01 Jan 2023

Continual Learning-Based Optimal Output Tracking of Nonlinear Discrete-Time Systems with Constraints: Application to Safe Cargo Transfer

Behzad Farzanegan

S. (Sarangapani) Jagannathan
Missouri Unviersity of Science and Technology, sarangap@mst.edu

Recommended Citation

B. Farzanegan and S. Jagannathan, "Continual Learning-Based Optimal Output Tracking of Nonlinear Discrete-Time Systems with Constraints: Application to Safe Cargo Transfer," *2023 IEEE Conference on Control Technology and Applications, CCTA 2023*, pp. 73 - 78, Institute of Electrical and Electronics Engineers, Jan 2023.
The definitive version is available at https://doi.org/10.1109/CCTA54093.2023.10253015

2023 IEEE Conference on Control Technology and Applications (CCTA)
August 16-18, 2023. Bridgetown, Barbados

2023 IEEE Conference on Control Technology and Applications (CCTA) | 979-8-3503-3544-6/23/$31.00 ©2023 IEEE | DOI: 10.1109/CCTA54093.2023.10253015

Continual Learning-based Optimal Output Tracking of Nonlinear Discrete-time Systems with Constraints: Application to Safe Cargo Transfer

Behzad Farzanegan[1], *Student Member, IEEE,* and S. Jagannathan[1]

***Abstract*— This paper addresses a novel lifelong learning (LL)-based optimal output tracking control of uncertain nonlinear affine discrete-time systems (DT) with state constraints. First, to deal with optimal tracking and reduce the steady state error, a novel augmented system, including tracking error and its integral value and desired trajectory, is proposed. To guarantee safety, an asymmetric barrier function (BF) is incorporated into the utility function to keep the tracking error in a safe region. Then, an adaptive neural network (NN) observer is employed to estimate the state vector and the control input matrix of the uncertain nonlinear system. Next, an NN-based actor-critic framework is utilized to estimate the optimal control input and the value function by using the estimated state vector and control coefficient matrix. To achieve LL for a multitask environment in order to avoid the catastrophic forgetting issue, the exponential weight velocity attenuation (EWVA) scheme is integrated into the critic update law. Finally, the proposed tracker is applied to a safe cargo/ crew transfer from a large cargo ship to a lighter surface effect ship (SES) in severe sea conditions.**

***Index Terms*— Barrier function, Neural networks, Uncertain nonlinear discrete-time system, Optimal tracking control, State constraints, Lifelong learning, Surface effect ship.**

I. Introduction

In the past decade, researchers have paid considerable attention to optimal tracking control for nonlinear discrete-time (DT) systems [1]–[3]. Conventionally, the objective of the tracking controller design is to minimize a well-defined cost function defined in terms of tracking error and control input in order to achieve an optimal performance. In practice, external disturbances and dynamic uncertainties impact the closed-loop system performance [4], [5]. The optimal adaptive control (OAC) has been studied to obtain the best performance even in the presence of external disturbances and unknown dynamics.

While available literature predominantly attempts regulation by using a variety of ADP techniques [6]–[8], a few have considered optimal tracking for nonlinear DT systems. To deal with the tracking problem, an additional feedforward term in the control policy has been provided through dynamic inversion [9]. In contrast, an augmented formulation has been proposed to eliminate the necessity for a feedforward control term by casting tracking into a regulation problem [1]. In the augmented system, it is assumed the reference trajectory generator converges to zero to guarantee stability. To relax this restriction, a discounted tracking problem has been presented in [1]. However, both the system dynamics inversion and augmented formulation techniques cannot ensure a zero steady-state tracking error [2].

On the other hand, safety assurances for a nonlinear dynamic system is addressed by asserting constraints on the state, input and outputs through Barrier functions (BF) [10]. The BFs enable designers to evaluate forward invariance without explicitly estimating the system reachable set [11]. To address safety, a BF has been added to the cost function for completely known nonlinear systems [12]. However, these studies [12], [13] cannot be applied to DT systems and require the state vector to be measurable.

Besides safety and optimality, in multitask scenarios, lifelong learning (LL) is a critical factor in satisfying robustness. In the case of gradient-based and online learning schemes, forgetting is a common phenomenon. When the number of tasks is not fixed in advance, it is important to ensure that learning new tasks does not result in catastrophic forgetting of previously learned tasks, in order to maintain robustness. The LL method has not been adapted for optimal tracking control problems of nonlinear DT systems to-date [7].

In this paper, a novel optimal tracking scheme is presented for uncertain nonlinear DT systems in affine form by using estimated state vector. First, the original system and its cost function are augmented with the desired trajectory and the integral value of tracking error to relax the steady state error of both internal dynamics and the control coefficient matrix in comparison with [2]. Then, to address the state constraints, an asymmetric BF is added to the utility function including the augmented state vector. The constrained HJB equation is constructed, and two NNs are employed to estimate the optimal control input and the safe value function. Next, since the state vector is unavailable, an adaptive NN-based observer is introduced to identify the nonlinear system dynamics and estimate the state vector by using the system outputs.

Then, an NN-based actor-critic framework is utilized to estimate both the optimal control input and the value function by using the estimated state vector and control coefficient matrix. The temporal difference (TD) error is defined as the difference between the estimated and actual value function by using the estimated state vector to tune the critic weights. The actor NN is tuned by using the control input errors. Apart

[1]B. Farzanegan and S. Jagannathan are with the Dept. of Elec. and Comp. Engg, Missouri University of Science and Technology, Rolla, MO, USA. b.farzanegan@mst.edu and sarangap@mst.edu .

The project or effort undertaken was or is sponsored by the Office of Naval Research Grant N00014-21-1-2232 and Army Research Office Cooperative Agreement W911NF-21-2-0260.

from that, in order to improve the LL capabilities of the critic network, the exponential weight velocity attenuation (EWVA) term is incorporated into the critic update law. Then, the stability of the overall closed-loop system is investigated (not included due to space consideration) by using Lyapunov theory. Finally, the proposed controller is implemented in a Surface Effect Ship (SES) to guarantee cargo transfer in severe sea conditions in a safe and optimal manner.

II. Problem Formulation and Background

First, the class of nonlinear DT systems is defined. Subsequently, the safe optimal control problem formulation and NN observer will be introduced.

A. Class of Nonlinear Discrete-time Systems

Let an affine nonlinear discrete-time system be described by

$$\begin{aligned} \zeta(k+1) &= f(\zeta(k)) + g(\zeta(k))u(k), \\ y(k) &= C\zeta(k), \end{aligned} \tag{1}$$

where $\zeta(k) \in \mathbb{R}^n$ denotes the system state vector, which is not measurable, $u(k) \in \mathbb{R}^m$ represents the control input, and $y(k) \in \mathbb{R}^l$ is the measured output. The matrix C is a known output matrix. The smooth functions $f(.) \in \mathbb{R}^n$ represents unknown internal dynamics, and $g(.) \in \mathbb{R}^{n\times m}$ represents the control coefficient matrix of the system which is considered uncertain but bounded above on a compact set such that $||g(\zeta(k))||_F < g_M$.

Define a command generator function to obtain the bounded desired trajectory as

$$\zeta_d(k+1) = \psi(\zeta_d(k)), \tag{2}$$

where $\zeta_d(k) \in \mathbb{R}^n$ presents the desired trajectory and $\psi(\zeta_d(k))$ is a continuously differentiable function with $\psi(0) = 0$. We assume that the desired trajectory in (2) is reachable. Using (1) and (2), one can define the tracking error as

$$e(k) = \zeta(k) - \zeta_d(k), \tag{3}$$

with dynamics

$$\begin{aligned} e(k+1) &= f(e(k)+\zeta_d(k)) + g(e(k)+\zeta_d(k))u(k) \\ &- \psi(\zeta_d(k)). \end{aligned} \tag{4}$$

To deal with steady state errors, a new state variable $\eta \in \mathbb{R}^n$, is introduced as

$$\eta(t) = \int (\zeta_d(t) - \zeta(t))dt. \tag{5}$$

By employing the Euler's approximation method, the integral value of the tracking error in (5) is discretized as

$$\eta(k+1) = \eta(k) - Te(k), \tag{6}$$

where $\eta(k) \subset \mathbb{R}^n$ and T is the sampling time. Now, by invoking the tracking error dynamics (4) and the reference trajectory (2), one can write the augmented system as

$$\zeta_a(k+1) = f_a(\zeta_a(k)) + g_a(\zeta_a(k))u(k), \tag{7}$$

where $\zeta_a(k) = [e(k)^\top, \zeta_d(k)^\top, \eta(k)^\top]^\top \in \mathbb{R}^{3n}$, $f_a(\zeta_a) := \begin{bmatrix} f(e(k)+\zeta_d(k)) - \psi(\zeta_d(k)) \\ \psi(\zeta_d(k)) \\ \eta(k) - Te(k) \end{bmatrix}$, and $g_a(\zeta_a) := \begin{bmatrix} g(e(k)+\zeta_d(k)), 0, 0 \end{bmatrix}^\top$.

Remark 1: The defined augmented system which comprises the error dynamics, the desired system dynamics, and the integral error term, expressed in (7) is different from [2] because the steady state value of the unknown internal dynamics, i.e., $f(\infty)$ is not required. Also, the augmented system in (7) is not explicitly a function of time since the reference trajectory is generated by the desired system dynamics in (2). Therefore, the value function is only a function of the augmented state, and the stationary condition still hold.

In the next subsection, a novel formulation for designing a safe optimal tracking control scheme is presented.

B. Optimal Control Formulation with State Constraints

To find the safe optimal control strategy, a quadratic value function in the presence of constraints is introduced as

$$\begin{aligned} J(\zeta_a(k)) &= \sum_{j=k}^{\infty} \gamma^{j-k} r(\zeta_a(j), u(j)), \\ s.t.\ \zeta_a(k) &\in \mathcal{C}_s, \end{aligned} \tag{8}$$

where $r(\zeta_a(k), u(k)) = \zeta_a^\top(k) Q \zeta_a(k) + u^\top(k) R u(k)$ is the utility function. The user-defined matrices, $Q \in \mathbb{R}^{n\times n}$ and $R \in \mathbb{R}^{m\times m}$, are symmetric positive definite. The parameter γ denotes a discount factor; $\mathcal{C}_s$ is a safe set defined as

$$\mathcal{C}_s = \{\zeta_a \in \mathcal{X} \subset \mathbb{R}^{3n} | \zeta_{i,min} < \zeta_{a,i} < \zeta_{i,max}, i \in \{1, \ldots, n\}\} \tag{9}$$

where $\zeta_{i,min} < 0$ and $\zeta_{i,max} > 0$ present the lower and upper bounds of the constraint, respectively. Next the following definition is needed.

Definition 1: The function $b(.) : \mathbb{R}^{3n} \to \mathbb{R}$ is a *barrier function* which is continuously differentiable, and positive on the safe set $\mathcal{C}$ and converges to infinity at the boundary $\partial\mathcal{C}$.

Next, a logarithmic barrier function (LBF) satisfying the aforementioned BF characteristic is defined. Then, the BF is integrated with the cost function (8) to ensure the state constraints are met. The designed safe control policy guarantees optimal performance when the trajectories do not violate the safe region.

Now define one such LBF candidate as

$$b_i(\zeta_{a,i}(k); L_i, U_i) = ln\left(\frac{L_i U_i}{(U_i - \zeta_{a,i}(k))(L_i - \zeta_{a,i}(k))}\right), \tag{10}$$

where $L_i < 0$ and $U_i > 0$ denote the lower and upper bounds, respectively. Then, the LBF term is defined as

$$b(\zeta_a(k); L, U) = \mu \sum_{i=1}^{n} b_i(\zeta_{a,i}(k); L_i, U_i) \tag{11}$$

where μ is a positive trade-off factor, $\zeta_a = [\zeta_{a,1}, \ldots, \zeta_{a,3n}]^\top$, $L = [L_1, \ldots, L_n]$ and $U = [U_1, \ldots, U_n]$. The utility function

in (8) can be modified as

$$r_b(\zeta_a(k),u(k)) = \zeta_a^\top(k)Q\zeta_a(k) + u^\top(k)Ru(k) + b(\zeta_a(k);L,U) \tag{12}$$

Then, the modified value function integrating the constraints through LBF is given by

$$J_b(\zeta_a(k)) = \sum_{j=k}^{\infty}\gamma^{j-k}\big(\zeta_a^\top(j)Q\zeta_a(j) + u^\top(j)Ru(j) + b(\zeta_a(j);L,U)\big). \tag{13}$$

Thus, by employing (13), the recursive Bellman equation is obtained as

$$J_b(\zeta_a(k)) = r_b(\zeta_a(k),u(k)) + \gamma J_b(\zeta_a(k+1)). \tag{14}$$

Now, the Hamiltonian function by integrating the LBF can be written as

$$\begin{aligned} H(\zeta_a,J_b,u) &= \gamma J_b\big(f_a(\zeta_a(k)) + g_a(\zeta_a(k))u(k)\big) \\ &- J_b(\zeta_a(k)) + \zeta_a(k)^\top Q\zeta_a(k) + u(k)^\top Ru(k) \\ &+ b(\zeta_a(k);L,U). \end{aligned} \tag{15}$$

The safe optimal cost function is expressed as

$$J_b^\star(\zeta_a(k)) = \min_{u(\zeta_a(k))}\Big(\sum_{j=k}^{\infty}\gamma^{j-k}\big(\zeta_a^\top(j)Q\zeta_a(j) + u^\top(j)Ru(j) + b(\zeta_a(j);L,U)\big)\Big). \tag{16}$$

and satisfies

$$\min_{u(\zeta_a(k))} H(\zeta_a(k),J_b^\star,u(k)) = 0. \tag{17}$$

Solving the stationary condition, $\partial H(\zeta_a,J_b,u)/\partial u(\zeta_a(k)) = 0$, the optimal tracking control policy is derived as

$$u^\star(\zeta_a(k)) = -\frac{\gamma}{2}R^{-1}g_a^\top(\zeta_a(k))\frac{\partial J_b^\star(\zeta_a(k+1))}{\partial\zeta_a(k+1)}. \tag{18}$$

It is worth noting that to generate the optimal control policy in (18), the future state vector $\zeta_a(k+1)$, which is unavailable, is required. In addition, the control coefficient matrix is needed which is also considered unknown and the state vector is not considered measurable. Therefore, an observer is designed next to generate the control coefficient matrix as well as the state vector under the assumption the output is measured. In the following subsection, the NN observer development is given.

C. NN-based Observer

By exploiting results in [14], the system dynamics in (1) can be rewritten as

$$\zeta(k+1) = A\zeta(k) + W_1^\top\Psi_1(\zeta(k))\bar{u}(k) + \bar{\varepsilon} \tag{19}$$

where $W_1 = [w_F\ w_g]^\top$, $\Psi_1(\zeta(k)) = diag([\Psi_F(\zeta(k))\,\Psi_g(\zeta(k))])$, $\bar{u} = [1\ u(k)]^\top$, and $\bar{\varepsilon} = [\varepsilon_F\,\varepsilon_g]\bar{u}(k)$. The matrix A is Schur stable and the pair (A,C) is observable. The following NN observer estimates the unknown dynamics and state vector [14] as

$$\begin{aligned} \hat{\zeta}(k+1) =& A\hat{\zeta}(k) + \hat{W}_1(k)^\top\Psi_1\left(\hat{\zeta}(k)\right)\bar{u}(k) \\ &+ L\left(y(k) - C\hat{\zeta}(k)\right) \\ \hat{y}(k) =& C\hat{\zeta}(k) \end{aligned} \tag{20}$$

where $\hat{x}(k)$, $\hat{y}(k)$, and $\hat{W}_1(k)$ are the estimated system state, output, observer weights, respectively. The matrix $L\in\mathbb{R}^{n\times l}$ denotes the designed observer gain. Invoking (19) and (20), the state estimation error, $\tilde{\zeta}(k) = \zeta(k) - \hat{\zeta}(k)$, is obtained as

$$\tilde{\zeta}(k+1) = A_c\tilde{\zeta}(k) + \tilde{W}_1(k)^\top\Psi_1\left(\hat{\zeta}(k)\right)\bar{u}(k) + \bar{\varepsilon}_{ok}, \tag{21}$$

where $\tilde{W}_1(k) = W_1 - \hat{W}_1(k)$ presents the NN observer weight estimation error. $\bar{\varepsilon}_{ok} = W_1^\top\tilde{\Psi}_1\left(\zeta(k),\hat{\zeta}(k)\right)\bar{u}(k) + \bar{\varepsilon}(k)$ is bounded with $\tilde{\Psi}_1\left(\zeta(k),\hat{\zeta}(k)\right) = \Psi_1(\zeta(k)) - \Psi_1\left(\hat{\zeta}(k)\right)$. The matrix A_c presents the closed-loop matrix defined as $A_c = A - LC$. The observer NN weight matrix is tuned as

$$\begin{aligned} \hat{W}_1(k+1) =& (1-\alpha_I)\hat{W}_1(k) \\ &+ \beta_I\Psi_1\left(\hat{\zeta}(k)\right)\bar{u}(k)\tilde{y}(k+1)^\top l^\top \end{aligned} \tag{22}$$

where $\alpha_I > 0$ and $\beta_I > 0$ are the damping and learning parameters, respectively. $\tilde{y}(k)$ is the output estimation error defined as $\tilde{y}(k) = y(k) - \hat{y}(k)$, and $l\in\mathbb{R}^{n\times l}$ is a user designed matrix. Next, by involving (22), the observer weight estimation error dynamic is obtained as

$$\begin{aligned} \tilde{W}_1(k+1) =& (1-\alpha_I)\tilde{W}_1(k) + \alpha_I W_1 \\ &- \beta_I\Psi_1(k)\bar{u}(k)\tilde{y}(k+1)^\top l^\top \end{aligned} \tag{23}$$

where $\tilde{W}_1(k) = W_1 - \hat{W}_1(k)$.

III. Lifelong Learning-based Safe Optimal Tracking

In this section, a safe optimal tracking control policy is derived for the uncertain nonlinear DT system in (7) with state constraints by using the estimated state vector from the NN observer. First, a critic NN is constructed to estimate the value function. Then, a novel update law integrating the weight consolidation method is proposed for the critic NNs to avoid catastrophic forgetting. Next, an actor NN is employed to approximate the optimal control policy by using the estimated state vector.

A. Temporal Difference Error using Estimated State

The value function in (8) can be represented as

$$J_b(\zeta_a(k)) = W_2^\top\Psi_2(\zeta_a(k)) + \varepsilon_j(\zeta_a(k)) \tag{24}$$

where $W_2\in\mathbb{R}^{N_c}$ presents the critic weights, $\Psi_2\in\mathbb{R}^{N_c}$ denotes the critic activation function, and $\varepsilon_j(\zeta_a(k))$ represents the NN error. The ideal weight W_2 and the NN error ε_j are assumed to be upper bounded by $\|W_2\| \leqslant w_{cM}$ and

$\|\varepsilon_j(k)\| \leqslant \varepsilon_{jM}$, respectively. Moreover, the safe optimal control policy in (18) can be approximated by

$$u(\zeta_a(k)) = W_3^\top \Psi_3(\zeta_a(k)) + \varepsilon_u(\zeta_a(k)), \tag{25}$$

where $\Psi_3 \in \mathbb{R}^{N_a}$ denotes the actor activation function, W_3 is the actor NN weights which is assumed upper bounded $\|W_3\| \leqslant w_{aM}$, and $\varepsilon_u(k)$ is the approximation error and it is assumed $\|\varepsilon_u(k)\| \leqslant \varepsilon_{uM}$. One can assume the gradient of the actor and critic approximation errors are bounded, i.e., $\|\nabla\varepsilon_u\|_F \leqslant \varepsilon'_{uM}$ and $\|\nabla\varepsilon_j\|_F \leqslant \varepsilon'_{jM}$. One can define the estimated value function $\hat{J}_b(\hat{\zeta}_a(k))$ as

$$\hat{J}_b(\hat{\zeta}_a(k)) = \hat{W}_2^\top \Psi_2(\hat{\zeta}_a(k)) \tag{26}$$

where the NN observer state is given by $\hat{\zeta}_a(k) = [(\hat{\zeta}(k) - \zeta_d(k))^\top, \zeta_d(k)^\top, \hat{\eta}(k)^\top]^\top$; $\hat{W}_2^\top$ denotes the estimated critic NN weight. Substituting the estimated value function (26) into (14) results in estimated TDE using observer state vector as

$$\begin{aligned}\mathcal{T}_{DE} = r_b(\hat{\zeta}_a(k-1), u(\hat{\zeta}_a(k-1))) \\ + \hat{W}_2^\top \Delta\Psi_2(\hat{\zeta}_a(k-1)),\end{aligned} \tag{27}$$

where $\mathcal{T}_{DE} \in \mathbb{R}$ is the TDE, and $\Delta\Psi_2(\hat{\zeta}_a(k-1)) = \gamma\Psi_2(\hat{\zeta}_a(k)) - \Psi_2(\hat{\zeta}_a(k-1))$. Notice that the TDE is now a function of the estimated state vector.

To tune the critic NN weights based on TDE, the performance measure is defined as

$$E_{c1} = \frac{1}{2}\mathcal{T}_{DE}(k)^2. \tag{28}$$

To avoid catastrophic forgetting in multi-task scenarios, the performance measure in (28) will be modified by integrating the continual learning term in the following subsection.

B. Continual Learning and Critic NN Update

The EWC is one of the more widely used regularization methods among continual learning approaches [15] which implements the preservation of previously learned information in successive learning by incorporating a regularizer to the loss function as

$$E_c = E_{c1} + \frac{\varrho}{2}||\hat{W}_2 - \hat{W}^\star_{2,\tau_i}||^2_{\mathcal{F}} \tag{29}$$

where $\mathcal{F}$ denotes the Fisher information matrix for the critic NN, ϱ presents the hyperparameter which indicates the connection of the previous tasks with the current one, and $\hat{W}^\star_{2,\tau_i}$ is the NN weight matrix after training to the task τ_i, i.e., the ith task. The term $||\hat{W}_2 - \hat{W}^\star_{2,\tau_i}||^2_{\mathcal{F}}$ denotes $(\hat{W}_2 - \hat{W}^\star_{2,\tau_i})^\top \mathcal{F}(\hat{W}_2 - \hat{W}^\star_{2,\tau_i})$. EWC keeps the most critical weights from deviating far from the consolidated values throughout the learning of the following tasks. In the EWC approach, it is necessary to store $\hat{W}^\star_{2,\tau_i}$ after each task. However, the Exponential Weight Velocity Attenuation (EWVA) approach presented in [16] saves only the important values of the weights. Here, based on the EWVA method, to tune the critic weights, the overall performance measure is defined as

$$E_c = E_{c1} + \frac{1}{2}(\hat{W}_2 - \hat{W}^\star_{2,\tau_i})^\top e^{-\varrho\Omega_{\tau_i}}(\hat{W}_2 - \hat{W}^\star_{2,\tau_i}) \tag{30}$$

where Ω_{τ_i} is the fisher information matrix, however we define a different value which shows the significance of connection in a NN. Therefore, we define Ω_{τ_i} as a diagonal matrix, where the diagonal elements corresponding to each task are defined as follows

$$\Omega_i = \frac{1}{N}\sum_k |\hat{\zeta}_{a,i}(k)\hat{W}_{2,i}| \tag{31}$$

where N is the number of samples in the each task. To minimize significant changes in $\hat{W}_2$ within learning a new task, the regularizer term in (29) is incorporated to the performance function. Therefore, by using the normalized gradient-descent method, the critic weight tuning law is obtained as

$$\begin{aligned}\hat{W}_2(k+1) = \hat{W}_2(k) - \frac{\alpha_J \Delta\Psi_2(\hat{\zeta}_a(k))\mathcal{T}_{DE}(k)}{\Delta\Psi_c^\top(\hat{\zeta}_a(k))\Delta\Psi_2(\hat{\zeta}_a(k)) + 1} \\ - \alpha_J e^{-\varrho\Omega_{\tau_i}}(\hat{W}_2(k) - \hat{W}^\star_{2,\tau_i}),\end{aligned} \tag{32}$$

where α_J is a designed parameter.

Remark 2: Note that computation of the Fisher information matrix, $\mathcal{F}$, is involved in EWC. Instead, EWVA [17] relies on the total absolute weighted input, $\hat{\zeta}_{a,i}(k)$, and a diagonal Fisher information matrix, Ω_{τ_i}, processed using an exponential function. In other words, in the EWVA method, the regularizer has access to a broad range of previous weight information.

C. Actor NN Update and Closed-loop Stability

In this subsection, the actor NN weight tuning law is investigated, and then, the boundedness of the overall closed-loop system is guaranteed through Lyapunov analysis. A simple schematic of the proposed controller is depicted in Fig. 1.

Now, by using (18) and (26), the estimated control input is constructed as

$$\hat{u}(k) = -\frac{\gamma}{2}R^{-1}\hat{g}_a(\hat{\zeta}_a(k))^\top \frac{\partial\Psi_2\left(\hat{\zeta}_a(k+1)\right)^\top}{\partial\hat{\zeta}_a(k+1)}\hat{W}_2\ , \tag{33}$$

where $\hat{g}_a(k) = [\hat{w}_g^T\Psi_g(\hat{\zeta}(k))\,0\,0]^\top$. A feedforward NN is employed to obtain the estimated optimal control policy as

$$\hat{u}(\hat{\zeta}_a(k)) = \hat{W}_3^\top \Psi_3(\hat{\zeta}_a(k)) \tag{34}$$

where $\hat{W}_3$ is the estimated actor weight. The difference between the approximated control input in (34) and the real control input (33) is the control input error presented as

$$\begin{aligned}\tilde{u}(k) = \hat{W}_3^\top \Psi_3\left(\hat{\zeta}_a(k)\right) \\ + \frac{\gamma}{2}R^{-1}\hat{g}_a\left(\hat{\zeta}_a(k)\right)^\top \frac{\partial\Psi_2\left(\hat{\zeta}_a(k+1)\right)^\top}{\partial\hat{\zeta}_a(k+1)}\hat{W}_2.\end{aligned} \tag{35}$$

Since the control input error $\tilde{u}(\hat{\zeta}_a(k))$ is measurable, one can write the actor update law as

$$\hat{W}_3(k+1) = \hat{W}_3(k) - \frac{\alpha_u\Psi_3\left(\hat{\zeta}_a(k)\right)\tilde{u}^\top(k)}{\Psi_3^\top\left(\hat{\zeta}_a(k)\right)\Psi_3\left(\hat{\zeta}_a(k)\right) + 1}, \tag{36}$$

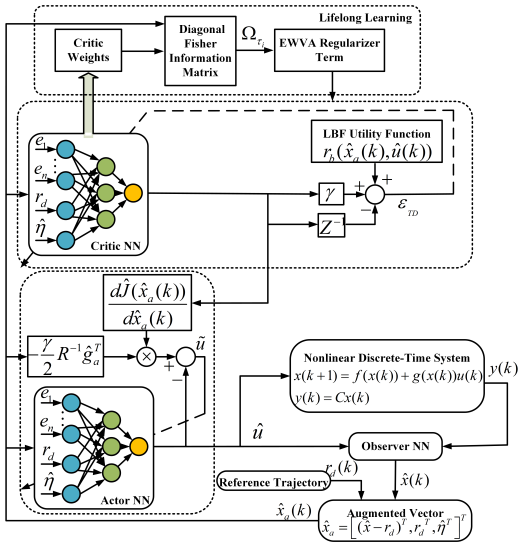

Fig. 1: A simple schematic of the proposed controller.

where α_u is the designed parameter.

Remark 3: It is worth noticing that this effort relaxes the control coefficient matrix by using the observer NN and the need for $\hat{\zeta}(k+1)$ through an actor NN. By using these two NNs, the control input error is computed which is utilized to tune the actor NN weights when the optimal control policy (34) is applied and the next observer state $\hat{\zeta}(k+1)$ becomes available.

IV. SIMULATION RESULTS

In this section, a safe cargo transfer between a heavy cargo ship and a small SES is used as an example to demonstrate the high performance of the proposed controller. The main goal is to cargo transfer at sea and track the desired ramp motion between an SES and a large cargo ship in an optimal and safe fashion. The heave motion of the cargo vessel is quite negligible compared to that of the SES due to the cargo ship's massive structure. Therefore, we only focus on the heave motion tracking of SES. One can write the mathematical model of the heave motion as [4]

$$(m+A_{33})\ddot{\zeta}_3(t)+B_{33}\dot{\zeta}_3(t)+C_{33}\zeta_3(t)-\mathcal{A}_C P_0\mu(t)=F_3^e \tag{37}$$

where ζ_3 is the heave, $\mathcal{A}_C$ represents the cushion area, P_0 denotes the equilibrium pressure, and m is the vessel mass. Besides, A_{33}, B_{33}, and C_{33} present the hydrodynamic, the radiation damping coefficient, the hydrostatic term for the heave, respectively. The air cushion pressure equation is presented as

$$\begin{aligned}\dot{\mu}(t)=&\frac{\gamma(P_0+P_a+\mu(t)P_0)}{P_0\,V_c(t)}\Big(Q_{in}(P_u)\\&-\mathcal{A}_C\dot{\zeta}_3(t)+\dot{V}_0(t)-c_n\sqrt{\frac{2(P_0+\mu(t)P_0)}{\rho_{c0}}}\,u(t)\Big).\end{aligned} \tag{38}$$

where $\mu(t)$ is the air cushion pressure. P_a denotes the atmospheric pressure. $V_c(t)$ and $\rho_c(t)$ present the volume of the cushion and the equilibrium point of the air density. Q_{in} is the cushion air inflow volume rate which is a function

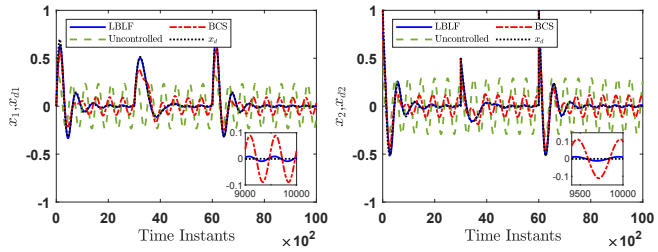

Fig. 2: The system state and reference trajectory.

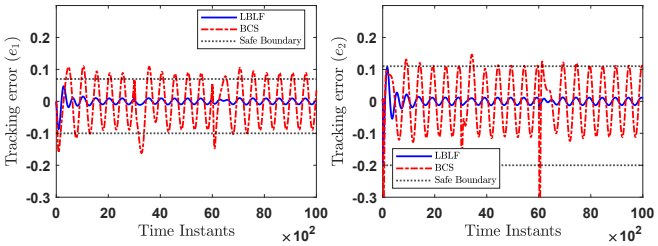

Fig. 3: Performance of our scheme in terms of tracking error.

of the excess air cushion pressure inside the cushion $P_u(t)$. $V_0(t)$ denotes the wave volume pumping disturbance. $c_n = 0.61$ denotes the orifice coefficient.

Therefore, the state space form of (37) and (38) can be written as

$$\dot{x}=f(x)+g(x)u+d(t)=\begin{bmatrix}x_2\\ a_1x_1+a_2x_2+f_3+a_3P_0x_3\\ \frac{\gamma(P_0+P_a+x_3P_0)}{P_0\,V_c(t)}\left(Q_{in}(P_u)-\mathcal{A}_Cx_2+\dot{V}_0-c_n\sqrt{\frac{2(P_0+x_3P_0)}{\rho_{c0}}}u\right)\end{bmatrix},$$

where

$$x=[x_1,x_2,x_3]^\top=[\zeta_3,\dot{\zeta}_3,\mu]^\top,\quad C=[0,1,0]^\top$$
$$a_1=\frac{-C_{33}}{(m+A_{33})},\ a_2=\frac{-B_{33}}{(m+A_{33})},\ a_3=\frac{\mathcal{A}_C}{(m+A_{33})},$$
$$f_3=\frac{F_3^e}{(m+A_{33})},\quad V_c(t)=\mathcal{A}_C(h_0+x_1)-V_0,$$
$$d(t)=\begin{bmatrix}0 & f_3 & V_c(t)\end{bmatrix}^\top$$

To apply the proposed method, the continuous model is discretized by the sample time $T = 0.01s$ and we use the SES parameters and hydrodynamic coefficients from [18] for simulation. To better show the impact of the proposed approach for constrained

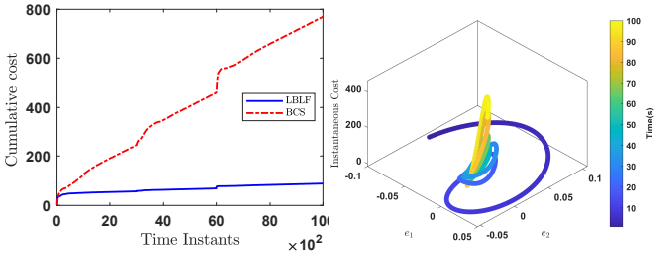

Fig. 4: Cumulative cost and the instantaneous cost comparison without and with continual learning term.

control through LL formulation, we define the reference trajectory as $x_d(k) = e^{(-0.25k)}[sin(k), cos(k) - 1/4sin(k), \frac{-a_1+\frac{a_2}{4}-\frac{15}{16}}{a_3P_0}sin(k) - \frac{a2+0.5}{a_3P_0}cos(k)]^\top$, $0 < k \leq 3000$ and $x_d(k) = e^{(-0.25k)}[sin(0.5k), 0.5cos(2k) - 1/4sin(0.5k), \frac{-a_1+\frac{a_2}{4}-\frac{3}{16}}{a_3P_0}sin(k) - \frac{0.5a2+0.25}{a_3P_0}cos(k)]^\top$, $3000 < k \leq 6000$ and $x_d(k) = e^{(-0.25k/)}[sin(k), cos(k) - 1/4sin(k), \frac{-a_1+\frac{a_2}{4}-\frac{15}{16}}{a_3P_0}sin(k) - \frac{a2+0.5}{a_3P_0}cos(k)]^\top$, $6000 < k \leq 10000$.

To apply the tracking error constraints on the heave and the heave rate error, the asymmetric LBF functions in (10) are chosen as $U_1 = 0.07$, $U_2 = 0.11$, $L_1 = -0.1$, and $L_2 = -0.2$. The penalty values of the augmented reward function are selected ad $Q = 5I$ and $R = 0.01$. The initial values for the state set as $x_0 = \begin{bmatrix} 0 & 1 & 0 \end{bmatrix}^T$. To verify the effectiveness of the proposed technique, we select a NN with 70 neurons for the critic NN. The output layer with polynomial activation functions are selected. The designed parameters are taken as $\gamma = 0.5$, $\alpha_u = 0.04$, and $\alpha_J = 0.1$. The NN weight initialization is chosen randomly selected.

Fig. 2 illustrates the state and reference trajectories using the lifelong barrier Lyapunov function (LBLF) based learning scheme, and Boarding Control System (BCS) in [19]. The case of without controller is also included. As can be seen, the system state vector and the reference trajectory are close to each other for the LBLF technique in all three tasks.

Moreover, the tracking errors are shown in Fig. 3 which shows the convergence of tracking error to near zero. Indeed, the proposed method helps in generating optimal control input and enables both faster convergence of tracking error near zero and NN weights. It is obvious that the tracking errors remain within the safe region by using the LBLF method for all three scenarios.

In Fig. 4, the cumulative cost and the instantaneous cost comparison are given. The cumulative cost plot shows that the optimal solution is not only attained but also it is lower than the existing method. The instantaneous cost converges to near zero. Note the color bar on this plot shows the time from zero to 100s as the 3D plot does not have the time.

V. CONCLUSIONS

In this paper, a LL-based optimal tracking control was presented for uncertain nonlinear discrete-time systems with constraints. Solving the tracking problem through an augmented system with the integral term considerably improved the tracking performance. Then, a BLF was integrated into the value function to deal with state constraints. It was shown that casting constraint problem with adding BLF to the value function ensures the safe region could not be violated by the control policy. Besides, adding a quadratic penalty function to the performance measure improved the LL functionality of the critic network to avoid catastrophic forgetting in multitask scenarios. Finally, the proposed method has been implemented in a safely cargo transfer example to control the vertical position of an SES in severe sea conditions. The simulation results have confirmed the validity of the proposed safe LL-based optimal tracking control.

REFERENCES

[1] B. Kiumarsi and F. L. Lewis, "Actor–critic-based optimal tracking for partially unknown nonlinear discrete-time systems," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 26, no. 1, pp. 140–151, 2014.

[2] I. Sanusi, A. Mills, T. Dodd, and G. Konstantopoulos, "Online optimal and adaptive integral tracking control for varying discrete-time systems using reinforcement learning," *International Journal of Adaptive Control and Signal Processing*, vol. 34, no. 8, pp. 971–991, 2020.

[3] D. Wang, M. Ha, and L. Cheng, "Neuro-optimal trajectory tracking with value iteration of discrete-time nonlinear dynamics," *IEEE Transactions on Neural Networks and Learning Systems*, 2021.

[4] E. Esmailian, B. Farzanegan, M. B. Menhaj, and H. Ghassemi, "A robust neuro-based adaptive control system design for a surface effect ship with uncertain dynamics and input saturation to cargo transfer at sea," *Applied Ocean Research*, vol. 74, pp. 59–68, 2018.

[5] S. Ameli and O. M. Anubi, "Hierarchical robust control for variable-pitch wind turbine with actuator faults," *International Journal of Robust and Nonlinear Control*, vol. 32, no. 12, pp. 7039–7056, 2022.

[6] J. Xu, J. Wang, J. Rao, Y. Zhong, and H. Wang, "Adaptive dynamic programming for optimal control of discrete-time nonlinear system with state constraints based on control barrier function," *International Journal of Robust and Nonlinear Control*, vol. 32, no. 6, pp. 3408–3424, 2022.

[7] R. Moghadam, B. Farzanegan, S. Jagannathan, and P. Natarajan, "Optimal adaptive output regulation of uncertain nonlinear discrete-time systems using lifelong concurrent learning," in *2022 IEEE 61st Conference on Decision and Control (CDC)*, pp. 2005–2010, IEEE, 2022.

[8] X. Zhong, H. He, H. Zhang, and Z. Wang, "Optimal control for unknown discrete-time nonlinear markov jump systems using adaptive dynamic programming," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 25, no. 12, pp. 2141–2155, 2014.

[9] H. Zhang, Q. Wei, and Y. Luo, "A novel infinite-time optimal tracking control scheme for a class of discrete-time nonlinear systems via the greedy hdp iteration algorithm," *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, vol. 38, no. 4, pp. 937–942, 2008.

[10] A. Taylor, A. Singletary, Y. Yue, and A. Ames, "Learning for safety-critical control with control barrier functions," in *Learning for Dynamics and Control*, pp. 708–717, PMLR, 2020.

[11] A. D. Ames, X. Xu, J. W. Grizzle, and P. Tabuada, "Control barrier function based quadratic programs for safety critical systems," *IEEE Transactions on Automatic Control*, vol. 62, no. 8, pp. 3861–3876, 2016.

[12] Z. Marvi and B. Kiumarsi, "Safe reinforcement learning: A control barrier function optimization approach," *International Journal of Robust and Nonlinear Control*, vol. 31, no. 6, pp. 1923–1940, 2021.

[13] Y. Yang, Y. Yin, W. He, K. G. Vamvoudakis, H. Modares, and D. C. Wunsch, "Safety-aware reinforcement learning framework with an actor-critic-barrier structure," in *2019 American Control Conference (ACC)*, pp. 2352–2358, IEEE, 2019.

[14] Q. Zhao, H. Xu, and S. Jagannathan, "Near optimal output feedback control of nonlinear discrete-time systems based on reinforcement neural network learning," *IEEE/CAA Journal of Automatica Sinica*, vol. 1, no. 4, pp. 372–384, 2014.

[15] B. Maschler, H. Vietz, N. Jazdi, and M. Weyrich, "Continual learning of fault prediction for turbofan engines using deep learning with elastic weight consolidation," in *2020 25th IEEE international conference on emerging technologies and factory automation (ETFA)*, vol. 1, pp. 959–966, IEEE, 2020.

[16] A. Kutalev and A. Lapina, "Empirical investigations on wva structural issues," *arXiv preprint arXiv:2208.05791*, 2022.

[17] A. Kutalev, "Natural way to overcome the catastrophic forgetting in neural networks," *arXiv preprint arXiv:2005.07107*, 2020.

[18] B. Farzanegan, E. Esmailian, and M. B. Menhaj, "A data-driven method for optimal control of ship motions for safe crew transfer to offshore wind turbines," *Applied Ocean Research*, vol. 90, p. 101847, 2019.

[19] Ø. F. Auestad, J. T. Gravdahl, T. Perez, A. J. Sørensen, and T. H. Espeland, "Boarding control system for improved accessibility to offshore wind turbines: Full-scale testing," *Control Engineering Practice*, vol. 45, pp. 207–218, 2015.