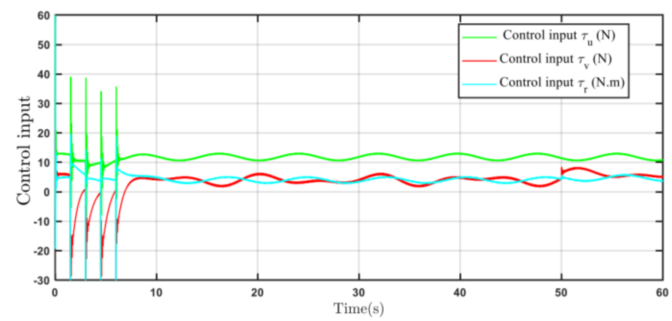

Fig. 17. The Control Inputs Actor/Critic algorithm (Case 2).

V. CONCLUSION

An ARL based optimal control was developed for nonlinear SV in the presence of unknown parameters and disturbances. A transform method was proposed for obtaining the corresponding autonomous tracking error model. This enables us to extend IRL algorithm for solving the optimal tracking control. The stability and optimality were guaranteed by analyzing with Lyapunov stability theory. Simulation results show the suitability of the proposed solution.

REFERENCES

[1] Wang, N., Xie, G., Pan, X., & Su, S. F. (2019). Full-state regulation control of asymmetric underactuated surface vehicles. *IEEE Transactions on Industrial Electronics*, *66*(11), 8741-8750.
[2] Wang, N., Karimi, H. R., Li, H., & Su, S. F. (2019). Accurate trajectory tracking of disturbed surface vehicles: A finite-time control approach. *IEEE/ASME Transactions on Mechatronics*, *24*(3), 1064-1074.
[3] Qin, H., Li, C., Sun, Y., Li, X., Du, Y., & Deng, Z. (2019). Finite-time trajectory tracking control of unmanned surface vessel with error constraints and input saturations. *Journal of the Franklin Institute*.

[4] Park, B. S., Kwon, J. W., & Kim, H. (2017). Neural network-based output feedback control for reference tracking of underactuated surface vessels. *Automatica*, *77*, 353-359

[5] Zheng, Z., Ruan, L., Zhu, M., & Guo, X. (2020). Reinforcement Learning Control for Underactuated Surface Vessel with Output Error Constraints and Uncertainties. *Neurocomputing*.

[6] Van, M. (2019). Adaptive neural integral sliding-mode control for tracking control of fully actuated uncertain surface vessels. *International Journal of Robust and Nonlinear Control*, *29*(5), 1537-1557.

[7] Wen, G., Ge, S. S., Chen, C. P., Tu, F., & Wang, S. (2018). Adaptive Tracking Control of Surface Vessel Using Optimized Backstepping Technique. *IEEE transactions on cybernetics*, *49*(9), 3420-3431.

[8] Wen, G., Chen, C. P., & Ge, S. S. (2020). Simplified Optimized Backstepping Control for a Class of Nonlinear Strict-Feedback Systems With Unknown Dynamic Functions. *IEEE Transactions on Cybernetics*.

[9] Guo, X., Yan, W., & Cui, R. (2019). Integral reinforcement learning-based adaptive NN control for continuous-time nonlinear MIMO systems with unknown control directions. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, *50*(11), 4068-4077.