Observer-Based Adaptive Robust Actor-Critic Learning Saturated *PID* Controller for a Class of Euler-Lagrange Robotic Systems with Guaranteed Performance: Theory and Practice

Omid Elhaki, Khoshnam Shojaei, Abbas Chatraei, and Allahyar Montazeri

***Abstract*—This article addresses the output-feedback reinforcement learning-based saturated proportional-integral-derivative (*PID*) control design for fully-actuated Euler–Lagrange (EL) systems which are uncertain subject to actuator saturation with prescribed performance. It is assumed that the actuator input nonlinearity, uncertain nonlinearities and unmeasurable external disturbances have a significant impact on the system. The presence of actuator saturation and complex uncertainties may inevitably give rise to the breakdown of the EL control system. The lack of prior knowledge of the system dynamics renders the presented technique to achieve a robust prescribed tracking performance without using velocity sensors. To conquer mentioned obstacles, a novel reinforcement learning saturated *PID* controller, which isn't dependent on the system's dynamics and only requires measurable output signals is designed via actor-critic structure to deeply estimate and compensate complex unknowns. An adaptive robust controller is used to reduce external disturbances effects adaptively. The prescribed performance funnel control way is considered to guarantee predetermined output constraints. The high-gain observer (HGO) is used to estimate velocities and derivatives free of system dynamics, and generalized saturation functions are utilized to efficiently decrease actuator saturation danger. It is proved that suggested technique ensures a robust prescribed performance with input constraints in the absence of velocity sensors and the existence of considerable complicated model uncertainties. A SGUUB (semi-global uniform ultimate boundedness) stability for tracking deviation errors and state estimation deviation is ensured through a Lyapunov stability study. Finally, experimental results on a real robotic arm is carried out to further demonstrate the effectiveness of all theoretical findings.**

***Index Terms*—Reinforcement learning, actor-critic neural network, performance bound, high-gain observer, generalized saturation function.**

I. Introduction

MULTI-LAYER neural network (NN) reinforcement learning (MLNNRL) solves the dimensional explosion problem by estimating the critic function and nonlinearities with multi-layer neural networks (MLNNs). In fact, MLNNRL method is a compound of MLNNs, which are able to deal with the curse of dimensionality problem [1], and reinforcement learning (RL) [2, 3]. RL-based tactics are used in various fields such as health-care and medicine [4], operations research and economics, robotics and autonomous systems [3], and optimal control design problems [5], due to their wide potential benefits on decision-making and independent learning, which can strengthen the adaptability and robustness of the systems. Actor–critic (AC) learning framework [6, 7] has been observed as a major structure for RL that unifies the value-based (like Q-learning) as well as action-based (suchlike gradients of policy) RL methods to improve the learning performance of the system. RL is mainly made up of two NNs, i.e. the critic-NN (CNN) and the actor-NN (ANN). The CNN is exerted to assess the ongoing system execution and manage the subsequent step of the ANN action to enhance the performance of learning for the system. Thus, AC learning design has been viewed as a key basis for RL in designing control systems for EL dynamic systems [8, 9]. Therefore, the fusion of MLNNs and RL has become a sensible solution for complex decision-making problems.

Currently, some scholars have researched the potential of intelligent control schemes to achieve high control performance for EL systems while others are using RL to improve autonomous operation of such dynamic systems. In [10], the output-feedback Q-learning problem was investigated for a finite-horizon linear-quadratic (LQ) optimal control. However, this method suffers from redefining Q-factors. Reference [11] has proposed a RL-based controller for the flexible arms. The output-feedback regulator problem has been investigated in [12] by presenting a new linear quadratic regulation (LQR) Q-function. Nevertheless, the solution for LQR must be provided with some data samples and it should be viewed as an important computational burden. From the presented review, it is clear that most of the RL methods face a considerable amount of the computational load and the curse of dimensionality due to the requirement of finding the solution to the Hamilton–Jacobi–Bellman (HJB) optimization principle, other optimization standards, and iterative algorithms. Actually, analytical results for the equation of HJB are mostly infeasible to be obtained and some optimal methods are devoted to solve the HJB problem [13]. For the particular cases

O. Elhaki, K. Shojaei and A. Chatraei are with Department of Electrical Engineering, Najafabad Branch, Islamic Azad University, Najafabad, Iran, e-mail: (omid.elhaki@gmail.com).

K. Shojaei and A. Chatraei are with Digital Processing and Machine Vision Research Center, Najafabad Branch, Islamic Azad University, Najafabad, Iran, e-mail: (khoshnam.shojaee@gmail.com and abbas.chatraei@gmail.com).

A. Montazeri is with School of Engineering, Lancaster University, Gillow Avenue, Lancaster, England, LA14YW, e-mail: (a.montazeri@lancaster.ac.uk). The work is supported in part by the ASUNDER - Adaptable Semiautonomous Underwater Decommissioning Sample Retrieval Robot, EP/V027379/1, funded by EPSRC and the EPSRC Centre for Doctoral Training, grant number EP/S022295/1.

of systems that are linear, the equation of HJB decreases to an algebraic Riccati equation (ARE). Traditional approaches to find an answer for the ARE are offline and need the complete knowledge of the system dynamics that may be impossible in practice. AREs are arduous to resolve because of their non-linear character. So, iterative approaches like value and policy iteration methods are employed to discover their solutions [14]. In the meantime, these solutions model-based and need a flawless system dynamics knowledge. Thus, finding an alternative method that can be fused with RL algorithms independent from the requirement of the solutions of the HJB, ARE, LQR, and other similar methods is of paramount importance. To deal with the aforesaid problem, in this paper, the CNN is used to estimate a strategic long-term cost function (LTCF), and the output of ANN is used to estimate unknown nonlinearities and to minimize the strategic LTCF [15].

In most of the present EL control systems, the full feedback of states is considered which is usually impossible in many practical applications. Thus, the output-feedback (OFB) approaches are more preferable due to the lack of velocity sensors for the EL systems. Noghreian et al. [16] have designed a robust OFB controller (OFBC) for robot arms, and it is assumed that the dynamics of the robot is linearly parameterized. However, this assumption may not be satisfied in all EL systems in different situations, such as the input saturation nonlinearity in which the controller may not be able to compensate the effects of the nonlinear-in-parameter (NLIP) unknown terms, which can deteriorate the controller's performance. Reference [17] has studied OFBC problem for robot manipulators by extended state observers. However, the complete dynamic model wasn't considered. In addition, many RL-based controllers didn't consider the OFBC problem due to the mathematical challenges to prove the closed-loop system stability [18]. Moreover, designing an OFBC which doesn't need model parameters and can be designed only on the basis of measurable outputs of the system has a more chance for the successful practical implementation rather than dynamic model-based OFBCs. Therefore, integrating a model-free OFB problem with a MLNNRL-based controller that can learn and compensate nonlinearities is deeply interesting and cumbersome.

Another important control viewpoint for the EL systems concerns with the actuator saturation phenomenon. It is well-known that the actuator saturation may lead to an undesirable transient performance, physical destruction and mechanical breakdown of the actuators, and the closed-loop system breakage. To cope with this serious problem, the literature has witnessed different efforts. In [19], a saturated tracking controller for n-link robot arms was designed, and a finite-time controller for rigid robots with input saturation has been developed in [20]. However, such valuable designs need all states for a practical implementation and they don't benefit from MLNNRL advantages.

The next viewpoint is related to the performance of tracking quality. In the existence of disturbances and NLIP uncertainties that are usual in the real-world systems, the above-mentioned methods obtain the tracking error convergence to an unknown-size residual set. Thus, the tracking performance with specified characteristics is needed in many important tracking missions for the EL systems. An approach named prescribed performance control (PPC) [21] enables users to predetermine some required performance specifications, including maximum and minimum peaking, velocity of convergence and steady-state error exactness by constraining tracking errors to the performance bounds. Zhang et al. [22] extended PPC-based design subjected to actuator faults for the EL systems. In [23], a practically finite-time PPC scheme has been developed for EL systems. A practical controller with PPC has been proposed in [24], and the distributed formation problem of multi EL systems that their control directions are unknown was studied in [25] with a guaranteed performance. However, such works don't propose any OFB-MLNNRL-based control design.

Motivated by the above literature review and to improve our previous works [26, 27] that were developed for particular cases of the EL systems, a novel output-feedback adaptive robust actor-critic learning saturated *PID* controller is proposed for a wider group of the EL systems with a guaranteed performance in the present article. The major contributions could be summed up as follows: 1) In contrast to [10–13], a novel MLNNRL approach is suggested in this article which is independent of HJB, ARE, LQR and iterative methods. Instead, a novel LTCF is considered and minimized by the proposed scheme. The nonlinear equations of MLNNs are used and expanded in an innovative way to introduce the suggested MLNNRL with a rigorous stability analysis. The proposed MLNNRL method diminishes the computational burden of the RL-based controllers and just relies on the output signals. 2) Unlike the existing results that address OFB problem for the EL systems [16, 17], a HGO, which is only dependent on measurable output signals, is efficiently employed to estimate derivatives and velocities of the EL system without the need of system's model information considering NLIP terms, unmodeled dynamics and a complete model for the EL systems. 3) Different from the existing saturated controllers [19, 20] for the EL systems, generalized saturation functions (GSFs) are effectively adapted and utilized here to minimize the possibility of the actuator saturation in fusion with MLNNRL and OFB methods which can learn and counteract the actuator saturation nonlinearity deeply while preventing any unwanted peaking. 4) In contrast to many previous works including [28–30] which usually rely on proportional controllers, a saturated *PID* controller is heuristically developed in this article. Its essential property is that the performance-tuning role of the proportional, integral and derivative gains is not restricted by using GSFs to produce smaller control efforts while preserving an excellent performance. 5) Prescribed funnel characteristics for the tracking errors are ensured in advance. 6) An efficient adaptive upper-bounding robust controller is proposed to deal with the unknown time-varying external disturbances and networks approximation errors. 7) Some auxiliary control terms are designed innovatively to increase the closed-loop performance and to facilitate the proof of stability. 8) The suggested control approach is experimentally assessed on SCARA IBM7547 robot manipulator via an Arduino Due control hardware, and the computational burden is evaluated to prove that the proposed controller is computationally effective in practical applications.

Although intelligent controllers are reported in [29, 30] and very novel RL-based controllers have been proposed in [5, 8, 31–38], they do not cover simultaneously all the above-stated novelties in items 1-7. Besides, in the recent literature [39–41], output feedback reinforcement-learning-based control methods have been developed. However, these controllers cannot ensure a prescribed performance with the actuators' saturation prevention. Moreover, in contrast to reference [42], the suggested control system is a full output feedback *PID* sliding-mode-type controller that uses GSFs to reduce the risk of actuators saturation more effectively and it is designed for a wide group of fully-actuated EL systems with practical implementation in mind. To sum up this expression, the robust OFB-MLNNRL-based saturated *PID* control technique with a funnel performance has not been studied for fully actuated EL systems in the literature to the best of our knowledge. It is the first time to propose an MLNNRL-based saturated *PID* OFBC for EL systems which doesn't need any information on model structure and measurements of higher-order output derivatives and doesn't require a solution for an HJB equation. This strongly contributes to the existing literature. Additionally, an integration of contributions 1-7 is crucial to deal with previously mentioned deficiencies of the present literature. This integration is highly challenging due to the existence of complex mathematical relationships between the mentioned control methods. This causes serious complexities and obstacles in the design procedure, especially in the proof of stability. In this paper, this integration is successfully handled, and the stability analysis is studied meticulously. Besides, the practical implementation of the suggested controller is another challenge stated by item 8 which is conducted with success in this paper.

The article is arranged as follows. The next section provides some prefaces. The main theoretical findings are presented in Section III. Experiments are carried out in Section IV and in Section V, conclusions are provided.

II. Preliminaries

A. System Description

The following set of equations can represent the general dynamic model of a fully-actuated EL system:

$$\begin{aligned} \dot{q} &= J_1(q)\nu, \\ M_1\dot{\nu} &= -C_1(q,\nu)\nu - D_1\nu - G(q) - H(\nu) + B(q)\tau_s + \tau_d, \end{aligned} \tag{1}$$

at which $q = [q_1, \ldots, q_n]^T$ is the generalized coordinates vector, $J_1(q) \in \Re^{n\times n}$ stands for a rotation matrix that is orthogonal, $\nu \in \Re^n$ shows the velocity vector, $M_1 \in \Re^{n\times n}$ is the inertia matrix such that $M_1(q) = M_1^T(q) > 0$, $C_1(q,\nu) \in \Re^{n\times n}$ denotes the centripetal and Coriolis forces matrix that has the skew-symmetric property, $D_1 \in \Re^{n\times n}$ indicates the damping matrix including viscous friction coefficients which is strictly positive, $G(q) \in \Re^n$ signifies the vector of gravity effects, $H(\nu) \in \Re^n$ shows the unmodeled dynamics vector and NLIP nonlinearities as a function of the velocities, $B(q) \in \Re^{n\times n}$ is an unknown matrix of input transformation, $\tau_s = [\tau_{1s}, \ldots, \tau_{ns}]^T$ indicates the saturated input vector, and $\tau_d \in \Re^n$ is the vector of bounded disturbances such that $|\tau_{di}| \leq B_{di}, \forall i = 1, \ldots, n$, where $B_{di} \in \Re^+$. The control input τ_s is characterized by

$$\tau_{is} = \begin{cases} \mathrm{sign}(\tau_i)\chi_{Mi} & , |\tau_i| \geq \chi_{Mi}, \forall i = 1, \ldots, n, \\ m_i\tau_i & , |\tau_i| < \chi_{Mi}, \end{cases} \tag{2}$$

where τ_{is}, τ_i, and χ_{Mi} are the saturated control signals, unsaturated control forces, and the actuators boundary. m_i shows a ratio among τ_{is} and τ_i. Hence, the nonlinearity of saturation $d_{is}(\tau_i) = \tau_{is} - \tau_i$, which is a NLIP function and cannot be applied by the actuators, is defined as

$$d_{is} = \begin{cases} \mathrm{sign}(\tau_i)\chi_{Mi} - \tau_i & , |\tau_i| \geq \chi_{Mi}, \\ (m_i - 1)\tau_i & , |\tau_i| < \chi_{Mi}. \end{cases} \tag{3}$$

B. Control Objectives and Conditions

The control objectives is to design a controller for fully-actuated EL systems so that tracking errors converge asymptotically to a small area comprising the origin under following conditions: *C1:* There are no sensors to measure output derivatives including velocities and accelerations, and only output is accessible for the feedback. *C2:* The parametric and non-parametric uncertainties should have minimum negative effects on tracking performance. *C3:* Actuator saturation should be avoided to keep desirable tracking performance and prevent mechanical failure of actuators. *C4:* An improved type of RL (providing lower computational load compared with conventional RL methods) should be designed to render a more accurate estimation of nonlinearities. *C5:* The controller robustness should be preserved against all types of bounded external disturbances. *C6:* The tracking errors should approach to vicinity of the origin exponentially with desired transient and steady-state performance features. *C7:* The performance of the suggested controller should be evaluated on a real EL system by assessing practicability and feasibility of the design.

C. Error Dynamics and PPC Transformation

In completion of the prescribed performance control, the tracking errors (i.e. $e = q - q_d$) must evolve strictly within a funnel set [24, 43]. The function $\eta_i : \Re^+ \to \Re^+$, which is limited and smooth, represents the performance bound, provided that η_i remains decreasing and $\lim_{t\to\infty} \eta_i(t) = \eta_{i\infty}$ [21]. Then, the prescribed performance property will be assured if the condition $\eta_{li}(t) \leq e_i(t) \leq \eta_{ui}(t), i = 1, \ldots, n$, stays true as $t \to \infty$. Here, e_i is the ith element of the vector e, η_{ui} and η_{li} are the predefined upper and lower bounds for e_i, respectively, and

$$\eta_i(t) := (\eta_{i0} - \eta_{i\infty})\exp(-a_i t) + \eta_{i\infty}, \tag{4}$$

where $a_i \in \Re^+$ is the convergence rate, $\eta_{i0} \gg \eta_{i\infty} \in \Re^+$ are the performance bound variables, in which $\eta_{i\infty}$ is a small value. Then, the prescribed performance limits could be guaranteed by setting $\eta_{li} = -\alpha_i\eta_i$, $\eta_{ui} = \beta_i\eta_i$, $\forall i = 1, \ldots, n$, in which $\beta_i, \alpha_i \in \Re^+$ are set by the user.

Assumption 1 [21]. The inequality $\eta_{li}(0) \le e_i(0) \le \eta_{ui}(0)$ must be satisfied to ensure PPC in the paper.
Next, the following transformation for PPC is utilized:

$$e_i = S_i(\epsilon_{e_i}) = \frac{\eta_{ui} - \eta_{li}}{\pi} \arctan(\epsilon_{e_i}) + \frac{\eta_{ui} + \eta_{li}}{2}. \quad (5)$$

From (5), the transformed error $\epsilon_{e_i}(t)$ is given by

$$\epsilon_{e_i} = \tan\Big(\frac{\pi(2e_i - \eta_{li} - \eta_{ui})}{2(\eta_{ui} - \eta_{li})}\Big). \quad (6)$$

One may verify that (5) has the following features:

$$\lim_{\epsilon_{e_i} \to +\infty} S_i(\epsilon_{e_i}) = \eta_{ui}, \quad \lim_{\epsilon_{e_i} \to -\infty} S_i(\epsilon_{e_i}) = \eta_{li}, \quad (7)$$

where (7) shows that tracking errors converge to performance bounds and tend to origin vicinity provided that Assumption 1 holds and $\epsilon_{e_i} \in \mathcal{L}_\infty$. The time derivative of (6) leads to

$$\dot{\epsilon}_{e_i} = (\partial \epsilon_{e_i} / \partial e_i)\dot{e}_i + \Phi_i, \quad (8)$$

where $\Phi_i = \frac{\partial \epsilon_{e_i}}{\partial \eta_{ui}} \dot{\eta}_{ui} + \frac{\partial \epsilon_{e_i}}{\partial \eta_{li}} \dot{\eta}_{li}$. Using (5) and (6) yields:

$$\frac{\partial \epsilon_{e_i}}{\partial e_i} = \frac{\pi}{\tilde{\eta}_i} \cos^{-2}\Big(\frac{\pi}{2} \times \frac{2e_i - \eta_{ui} - \eta_{li}}{\tilde{\eta}_i}\Big) > 0, \quad (9)$$

$$\frac{\partial e_i}{\partial \epsilon_{e_i}} = \frac{\tilde{\eta}_i}{\pi(1 + \epsilon_{e_i}^2)} > 0, \quad (10)$$

in which $\tilde{\eta}_i = \eta_{ui} - \eta_{li}$. By substituting $\dot{e} = \dot{q} - \dot{q}_d$ in (8) and using (1), one has

$$\dot{\epsilon}_e = R\nu + \hbar, \quad (11)$$

in which $R = TJ_1$, $\hbar = \Phi - T\dot{q}_d$ is a function of the performance bounds and transformed errors $\big($i.e. $\hbar(\epsilon_e, \eta_u, \eta_l, \dot{\eta}_u, \dot{\eta}_l, q_d, \dot{q}_d)\big)$, $T = diag[{}^{\partial \epsilon_{e_1}}/_{\partial e_1}, \ldots, {}^{\partial \epsilon_{e_n}}/_{\partial e_n}]$, $\Phi = [\Phi_1, \ldots, \Phi_n]^T$, $\eta_u = [\eta_{u1}, \ldots, \eta_{un}]^T$, $\eta_l = [\eta_{l1}, \ldots, \eta_{ln}]^T$, $\dot{\eta}_u = [\dot{\eta}_{u1}, \ldots, \dot{\eta}_{un}]^T$, and $\dot{\eta}_l = [\dot{\eta}_{l1}, \ldots, \dot{\eta}_{ln}]^T$.

D. Mathematical Preliminaries

Since we assume that output derivatives are unavailable, the HGO is employed here.

Lemma 1 [1]. Assume the system output signals vector up to $n-1$ derivatives is finite such that $\|y^{(k)}\| \le B_k$, in which $B_k \in \Re^+$. Then, the next system is presented:

$$\begin{aligned} \epsilon\dot{\ell}_k &= \ell_{(k+1)}, \ k = 1, \ldots, n-1, \\ \epsilon\dot{\ell}_n &= -\lambda_1 \ell_n - \lambda_2 \ell_{(n-1)} - \cdots - \lambda_{(n-1)}\ell_2 - \ell_1 + y, \quad (12) \end{aligned}$$

in which ℓ_k are the observer states, $\epsilon \in \Re^+$ is chosen by designer, and variables $\lambda_1, \ldots, \lambda_{(n-1)}$ must be specified so that the term $\rho^n + \lambda_1 \rho^{n-1} + \cdots + \lambda_{(n-1)}\rho + 1$ becomes Hurwitz. Thus, the next items are true. (i) $\ell_{(j+1)}/\epsilon^j - y^{(j)} = -\epsilon\kappa^{(j+1)}$, $j = 0, 1, \ldots, n-1$, in which $\kappa = \ell_n + \lambda_1 \ell_{(n-1)} + \cdots + \lambda_{(n-1)}\ell_1$, the j-th derivative of κ is $\kappa^{(j)}$. By invoking [1], $\ell_{(j+1)}/\epsilon^j$ inclines to $y^{(j)}$ with a limited small error when $y(t)$ and jth derivatives of $y(t)$ are limited; (ii) there exist $t_1, G_j \in \Re^+$ such that $\forall t > t_1$, the term $\|\kappa^{(j)}\| \le G_j$ is true.

Definition 1. The function $\varpi_i : \Re \to \Re : \xi \to \varpi_i(\xi)$ with a bound $M_i \in \Re^+$ is a GSF, if ϖ_i is locally Lipschitz, non-decreasing, and verifies (i) $\xi\varpi_i(\xi) > 0, \forall \xi \ne 0$; and (ii) $|\varpi_i(\xi)| \le M_i, \forall \xi \in \Re$.

Lemma 2 [44]. Allow $\varpi_i : \Re \to \Re : \xi \to \varpi_i(\xi)$ with bound M_i be a strictly increasing continuously GSF, $k_1 \in \Re^+$ is a constant and $\varrho_i : \xi \to d\varpi_i/d\xi$. Then, we have

(i) $\varrho_i(\xi) \in \Re^+$ is bounded in a way that there is a constant $\varrho_{iM} \in (0, \infty)$ so that $0 < \varrho_i(\xi) \le \varrho_{iM}, \forall \xi \in \Re$,
(ii) $\varpi_i^2(k_1\xi)/(2k_1\varrho_{iM}) \le \int_0^\xi \varpi_i(k_1 r)dr \le k_1 \varrho_{iM}\xi^2/2$,
(iii) $\int_0^\xi \varpi_i(k_1 r)dr > 0, \forall \xi \ne 0$,
(iv) $\int_0^\xi \varpi_i(k_1 r)dr \to \infty$ as $|\xi| \to \infty$,
(v) $\varpi_i^2(\xi) \le \varpi_i^2(k_1\xi)$, $\forall \xi \in \Re, k_1 \ge 1$.

Lemma 3 [45]. $ab \le \frac{a^2}{2\epsilon} + \frac{\epsilon b^2}{2}$, $\forall a, b \in \Re$, and $\forall \epsilon \in \Re^+$.

Lemma 4 [45]. If function $f(y,t) : \Re^n \to \Re^n$ is a continuous differentiable function in $\mathcal{D} \subset \Re^n$, there exists a Lipschitz parameter $l \in \Re^+$ such that $\|f(y,t) - f(\hat{y},t)\| \le l\|y - \hat{y}\|$.

Lemma 5 [46]. There exists a saturation functions group $\varpi_h := [\varpi_{h1}, \ldots, \varpi_{hn}]^T$ such that $|\varpi_{hi}| \le M_{hi}$, where M_{hi} is an upper bound so that $M_{hi} \in \Re^+$ and $M_{hi} < \infty$, and fulfills the inequality $0 \le |\chi| - \chi\varpi_{hi}(\chi/a) \le 0.2785a, \forall a \in \Re^+, \chi \in \Re$.

Lemma 6 [45]. $\lambda_{\min}\{N\}\|w\|^2 \le w^T N w \le \lambda_{\max}\{N\}\|w\|^2$, $\forall w \in \Re^n$ and $\forall N = N^T \in \Re^{n \times n}$.

III. MLNNRL-Based Saturated *PID* Controller Design

In this section, main results of this paper are presented. At first, the concept of MLNNs is given in Section A which is used to design MLNNRL-based controller in the next sections. The open-loop error dynamics, saturated control and observer design are provided in Section B. Sections C and D respectively introduce the ANN and CNN components of the proposed RL whose learning rules and proposed controller are given in Section E. Some gain tuning instructions are depicted in Section F, and finally, Section III is warped up through a stability analysis of the proposed controller in Section G.

A. Multi-layer Neural Network

MLNNs have been efficiently used to estimate the unknown NLIP parts of EL systems as follows $\forall i = 1, \ldots, N_o$ [1, 47]:

$$y_i = \sum_{j=1}^{N_h} \Big[w_{ij}\bar{\sigma}\Big(\sum_{k=1}^{N_i} v_{jk}x_k + \theta_{vj}\Big) + \theta_{wi}\Big], \quad (13)$$

in which x_k denotes the kth input of the NN, y_i presents the ith output, N_o, N_i and N_h are the numbers of output, input and hidden layers cells, respectively, w_{ij} and v_{jk} show the network weights, $\bar{\sigma}(x) = 1/(1 + e^{-x})$ is a sigmoid activation function and θ_{vj} and θ_{wi} denote the threshold offsets. Equation (13) can be written as $y = W^T\sigma(V^T x)$ in which $W \in \Re^{(N_h+1)\times N_o}$ and $V \in \Re^{(N_i+1)\times N_h}$ denote the weight matrices whose first columns include the thresholds θ_{vj} and θ_{wi}, $x = [1, x_1, \ldots, x_{N_i}]^T \in \Re^{N_i+1}$, $y = [y_1, \ldots, y_{N_o}]^T \in \Re^{N_o}$, and $\sigma(V^T x) = \big[1, \bar{\sigma}(V_{r_1}^T x), \ldots, \bar{\sigma}(V_{r_{N_h}}^T x)\big]^T \in \Re^{N_h+1}$, where $V_{r_j}^T, j = 1, ..., N_h$, is the jth row of V^T. For a given continuous function $f(x) : U \to \Re^{N_o}$ where $U \subset \Re^{N_i+1}$ indicates a set which is compact, one can find the optimal thresholds, weights and hidden-layer neurons such that $f(x) =$

$W^{*T}\sigma(V^{*T}x) + e_x(x)$, in which $e_x(x) \in \Re^{N_o}$ denotes the error of functional estimation which is limited on the set U such that $|e_{xi}| \leq B_{xi}, i = 1, \ldots, N_o, \forall x \in U$, in which $B_{xi} \in \Re^{+}$. Ideal NN matrices of weights $W^* \in \Re^{(N_h+1)\times N_o}$ and $V^* \in \Re^{(N_i+1)\times N_h}$ are defined by

$$(W^*, V^*) := \underset{(W,V)}{\operatorname{argmin}}\Big\{ \sup_{x\in U} \big\| W^T\sigma(V^Tx) - f(x)\big\|\Big\}. \quad (14)$$

Because W^* and V^* are unknown, $f(x)$ is substituted with its approximation, i.e. $\hat{f}(x) = \hat{W}^T\sigma(\hat{V}^Tx)$, in which $\hat{W}$ and $\hat{V}$ are the approximated matrices.

Assumption 2. W^* and V^* are limited on the set U such that $\|W^*\|_F \leq B_w$, $\|V^*\|_F \leq B_v$, in which $B_w, B_v \in \Re^+$.

B. Saturated Controller-Observer Design

By using (11) and its time derivative, one has

$$\nu = R^{-1}\dot{\epsilon}_e - R^{-1}\hbar, \quad (15)$$

$$\dot{\nu} = R^{-1}\ddot{\epsilon}_e - R^{-1}\dot{R}R^{-1}\dot{\epsilon}_e + R^{-1}\dot{R}R^{-1}\hbar - R^{-1}\dot{\hbar}. \quad (16)$$

Replacing (15)-(16) in the second equation of (1) and multiplying both sides by R^{-T} results in

$$M(\epsilon_e)\ddot{\epsilon}_e + D(\epsilon_e)\dot{\epsilon}_e + C(\epsilon_e,\dot{\epsilon}_e)\dot{\epsilon}_e - \varsigma = R^{-T}\tau + \delta, \quad (17)$$

in which $M(\epsilon_e) = R^{-T}M_1R^{-1}$, $C(\epsilon_e,\dot{\epsilon}_e) = R^{-T}\big(C_1(q,\nu) - M_1R^{-1}\dot{R}\big)R^{-1}$, $D(\epsilon_e) = R^{-T}D_1R^{-1}$, $\delta = R^{-T}\tau_d \in \Re^n$, where $|\delta_i| \leq B_{\delta_i}, i = 1, \ldots, n$, and ς, which includes lumped NLIP uncertainties, is given by

$$\varsigma = C(\epsilon_e,\dot{\epsilon}_e)\hbar + M(\epsilon_e)\dot{\hbar} + D(\epsilon_e)\hbar - R^{-T}G(q) - R^{-T}H(\nu) + R^{-T}\big(B(q) - I_{n\times n}\big)\tau + R^{-T}B(q)d_s(\tau). \quad (18)$$

Property 1. Since R is full-rank by referring to the strictly positive relation (9), Eq. (11) and the properties of rotation matrix J_1, the followings are true for (17) $\forall w_1, w_2, w \in \Re^n$:

P1.1: $M(\epsilon_e) = M^T(\epsilon_e) > 0$, $\lambda_m\|w\|^2 \leq w^TMw \leq \lambda_M\|w\|^2$, $\forall w \in \Re^n$, $0 < \lambda_m < \lambda_M < \infty$, $\lambda_m := \min_{\forall\epsilon_e\in\Re^n} \lambda_{\min}\big(M(\epsilon_e)\big)$, $\lambda_M := \max_{\forall\epsilon_e\in\Re^n} \lambda_{\max}\big(M(\epsilon_e)\big)$.

P1.2: $D(\epsilon_e) = D^T(\epsilon_e) > 0$, $\lambda_d\|w\|^2 \leq w^TDw \leq \lambda_D\|w\|^2$, $\forall w \in \Re^n$, $0 < \lambda_d < \lambda_D < \infty$, $\lambda_d := \min_{\forall\epsilon_e\in\Re^n} \lambda_{\min}\big(D(\epsilon_e)\big)$, $\lambda_D := \max_{\forall\epsilon_e\in\Re^n} \lambda_{\max}\big(D(\epsilon_e)\big)$.

P1.3: Matrix $C(\epsilon_e,\dot{\epsilon}_e)$ has the following properties [47]:

(i) $w^T\big(\dot{M}(\epsilon_e) - 2C(\epsilon_e,\dot{\epsilon}_e)\big)w = 0, \forall w \in \Re^n$,
(ii) $C(\epsilon_e, w_1)w_2 = C(\epsilon_e, w_2)w_1$,
(iii) $C(\epsilon_e, w_1 + w_2)y = C(\epsilon_e, w_1)y + C(\epsilon_e, w_2)y$,
(iv) $\big\|C(\epsilon_e, w_1)w_2\big\| \leq U_c\|w_1\|\|w_2\|$ where $U_c \in \Re^+$.

Now, a saturated *PID* error surface is defined as follows:

$$z_f = \dot{\epsilon}_e + \Lambda_P\varpi(\epsilon_e) + \Lambda_I\mu_I, \quad (19)$$

where $\Lambda_P = \Lambda_P^T > 0$ and $\Lambda_I = \Lambda_I^T > 0$ are gain matrices, $\varpi(\bullet)$ is the vector of GSFs [44] and μ_I is updated by

$$\dot{\mu}_I = -c_f\mu_I + \beta_f\big(\dot{\epsilon}_e + \Lambda_P\varpi(\epsilon_e) + \Lambda_I\mu_I\big), \quad (20)$$

where c_f and β_f are design parameters. Using (17), (19), properties (ii) and (iii) from P1.3 yields:

$$M(\epsilon_e)\dot{z}_f = -C(\epsilon_e,\dot{\epsilon}_e)z_f - D(\epsilon_e)z_f + \varsigma(x) + \vartheta + R^{-T}\tau + \delta, \quad (21)$$

and ϑ is calculated as follows:

$$\vartheta = M(\epsilon_e)\Lambda_P\varrho(\epsilon_e)\dot{\epsilon}_e + M(\epsilon_e)\Lambda_I\dot{\mu}_I + C\big(\epsilon_e,\dot{\epsilon}_e\big)\Lambda_P\varpi(\epsilon_e) + C\big(\epsilon_e,\dot{\epsilon}_e\big)\Lambda_I\mu_I + D(\epsilon_e)\Lambda_P\varpi(\epsilon_e) + D(\epsilon_e)\Lambda_I\mu_I \quad (22)$$

that is limited through employing Property 1 so that

$$\|\vartheta\| \leq \iota_1\|x_f\| + \iota_2\|x_f\|^2, \quad (23)$$

in which x_f is the augmented state vector that is given by

$$x_f = \big[\varpi^T(\epsilon_e), \mu_I^T, z_f^T\big]^T, \quad (24)$$

and $\iota_1, \iota_2 \in \Re^+$ are unknown. Since velocities are unavailable, an approximation of z_f is generated as follows:

$$\hat{z}_f = \ell_2/\epsilon + \Lambda_P\varpi(\epsilon_e) + \Lambda_I\mu_I \quad (25)$$

by using $\dot{\hat{\epsilon}}_e = \ell_2/\epsilon$ based on following HGO and Lemma 1:

$$\epsilon\dot{\ell}_1 = \ell_2, \;\; \epsilon\dot{\ell}_2 = -\lambda_1\ell_2 - \ell_1 + \epsilon_e(t), \quad (26)$$

and μ_I will be updated as

$$\dot{\mu}_I = -c_f\mu_I + \beta_f\big(\ell_2/\epsilon + \Lambda_P\varpi(\epsilon_e) + \Lambda_I\mu_I\big). \quad (27)$$

Then, by employing item (ii) of Lemma 1 and $\tilde{z}_f = \hat{z}_f - z_f = \dot{\hat{\epsilon}}_e - \dot{\epsilon}_e$, we have $\|\tilde{z}_f\| = \|\epsilon\ddot{\kappa}\| \leq \epsilon G_2 := B_f$ where $B_f \in \Re^+$. Based on the above results, the MLNNRL-based control law is proposed in the sequel.

C. The Actor Neural Network

Since $\varsigma(x)$ in (21) with its exact definition in (18) is a complex NLIP uncertain term, a multi-layer ANN is necessary in this paper to estimate it as follows:

$$\varsigma(x) = W_a^{*T}\sigma_1(V_a^{*T}x) + e_x(x), \; \forall x \in U \subset \Re^{N_i+1}, \quad (28)$$

where $x = [1, \epsilon_e^T, \dot{\epsilon}_e^T, \nu^T, \tau^T, \eta_l^T, \dot{\eta}_l^T, \ddot{\eta}_l^T, \eta_u^T, \dot{\eta}_u^T, \ddot{\eta}_u^T, q^T, \dot{q}_d^T, \ddot{q}_d^T]^T \in \Re^{N_i+1}$ denotes the ANN vector of inputs and $e_x(x) = [e_{x_1}, \ldots, e_{x_n}]^T$ is the NN error. Since V_a^* and W_a^* are not known and the input vector x contains unmeasurable states, we design an actor network as $\hat{\varsigma} = \hat{W}_a^T\sigma_1(\hat{V}_a^T\hat{x})$ where $\hat{x} = [1, \epsilon_e^T, \dot{\hat{\epsilon}}_e^T, \hat{\nu}^T, \tau^T, \eta_l^T, \dot{\eta}_l^T, \ddot{\eta}_l^T, \eta_u^T, \dot{\eta}_u^T, \ddot{\eta}_u^T, q^T, \dot{q}_d^T, \ddot{q}_d^T]^T$, in which $\dot{\hat{\epsilon}}_e := \ell_2/\epsilon$ and $\hat{\nu} = R^{-1}\dot{\hat{\epsilon}}_e - R^{-1}\hbar$ are computed by (15) and the HGO in (26), $\hat{W}_a$ and $\hat{V}_a$ show the estimates of W_a^* and V_a^*, respectively, and $\tilde{W}_a = W_a^* - \hat{W}_a$ and $\tilde{V}_a = V_a^* - \hat{V}_a$ denote ANN estimation errors. Thus, it is possible to write [1]:

$$\varsigma - \hat{\varsigma} = \tilde{W}_a^T\Big(- \sigma_1'(\hat{V}_a^T\hat{x})\hat{V}_a^T\hat{x} + \sigma_1(\hat{V}_a^T\hat{x})\Big) + e_x(x) + r_t + \hat{W}_a^T\sigma_1'(\hat{V}_a^T\hat{x})\tilde{V}_a^T\hat{x}, \quad (29)$$

in which $\sigma_1'(\hat{V}_a^T\hat{x}) = \big[0_{N_h\times 1}, diag[\sigma_{11}', \ldots, \sigma_{1N_h}']\big]^T \in \Re^{(N_h+1)\times N_h}$, with $\sigma_{1i}' = d\bar{\sigma}_1(z)/dz|_{z=\hat{V}_{ar_i}^T\hat{x}}$, $i = 1, \ldots, N_h$, and r_t is limited as

$$\begin{aligned}\|r_t\| \leq& \|W_a^*\|_F \Big(\|\sigma_1'(\hat{V}_a^T \hat{x})\hat{V}_a^T \hat{x}\| + \|\sigma_1(\hat{V}_a^T \hat{x})\| \Big) \\ &+ \|V_a^*\|_F \|\hat{x}\| \|\hat{W}_a^T \sigma_1'(\hat{V}_a^T \hat{x})\|_F. \end{aligned} \tag{30}$$

Substituting (29) into (21) leads to

$$\begin{aligned} M(\epsilon_e)\dot{z}_f =& - C(\epsilon_e, \dot{\epsilon}_e) z_f - D(\epsilon_e) z_f + \chi_p + \hat{\varsigma} \\ &+ \tilde{W}_a^T \Big(-\sigma_1'(\hat{V}_a^T \hat{x})\hat{V}_a^T \hat{x} + \sigma_1(\hat{V}_a^T \hat{x}) \Big) \\ &+ \hat{W}_a^T \sigma_1'(\hat{V}_a^T \hat{x}) \tilde{V}_a^T \hat{x} + r_t + \vartheta + R^{-T}\tau, \end{aligned} \tag{31}$$

where $\chi_p = \delta + e_x(x) \in \Re^n$ which is bounded so that $|\chi_{p_i}| \leq p_i, i = 1, \dots, n$, in which $p_i \in \Re^+$ and $p = [p_1, \dots, p_n]^T$.

D. The Critic Function

The following LTCF or ideal critic function which includes a multi-layer CNN, is introduced in this paper:

$$\begin{aligned} F_c =& \dot{\epsilon}_e + \Lambda_P \varpi(\epsilon_e) + \Lambda_I \mu_I \\ &+ \|\dot{\epsilon}_e + \Lambda_P \varpi(\epsilon_e) + \Lambda_I \mu_I \| W_c^{*T} \sigma_2(V_c^{*T} x), \end{aligned} \tag{32}$$

where $W_c^{*T} \in \Re^{N_o \times (N_h+1)}$, $V_c^{*T} \in \Re^{N_h \times (N_i+1)}$ and $\sigma_2(V_c^{*T} x) = \big[1, \bar{\sigma}_2(V_{cr_1}^{*T} x), \dots, \bar{\sigma}_2(V_{cr_{N_h}}^{*T} x)\big]^T \in \Re^{N_h+1}$ where $V_{cr_i}^{*T}, i = 1, ..., N_h$, is the ith row of V_c^{*T}. Since W_c^*, V_c^* and the velocities are not known, their approximations are utilized. Thus, the critic function could be approximated as

$$\begin{aligned} \hat{F}_c =& \ell_2/\epsilon + \Lambda_P \varpi(\epsilon_e) + \Lambda_I \mu_I \\ &+ \|\ell_2/\epsilon + \Lambda_P \varpi(\epsilon_e) + \Lambda_I \mu_I \| \hat{W}_c^T \sigma_2(\hat{V}_c^T \hat{x}). \end{aligned} \tag{33}$$

Remark 1. Note that in reference [48], MLNNs are used as the structure of the actor and critic. MLNNs have two layers of adjustable weights W^T and V^T with one hidden layer that can be seen in Eq. (13). W^T represents the weight matrix for the layer of outputs and V^T denotes the weight matrix of the hidden layer [47]. In [48], the concept of using MLNNs as the structure of ANN and CNN is introduced, but the design of the nets was simplified to design W_a^T and W_c^T adaptively, and V_a^T and V_c^T are firstly picked at random and kept constant. However, in the proposed MLNNRL method, we managed to design V_a^T and V_c^T adaptively, and we overcame the difficulties arisen from designing proper adaptive laws for V_a^T and V_c^T and the proof of their convergence by the Lyapunov theory, which is an essential contribution of this paper. The corresponding adaptive laws for determining $\hat{V}_a^T$ and $\hat{V}_c^T$ and their convergence analysis are given in the sequel.

Remark 2. Because the LTCF (33) could be considered like a sort of reinforcement variable [48], $\hat{F}_c$ has more processed data compared with the system variables. So, a greater outcome will be achieved for the controller [49].

E. Actor-Critic NN Update Rules and Control Law

In this paper, the next adaptive rules are provided by utilizing $\sigma-$modification to create $\hat{W}_a, \hat{V}_a, \hat{W}_c$ $\hat{V}_c$ and $\hat{p}$:

$$\begin{aligned} \dot{\hat{W}}_a =& -\sigma_{w_a} + \Gamma_{w_a} \Big(\sigma_1(\hat{V}_a^T \hat{x}) - \sigma_1'(\hat{V}_a^T \hat{x})\hat{V}_a^T \hat{x} \Big) \times \\ & \Big(\frac{\ell_2}{\epsilon} + \Lambda_P \varpi(\epsilon_e) + \Lambda_I \mu_I + \\ & \| \frac{\ell_2}{\epsilon} + \Lambda_P \varpi(\epsilon_e) + \Lambda_I \mu_I \| \hat{W}_c^T \sigma_2(\hat{V}_c^T \hat{x}) \Big)^T, \end{aligned} \tag{34}$$

$$\begin{aligned} \dot{\hat{V}}_a =& \Gamma_{v_a} \hat{x} \Big(\frac{\ell_2}{\epsilon} + \Lambda_P \varpi(\epsilon_e) + \Lambda_I \mu_I \\ & + \| \frac{\ell_2}{\epsilon} + \Lambda_P \varpi(\epsilon_e) + \Lambda_I \mu_I \| \hat{W}_c^T \sigma_2(\hat{V}_c^T \hat{x}) \Big)^T \\ & \times \hat{W}_a^T \sigma_1'(\hat{V}_a^T \hat{x}) - \delta_{v_a} \Gamma_{v_a} \hat{V}_a, \end{aligned} \tag{35}$$

$$\begin{aligned} \dot{\hat{W}}_c =& \Gamma_{w_c} \| \frac{\ell_2}{\epsilon} + \Lambda_P \varpi(\epsilon_e) + \Lambda_I \mu_I \| \sigma_2(\hat{V}_c^T \hat{x}) \big(\hat{W}_a^T \sigma_1(\hat{V}_a^T \hat{x}) \big)^T \\ & - \sigma_{w_c}, \end{aligned} \tag{36}$$

$$\dot{\hat{V}}_c = \Gamma_{v_c} \| \frac{\ell_2}{\epsilon} + \Lambda_P \varpi(\epsilon_e) + \Lambda_I \mu_I \| \hat{V}_a - \delta_{v_c} \Gamma_{v_c} \hat{V}_c, \tag{37}$$

$$\dot{\hat{p}} = Q \big[H(\hat{z}_f) (\frac{\ell_2}{\epsilon} + \Lambda_P \varpi(\epsilon_e) + \Lambda_I \mu_I) - \Theta(\hat{p} - p^0) \big], \tag{38}$$

where $\sigma_{w_a} = \delta_{w_a} \Gamma_{w_a} \hat{W}_a$, $\sigma_{w_c} = \delta_{w_c} \Gamma_{w_c} \hat{W}_c$, $\Gamma_{w_a}, \Gamma_{w_c} \in \Re^{(N_h+1) \times (N_h+1)}$, $\Gamma_{v_a}, \Gamma_{v_c} \in \Re^{(N_i+1) \times (N_i+1)}$ are adaptation gains, $\delta_{w_a}, \delta_{v_a}, \delta_{w_c}, \delta_{v_c} \in \Re^+$ denote design variables, $\Theta, Q \in \Re^{n \times n}$ are adaptive gains, $p^0 \in \Re^n$ is a vector to be designed and $H = diag\big[\varpi_{h1}(\hat{z}_{f_1}/c_{1r}) \ \ , \dots, \varpi_{hn}(\hat{z}_{f_n}/c_{nr})\big]$ where $c_{1r}, \dots, c_{nr} \in \Re^+$ indicate some control constants. Then, the next saturated *PID* control framework is suggested here:

$$\begin{aligned} \tau = R^T \Big(& -\varpi(K_p \epsilon_e) - \varpi(K_v \hat{z}_f) - \hat{W}_a^T \sigma_1(\hat{V}_a^T \hat{x}) \\ & - \sum_{j=1}^{5} k_j h_j \hat{z}_f - H\hat{p} \Big), \end{aligned} \tag{39}$$

where $K_p = diag[k_{p_i}], i = 1, \dots, n$, with $\min\{k_{p_1}, \dots, k_{p_n}\} \geq 1$, $K_v = diag[k_{v_i}], i = 1, \dots, n$, $k_{v_i} \geq 1$, $k_1, k_2, k_3, k_4, k_5 \in \Re^+$ indicate design gains, and h_j are in the next form:

$$\begin{cases} h_1 = \|\hat{V}_a\|_F^2, h_2 = \|\sigma_2(\hat{V}_c^T \hat{x}) \sigma_1^T(\hat{V}_a^T \hat{x}) \hat{W}_a\|_F^2, \\ h_3 = \big(\|\sigma_1'(\hat{V}_a^T \hat{x}) \hat{V}_a^T \hat{x}\| + \|\sigma_1(\hat{V}_a^T \hat{x})\| \big)^2 \|\sigma_2^T(\hat{V}_c^T \hat{x}) \hat{W}_c\|^2, \\ h_4 = \|\hat{x}\|^2 \big\| \sigma_2^T(\hat{V}_c^T \hat{x}) \hat{W}_c \hat{W}_a^T \sigma_1'(\hat{V}_a^T \hat{x}) \big\|^2, \\ h_5 = \big(\|\sigma_1'(\hat{V}_a^T \hat{x}) \hat{V}_a^T \hat{x}\| + \|\sigma_1(\hat{V}_a^T \hat{x})\| \big)^2 \\ \quad + \|\hat{x}\|^2 \|\hat{W}_a^T \sigma_1'(\hat{V}_a^T \hat{x})\|_F^2. \end{cases} \tag{40}$$

The third, fourth and last terms in (39) are responsible for the objective $C2$ in Section II.B. Finally, through replacing (39) into (31), one gets the following closed-loop error dynamics:

$$\begin{aligned} M(\epsilon_e)\dot{z}_f =& - C(\epsilon_e, \dot{\epsilon}_e) z_f - D(\epsilon_e) z_f + \vartheta - \varpi(K_p \epsilon_e) + \chi_p \\ & - \varpi(K_v \hat{z}_f) + \tilde{W}_a^T \Big(\sigma_1(\hat{V}_a^T \hat{x}) - \sigma_1'(\hat{V}_a^T \hat{x}) \hat{V}_a^T \hat{x} \Big) \\ & + \hat{W}_a^T \sigma_1'(\hat{V}_a^T \hat{x}) \tilde{V}_a^T \hat{x} + r_t - H\hat{p} - k_1 h_1 \hat{z}_f \\ & - k_2 h_2 \hat{z}_f - k_3 h_3 \hat{z}_f - k_4 h_4 \hat{z}_f - k_5 h_5 \hat{z}_f. \end{aligned} \tag{41}$$

Fig. 1 shows a detailed block diagram of the controller.

Remark 3. The goal of utilizing the LTCF in equation (33) is to find a better control signal. It means the ANN cancels uncertainties and reduces the LTCF along with the tracking errors in parallel. Concomitantly, the CNN estimates the LTCF to adjust the ANN weights. It means supplementary data is provided in order to improve the performance of learning for the ANN in (34) and (35). Therefore, through learning, the overall performance of the estimator $\hat{W}_a^T \sigma_1(\hat{V}_a^T \hat{x})$ could

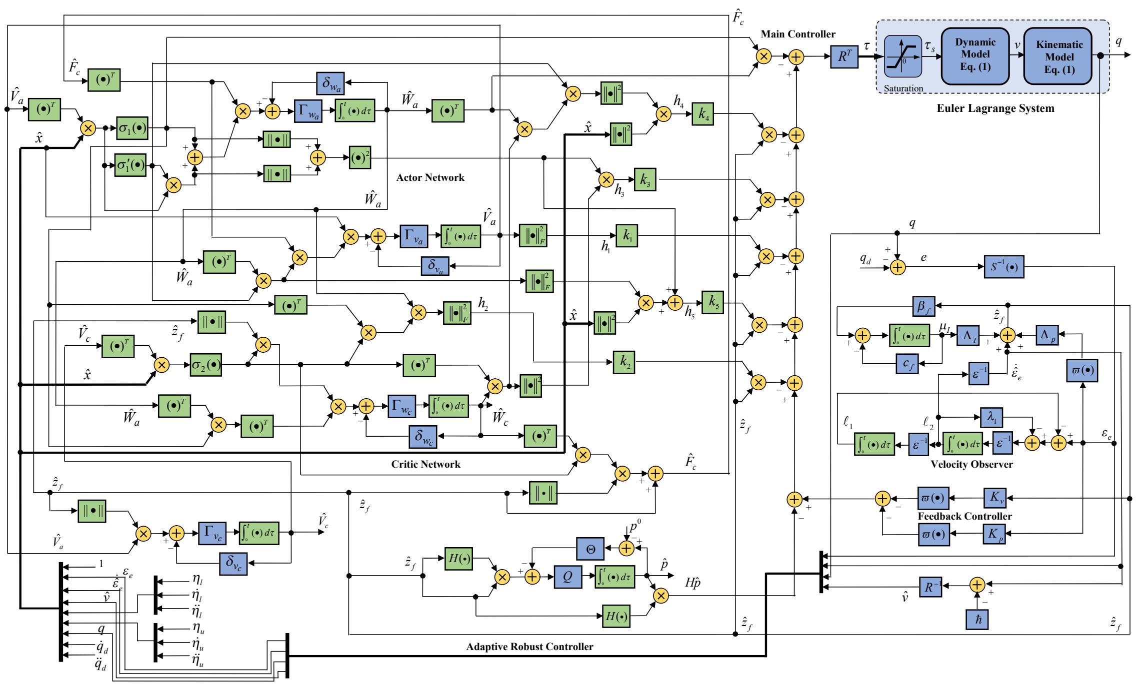

Fig. 1: Block diagram of the proposed control system.

be enhanced, and the performance of controller could be reinforced accordingly [15, 49, 50].

Remark 4. The CNN receives the output of the ANN in (36) and (37) in order to estimate the LTCF in (33). Then, by using the estimation of LTCF in (33) which could be perceived like a kind of strengthening signal stated in Remark 2, the ANN weights are better updated in (34) and (35) to estimate ς, which is the unknown part of the system dynamic in (17), in control law (39). The convergence of update rules (34)-(37) is studied in the sequel.

Remark 5. An interesting property of the feedback part of the proposed controller is that the performance-tuning role of the proportional, integral, and derivative gains is not restricted by using GSFs to produce smaller control efforts.

F. Proposed Controller Gains Selection

1) Parameter selection for the update rules (34)-(37): by increasing the learning gains $\Gamma_{w_a}, \Gamma_{v_a}, \Gamma_{w_c}$ and Γ_{v_c}, the learning speed of the AC is increased. Choosing a large value for these gains leads to divergence of the learning laws, and having small values for the gains will considerably reduce the learning procedure. A proper selection of $\delta_{w_a}, \delta_{v_a}, \delta_{w_c}$ and δ_{v_c} can provide a trade-off among the robustness of these adaptation rules and accuracy of tracking. In other words, large values for these parameters increase the robustness of the learning laws (34)-(37), but note that the ultimate accuracy for tracking could be reduced as a result of the large values of $\delta_{w_a}, \delta_{v_a}, \delta_{w_c}$ and δ_{v_c}. So, operators should equilibrate the final tracking precision and learning robustness attentively by choosing proper values for $\delta_{w_a}, \delta_{v_a}, \delta_{w_c}$ and δ_{v_c}.

2) Parameter selection for the update rule (38): increasing the value of Q results in a quick parameters estimation, and leads to improving the robustness for the control system. Additionally, in order to prevent the divergence of the adaptive law (38), users should avoid picking Q very large. The parameters Θ and p^0 are assigned according to the experimental setup conditions. For the example, Θ and p^0 are appropriately selected by observing the level of external disturbances to secure the robustness for system. Nevertheless, increasing Θ and p^0 decreases the final tracking accuracy. One of the most important impacts with respect to the regulation of (38) and the closed-loop performance concerns the values $c_{ir}, i = 1, \ldots, n$, which are the boundary layer thicknesses. One should make a trade-off among the accuracy of tracking and smoothness of the system variables by correct selection of the c_{ir} values.

G. Stability Analysis

Theorem 1. Assume a class of EL systems for which the kinematics and dynamics are represented by (1). Under the Assumptions 1-2, the suggested saturated *PID* controller (39) along with the strengthening variable designed in (33) and the rules of adaptation (34)-(38) with the HGO (26), secures that variables in the system of control stay limited and the errors of tracking are SGUUB which tend to a region comprising origin. In addition, it can be proved that the output constraints are never transgressed and the region of attraction

$$R_A = \left\{ x_u \in \Re^{3n+c_n} \Big| \|x_u\|^2 < \frac{\lambda_f(2\chi_f - \iota_1)}{\iota_2 \lambda_{x_u}} \right\} \tag{42}$$

could be created large in order to cover all initial conditions by choosing the control gains properly. Here, $c_n = 2(N_h + 1)N_o + 2(N_i + 1)N_h$, $\chi_f \in \Re^+$ is a gain-dependent constant, λ_f and λ_{x_u} are constants defined later.

Proof. The proof of Theorem 1 is given in Appendix A.

Remark 6. Since all signals for the control system, including (33), are analyzed by the Lyapunov theory, it is concluded that $\dot{E}$ is strictly negative outside $\Omega_{x_l} = \{x_l | 0 \leq \|x_l\| \leq$

$\sqrt{\Xi/c}\}$ according to the proof in Appendix A, where Ξ and c are gain dependent parameters and can be adjusted by selecting proper control gains to reduce Ω_{x_l}. Consequently, the LTCF in (33) converges to a zone around zero, and this zone could be decreased by selecting proper control parameters. So, the LTCF in (33) is minimized about the origin by the proposed AC structure and the overall control performance is improved in the sense of minimizing LTCF in (33).

Remark 7. In this paper, the actuator saturation is prevented by a successful combination of the GSFs and actuators saturation nonlinearity compensation. As long as the actuator saturation is avoided by the proposed controller, there is no concern about the PPC operation. However, if the actuator saturation occurs in the worst case, the PPC fragility problem may take place. Toward this end, a non-fragile PPC approach is suggested in [51]. However, this issue does not lie in the scope of this paper and it is devoted to our future works.

IV. EXPERIMENTAL VERIFICATION

To verify the suggested controller efficacy in a real-world scenario, experimental results are reported here. To this end, the suggested intelligent controller is implemented on a SCARA IBM7547 robot arm whose image is shown by Fig 2.

The proposed controller is programmed into Arduino Due control board to test its performance empirically on SCARA IBM7547 arm. The board utilizes Atmel SAM3X8E ARM micro-controller with 84 MHz clock frequency, 96 KB SRAM, 512 KB flash memory, 12 PWM (pulse width modulation) outputs, 54 digital input-output pins, 12 analog inputs, 4 serial ports, and two digital-to-analog converters. A 12-bit resolution decoder latch is also used in this study. The joint limits for the first and second joints are $0 \leq q_1 \leq 3.5rad$ and $0 \leq q_2 \leq 2.8rad$, respectively. The control signals are generated by the controller-observer block programmed into the Arduino Due controller and changed to 40 KHz PWM signals with a 13-bit resolution and amplified through IRF540N-MOSFET power amplifiers before being applied to the direct current (DC) motors mounted on each joint. Subsequently, the sensor signals measured by the shaft encoder at the sampling rate of 0.02 s are transmitted to a decoder-counter latch and the angular position $q(t)$ is fed back to the controller program. Thus, the proposed control laws (25), (26), (33), (34)-(38) and (39), programmed on the Arduino board digitally, calculate the control effort $\tau(t)$. The torque applied to each motor is based on measuring the position error $e = q - q_d$ for each joint at each sampling time so that the position errors converge toward zero. The following function is used as a GSF for (19) and (39):

$$\varpi_i(z_i) = \begin{cases} -L_i + (M_i - L_i)\tanh(\frac{z_i+L_i}{M_i-L_i}) & ,\forall z_i < -L_i \\ z_i & ,\forall |z_i| \leq L_i \\ L_i + (M_i - L_i)\tanh(\frac{z_i-L_i}{M_i-L_i}) & ,\forall z_i > L_i \end{cases} \tag{43}$$

where $L_i < M_i, \forall i = 1, 2, 3$ with $M_i = 10$ and $L_i = 0.9M_i$, and the control inputs are saturated so that $|\tau_i| \leq 30$. The controller parameters are chosen as follows for a best tracking performance: $\eta_{10} = \eta_{20} = 10, \eta_{1\infty} = \eta_{2\infty} = 0.1, a_1 = 0.2, a_2 = 0.2, \alpha_1 = 2, \alpha_2 = 2, \beta_1 = \beta_2 = 2, \Lambda_P = diag[1, 1], \epsilon = 0.1, \lambda_1 = 5, N_h = 5, N_o = 2, \Gamma_{w_a} = 0.001I_6, \delta_{w_a} = 2, \Gamma_{v_a} = 0.001I_{N_i+1}, \delta_{v_a} = 2, \Gamma_{w_c} = 0.1I_6, \delta_{w_c} = 2, \Gamma_{v_c} = 0.1I_{N_i+1}, \delta_{v_c} = 2, c_{1r} = c_{2r} = 0.01, \Theta = diag[5, 5], k_1 = k_2 = k_3 = k_4 = 1.5, k_5 =$

Algorithm 1: The proposed controller pseudocode

Initialization;
Set $t = 0$;
Compute the tracking errors at $t = 0$ as $e(0) = q(0) - q_d(0)$;
If the condition $\eta_{li}(0) \leq e_i(0) \leq \eta_{ui}(0)$ is true, start the control loop. Otherwise, increase η_{i0}, β_i and α_i;
while $t < t_{\max}$ **do**

1) Compute the tracking errors as $e(t) = q(t) - q_d(t)$;
2) Transform the tracking errors using (6);
3) Form the vector of the transformed tracking errors as $\epsilon_e = [\epsilon_{e_1}, \ldots, \epsilon_{e_n}]^T$;
4) Estimate the system derivatives in (19) by (26);
5) Calculate the saturated filtered error $\hat{z}_f = \dot{\hat{\epsilon}}_e + \Lambda_P \varpi(\epsilon_e) + \Lambda_I \mu_I$;
6) Get the input vector $\hat{x}$ for the multilayer ANN and CNN;
7) Update $\hat{W}_a$, $\hat{V}_a$ and $\hat{p}$ by (34), (35) and (38), respectively;
8) Update $\hat{W}_c$ and $\hat{V}_c$ by (36) and (37);
9) Estimate the reinforcement signal in (33);
10) Calculate the input control signal τ in (39);
11) Measure the system states q for the new execution loop;

end
return outcomes;
End

Fig. 2: SCARA IBM7547 for the experimental setup.

TABLE I: Numerical evaluation of the computational complexity

Variable in Algorithm 1	Elapsed time (ms)	Number of math operations
e_1, e_2	0.048	16
$\epsilon_{e_1}, \epsilon_{e_2}$	0.144	50
$\hat{z}_{f,1}, \hat{z}_{f,2}$	0.231	88
$\hat{p}_1, \hat{p}_2$	0.292	110
$\hat{W}_a$	7.68	3072
$\hat{V}_a$	8.9225	3569
$\hat{W}_c$	12.05	4420
$\hat{V}_c$	14.7175	5887
$\hat{F}_{c_1}, \hat{F}_{c_2}$	16.0875	6036
τ_1, τ_2	19.4525	9384

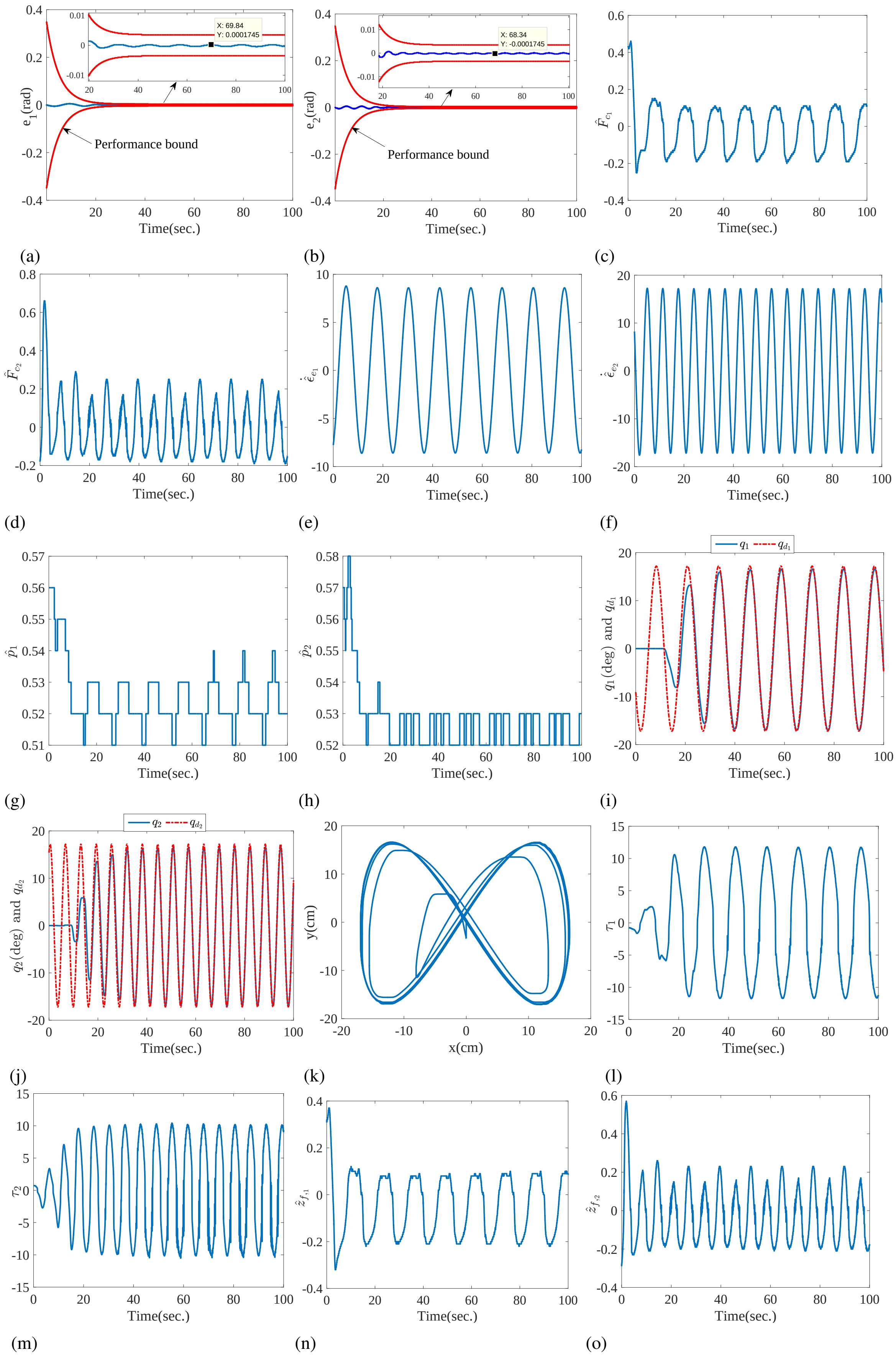

Fig. 3: The experimental setup outcomes: (a) tracking error e_1, (b) tracking error e_2, (c) estimated RL signal $\hat{F}_{c_1}$, (d) estimated RL signal $\hat{F}_{c_2}$, (e) $\dot{\hat{e}}_{e_1}$, (f) $\dot{\hat{e}}_{e_2}$, (g) $\hat{p}_1$, (h) $\hat{p}_2$, (i) time evolution of q_1, (j) time evolution of q_2, (k) end-effector trajectory, (l) control signal τ_1, (m) generated control signal τ_2, (n) estimation of the saturated filtered transformed error $\hat{z}_{f,1}$, and (o) $\hat{z}_{f,2}$.

$0.15, K_p = 2diag[1,1]$ and $K_v = 0.5diag[1,1]$. The results of this experimental setup are shown in Figs. 3. As it is clear in Figs. 3(a)-3(b), the position errors converge to zero in the performance bounds, and the joints are tracking the desired angles well as shown in Figs. 3(i)-3(j) by the smooth, well-bounded control signals generated by the suggested control method in Figs. 3(l)-(m). Besides, the other bounded signals of the system, plotted in Fig. 3, demonstrate a nice performance for the proposed intelligent controller. Thus, the practical implementation displays the merit of the suggested controller. The computational complexity of the proposed algorithm is also evaluated experimentally and listed in Table I. The table confirms the practicability of the proposed controller on the presented hardware setup for real-time implementation.

V. Conclusion and Future Research Direction

This paper presented a novel OFB-MLNNRL-based saturated *PID* control strategy for output constrained EL systems with unknown NLIP dynamics and actuator saturation. The need for a prior knowledge on system dynamics are removed for both OFB and MLNNRL methods successfully. The MLNNRL has been devised to adaptively detect complex NLIP unknowns and a critic function for supervising control performance. In combination with the critic function and OFB problem, an AC-MLNNRL structure has been established. Extra advantages of our method include more robustness, low design complexity, low computational burden, saturated inputs, prescribed output tracking and removing velocity sensors in the presence of NLIP uncertainties and disturbances. Moreover, by an effective combination of compensating the nonlinearity of saturation through the suggested MLNNRL-based controller and employing GSFs, the actuator saturation risk was effectively avoided. Also, the stability analysis proves that the entire OFB-MLNNRL control scheme guarantees the errors convergence to an arbitrary small region in the origin vicinity with a prescribed performance. Although innovative aspects of the proposed controller have been reviewed, there are yet some open problems that clarify our future research directions including relaxation of Assumption 1, considering the actuator dynamics, and bounding the estimated HGO states to avoid the observer peaking phenomenon.

Appendix A
Proof of Theorem 1

The proof of Theorem 1 is presented in this section by using a Lyapunov's direct method. For this purpose, consider the following Lyapunov function candidate:

$$\begin{aligned} E =& \sum_{i=1}^{n}\int_0^{\epsilon_{e_i}} \varpi_i(k_{pi}\xi)d\xi + \frac{1}{2}z_f^T M(\epsilon_e)z_f + \frac{1}{2}\tilde{p}^T Q^{-1}\tilde{p} \\ &+ \frac{1}{2}tr\{\tilde{W}_a^T\Gamma_{w_a}^{-1}\tilde{W}_a\} + \frac{1}{2}tr\{\tilde{V}_a^T\Gamma_{v_a}^{-1}\tilde{V}_a\} + \frac{1}{2}\mu_I^T\mu_I \\ &+ \frac{1}{2}tr\{\tilde{W}_c^T\Gamma_{w_c}^{-1}\tilde{W}_c\} + \frac{1}{2}tr\{\tilde{V}_c^T\Gamma_{v_c}^{-1}\tilde{V}_c\}, \end{aligned} \tag{44}$$

where $\tilde{W}_a = W_a^* - \hat{W}_a, \tilde{V}_a = V_a^* - \hat{V}_a, \tilde{W}_c = W_c^* - \hat{W}_c, \tilde{V}_c = V_c^* - \hat{V}_c$ and $\tilde{p} = \hat{p} - p$. From (44) and using items (iii) and (iv) from Lemma 2, it is obvious that (44) is an unbounded, positive-definite, and decrescent function. Recalling property (ii) from Lemma 2 gives:

$$\lambda_f\|x_f\|^2 \le \lambda_{x_l}\|x_l\|^2 \le E(t) \le \lambda_{x_u}\|x_u\|^2, \tag{45}$$

where λ_f, λ_{x_l}, λ_{x_u}, x_f, x_l and x_u are given by

$$\begin{aligned} \lambda_f =& 0.5\min\{1, (k_{p_1}\varrho_{1M})^{-1}, \dots, (k_{p_n}\varrho_{nM})^{-1}, \lambda_m\}, \\ \lambda_{x_l} =& 0.5\min\{(k_{p_1}\varrho_{1M})^{-1}, \dots, (k_{p_n}\varrho_{nM})^{-1}, \lambda_m, \lambda_{\min}\{\Gamma_{w_a}^{-1}\} \\ &, \lambda_{\min}\{\Gamma_{v_a}^{-1}\}, \lambda_{\min}\{\Gamma_{w_c}^{-1}\}, \lambda_{\min}\{\Gamma_{v_c}^{-1}\}, \lambda_{\min}\{Q^{-1}\}, 1\}, \\ \lambda_{x_u} =& 0.5\max\{k_{p_1}\varrho_{1M}, \dots, k_{p_n}\varrho_{nM}, \lambda_M, \lambda_{\max}\{\Gamma_{w_a}^{-1}\} \\ &, \lambda_{\max}\{\Gamma_{v_a}^{-1}\}, \lambda_{\max}\{\Gamma_{w_c}^{-1}\}, \lambda_{\max}\{\Gamma_{v_c}^{-1}\}, \lambda_{\max}\{Q^{-1}\}, 1\}, \\ x_l =& \Big[\varpi^T(\epsilon_e), z_f^T, \tilde{w}_{a11}, \dots, \tilde{w}_{a(N_h+1)N_o}, \tilde{v}_{a11}, \dots, \tilde{v}_{a(N_i+1)N_h} \\ &, \tilde{w}_{c11}, \dots, \tilde{w}_{c(N_h+1)N_o}, \tilde{v}_{c11}, \dots, \tilde{v}_{c(N_i+1)N_h}, \tilde{p}^T, \mu_I^T\Big]^T, \\ x_u =& \Big[\epsilon_e^T, z_f^T, \tilde{w}_{a11}, \dots, \tilde{w}_{a(N_h+1)N_o}, \tilde{v}_{a11}, \dots, \tilde{v}_{a(N_i+1)N_h}, \\ & \tilde{w}_{c11}, \dots, \tilde{w}_{c(N_h+1)N_o}, \tilde{v}_{c11}, \dots, \tilde{v}_{c(N_i+1)N_h}, \tilde{p}^T, \mu_I^T\Big]^T. \end{aligned}$$

Time derivative of (44) along (41), adding and subtracting $z_f^T\varpi(K_v z_f)$, employing (19), (27) and item (i) of P1.3 yields:

$$\begin{aligned} \dot{E} =& -\varpi^T(K_p\epsilon_e)\Lambda_P\varpi(\epsilon_e) - \varpi^T(K_p\epsilon_e)\Lambda_I\mu_I \\ &- c_f\mu_I^T\mu_I + \beta_f\mu_I^T\hat{z}_f - z_f^T D(\epsilon_e)z_f + z_f^T\vartheta \\ &- z_f^T\varpi(K_v\hat{z}_f) + z_f^T\varpi(K_v z_f) - z_f^T\varpi(K_v z_f) \\ &+ z_f^T(\chi_p - H\hat{p}) + z_f^T\tilde{W}_a^T\Big(\sigma_1(\hat{V}_a^T\hat{x}) - \sigma_1'(\hat{V}_a^T\hat{x})\hat{V}_a^T\hat{x}\Big) \\ &+ z_f^T\hat{W}_a^T\sigma_1'(\hat{V}_a^T\hat{x})\tilde{V}_a^T\hat{x} + z_f^T r_t - z_f^T\sum_{j=1}^{5}k_j h_j\hat{z}_f \\ &- tr\{\tilde{W}_a^T\Gamma_{w_a}^{-1}\dot{\hat{W}}_a\} - tr\{\tilde{V}_a^T\Gamma_{v_a}^{-1}\dot{\hat{V}}_a\} - tr\{\tilde{W}_c^T\Gamma_{w_c}^{-1}\dot{\hat{W}}_c\} \\ &- tr\{\tilde{V}_c^T\Gamma_{v_c}^{-1}\dot{\hat{V}}_c\} + \tilde{p}^T Q^{-1}\dot{\hat{p}}. \end{aligned} \tag{46}$$

Now, by employing Lemma 6 and item (v) of Lemma 2 and using the followings:

$$\begin{aligned} -z_f^T\varpi(K_v z_f) \le& -z_f^T\varpi(z_f), \\ -\varpi^T(K_p\epsilon_e)\Lambda_P\varpi(\epsilon_e) \le& -\varpi^T(\epsilon_e)\Lambda_P\varpi(\epsilon_e), \\ -\varpi^T(\epsilon_e)\Lambda_P\varpi(\epsilon_e) \le& -\lambda_{\min}\{\Lambda_P\}\|\varpi(\epsilon_e)\|^2, \\ -\varpi^T(K_p\epsilon_e)\Lambda_I\mu_I \le& -\varpi^T(\epsilon_e)\Lambda_I\mu_I, \end{aligned} \tag{47}$$

and employing (34)-(38) based on (33), the fact that $-z_f^T\varpi(z_f) \le 0$ and utilizing Property P1.2, we get

$$\begin{aligned} \dot{E} \le& -\lambda_{\min}\{\Lambda_P\}\|\varpi(\epsilon_e)\|^2 - \lambda_d\|z_f\|^2 + z_f^T\vartheta \\ &- \varpi^T(\epsilon_e)\Lambda_I\mu_I - c_f\|\mu_I\|^2 + \beta_f\mu_I^T z_f + \beta_f\mu_I^T\tilde{z}_f \\ &- tr\{\tilde{W}_a^T\big(\sigma_1(\hat{V}_a^T\hat{x}) - \sigma_1'(\hat{V}_a^T\hat{x})\hat{V}_a^T\hat{x}\big)\tilde{z}_f^T\} + z_f^T r_t \\ &- tr\{\tilde{W}_a^T\big(\sigma_1(\hat{V}_a^T\hat{x}) - \sigma_1'(\hat{V}_a^T\hat{x})\hat{V}_a^T\hat{x}\big)\|\hat{z}_f\|\sigma_2^T(\hat{V}_c^T\hat{x})\hat{W}_c\} \\ &- tr\{\tilde{V}_a^T\hat{x}\tilde{z}_f^T\hat{W}_a^T\sigma_1'(\hat{V}_a^T\hat{x})\} + \delta_{v_a}tr\{\tilde{V}_a^T\hat{V}_a\} \\ &- tr\{\tilde{V}_a^T\hat{x}\|\hat{z}_f\|\sigma_2^T(\hat{V}_c^T\hat{x})\hat{W}_c\hat{W}_a^T\sigma_1'(\hat{V}_a^T\hat{x})\} \\ &- tr\{\tilde{W}_c^T\|\hat{z}_f\|\sigma_2(\hat{V}_c^T\hat{x})\sigma_1^T(\hat{V}_a^T\hat{x})\hat{W}_a\} + z_f^T(\chi_p - H\hat{p}) \\ &- tr\{\tilde{V}_c^T\|\hat{z}_f\|\hat{V}_a\} + \delta_{v_c}tr\{\tilde{V}_c^T\hat{V}_c\} + \delta_{w_c}tr\{\tilde{W}_c^T\hat{W}_c\} \end{aligned}$$

$$+\tilde{p}^T[H(\hat{z}_f)\hat{z}_f-\Theta(\hat{p}-p^0)]-z_f^T\sum_{j=1}^{5}k_jh_j\hat{z}_f$$
$$-z_f^T\varpi(K_v\hat{z}_f)+z_f^T\varpi(K_vz_f)+\delta_{w_a}tr\{\tilde{W}_a^T\hat{W}_a\}. \quad (48)$$

By using the following inequalities

$$\|z_f^Tr_t\|\le 0.5\|z_f\|^2\big(\|\sigma_1'(\hat{V}_a^T\hat{x})\hat{V}_a^T\hat{x}\|+\|\sigma_1(\hat{V}_a^T\hat{x})\|\big)^2$$
$$+0.5\|W_a^*\|_F^2+0.5\|V_a^*\|_F^2+0.5\|z_f\|^2\|\hat{x}\|^2\|\hat{W}_a^T\sigma_1'(\hat{V}_a^T\hat{x})\|_F^2,$$
$$-tr\{\tilde{W}_a^T\big(\sigma_1(\hat{V}_a^T\hat{x})-\sigma_1'(\hat{V}_a^T\hat{x})\hat{V}_a^T\hat{x}\big)\tilde{z}_f^T\}\le 0.5\|\tilde{W}_a\|_F^2$$
$$+0.5B_f^2\big(\|\sigma_1'(\hat{V}_a^T\hat{x})\hat{V}_a^T\hat{x}\|+\|\sigma_1(\hat{V}_a^T\hat{x})\|\big)^2,$$
$$-tr\{\tilde{V}_a^T\hat{x}\tilde{z}_f^T\hat{W}_a^T\sigma_1'(\hat{V}_a^T\hat{x})\}\le 0.5\|\tilde{V}_a\|_F^2$$
$$+0.5B_f^2\|\hat{x}\|^2\|\hat{W}_a^T\sigma_1'(\hat{V}_a^T\hat{x})\|_F^2,$$
$$-tr\{\tilde{W}_a^T\big(\sigma_1(\hat{V}_a^T\hat{x})-\sigma_1'(\hat{V}_a^T\hat{x})\hat{V}_a^T\hat{x}\big)\|\hat{z}_f\|\sigma_2^T(\hat{V}_c^T\hat{x})\hat{W}_c\}\le$$
$$0.25\|\hat{z}_f\|^2\big(\|\sigma_1'(\hat{V}_a^T\hat{x})\hat{V}_a^T\hat{x}\|+\|\sigma_1(\hat{V}_a^T\hat{x})\|\big)^2\|\sigma_2^T(\hat{V}_c^T\hat{x})\hat{W}_c\|^2$$
$$+\|\tilde{W}_a\|_F^2,$$
$$-tr\{\tilde{V}_a^T\hat{x}\|\hat{z}_f\|\sigma_2^T(\hat{V}_c^T\hat{x})\hat{W}_c\hat{W}_a^T\sigma_1'(\hat{V}_a^T\hat{x})\}\le\|\tilde{V}_a\|_F^2$$
$$+0.25\|\hat{z}_f\|^2\|\hat{x}\|^2\|\sigma_2^T(\hat{V}_c^T\hat{x})\hat{W}_c\hat{W}_a^T\sigma_1'(\hat{V}_a^T\hat{x})\|^2,$$

$$-tr\{\tilde{W}_c^T\|\hat{z}_f\|\sigma_2(\hat{V}_c^T\hat{x})\sigma_1^T(\hat{V}_a^T\hat{x})\hat{W}_a\}\le\|\tilde{W}_c\|_F^2$$
$$+0.25\|\hat{z}_f\|^2\|\sigma_2(\hat{V}_c^T\hat{x})\sigma_1^T(\hat{V}_a^T\hat{x})\hat{W}_a\|_F^2,$$
$$-tr\{\tilde{V}_c^T\|\hat{z}_f\|\hat{V}_a\}\le\|\tilde{V}_c\|_F^2+0.25\|\hat{z}_f\|^2\|\hat{V}_a\|_F^2,$$

inequality (48) is simplified to

$$\begin{aligned}\dot{E}\le&-(\lambda_{\min}\{\Lambda_P\}-0.5\lambda_{\max}\{\Lambda_I\})\|\varpi(\epsilon_e)\|^2\\&-(c_f-\beta_f-0.5\lambda_{\max}\{\Lambda_I\})\|\mu_I\|^2+0.5\beta_f\|\tilde{z}_f\|^2\\&-(\lambda_d-0.5\beta_f)\|z_f\|^2+z_f^T\vartheta+z_f^T(\chi_p-H\hat{p})\\&+\sum_{j=1}^{5}\xi_jh_j-z_f^T\varpi(K_v\hat{z}_f)+z_f^T\varpi(K_vz_f)+\|\tilde{W}_c\|_F^2\\&+\delta_{w_a}tr\{\tilde{W}_a^T\hat{W}_a\}+\delta_{v_a}tr\{\tilde{V}_a^T\hat{V}_a\}+\delta_{w_c}tr\{\tilde{W}_c^T\hat{W}_c\}\\&+\delta_{v_c}tr\{\tilde{V}_c^T\hat{V}_c\}+0.5\|W_a^*\|_F^2+0.5\|V_a^*\|_F^2+1.5\|\tilde{W}_a\|_F^2\\&+1.5\|\tilde{V}_a\|_F^2+\|\tilde{V}_c\|_F^2+\tilde{p}^T[H(\hat{z}_f)\hat{z}_f-\Theta(\hat{p}-p^0)],\end{aligned} \quad (49)$$

in which $\xi_jh_j=\big(-\bar{\rho}\|z_f\|^2+\underline{\rho}B_f^2\big)h_j$, where $\bar{\rho}=0.5k_j-0.5$ and $\underline{\rho}=0.5k_j+0.5$. Thus, if $\|z_f\|\ge B_f\sqrt{\rho_u/\rho_l}$, where $\rho_u=k_j+1$ and $\rho_l=k_j-1$, holds true with $k_j>1$, we get $\xi_jh_j\le 0$. Then, by utilizing Lemmas 3 and 4 and the following inequalities [45]:

$$\tilde{p}^T\big[H\hat{z}_f-\Theta(\hat{p}-p^0)\big]+z_f^T(-H\hat{p}+\chi_p)\le 1.5B_f^2$$
$$+0.2785[c_{1r},\ldots,c_{nr}]p-0.5\lambda_{\min}\{\Theta\}\|\tilde{p}\|^2+\|p\|^2$$
$$+0.5(p-p^0)^T\Theta(p-p^0)+0.5\|\tilde{p}\|^2,$$
$$\|z_f\|\|-\varpi(K_v\hat{z}_f)+\varpi(K_vz_f)\|\le$$
$$l\|z_f\|\|-K_v\hat{z}_f+K_vz_f\|\le l\|z_f\|\|K_v\|\|\tilde{z}_f\|,$$
$$l\|z_f\|\|K_v\|\|\tilde{z}_f\|\le\frac{l}{2}\lambda_{\max}\{K_v\}\|z_f\|^2+\frac{l}{2}\lambda_{\max}\{K_v\}B_f^2,$$

where $k>\sqrt{2}/2$, inequality (49) can be rewritten as

$$\dot{E}\le-\bar{c}_1\|x_f\|^2-\bar{c}_2\|\tilde{W}_a\|_F^2-\bar{c}_3\|\tilde{V}_a\|_F^2-\bar{c}_4\|\tilde{W}_c\|_F^2$$
$$-\bar{c}_5\|\tilde{V}_c\|_F^2-\bar{c}_6\|\tilde{p}\|^2+\Xi, \quad (50)$$

in which $\bar{c}_i>0$ are defined as $\bar{c}_1=\chi_f-0.5\iota_2\|x_f\|^2-0.5\iota_1, \bar{c}_2=\delta_{w_a}(-1/2k^2+1)-1.5, \bar{c}_3=\delta_{v_a}(-1/2k^2+1)-1.5$, $\bar{c}_4=\delta_{w_c}(-1/2k^2+1)-1, \bar{c}_5=\delta_{v_c}(-1/2k^2+1)-1$ and $\bar{c}_6=0.5\lambda_{\min}\{\Theta\}-0.5$, $\Xi=0.2785[c_{1r},\ldots,c_{nr}]p+\|p\|^2+0.5(p-p^0)^T\Theta(p-p^0)+0.5l\lambda_{\max}\{K_v\}B_f^2+0.5\delta_{w_a}k^2\|W_a^*\|_F^2+0.5\delta_{v_a}k^2\|V_a^*\|_F^2+0.5\delta_{w_c}k^2\|W_c^*\|_F^2+0.5\delta_{v_c}k^2\|V_c^*\|_F^2+0.5\|W_a^*\|_F^2+0.5\|V_a^*\|_F^2+0.5\beta_fB_f^2+1.5B_f^2$, and $\chi_f=\min\{(\lambda_{\min}\{\Lambda_P\}-0.5\lambda_{\max}\{\Lambda_I\}),(\lambda_d-0.5\beta_f-0.5\iota_1-0.5\iota_2-0.5l\lambda_{\max}\{K_v\}),(c_f-\beta_f-0.5\lambda_{\max}\{\Lambda_I\})\}$. Next, by choosing χ_f such that

$$\chi_f>0.5\iota_1+0.5\iota_2\|x_f\|^2, \quad (51)$$

one gets $\dot{E}\le-\bar{c}\|x_l\|^2+\Xi$, where $\bar{c}=\min\{\bar{c}_1,\ldots,\bar{c}_6\}$. That implies $\dot{E}$ will be strictly negative out of $\Omega_{x_l}=\{x_l|0\le\|x_l\|\le\sqrt{\Xi/\bar{c}}\}$, $\forall t\ge 0$, which means that E is decreasing out of Ω_{x_l}. Then, one has $E(t)\le E(0)\le\lambda_{x_u}\|x_u(0)\|^2$. Next, one gets $\|x_f\|^2\le\lambda_{x_u}/\lambda_f\|x_u(0)\|^2$. Therefore, a sufficient condition for (51) is $2\chi_f>\iota_1+\iota_2(\lambda_{x_u}/\lambda_f)\|x_u(0)\|^2$, which implies zone (42) can be increased adequately to include all initial conditions by designating control gains properly. Thus, if (51) is satisfied, then $\dot{E}(t)<0$ out of Ω_{x_l}. Then, $x_l(t)$ is SGUUB when it is out of Ω_{x_l}. Also, $\dot{\epsilon}_e\in\mathcal{L}_\infty$ by recalling (19). Therefore, tracking errors, NN weights and parameters approximation errors are SGUUB too. Also, by using Assumption 2, $\|\tilde{z}_f\|\le B_f$, and recalling GSFs properties, one deduces that $\epsilon_e,\hat{z}_f,\hat{w}_{a11},\ldots,\hat{w}_{a(N_h+1)N_o},\hat{v}_{a11},\ldots,\hat{v}_{a(N_i+1)N_h},\hat{w}_{c11},\ldots,\hat{w}_{c(N_h+1)N_o},\hat{v}_{c11},\ldots,\hat{v}_{c(N_i+1)N_h},\hat{p}\in\mathcal{L}_\infty$. Since $\epsilon_e\in\mathcal{L}_\infty$, $\epsilon_e\to 0$ and by considering Assumption 1, $e_i\to 0$ with a prescribed performance as $t\to\infty$. □

References

[1] S. S. Ge, C. C. Hang, T. H. Lee, and T. Zhang, *Stable adaptive neural network control*. Springer Science & Business Media, 2013, vol. 13.

[2] J. F. Fisac, A. K. Akametalu, M. N. Zeilinger, S. Kaynama, J. Gillula, and C. J. Tomlin, "A general safety framework for learning-based control in uncertain robotic systems," *IEEE Transactions on Automatic Control*, vol. 64, no. 7, pp. 2737–2752, 2019.

[3] L. Busoniu, R. Babuska, B. De Schutter, and D. Ernst, *Reinforcement learning and dynamic programming using function approximators*. CRC press, 2017.

[4] M. Naeem, S. T. H. Rizvi, and A. Coronato, "A gentle introduction to reinforcement learning and its application in different fields," *IEEE Access*, vol. 8, pp. 209 320–209 344, 2020.

[5] F. Adib Yaghmaie, F. Gustafsson, and L. Ljung, "Linear quadratic control using model-free reinforcement learning," *IEEE Transactions on Automatic Control*, pp. 1–1, 2022, DOI: 10.1109/TAC.2022.3145632.

[6] L. Xi, J. Wu, Y. Xu, and H. Sun, "Automatic generation control based on multiple neural networks with actor-critic strategy," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 32, no. 6, pp. 2483–2493, 2021.

[7] L. Li, D. Li, T. Song, and X. Xu, "Actor–critic learning control with regularization and feature selection in policy gradient estimation," *IEEE transactions on neural networks and learning systems*, vol. 32, no. 3, pp. 1217–1227, 2021.

[8] N. Wang, Y. Gao, H. Zhao, and C. K. Ahn, "Reinforcement learning-based optimal tracking control of an unknown unmanned surface vehicle," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 32, no. 7, pp. 3034–3045, 2021.

[9] Y. Hu, W. Wang, H. Liu, and L. Liu, "Reinforcement learning tracking control for robotic manipulator with kernel-based dynamic model," *IEEE transactions on neural networks and learning systems*, vol. 31, no. 9, pp. 3570–3578, 2020.

[10] G. C. Calafiore and C. Possieri, "Output feedback Q-learning for linear-quadratic discrete-time finite-horizon control problems," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 32, no. 7, pp. 3274–3281, 2021.

[11] T. Long, E. Li, Y. Hu, L. Yang, J. Fan, Z. Liang, and R. Guo, "A vibration control method for hybrid-structured flexible manipulator based on sliding mode control and reinforcement learning," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 32, no. 2, pp. 841–852, 2021.

[12] S. A. A. Rizvi and Z. Lin, "Output feedback Q-learning control for the discrete-time linear quadratic regulator problem," *IEEE transactions on neural networks and learning systems*, vol. 30, no. 5, pp. 1523–1536, 2019.

[13] P. Deptula, Z. I. Bell, E. A. Doucette, J. W. Curtis, and W. E. Dixon, "Data-based reinforcement learning approximate optimal control for an uncertain nonlinear system with control effectiveness faults," *Automatica*, vol. 116, p. 108922, 2020.

[14] G. Hewer, "An iterative technique for the computation of the steady state gains for the discrete optimal regulator," *IEEE Transactions on Automatic Control*, vol. 16, no. 4, pp. 382–384, 1971.

[15] 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.

[16] E. Noghreian and H. R. Koofigar, "Robust tracking control of robot manipulators with friction and variable loads without velocity measurement: A switched control strategy," *Proceedings of the Institution of Mechanical Engineers, Part I: Journal of Systems and Control Engineering*, vol. 235, no. 4, pp. 532–539, 2021.

[17] A. Saleki and M. M. Fateh, "Model-free control of electrically driven robot manipulators using an extended state observer," *Computers & Electrical Engineering*, vol. 87, p. 106768, 2020.

[18] S. A. A. Rizvi and Z. Lin, "Reinforcement learning-based linear quadratic regulation of continuous-time systems using dynamic output feedback," *IEEE Transactions on Cybernetics*, vol. 50, no. 11, pp. 4670–4679, 2020.

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

[20] J. Pliego-Jiménez, M. A. Arteaga-Pérez, and M. López-Rodríguez, "Finite-time control for rigid robots with bounded input torques," *Control Engineering Practice*, vol. 102, p. 104556, 2020.

[21] C. P. Bechlioulis and G. A. Rovithakis, "Robust adaptive control of feedback linearizable MIMO nonlinear systems with prescribed performance," *IEEE Transactions on Automatic Control*, vol. 53, no. 9, pp. 2090–2099, 2008.

[22] J.-X. Zhang and G.-H. Yang, "Fault-tolerant output-constrained control of unknown euler–lagrange systems with prescribed tracking accuracy," *Automatica*, vol. 111, p. 108606, 2020.

[23] Z. Yin, J. Luo, and C. Wei, "Robust prescribed performance control for euler–lagrange systems with practically finite-time stability," *European Journal of Control*, vol. 52, pp. 1–10, 2020.

[24] J. Li, J. Li, Z. Wu, and Y. Liu, "Practical tracking control with prescribed transient performance for euler-lagrange equation," *Journal of the Franklin Institute*, vol. 357, no. 10, pp. 5809–5830, 2020.

[25] L. Xu, C. Wang, X. Cai, Y. Xu, C. Jing, and L. Hao, "Distributed formation-containment control for multiple euler–lagrange systems with guaranteed performance and unknown control directions," *International Journal of Systems Science*, vol. 51, no. 15, pp. 2781–2792, 2020.

[26] O. Elhaki and K. Shojaei, "A robust neural network approximation-based prescribed performance output-feedback controller for autonomous underwater vehicles with actuators saturation," *Engineering Applications of Artificial Intelligence*, vol. 88, p. 103382, 2020.

[27] K. Shojaei, "Three-dimensional neural network tracking control of a moving target by underactuated autonomous underwater vehicles," *Neural Computing and Applications*, vol. 31, no. 2, pp. 509–521, 2019.

[28] W. Bai, Q. Zhou, T. Li, and H. Li, "Adaptive reinforcement learning neural network control for uncertain nonlinear system with input saturation," *IEEE Transactions on Cybernetics*, vol. 50, no. 8, pp. 3433–3443, 2020.

[29] Y. Zeng, H.-K. Lam, B. Xiao, and L. Wu, "Tracking control for nonlinear systems with actuator saturation via interval type-2 t-s fuzzy framework," *IEEE Transactions on Cybernetics*, pp. 1–10, 2022, DOI: 10.1109/TCYB.2022.3167917.

[30] ——, "ℓ_2-ℓ_∞ control of discrete-time state-delay interval type-2 fuzzy systems via dynamic output feedback," *IEEE Transactions on Cybernetics*, vol. 52, no. 6, pp. 4198–4208, 2022.

[31] X. Yang and H. He, "Decentralized event-triggered control for a class of nonlinear-interconnected systems using reinforcement learning," *IEEE Transactions on Cybernetics*, vol. 51, no. 2, pp. 635–648, 2021.

[32] C. Mu, K. Wang, and T. Qiu, "Dynamic event-triggering neural learning control for partially unknown nonlinear systems," *IEEE Transactions on Cybernetics*, vol. 52, no. 4, pp. 2200–2213, 2022.

[33] X. Guo, W. Yan, and R. Cui, "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*, vol. 50, no. 11, pp. 4068–4077, 2020.

[34] S. Li, L. Ding, H. Gao, Y.-J. Liu, N. Li, and Z. Deng, "Reinforcement learning neural network-based adaptive control for state and input time-delayed wheeled mobile robots," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 50, no. 11, pp. 4171–4182, 2020.

[35] G. Liu, N. Sun, T. Yang, and Y. Fang, "Reinforcement learning-based prescribed performance motion control of pneumatic muscle actuated robotic arms with measurement noises," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 53, no. 3, pp. 1801–1812, 2023.

[36] G. Wen, W. Hao, W. Feng, and K. Gao, "Optimized backstepping tracking control using reinforcement learning for quadrotor unmanned aerial vehicle system," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 52, no. 8, pp. 5004–5015, 2022.

[37] Z. Huang, X. Xu, H. He, J. Tan, and Z. Sun, "Parameterized batch reinforcement learning for longitudinal control of autonomous land vehicles," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 49, no. 4, pp. 730–741, 2019.

[38] Z. Wang, L. Liu, H. Zhang, and G. Xiao, "Fault-tolerant controller design for a class of nonlinear MIMO discrete-time systems via online reinforcement learning algorithm," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 46, no. 5, pp. 611–622, 2016.

[39] A. Dalla Libera, D. Romeres, D. K. Jha, B. Yerazunis, and D. Nikovski, "Model-based reinforcement learning for physical systems without velocity and acceleration measurements," *IEEE Robotics and Automation Letters*, vol. 5, no. 2, pp. 3548–3555, 2020.

[40] Y. Yang, Z. Ding, R. Wang, H. Modares, and D. C. Wunsch, "Data-driven human-robot interaction without velocity measurement using off-policy reinforcement learning," *IEEE/CAA Journal of Automatica Sinica*, vol. 9, no. 1, pp. 47–63, 2022.

[41] S. Zhao, J. Wang, H. Xu, and B. Wang, "Composite observer-based optimal attitude-tracking control with reinforcement learning for hypersonic vehicles," *IEEE Transactions on Cybernetics*, vol. 53, no. 2, pp. 913–926, 2023.

[42] O. Elhaki and K. Shojaei, "Output-feedback robust saturated actor–critic multi-layer neural network controller for multi-body electrically driven tractors with n-trailer guaranteeing prescribed output constraints," *Robotics and Autonomous Systems*, vol. 154, p. 104106, 2022.

[43] 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.

[44] K. Shojaei and A. Chatraei, "A saturating extension of an output feedback controller for internally damped euler-lagrange systems," *Asian Journal of Control*, vol. 17, no. 6, pp. 2175–2187, 2015.

[45] H. K. Khalil, *Nonlinear systems, 3rd ed*. Upper Saddle River, NJ, USA: Prentice-Hall Press, 1996.

[46] M. M. Polycarpou, "Stable adaptive neural control scheme for nonlinear systems," *IEEE Transactions on Automatic control*, vol. 41, no. 3, pp. 447–451, 1996.

[47] C. T. A. F. L. Lewis, D. M. Dawson, *Robot Manipulator Control: Theory and Practice.* 2nd ed. Revised and Expanded, Marcel Dekker: New York, 2004.

[48] Q. Yang and S. Jagannathan, "Reinforcement learning controller design for affine nonlinear discrete-time systems using online approximators," *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, vol. 42, no. 2, pp. 377–390, 2012.

[49] Y. Luo, Q. Sun, H. Zhang, and L. Cui, "Adaptive critic design-based robust neural network control for nonlinear distributed parameter systems with unknown dynamics," *Neurocomputing*, vol. 148, pp. 200–208, 2015.

[50] Z. Zheng, L. Ruan, M. Zhu, and X. Guo, "Reinforcement learning control for underactuated surface vessel with output error constraints and uncertainties," *Neurocomputing*, vol. 399, pp. 479–490, 2020.

[51] X. Bu, B. Jiang, and H. Lei, "Nonfragile quantitative prescribed performance control of waverider vehicles with actuator saturation," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 58, no. 4, pp. 3538–3548, 2022.

Omid Elhaki was born in Esfahan, Iran, on May 19, 1991. He successfully received his B.S. degree in electrical engineering from Iran University of Applied Sciences and Technology (UAST) in 2015. Then, he received his M.S. degree in electrical engineering in automatic control from Najafabad Branch, Islamic Azad University (IAUN) in winter 2018. His research fields are adaptive, robust and intelligent control of nonlinear systems and control of autonomous mobile robots including wheeled mobile robots, unmanned surface vessels, unmanned aerial vehicles, and autonomous underwater vehicles.

Khoshnam Shojaei was born in Esfahan, Iran. He received the B.S., M.S. and Ph.D. degrees (with distinction) from the Iran University of Science and Technology (IUST), Tehran, Iran, in 2004, 2007, and 2011, respectively, all in electrical engineering. Currently, he is an Associate Professor with the Najafabad Branch, Islamic Azad University, Najafabad, Iran. His research interests include the adaptive control of nonlinear systems, control of autonomous robots including land, air and ocean vehicles, navigation of mobile robots, and multiagent systems.

Abbas Chatraei received the M.S. degree in electrical engineering from Islamic Azad University, Najafabad, Iran in 2002. In 2011, he received the Ph.D. degree in optimal control of industrial robots from the Faculty of Mechatronics, Informatics, and Interdisciplinary Studies, Technical University of Liberec (TUL), Liberec, Czech Republic. He is currently an Assistant Professor with the Department of Electrical Engineering, Najafabad Branch, Islamic Azad University, Iran. His research interests include robot's nonlinear control and system identification.

Allahyar Montazeri received the B.S. degree in electrical engineering from Tehran University, Tehran, Iran in 2000, and the M.Sc. and Ph.D. degrees in electrical engineering from Iran University of Science and Technology, Tehran, Iran, in 2002, and 2009, respectively. Between 2010 and 2011, he joined the Fraunhofer Institute, Germany as a research fellow, and then carried on his research with Fraunhofer IDMT and control engineering group, Ilmenau University, Germany during the years 2011 to 2013. He has been the visiting research scholar with the control engineering group at ETH Zurich, Switzerland and the chemical engineering group in Norwegian University of Science and Technology (NTNU), Trondheim, Norway. Since 2013, he has been appointed as assistant professor at School of Engineering at Lancaster University, United Kingdom and currently he is working there as an associate professor. His research interests cover a wide range of areas on control theory and digital signal processing. Particularly he is interested in Adaptive Signal Processing and Control, Robust Control, Linear and Nonlinear System Identification, Estimation Theory, and Evolutionary Computing and Optimization with applications in Robotics and Autonomous Systems as well as Active Noise and Vibration Control Systems. Dr. Montazeri is recipient of the European research consortium on informatic and mathematics (ERCIM) and Humboldt research awards in 2010 and 2011, respectively. He is also Fellow of Higher Education Academy. His research is funded by different councils and industries in UK such as Engineering and Physical Research Council, Sellafield Ltd, National Nuclear laboratory, and Nuclear Decommissioning Authority. He is currently Associate Editor of the journal Frontiers in Robotics and AI and serves as IFAC Technical Committees 'Adaptive and Learning Systems' and 'Modelling, Identification, and Signal Processing'.