Reinforcement Learning architecture Based Fixed-Time Optimal Control for Hydraulic Support Cylinder System

Yujun Zhang*, Yunyun Chai, Dongxiang Gao

Abstract—**This paper takes the hydraulic support cylinder system (HSCS) as the research object and conducts an in-depth discussion on it. Firstly, based on the working principle and physical characteristics of the HSCS, a model of the electro-hydraulic control system of the hydraulic support under the control of a proportional valve is established. Then, by combining reinforcement learning technology and fixed-time algorithm theory, an adaptive fixed-time optimal control strategy is proposed, aiming to enhance the robustness and convergence speed of the system. Through the adoption of a simplified optimal backstepping design method, an adaptive fixed-time optimal controller is constructed to ensure that the performance of each subsystem reaches the optimum and that all signals of the closed-loop system achieve stability within a fixed time. Finally, the effectiveness and feasibility of the proposed method are verified through a simulation example.**

Index Terms—**Hydraulic support cylinder, Proportional valve, Reinforcement learning, Fixed-time optimal control**

I. Introduction

IN the intelligent transformation of coal mines, the collaborative intelligent control system for the “three machines” in fully mechanized mining faces has emerged as a core technology to ensure safety and enhance efficiency. Among its key components, the electro-hydraulic control cylinder of the hydraulic support exhibits strong nonlinear hysteresis characteristics and time-varying parameters, posing challenges such as complex dynamic modeling and insufficient servo tracking accuracy [1, 2]. The advancement of adaptive nonlinear control theory offers an effective solution for achieving intelligence in such industrial systems [3].

Optimal control for nonlinear systems is one of the core aspects of modern control theory, focusing on optimizing the performance indicators of control systems [4]. It integrates fundamental conditions and methods derived from practical problems, with the research object being controlled dynamic systems or motion processes. The goal is to identify the best control strategy among the allowable ones, ensuring the system achieves optimal performance when transitioning from the initial state to the target state [5]. With the rapid advancement of digital technology and electronic computers, optimal control has been extensively applied in production, military, and economic activities, playing a crucial role in the national economy and national defense. Theoretically, solving the optimal control problem is equivalent to solving the Hamilton-Jacobi-Bellman (HJB) equation [6], but due to its strong nonlinearity and dynamic uncertainty, direct analytical solutions are challenging. To address this issue, reinforcement learning (RL) and adaptive dynamic programming (ADP) have emerged as effective approaches. Initially proposed by Werbos for discrete systems [7], RL and ADP were later extended to continuous systems [8, 9], though they remain limited to affine nonlinear systems. For the control of nonlinear mismatched systems, an optimal control method based on the backstepping framework was proposed in [10], ensuring the optimization of each subsystem. To reduce complexity and relax the continuous excitation condition, the optimal backstepping control strategy was further simplified in [11–13].

Although previous studies have made significant progress, they primarily focused on scenarios involving infinite time intervals. However, convergence time remains a critical issue in controller design [14]. To improve the convergence speed of system stability, [15] proposed a criterion for finite-time stability and applied it to various control systems. Nevertheless, the convergence time of finite-time control depends on the system’s initial conditions, which are often difficult to obtain in practical applications, making it challenging to estimate the convergence time accurately [16, 17]. Consequently, [18] introduced the theory of fixed-time stability and developed numerous fixed-time control methods that do not rely on the system’s initial values [19–21]. Based on our research, there are currently limited studies on fixed-time optimal control.

This paper considers the working principle of the electro-hydraulic control cylinder of the hydraulic support as the control object and integrates reinforcement learning algorithms with adaptive control theory to optimize the overall stability and robustness of system operation. By conducting modeling analysis, controller design, and simulation verification for the working process of the electro-hydraulic control cylinder system of the hydraulic support, an adaptive control strategy is proposed to ensure that the system state reaches optimality.

II. Model Description and Preliminaries

Assuming that the electro-hydraulic control cylinder system of the hydraulic support operates in the direction

Manuscript received March 7, 2025; revised May 7, 2025.

This work was supported by the Key Laboratory of Internet of Things Application Technology on Intelligent Construction, Liaoning Province (2021JH13/10200051)

Yujun Zhang is a Professor of School of Computer and Software Engineering, University of Science and Technology Liaoning, Anshan 114051, China. (Corresponding author, e-mail: 1997zyj@163.com).

Yunyun Chai is a graduate student of School of Computer and Software Engineering, University of Science and Technology Liaoning, Anshan 114051, China. (e-mail: 1633006219@qq.com).

Dongxiang Gao is a doctoral student at the School of Electronic and Information Engineering, University of Science and Technology Liaoning, Anshan 114051, China. (e-mail: lkdgao1997@163.com).

shown in Figure 1 during operation, according to the force balance equation and the flow balance equation, its dynamic equation can be expressed as:

$$M\ddot{X}_p = p_1 A_1 - p_2 A_2 - B\dot{X}_p + F_T \tag{1}$$

$$\begin{cases} 2p_1 = p_s + p_r \\ 2p_2 = p_s - p_r \end{cases} \tag{2}$$

where m is the load mass, X_p is the displacement of the hydraulic support cylinder, B is the damping coefficient, and F_T is the external force acting on the hydraulic support cylinder. A_1 and A_2 are the effective areas of the non-symmetric cylinder's rodless chamber and rod chamber respectively. p_1 and p_2 are the pressures at the oil cylinder's inlet and outlet respectively. p_s and p_r are the supply and return oil pressures respectively.

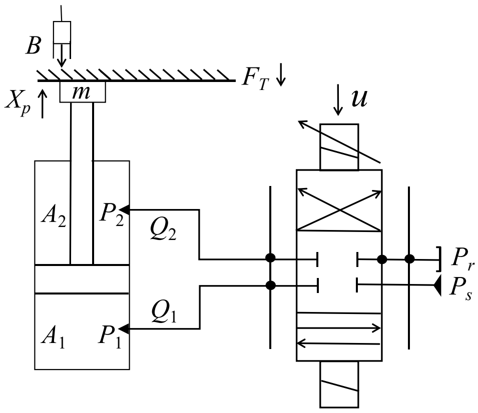

Fig. 1: Model diagram of hydraulic support cylinder.

Subsequently, by introducing a state space transformation $\xi_1 = mX_p$, $\xi_2 = m\dot{X}_p$, and $\xi_3 = p_1A_1 - p_2A_2$, (1) can be transformed into a nonlinear system of the following form:

$$\begin{cases} \dot{\xi}_1 = m\xi_2 \\ \dot{\xi}_2 = \xi_3 - \dfrac{B}{m}\xi_2 - F_T - mg \\ \dot{\xi}_3 = \gamma_1 u - \gamma_2 \xi_2 - \gamma_3 (p_1 - p_2) \\ y = \xi_1 \end{cases} \tag{3}$$

where $\xi = [\xi_1, \xi_2, \xi_3]^T \in \mathbb{R}^3$ denotes the state variables. u and y are the control input and output, respectively. The u is a voltage signal ranging from $0 - 10\ V$, which satisfies the linear relationship $x_v = k_v u$. In this context, x_v denotes the displacement of the spool in the proportional valve, and k_v represents a positive constant. Additionally, the parameters γ_1, γ_2, and γ_3 are defined as follows:

$$\begin{cases} \gamma_1 = (\dfrac{A_1R_1}{V_1 + \frac{A_1}{m}\xi_1} + \dfrac{A_2R_2}{V_2 - \frac{A_2}{m}\xi_1})\beta_e k_q k_v \\ \gamma_2 = (\dfrac{A_1^2}{mV_1 + A_1\xi_1} + \dfrac{A_2^2}{mV_2 - A_2\xi_1})\beta_e \\ \gamma_3 = (\dfrac{A_1}{V_1 + \frac{A_1}{m}\xi_1} + \dfrac{A_2}{V_2 - \frac{A_2}{m}\xi_1})\beta_e C_t \end{cases} \tag{4}$$

where $R_1 = \sqrt{p_s + sign(x_v)(p_s - 2p_1)}$ and $R_2 = \sqrt{p_s + sign(x_v)(2p_2 - p_s)}$. β_e represents the effective volume elastic modulus of the hydraulic system, C_t is the leakage coefficient within the hydraulic cylinder, K_q is the flow gain of the proportional valve, and V_1 and V_2 are the initial volumes of the two chambers of the hydraulic cylinder.

Definition 1 For system (3), if the control protocol u is continuous and satisfies $u(0) = 0$, then $u \in \Omega$ constitutes an admissible control strategy. This control protocol not only stabilizes the controlled system but also ensures that the performance cost function remains finite. In this context, Ω denotes the set of all admissible controls.

Control Objective: For HSCS, a fixed-time optimal control strategy based on reinforcement learning is proposed to ensure that: 1) the output signal y can precisely track the reference signal y_r; 2) while conserving communication resources, all signals in the system remain bounded within a fixed time.

Assumption 1. The reference signal y_r and its derivative $\dot{y}_r$ are bounded.

Lemma 1 [22] For $0 < p < 1$ and $q > 1$, there is

$$-\sum_{i=1}^{n}\tilde{W}_i^2 \le -\left(\sum_{i=1}^{n}\frac{1}{2}\tilde{W}_i^2\right)^p - n^{1-q}\left(\sum_{i=1}^{n}\frac{1}{2}\tilde{W}_i^2\right)^q + \varrho_W \tag{5}$$

where $\varrho_W = (1-p)p^{\frac{p}{1-p}} + \sum_{i=1}^{n}\left(\frac{\mu_{Wi}^2}{2}\right)^q$, with unknown constant μ_{Wi} exists to ensure that $|\tilde{W}_i| < \mu_{Wi}$.

Lemma 2 [23] Let $f(x)$ be a continuous function defined on a compact set Ω_x. Then for $\forall \varepsilon > 0$, there exist the NN $W^T\Psi(x)$ such that

$$\sup_{x\in\Omega_x} |f(x) - W^T\Psi(x)| \le \varepsilon \tag{6}$$

where $W = [W_1, W_2, \ldots, W_m]^T \in \mathbb{R}^m$ is the weight vector and $\Psi(x) = [\psi_1(x), \psi_2(x), \ldots, \psi_m(x)]^T$ is the NN basis function with $m > 1$ is the number of NN rules. $\psi_i(x) = exp[-\|x - \xi_i\|^2/\vartheta_i^2], i = 1, 2, \ldots, m$ is the Gaussian function, where ϑ_i and $\xi_i = [\xi_{i1}, \xi_{i2}, \ldots, \xi_{im}]^T$ represent the width and center, respectively. The optimal parameter vector W^* of NN is defined as

$$W^* = \arg\min_{W\in\mathbb{R}^m}\{\sup_{x\in\Omega_x} |f(x) - W^T\Psi(x)|\} \tag{7}$$

Therefore, the continuous function $f(x)$ can be expressed as

$$f(x) = W^{*T}\Psi(x) + \varepsilon(x) \tag{8}$$

where $\varepsilon(x)$ is the NN approximation error, which can be bounded by $|\varepsilon(x)| \le \bar{\varepsilon}$, where $\bar{\varepsilon}$ is a positive constant. It should be pointed out that since W^* is an analytical quantity, it needs to be estimated later for practical use.

Lemma 3 [24] For the system (3), if there is a positive definite and radially unbounded function $V(\xi(t))$ such that

$$\begin{gathered} \dot{V}(\xi(t)) \le -aV^p(\xi(t)) - bV^q(\xi(t)) + c, t \ge 0 \\ \sigma < \min\{(1-\varsigma)a, (1-\varsigma)b\} \end{gathered} \tag{9}$$

where $a > 0$, $b > 0$, $0 < p < 1$, $q > 1$, $0 < \varsigma < 1$ and $c > 0$ are design parameters, $\underline{\Lambda}$ and $\overline{\Lambda}$ are k_∞ functions, and $V(\xi(t))$ satisfies condition $\underline{\Lambda}\|\xi(t)\| \le V(\xi(t)) \le \overline{\Lambda}\|\xi(t)\|$. If these conditions are met, it means that the nonlinear system (3) has fixed-time stability, and the upper bound of the convergence time T_{max} can be expressed in the following form.

$$T \le T_{\max} = \frac{1}{a(1-p)\varsigma} + \frac{1}{b(q-1)\varsigma} \tag{10}$$

III. MAIN RESULT

In this section, we will combine reinforcement learning algorithms and fixed-time theory to design an optimal backstepping control strategy under the critic-actor architecture, thereby constructing an optimal controller.

A. Fixed-time optimized backstepping controller design

The following introduces an critic-actor architecture based on the reinforcement learning algorithm, which adopts a simplified fixed-time optimal backstepping method design to construct an optimal controller. First, consider the following tracking error coordinate transformation:

$$\begin{aligned} z_1 &= \xi_1 - y_r \\ z_2 &= m\xi_2 - \hat{\alpha}_1^* \\ z_3 &= \xi_3 - \hat{\alpha}_2^* \end{aligned} \tag{11}$$

where y_r is selected as the reference signal and set to $0.2\sin(t)$. α_{i-1} and $\hat{\alpha}_{i-1}^*$ represent the virtual control and actual optimal virtual control correspondingly.

Step 1: From (3) and (11), the derivative of z_1 can be calculated

$$\dot{z}_1 = m\xi_2 - \dot{y}_r \tag{12}$$

The optimal performance index function is chosen as

$$J_1(z_1) = \int_t^\infty h_1\Big(z_1(v), \alpha_1\big(z_1(v)\big)\Big)dv \tag{13}$$

where $h_1(z_1, \alpha_1) = z_1^2 + \alpha_1^2$ is the cost function, and let the optimal virtual control α_1^* replace α_1 in (13), the optimal performance index function can be obtained

$$\begin{aligned} J_1^*(z_1) &= \int_t^\infty h_1(z_1(v), \alpha_1^*(z_1(v)))dv \\ &= \min_{\alpha_1 \in \Omega_{z_1}} \{\int_t^\infty h_1\Big(z_1(v), \alpha_1\big(z_1(v)\big)\Big)dv\} \end{aligned} \tag{14}$$

Replace ξ_2 in (12) with the optimal virtual control α_1^*, and subsequently define the HJB equation associated with (12) and (14) as

$$H_1(z_1, \alpha_1^*, \frac{\mathrm{d}J_1^*}{\mathrm{d}z_1}) = z_1^2 + {\alpha_1^*}^2 + \frac{\mathrm{d}J_1^*}{\mathrm{d}z_1}(\alpha_1^* - \dot{y}_r) = 0 \tag{15}$$

The optimal virtual control α_1^* can be computed by solving $\partial H_1/\partial\alpha_1^* = 0$ as

$$\alpha_1^* = -\frac{1}{2}\frac{\mathrm{d}J_1^*(z_1)}{\mathrm{d}z_1} \tag{16}$$

Then, $\frac{\mathrm{d}J_1^*(z_1)}{\mathrm{d}z_1}$ is decomposed into

$$\frac{\mathrm{d}J_1^*(z_1)}{\mathrm{d}z_1} = 2c_1 z_1^{2p-1} + 2k_1 z_1^{2q-1} + \frac{5}{2}z_1 + J_1^o(\xi_1, z_1) \tag{17}$$

where $0 < p < 1$, $q > 1$, $c_1 > 0$ and $k_1 > 0$ are design parameters. $J_1^o(\xi_1, z_1) = -2c_1 z_1^{2p-1} - 2k_1 z_1^{2q-1} - \frac{5}{2}z_1 + \frac{\mathrm{d}J_1^*(z_1)}{\mathrm{d}z_1} \in \mathbb{R}$ is a continuous function, and substituting (17) into (16) has

$$\alpha_1^* = -c_1 z_1^{2p-1} - k_1 z_1^{2q-1} - \frac{5}{4}z_1 - \frac{1}{2}J_1^o(\xi_1, z_1) \tag{18}$$

Since $J_1^o(\xi_1, z_1)$ is continuous unknown function, it can be approximated by NN as follows:

$$J_1^o(\xi_1, z_1) = W_{J1}^{*T}\Psi_{J1}(\xi_1, z_1) + \varepsilon_{J1}(\xi_1, z_1) \tag{19}$$

where W_{J1}^* represents the ideal weight vector, $\Psi_{J1}(\xi_1, z_1)$ is the basis function vector, and $\varepsilon_{J1}(\xi_1, z_1)$ represents the approximation error bounded by $\|\varepsilon_{J1}(\xi_1, z_1)\| \le \bar{\varepsilon}_{J1}$ as arbitrarily small. Then, (17) and (18) can be reorganized as

$$\begin{aligned} \frac{\mathrm{d}J_1^*(z_1)}{\mathrm{d}z_1} =& 2c_1 z_1^{2p-1} + 2k_1 z_1^{2q-1} + \frac{5}{2}z_1 \\ &+ W_{J1}^{*T}\Psi_{J1} + \varepsilon_{J1} \end{aligned} \tag{20}$$

$$\begin{aligned} \alpha_1^* =& -c_1 z_1^{2p-1} - k_1 z_1^{2q-1} - \frac{5}{4}z_1 \\ &- \frac{1}{2}W_{J1}^{*T}\Psi_{J1} - \frac{1}{2}\varepsilon_{J1} \end{aligned} \tag{21}$$

Since W_{J1}^* is unknown constant vector, the optimal virtual control (21) is not available for the controlled system. To derive the effective optimized virtual control, the following RL algorithm with critic and actor is performed.

$$\frac{\mathrm{d}\hat{J}_1^*(z_1)}{\mathrm{d}z_1} = 2c_1 z_1^{2p-1} + 2k_1 z_1^{2q-1} + \frac{5}{2}z_1 + \hat{W}_{c1}^T\Psi_{J1} \tag{22}$$

$$\hat{\alpha}_1^* = -c_1 z_1^{2p-1} - k_1 z_1^{2q-1} - \frac{5}{4}z_1 - \frac{1}{2}\hat{W}_{a1}^T\Psi_{J1} \tag{23}$$

where $\frac{\mathrm{d}\hat{J}_1^*(z_1)}{\mathrm{d}z_1}$ and $\hat{\alpha}_1^*$ are the estimates of $\frac{\mathrm{d}J_1^*(z_1)}{\mathrm{d}z_1}$ and α_1^*, respectively. $\hat{W}_{c1}^T\Psi_{J1}$ and $\hat{W}_{a1}^T\Psi_{J1}$ are the NN weight vectors of critic and actor, respectively.

Following this, the weight vectors of the neural networks for both the critic and actor are trained according to the respective adaptive laws outlined below.

$$\dot{\hat{W}}_{c1} = -\kappa_{c1}\Psi_{J1}(z_1)\Psi_{J1}^T(z_1)\hat{W}_{c1} \tag{24}$$

$$\begin{aligned} \dot{\hat{W}}_{a1} =& -\Psi_{J1}(z_1)\Psi_{J1}^T(z_1)\big(\kappa_{a1}(\hat{W}_{a1} \\ &- \hat{W}_{c1}) + \kappa_{c1}\hat{W}_{c1}\big) \end{aligned} \tag{25}$$

where $\kappa_{c1} > 0$ and $\kappa_{a1} > 0$ represent critic and actor design parameters, while κ_{c1} and κ_{a1} satisfy $\kappa_{a1} > \frac{1}{2}$, $\kappa_{a1} > \frac{\kappa_{c1}}{2}$.

Using (23), (12) can be rewritten as

$$\begin{aligned} \dot{z}_1 =& -c_1 z_1^{2p-1} - k_1 z_1^{2q-1} + z_2 - \frac{1}{2}\hat{W}_{a1}^T\Psi_{J1} \\ &- \frac{5}{4}z_1 - \dot{y}_r \end{aligned} \tag{26}$$

For the first backstepping step, the Lyapunov function V_1 is designed as follows:

$$V_1 = \frac{1}{2}z_1^2 + \frac{1}{2}\tilde{W}_{c1}^T\tilde{W}_{c1} + \frac{1}{2}\tilde{W}_{a1}^T\tilde{W}_{a1} \tag{27}$$

where $\tilde{W}_{c1} = W_{J1}^* - \hat{W}_{c1}$ and $\tilde{W}_{a1} = W_{J1}^* - \hat{W}_{a1}$ are the estimation errors of the critic and the actor, respectively.

Then, the derivative of V_1 is

$$\begin{aligned} \dot{V}_1 =& z_1\big(-c_1 z_1^{2p-1} - k_1 z_1^{2q-1} + z_2 - \frac{1}{2}\hat{W}_{a1}^T\Psi_{J1} - \dot{y}_r\big) \\ &+ \kappa_{c1}\tilde{W}_{c1}^T\Psi_{J1}\Psi_{J1}^T\hat{W}_{c1} + \tilde{W}_{a1}^T\Psi_{J1}\Psi_{J1}^T\big(\kappa_{a1}(\hat{W}_{a1} \\ &- \hat{W}_{c1}) + \kappa_{c1}\hat{W}_{c1}\big) \end{aligned} \tag{28}$$

The Young's inequality yields the following results

$$\begin{aligned} z_1 z_2 &\le \frac{1}{2}z_1^2 + \frac{1}{2}z_2^2 \\ -z_1\dot{y}_r &\le \frac{1}{2}z_1^2 + \frac{1}{2}\dot{y}_r^2 \\ -\frac{1}{2}z_1\hat{W}_{a1}^T\Psi_{J1} &\le \frac{1}{4}z_1^2 + \frac{1}{4}\hat{W}_{a1}^T\Psi_{J1}\Psi_{J1}^T\hat{W}_{a1} \end{aligned} \tag{29}$$

Along with (28) and (29), we can calculate:

$$\begin{aligned}\dot{V}_1 \leq &-c_1 z_1^{2p} - k_1 z_1^{2q} + \kappa_{c1}\tilde{W}_{c1}^T\Psi_{J1}\Psi_{J1}^T\hat{W}_{c1} \\ &+ \kappa_{a1}\tilde{W}_{a1}^T\Psi_{J1}\Psi_{J1}^T\hat{W}_{a1} + \frac{1}{2}z_2^2 + \frac{1}{2}\dot{y}_r^2 \\ &+ (\kappa_{c1}-\kappa_{a1})\tilde{W}_{a1}^T\Psi_{J1}\Psi_{J1}^T\hat{W}_{c1} \\ &+ \frac{1}{4}\hat{W}_{a1}^T\Psi_{J1}\Psi_{J1}^T\hat{W}_{a1}\end{aligned} \tag{30}$$

Based on $\tilde{W}_{c1} = W_{J1}^* - \hat{W}_{c1}$, $\tilde{W}_{a1} = W_{J1}^* - \hat{W}_{a1}$ and Young's inequality, we have

$$\begin{aligned}\tilde{W}_{c1}^T\Psi_{J1}\Psi_{J1}^T\hat{W}_{c1} =&\frac{1}{2}W_{J1}^{*T}\Psi_{J1}\Psi_{J1}^T W_{J1}^* - \frac{1}{2}\tilde{W}_{c1}^T\Psi_{J1} \\ &\times \Psi_{J1}^T\tilde{W}_{c1} - \frac{1}{2}\hat{W}_{c1}^T\Psi_{J1}\Psi_{J1}^T\hat{W}_{c1} \\ \tilde{W}_{a1}^T\Psi_{J1}\Psi_{J1}^T\hat{W}_{a1} =&\frac{1}{2}W_{J1}^{*T}\Psi_{J1}\Psi_{J1}^T W_{J1}^* - \frac{1}{2}\tilde{W}_{a1}^T\Psi_{J1} \\ &\times \Psi_{J1}^T\tilde{W}_{a1} - \frac{1}{2}\hat{W}_{a1}^T\Psi_{J1}\Psi_{J1}^T\hat{W}_{a1} \\ \tilde{W}_{a1}^T\Psi_{J1}\Psi_{J1}^T\hat{W}_{c1} \leq& -\frac{1}{2}\tilde{W}_{a1}^T\Psi_{J1}\Psi_{J1}^T\tilde{W}_{a1} \\ &- \frac{1}{2}\hat{W}_{c1}^T\Psi_{J1}\Psi_{J1}^T\hat{W}_{c1}\end{aligned} \tag{31}$$

Subsequently, we can acquire

$$\begin{aligned}\dot{V}_1 \leq &-c_1 z_1^{2p} - k_1 z_1^{2q} - \frac{\kappa_{c1}}{2}\tilde{W}_{c1}^T\Psi_{J1}\Psi_{J1}^T\tilde{W}_{c1} \\ &- (\kappa_{a1} - \frac{\kappa_{c1}}{2})\tilde{W}_{a1}^T\Psi_{J1}\Psi_{J1}^T\tilde{W}_{a1} \\ &- \frac{\kappa_{a1}}{2}\hat{W}_{c1}^T\Psi_{J1}\Psi_{J1}^T\hat{W}_{c1} - (\frac{\kappa_{a1}}{2} - \frac{1}{4}) \\ &\times \hat{W}_{a1}^T\Psi_{J1}\Psi_{J1}^T\hat{W}_{a1} + \frac{1}{2}z_2^2 + \frac{1}{2}\dot{y}_r^2 \\ &+ \frac{\kappa_{c1}+\kappa_{a1}}{2}W_{J1}^{*T}\Psi_{J1}\Psi_{J1}^T W_{J1}^*\end{aligned} \tag{32}$$

The following inequality holds when $\lambda_{\Psi_{J1}}^{\min}$ is the minimum eigenvalue of $\Psi_{J1}\Psi_{J1}^T$.

$$\begin{aligned}-\tilde{W}_{c1}^T\Psi_{J1}\Psi_{J1}^T\tilde{W}_{c1} \leq -\lambda_{\Psi_{J1}}^{\min}\tilde{W}_{c1}^T\tilde{W}_{c1} \\ -\tilde{W}_{a1}^T\Psi_{J1}\Psi_{J1}^T\tilde{W}_{a1} \leq -\lambda_{\Psi_{J1}}^{\min}\tilde{W}_{a1}^T\tilde{W}_{a1}\end{aligned} \tag{33}$$

According to the design parameters $\kappa_{a1} > \frac{\kappa_{c1}}{2}$ and $\kappa_{a1} > \frac{1}{2}$, as well as (33), it can yield

$$\begin{aligned}\dot{V}_1 \leq &-c_1 z_1^{2p} - k_1 z_1^{2q} - \frac{\kappa_{c1}}{2}\lambda_{\Psi_{J1}}^{\min}\tilde{W}_{c1}^T\tilde{W}_{c1} \\ &- (\kappa_{a1} - \frac{\kappa_{c1}}{2})\lambda_{\Psi_{J1}}^{\min}\tilde{W}_{a1}^T\tilde{W}_{a1} + \frac{1}{2}z_2^2 + \sigma_1\end{aligned} \tag{34}$$

where $\sigma_1 = \frac{1}{2}\dot{y}_r^2 + \frac{\kappa_{c1}+\kappa_{a1}}{2}W_{J1}^{*T}\Psi_{J1}\Psi_{J1}^T W_{J1}^*$. Since all the terms in σ_1 are bounded, there exists a positive constant $\overline{\sigma}_1$ such that $|\sigma_1| \leq \overline{\sigma}_1$.

Step 2 : The derivative of z_2 is calculated in a similar manner.

$$\begin{aligned}\dot{z}_2 &= \dot{\xi}_2 - \dot{\alpha}_1^* \\ &= \xi_3 - \frac{B}{m}\xi_2 - F_T - mg - \dot{\alpha}_1^*\end{aligned} \tag{35}$$

Among them, $-\frac{B}{m}\xi_2 - F_T - mg$ can be approximated by NN as $W_{f2}^{*T}\Psi_{f2}(\xi) + \varepsilon_{f2}(\xi)$, there exists a positive constant $\overline{\varepsilon}_{f2}$ such that $|\varepsilon_{f2}(\xi)| \leq \overline{\varepsilon}_{f2}$. Then, the selection of the most suitable integral cost function is detailed as follows:

$$\begin{aligned}J_2^*(z_2) &= \int_t^\infty h_2\Big(z_2(v), \alpha_2^*\big(z_2(v)\big)\Big)dv \\ &= \min_{\alpha_2 \in \Omega_{z_2}}\{\int_t^\infty h_2\Big(z_2(v), \alpha_2\big(z_2(v)\big)\Big)dv\}\end{aligned} \tag{36}$$

where $h_2(z_2, \alpha_2) = z_2^2 + \alpha_2^2$ is the cost function, α_2^* represents the optimal controller.

Based on (36), the HJB equation is constructed as

$$\begin{aligned}H_2(z_2, \alpha_2^*, \frac{\mathrm{d}J_2^*}{\mathrm{d}z_2}) =& z_2^2 + \alpha_2^{*2} + \frac{\mathrm{d}J_2^*}{\mathrm{d}z_2}\big(\alpha_2^* + W_{f2}^{*T}\Psi_{f2}(\xi) \\ &+ \varepsilon_f(\xi) - \dot{\alpha}_1^*\big) = 0\end{aligned} \tag{37}$$

The same as before, we can solve for $\partial H_2/\partial\alpha_2^* = 0$ as

$$\alpha_2^* = -\frac{1}{2}\frac{\mathrm{d}J_2^*(z_2)}{\mathrm{d}z_2} \tag{38}$$

Then, $\frac{\mathrm{d}J_2^*(z_2)}{\mathrm{d}z_2}$ can be factored as

$$\begin{aligned}\frac{\mathrm{d}J_2^*(z_2)}{\mathrm{d}z_2} =& 2c_2 z_2^{2p-1} + 2k_2 z_2^{2q-1} + 2W_{f2}^{*T}\Psi_{f2} + 2\varepsilon_{f2} \\ &+ \frac{9}{2}z_2 + J_2^o(\xi_2, z_2)\end{aligned} \tag{39}$$

where $0 < p < 1$, $q > 1$, $c_2 > 0$ and $k_2 > 0$ are design parameters. $J_2^o(\xi_2, z_2) = -2c_2 z_2^{2p-1} - 2k_2 z_2^{2q-1} - 2W_{f2}^{*T}\Psi_{f2} - 2\varepsilon_{f2} - \frac{9}{2}z_2 + \frac{\mathrm{d}J_2^*(z_2)}{\mathrm{d}z_2}$ is a continuous function, and the α_2^* can be expressed as

$$\begin{aligned}\alpha_2^* =& -c_2 z_2^{2p-1} - k_2 z_2^{2q-1} - W_{f2}^{*T}\Psi_{f2} - \varepsilon_{f2} \\ &- \frac{9}{4}z_2 - \frac{1}{2}J_2^o(\xi_2, z_2)\end{aligned} \tag{40}$$

Since $J_2^o(\xi_2, z_2)$ is unknown continuous term, it can also be approximated using NN as follows:

$$J_2^o(\xi_2, z_2) = W_{J2}^{*T}\Psi_{J2}(z_2) + \varepsilon_{J2}(\xi_2, z_2) \tag{41}$$

where W_{J2}^* is the ideal weight vector, $\Psi_{J2}(z_2)$ is the NN basis function vector, and the NN approximation error $\varepsilon_{J2}(z_2)$ is bounded.

Similarly, we can derive the following conclusion

$$\begin{aligned}\frac{\mathrm{d}J_2^*(z_2)}{\mathrm{d}z_2} =& 2c_2 z_2^{2p-1} + 2k_2 z_2^{2q-1} + 2W_{f2}^{*T}\Psi_{f2} \\ &+ \frac{9}{2}z_2 + W_{J2}^{*T}\Psi_{J2} + \varepsilon_2\end{aligned} \tag{42}$$

$$\begin{aligned}\alpha_2^* =& -c_2 z_2^{2p-1} - k_2 z_2^{2q-1} - W_{f2}^{*T}\Psi_{f2} \\ &- \frac{9}{4}z_2 - \frac{1}{2}W_{J2}^{*T}\Psi_{J2} - \frac{1}{2}\varepsilon_2\end{aligned} \tag{43}$$

where $\varepsilon_2 = 2\varepsilon_{f2} + \varepsilon_{J2}$.

The optimal control (43), however, remains unattainable, necessitating the execution of an RL algorithm featuring both a critic and an actor to acquire viable control signal.

$$\begin{aligned}\frac{\mathrm{d}\hat{J}_2^*(z_2)}{\mathrm{d}z_2} =& 2c_2 z_2^{2p-1} + 2k_2 z_2^{2q-1} + 2\hat{W}_{f2}^T\Psi_{f2} \\ &+ \frac{9}{2}z_2 + \hat{W}_{c2}^T\Psi_{J2}\end{aligned} \tag{44}$$

$$\begin{aligned}\hat{\alpha_2}^* =& -c_2 z_2^{2p-1} - k_2 z_2^{2q-1} - \hat{W}_{f2}^T\Psi_{J2} \\ &- \frac{9}{4}z_2 - \frac{1}{2}\hat{W}_{a2}^T\Psi_{J2}\end{aligned} \tag{45}$$

where $\frac{\mathrm{d}\hat{J}_2^*(z_2)}{\mathrm{d}z_2}$ and $\hat{\alpha_2}^*$ are the estimate of $\frac{\mathrm{d}J_2^*(z_2)}{\mathrm{d}z_2}$ and α_2^*, respectively. $\hat{W}_{c2}^T\Psi_{J2}$ and $\hat{W}_{a2}^T\Psi_{J2}$ are the NN weight vectors of critic and actor, respectively.

Same as the first step, the corresponding three adaptive update laws are designed as follows:

$$\dot{\hat{W}}_{f2} = \Gamma_{f2}\Psi_{J2} - \kappa_{f2}\hat{W}_{f2} \tag{46}$$

$$\dot{\hat{W}}_{c2} = -\kappa_{c2}\Psi_{J2}\Psi_{J2}^T\hat{W}_{c2} \tag{47}$$

$$\dot{\hat{W}}_{a2} = -\Psi_{J2}\Psi_{J2}^T\big(\kappa_{a2}(\hat{W}_{a2} - \hat{W}_{c2}) + \kappa_{c2}\hat{W}_{c2}\big) \tag{48}$$

where $\Gamma_{f2} > 0$, $\kappa_{f2} > 0$, $\kappa_{c2} > 0$ and $\kappa_{a2} > 0$ are design parameters, while κ_{c2} and κ_{a2} satisfy $\kappa_{a2} > \frac{1}{2}$, $\kappa_{a2} > \frac{\kappa_{c2}}{2}$.

According to (45), the $\dot{z}_2$ can be expressed as follows

$$\begin{aligned}\dot{z}_2 =& - c_2 z_2^{2p-1} - k_2 z_2^{2q-1} + z_3 - \frac{1}{2}\hat{W}_{a2}^T\Psi_{J2}\\ &+ \tilde{W}_{f2}^T\Psi_{f2} + \varepsilon_{f2} - \frac{9}{4}z_2 - \dot{\alpha}_1^*\end{aligned} \tag{49}$$

Subsequently, the Lyapunov function V_2 is established as

$$V_2 = \frac{1}{2}z_2^2 + \frac{1}{2\Gamma_{f2}}\tilde{W}_{f2}^T\tilde{W}_{f2} + \frac{1}{2}\tilde{W}_{c2}^T\tilde{W}_{c2} + \frac{1}{2}\tilde{W}_{a2}^T\tilde{W}_{a2} \tag{50}$$

where $\tilde{W}_{f2} = W_{f2}^* - \hat{W}_{f2}$, $\tilde{W}_{c2} = W_{J2}^* - \hat{W}_{c2}$ and $\tilde{W}_{a2} = W_{J2}^* - \hat{W}_{a2}$.

Then, the $\dot{V}_2$ can be calculated as

$$\begin{aligned}\dot{V}_2 =& z_2(-c_2 z_2^{2p-1} - k_2 z_2^{2q-1} + z_3 - \frac{1}{2}\hat{W}_{a2}^T\Psi_{J2} + \tilde{W}_{f2}^T\Psi_{f2}\\ &+ \varepsilon_{f2} - \dot{\alpha}_1^*\big) + \kappa_{f2}\tilde{W}_{f2}^T\hat{W}_{f2} + \kappa_{c2}\tilde{W}_{c2}^T\Psi_{J2}\Psi_{J2}^T\hat{W}_{c2}\\ &+ \tilde{W}_{a2}^T\Psi_{J2}\Psi_{J2}^T\big(\kappa_{a2}(\hat{W}_{a2} - \hat{W}_{c2}) + \kappa_{c2}\hat{W}_{c2}\big) - \frac{9}{4}z_2^2\end{aligned} \tag{51}$$

Using the Young's inequality, we have

$$\begin{aligned} z_2 z_3 &\le \frac{1}{2}z_2^2 + \frac{1}{2}z_3^2\\ z_2\varepsilon_{f2} &\le \frac{1}{2}z_2^2 + \frac{1}{2}\bar{\varepsilon}_{f2}^2\\ -z_2\dot{\alpha}_1^* &\le \frac{1}{2}z_2^2 + \frac{1}{2}\dot{\alpha}_1^{*2}\\ -\frac{1}{2}z_2\hat{W}_{a2}^T\Psi_{J2} &\le \frac{1}{4}z_2^2 + \frac{1}{4}\hat{W}_{a2}^T\Psi_{J2}\Psi_{J2}^T\hat{W}_{a2}\end{aligned} \tag{52}$$

Substituting (52) into (51) yields

$$\begin{aligned}\dot{V}_2 \le& -c_2 z_2^{2p} - k_2 z_2^{2q} - \frac{\kappa_{f2}}{2}\tilde{W}_{f2}^T\tilde{W}_{f2} - \frac{\kappa_{c2}}{2}\tilde{W}_{c2}^T\Psi_{J2}\Psi_{J2}^T\tilde{W}_{c2}\\ &-(\kappa_{a2} - \frac{\kappa_{c2}}{2})\tilde{W}_{a2}^T\Psi_{J2}\Psi_{J2}^T\tilde{W}_{a2} - \frac{\kappa_{a2}}{2}\hat{W}_{c2}^T\Psi_{J2}\Psi_{J2}^T\hat{W}_{c2}\\ &- (\frac{\kappa_{a2}}{2} - \frac{1}{4})\hat{W}_{a2}^T\Psi_{J2}\Psi_{J2}^T\hat{W}_{a2} + \frac{\kappa_{c2} + \kappa_{a2}}{2}W_{J2}^{*T}\Psi_{J2}\\ &\times \Psi_{J2}^T W_{J2}^* + \frac{1}{2}\bar{\varepsilon}_{f2}^2 + \frac{1}{2}\dot{\alpha}_1^{*2} + \frac{\kappa_{f2}}{2}W_{f2}^{*T}W_{f2}^* + \frac{1}{2}z_3^2\\ \le& - c_2 z_2^{2p} - k_2 z_2^{2q} - \frac{\kappa_{f2}}{2}\tilde{W}_{f2}^T\tilde{W}_{f2} - \frac{\kappa_{c2}}{2}\lambda_{\Psi_{J2}}^{\min}\tilde{W}_{c2}^T\tilde{W}_{c2}\\ &- (\kappa_{a2} - \frac{\kappa_{c2}}{2})\lambda_{\Psi_{J2}}^{\min}\tilde{W}_{a2}^T\tilde{W}_{a2} - \frac{1}{2}z_2^2 + \frac{1}{2}z_3^2 + \sigma_2\end{aligned} \tag{53}$$

where $\sigma_2 = \frac{1}{2}\bar{\varepsilon}_{f2}^2 + \frac{\kappa_{c2}+\kappa_{a2}}{2}W_{J2}^{*T}\Psi_{J2}\Psi_{J2}^T W_{J2}^* + \frac{\kappa_{f2}}{2}W_{f2}^{*T}W_{f2}^* + \frac{1}{2}\dot{\alpha}_1^{*2}$ is bounded, and there exists a positive constant $\overline{\sigma}_2$ that ensures the existence of $|\sigma_2| \le \overline{\sigma}_2$. Additionally, $\lambda_{\Psi_{J2}}^{\min}$ represents the minimum eigenvalue of $\Psi_{J2}\Psi_{J2}^T$.

Step 3 : Similarly, the derivative of z_3 is

$$\begin{aligned}\dot{z}_3 &= \dot{\xi}_3 - \dot{\alpha}_2^*\\ &= \gamma_1 u - \gamma_2\xi_2 - \gamma_3(p_1 - p_2) - \dot{\alpha}_2^*\end{aligned} \tag{54}$$

where $-\gamma_2\xi_2 - \gamma_3(p_1 - p_2)$ can be approximated by NN as $W_{f3}^{*T}\Psi_{f3} + \varepsilon_{f3}$, there exists a positive constant $\overline{\varepsilon}_{f3}$ such that $|\varepsilon_{f3}(\xi)| \le \overline{\varepsilon}_{f3}$. Then, the selection of the most suitable integral cost function is detailed as follows:

$$\begin{aligned}J_3^*(z_3) &= \int_t^\infty h_3\Big(z_3(v), u^*\big(z_3(v)\big)\Big)dv\\ &= \min_{u\in\Omega_{z_3}}\{\int_t^\infty h_3\Big(z_3(v), u\big(z_3(v)\big)\Big)dv\}\end{aligned} \tag{55}$$

where $h_3(z_3, u) = z_3^2 + u^2$ is the cost function, u^* represents the optimal controller.

Based on (54), the HJB equation is constructed as

$$\begin{aligned}H_3(z_3, u^*, \frac{\mathrm{d}J_3^*}{\mathrm{d}z_3}) =& z_3^2 + u^{*2} + \frac{\mathrm{d}J_3^*}{\mathrm{d}z_3}\big(u^* + W_{f3}^{*T}\Psi_{f3}\\ &+ \varepsilon_{f3} - \dot{\alpha}_2^*\big) = 0\end{aligned} \tag{56}$$

The same as before, we can solve for $\partial H_3/\partial u^* = 0$ as

$$u^* = -\frac{1}{2}\frac{\mathrm{d}J_3^*(z_3)}{\mathrm{d}z_3} \tag{57}$$

Then, $\frac{\mathrm{d}J_3^*(z_3)}{\mathrm{d}z_3}$ can be factored as

$$\begin{aligned}\frac{\mathrm{d}J_3^*(z_3)}{\mathrm{d}z_3} =& \frac{1}{\gamma_1}\big(2c_3 z_3^{2p-1} + 2k_3 z_3^{2q-1} + 2W_{f3}^{*T}\Psi_{f3} + 2\varepsilon_{f3}\\ &+ \frac{7}{2}z_3 + J_3^o(\xi_3, z_3)\big)\end{aligned} \tag{58}$$

where $0 < p < 1$, $q > 1$, $c_3 > 0$ and $k_3 > 0$ are design parameters. $J_3^o(\xi_3, z_3) = -2c_3 z_3^{2p-1} - 2k_3 z_3^{2q-1} - 2W_{f3}^{*T}\Psi_{f3} - 2\varepsilon_{f3} - \frac{7}{2}z_3 + \frac{\mathrm{d}J_3^*(z_3)}{\mathrm{d}z_3}$ is a continuous function, and the u^* can be expressed as

$$\begin{aligned}u^* =& \frac{1}{\gamma_1}\big(- c_3 z_3^{2p-1} - k_3 z_3^{2q-1} - W_{f3}^{*T}\Psi_{f3} - \varepsilon_{f3}\\ &- \frac{7}{4}z_3 - \frac{1}{2}J_3^o(\xi_3, z_3)\big)\end{aligned} \tag{59}$$

Since $J_3^o(\xi_3, z_3)$ is unknown continuous term, it can also be approximated using NN as follows:

$$J_3^o(\xi_3, z_3) = W_{J3}^{*T}\Psi_{J3} + \varepsilon_{J3} \tag{60}$$

where W_{J3}^* is the ideal weight vector, Ψ_{J3} is the NN basis function vector, and the NN approximation error ε_{J3} is bounded.

Similarly, we can derive the following conclusion

$$\begin{aligned}\frac{\mathrm{d}J_3^*(z_3)}{\mathrm{d}z_3} =& \frac{1}{\gamma_1}(2c_3 z_3^{2p-1} + 2k_3 z_3^{2q-1} + 2W_{f3}^{*T}\Psi_{f3}\\ &+ \frac{7}{2}z_3 + W_{J3}^{*T}\Psi_{J3} + \varepsilon_3)\end{aligned} \tag{61}$$

$$\begin{aligned}u^* =& \frac{1}{\gamma_1}(-c_3 z_3^{2p-1} - k_3 z_3^{2q-1} - W_{f3}^{*T}\Psi_{f3}\\ &- \frac{7}{4}z_3 - \frac{1}{2}W_{J3}^{*T}\Psi_{J3} - \frac{1}{2}\varepsilon_3)\end{aligned} \tag{62}$$

where $\varepsilon_3 = 2\varepsilon_{f3} + \varepsilon_{J3}$. For (62), however, remains unattainable, necessitating the execution of an RL algorithm featuring both a critic and an actor to acquire viable control signal.

$$\begin{aligned}\frac{\mathrm{d}\hat{J}_3^*(z_3)}{\mathrm{d}z_3} =& \frac{1}{\gamma_1}(2c_3 z_3^{2p-1} + 2k_3 z_3^{2q-1} + 2\hat{W}_{f3}^T\Psi_{f3}\\ &+ \frac{7}{2}z_3 + \hat{W}_{c3}^T\Psi_{J3})\end{aligned} \tag{63}$$

$$\hat{u}^* = \frac{1}{\gamma_1}(-c_3 z_3^{2p-1} - k_3 z_3^{2q-1} - \hat{W}_{f3}^T \Psi_{J3} - \frac{7}{4} z_3 - \frac{1}{2}\hat{W}_{a3}^T \Psi_{J3}) \tag{64}$$

where $\frac{\mathrm{d}\hat{J}_3^*(z_3)}{\mathrm{d}z_3}$ and $\hat{u}^*$ are the estimate of $\frac{\mathrm{d}J_3^*(z_3)}{\mathrm{d}z_3}$ and u^*, respectively. $\hat{W}_{c3}^T \Psi_{J3}$ and $\hat{W}_{a3}^T \Psi_{J3}$ are the NN weight vectors of critic and actor, respectively.

Then, the corresponding three adaptive update laws are designed as follows:

$$\dot{\hat{W}}_{f3} = \Gamma_{f3} \Psi_{J3} - \kappa_{f3} \hat{W}_{f3} \tag{65}$$

$$\dot{\hat{W}}_{c3} = -\kappa_{c3} \Psi_{J3} \Psi_{J3}^T \hat{W}_{c3} \tag{66}$$

$$\dot{\hat{W}}_{a3} = -\Psi_{J3} \Psi_{J3}^T \big(\kappa_{a3}(\hat{W}_{a3} - \hat{W}_{c3}) + \kappa_{c3} \hat{W}_{c3}\big) \tag{67}$$

where $\Gamma_{f3} > 0$, $\kappa_{f3} > 0$, $\kappa_{c3} > 0$ and $\kappa_{a3} > 0$ are design parameters, while κ_{c3} and κ_{a3} satisfy $\kappa_{a3} > \frac{1}{2}$, $\kappa_{a3} > \frac{\kappa_{c3}}{3}$.

Following (54) and (64), we obtain $\dot{z}_3$

$$\dot{z}_3 = -c_3 z_3^{2p-1} - k_3 z_3^{2q-1} + z_3 - \frac{1}{2}\hat{W}_{a3}^T \Psi_{J3} + \tilde{W}_{f3}^T \Psi_{f3} + \varepsilon_{f3} - \frac{7}{4} z_3 - \dot{\hat{\alpha}}_1^* \tag{68}$$

Subsequently, the Lyapunov function V_3 is established as

$$V_3 = \frac{1}{2} z_3^2 + \frac{1}{2\Gamma_{f3}} \tilde{W}_{f3}^T \tilde{W}_{f3} + \frac{1}{2}\tilde{W}_{c3}^T \tilde{W}_{c3} + \frac{1}{2}\tilde{W}_{a3}^T \tilde{W}_{a3} \tag{69}$$

where $\tilde{W}_{f3} = W_{f3}^* - \hat{W}_{f3}$, $\tilde{W}_{c3} = W_{J3}^* - \hat{W}_{c3}$ and $\tilde{W}_{a3} = W_{J3}^* - \hat{W}_{a3}$.

Then, the $\dot{V}_3$ can be calculated as

$$\begin{aligned}\dot{V}_3 =& z_3(-c_3 z_3^{2p-1} - k_3 z_3^{2q-1} + z_3 - \frac{1}{2}\hat{W}_{a3}^T \Psi_{J3} + \tilde{W}_{f3}^T \Psi_{f3} \\ &+ \varepsilon_{f3} - \dot{\hat{\alpha}}_2^*) + \kappa_{f3} \tilde{W}_{f3}^T \hat{W}_{f3} + \kappa_{c3} \tilde{W}_{c3}^T \Psi_{J3} \Psi_{J3}^T \hat{W}_{c3} \\ &+ \tilde{W}_{a3}^T \Psi_{J3} \Psi_{J3}^T \big(\kappa_{a3}(\hat{W}_{a3} - \hat{W}_{c3}) + \kappa_{c3} \hat{W}_{c3}\big) - \frac{7}{4} z_3^2\end{aligned} \tag{70}$$

Using the Young's inequality, we have

$$\begin{aligned} z_3 \varepsilon_{f3} &\le \frac{1}{2} z_3^2 + \frac{1}{2}\bar{\varepsilon}_{f3}^2 \\ -z_3 \dot{\hat{\alpha}}_2^* &\le \frac{1}{2} z_3^2 + \frac{1}{2}\dot{\hat{\alpha}}_2^{*2} \\ -\frac{1}{2} z_3 \hat{W}_{a3}^T \Psi_{J3} &\le \frac{1}{4} z_3^2 + \frac{1}{4}\hat{W}_{a3}^T \Psi_{J3} \Psi_{J3}^T \hat{W}_{a3} \end{aligned} \tag{71}$$

Substituting (52) into (51) yields

$$\begin{aligned}\dot{V}_3 \le& -c_3 z_3^{2p} - k_3 z_3^{2q} - \frac{\kappa_{f3}}{2}\tilde{W}_{f3}^T \tilde{W}_{f3} - \frac{\kappa_{c3}}{2}\tilde{W}_{c3}^T \Psi_{J3} \Psi_{J3}^T \tilde{W}_{c3} \\ &- (\kappa_{a3} - \frac{\kappa_{c3}}{2})\tilde{W}_{a3}^T \Psi_{J3} \Psi_{J3}^T \tilde{W}_{a3} - \frac{\kappa_{a3}}{2}\hat{W}_{c3}^T \Psi_{J3} \Psi_{J3}^T \hat{W}_{c3} \\ &- (\frac{\kappa_{a3}}{2} - \frac{1}{4})\hat{W}_{a3}^T \Psi_{J3} \Psi_{J3}^T \hat{W}_{a3} + \frac{\kappa_{c3} + \kappa_{a3}}{2} W_{J3}^{*T} \Psi_{J3} \\ &\times \Psi_{J3}^T W_{J3}^* + \frac{1}{2}\bar{\varepsilon}_{f3}^2 + \frac{1}{2}\dot{\hat{\alpha}}_2^{*2} + \frac{\kappa_{f3}}{2} W_{f3}^{*T} W_{f3}^* + \frac{1}{2} z_3^2 \\ \le& -c_3 z_3^{2p} - k_3 z_3^{2q} - \frac{\kappa_{f3}}{2}\tilde{W}_{f3}^T \tilde{W}_{f3} - \frac{\kappa_{c3}}{2}\lambda_{\Psi_{J3}}^{\min} \tilde{W}_{c3}^T \tilde{W}_{c3} \\ &- (\kappa_{a3} - \frac{\kappa_{c3}}{2})\lambda_{\Psi_{J3}}^{\min} \tilde{W}_{a3}^T \tilde{W}_{a3} - \frac{1}{2} z_3^2 + \sigma_3\end{aligned} \tag{72}$$

where $\sigma_3 = \frac{1}{2}\bar{\varepsilon}_{f3}^2 + \frac{\kappa_{c3}+\kappa_{a3}}{2} W_{J3}^{*T} \Psi_{J3} \Psi_{J3}^T W_{J3}^* + \frac{1}{2}\dot{\hat{\alpha}}_2^{*2} + \frac{\kappa_{f3}}{2} W_{f3}^{*T} W_{f3}^*$ is bounded, and there exists a positive constant $\bar{\sigma}_3$ that ensures the existence of $|\sigma_3| \le \bar{\sigma}_3$. Additionally, $\lambda_{\Psi_{J3}}^{\min}$ represents the minimum eigenvalue of $\Psi_{J3} \Psi_{J3}^T$.

B. Stability analysis

Theorem 1 The fixed-time optimal control strategy proposed in this paper is applied to HSCS (3), where the adaptive laws of neural parameters, critic and actor are (46), (65) and (24), (47), (66), and (25), (48), (67), respectively. The optimal virtual controller are (23), (45), and the fixed-time optimal control actuator is (64). Thus, this control strategy can ensure that all control signals in the closed-loop system are bounded in fixed time, and simultaneously achieve the optimization of each subsystem.

Proof: Construct a Lyapunov function $V = \sum_{i=1}^{3} V_i$, and by integrating the preceding steps, we can compute

$$\begin{aligned}\dot{V} \le& -\sum_{i=1}^{3} c_i z_i^{2p} - \sum_{i=1}^{3} k_i z_i^{2q} - \sum_{i=2}^{3}\frac{\kappa_{fi}}{2}\tilde{W}_{fi}^T \tilde{W}_{fi} - \sum_{i=1}^{3}\frac{\kappa_{ci}}{2} \\ &\times \lambda_{\Psi_{Ji}}^{\min} \tilde{W}_{ci}^T \tilde{W}_{ci} - \sum_{i=1}^{3}(\kappa_{ai} - \frac{\kappa_{ci}}{2})\lambda_{\Psi_{Ji}}^{\min} \tilde{W}_{ai}^T \tilde{W}_{ai} + \sum_{i=1}^{3}\bar{\sigma}_i\end{aligned} \tag{73}$$

By virtue of Lemma 1, the following operations can be carried out

$$\begin{aligned}-\sum_{i=2}^{3}\frac{\kappa_{fi}}{2}\tilde{W}_{fi}^T \tilde{W}_{fi} \le& -\frac{\breve{\kappa}_f}{2}\left(\sum_{i=2}^{3}\frac{1}{2}\tilde{W}_{fi}^T \tilde{W}_{fi}\right)^p - \frac{\breve{\kappa}_f}{2} 3^{1-q} \\ &\times \left(\sum_{i=2}^{3}\frac{1}{2}\tilde{W}_{fi}^T \tilde{W}_{fi}\right)^q + \varrho_{W_f}\end{aligned} \tag{74}$$

$$\begin{aligned}-\sum_{i=1}^{3}\frac{\kappa_{ci}}{2}\lambda_{\Psi_{Ii}}^{\min} \tilde{W}_{ci}^T \tilde{W}_{ci} \le& -\frac{\breve{\kappa}_c}{2}\left(\sum_{i=1}^{3}\frac{1}{3}\tilde{W}_{ci}^T \tilde{W}_{ci}\right)^p - \frac{\breve{\kappa}_c}{2} 3^{1-q} \\ &\times \left(\sum_{i=1}^{3}\frac{1}{2}\tilde{W}_{ci}^T \tilde{W}_{ci}\right)^q + \varrho_{W_c}\end{aligned} \tag{75}$$

$$\begin{aligned}-\sum_{i=1}^{3}(\kappa_{ai} - \frac{\kappa_{ci}}{2})\lambda_{\Psi_{Ii}}^{\min} \tilde{W}_{ai}^T \tilde{W}_{ai} \le& -\frac{\breve{\kappa}_a}{2}\left(\sum_{i=1}^{3}\frac{1}{2}\tilde{W}_{ai}^T \tilde{W}_{ai}\right)^p \\ &- \frac{\breve{\kappa}_a}{2} 3^{1-q}\left(\sum_{i=1}^{3}\frac{1}{2}\tilde{W}_{ai}^T \tilde{W}_{ai}\right)^q + \varrho_{W_a}\end{aligned} \tag{76}$$

where $\breve{\kappa}_f = \min\{\kappa_{fi}, i = 2, 3\}$, $\breve{\kappa}_c = \min\{\kappa_{ci}\lambda_{\Psi_{Ii}}^{\min}, i = 1, 2, 3\}$ and $\breve{\kappa}_a = \min\{(2\kappa_{ai} - \kappa_{ci})\lambda_{\Psi_{Ii}}^{\min}, i = 1, 2, 3\}$. Furthermore, there are three unknown constants μ_{fi}, μ_{ci} and μ_{ai}, with $|\tilde{W}_{fi}| < \mu_{fi}$, $|\tilde{W}_{ci}| < \mu_{ci}$ and $|\tilde{W}_{ai}| < \mu_{ai}$.

Substituting (74)-(76) into (73) yields

$$\dot{V} \le -aV^p - bV^q + c \tag{77}$$

where $a = \min\{2^p c_i, \frac{\breve{\chi}_f}{2}, \frac{\breve{\chi}_c}{2}, \frac{\breve{\chi}_a}{2}, i = 1, 2, 3\}, b = \min\{2^q k_i, \frac{\breve{\chi}_f}{2} 3^{1-q}, \frac{\breve{\chi}_c}{2} 3^{1-q}, \frac{\breve{\chi}_a}{2} 3^{1-q}, i = 1, 2, 3\}, c = \varrho_{W_f} + \varrho_{W_c} + \varrho_{W_a} + \sum_{i=0}^{3}\bar{\sigma}_i$.

■

The proof of Theorem 1 is completed.

IV. SIMULATION EXAMPLE

To verify the effectiveness of the control algorithm proposed in this paper, numerical simulation verification was carried out with the aid of MATLAB. The parameters used in the simulation process are summarized as follows:

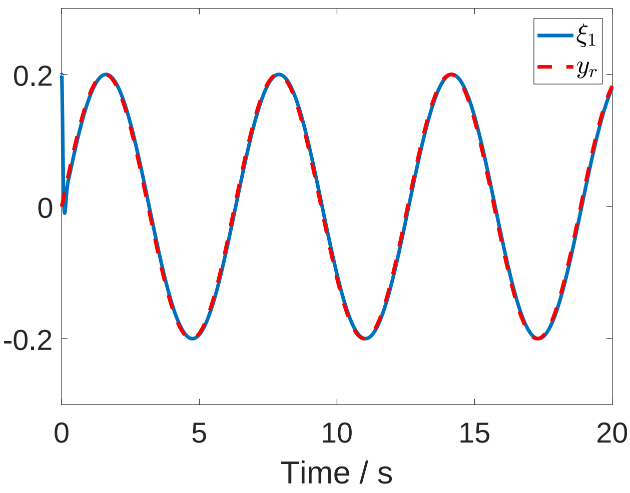

Fig. 2: The trajectories of ξ_1 and y_r.

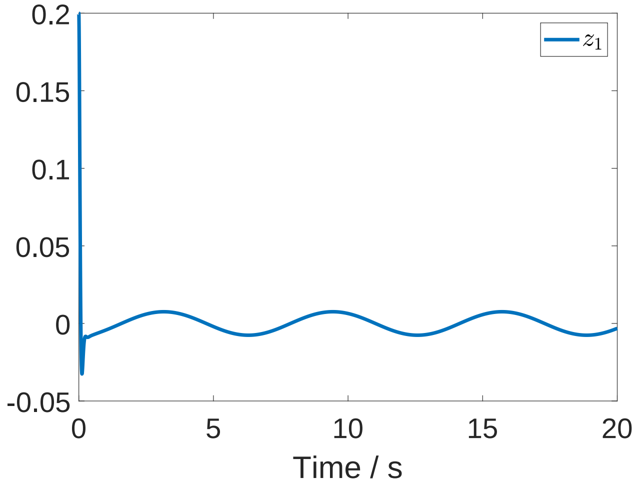

Fig. 3: The trajectory of z_1.

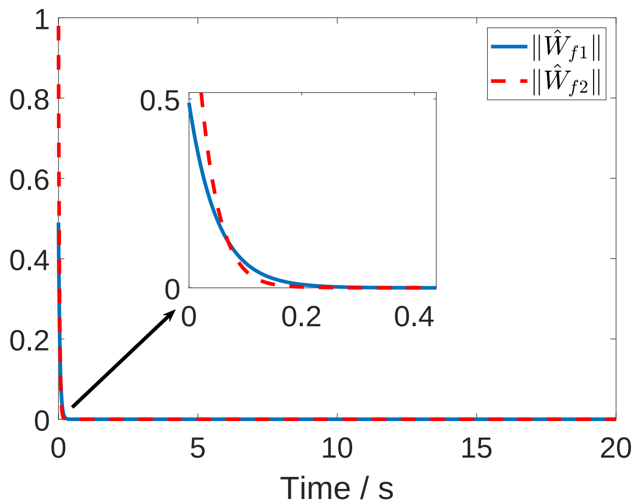

Fig. 4: The norms of the $\hat{W}_{f2}$ and $\hat{W}_{f3}$.

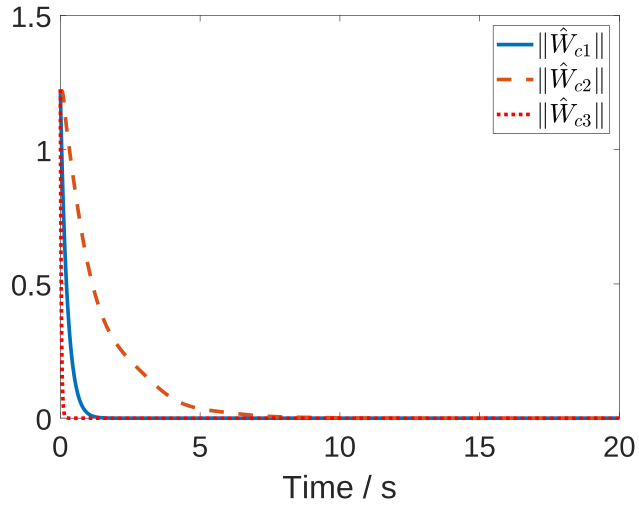

Fig. 5: The norms of the $\hat{W}_{c1}$, $\hat{W}_{c2}$ and $\hat{W}_{c3}$.

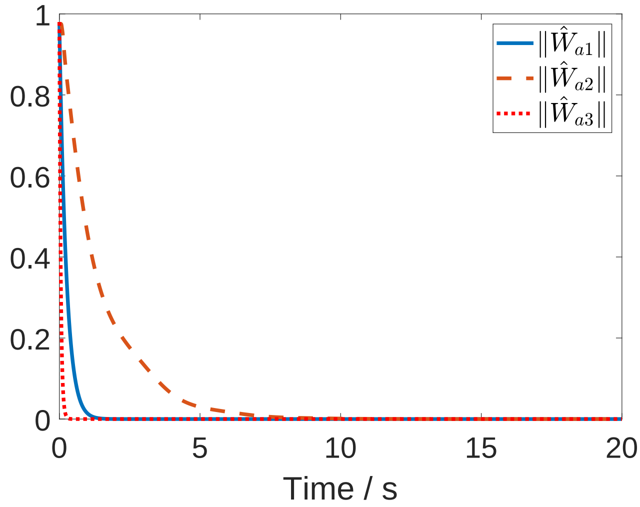

Fig. 6: The norms of the $\hat{W}_{a1}$, $\hat{W}_{a2}$ and $\hat{W}_{a3}$.

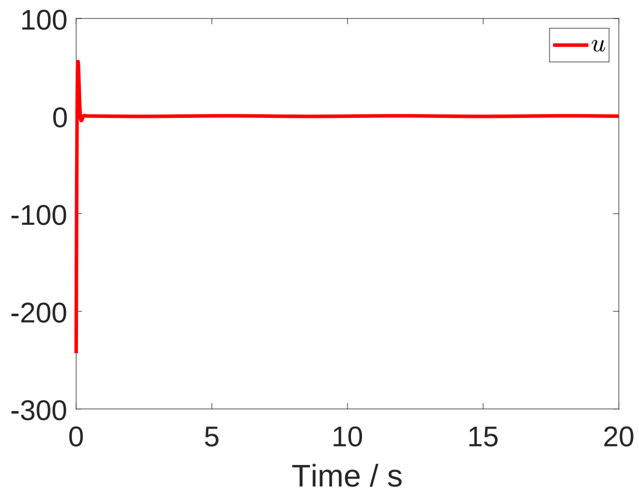

Fig. 7: Control input u.

The corresponding process parameters in the electro-hydraulic control cylinder system of the hydraulic support are $m = 300\,kg$, $B = 1000\,N/(m \cdot S^{-1})$, $A_1 = 1.92625 \times 10^{-3}\,m^2$, $A_2 = 9.4514 \times 10^{-4}\,m^2$, $p_s = 2 \times 10^7\,Pa$, $p_r = 0$, $k_q k_v = 8.9 \times 10^{-8}\,m^3/(s \cdot V \cdot \sqrt{Pa})$, $\beta_e = 7 \times 10^8\,Pa$, $C_t = 4 \times 10^{-13}\,m^3/(s \cdot Pa)$.

The control parameters are designed as $c_1 = 14$, $c_2 = 16$, $c_3 = 18$, $k_1 = 20$, $k_2 = 18$, $k_3 = 16$, $\kappa_{f2} = 15$, $\kappa_{f3} = 20$, $\kappa_{c1} = \kappa_{c2} = \kappa_{c3} = 10$, $\kappa_{a1} = \kappa_{a2} = \kappa_{a3} = 12$, $p = 99/101$, $q = 102/99$.

The initial values are set as $\xi_1(0) = \xi_2(0) = \xi_3(0) = 0.2$, $\hat{W}_{f2}(0) = \hat{W}_{f3}(0) = [0.2, \ldots, 0.2]^T \in \mathbb{R}^{6\times1}$, $\hat{W}_{c1}(0) = \hat{W}_{a1}(0) = [0.5, \ldots, 0.5]^T \in \mathbb{R}^{6\times1}$, $\hat{W}_{c2}(0) = \hat{W}_{a2}(0) = [0.4, \ldots, 0.4]^T \in \mathbb{R}^{6\times1}$, $\hat{W}_{c3}(0) = \hat{W}_{a3}(0) = [0.4, \ldots, 0.4]^T \in \mathbb{R}^{6\times1}$.

The simulation results show that the dual neural network structure based on the actor-critic framework proposed in this paper can quickly evaluate the value function of the current control strategy, generate adaptive control law compensation terms, and dynamically adjust the control gain of the system online, while achieving fixed-time stability of the hydraulic support cylinder system.

Figures 2 and 3 indicate that this control strategy can ensure that the electro-hydraulic control cylinder system of the hydraulic support has excellent tracking performance.

Figures 4-6 demonstrate that the critic adaptive law, actor adaptive law, and optimal controller designed in this paper can all converge rapidly and remain stable to achieve the optimal state of the system.

V. Conclusion

This paper constructs the dynamic system of the hydraulic support electro-hydraulic control cylinder, integrates reinforcement learning technology with fixed-time algorithm theory, and proposes an adaptive fixed-time optimal control strategy. This strategy designs an adaptive controller using the simplified optimal backstepping method, ensuring performance optimization for each subsystem while guaranteeing that all signals in the closed-loop system stably converge within a fixed time. Simulation results demonstrate that the proposed method significantly enhances the robustness and convergence efficiency of the system, thereby verifying the effectiveness and engineering feasibility of the control strategy. This research provides theoretical support and practical solutions for optimizing the performance of the Hydraulic Support Control System (HSCS) and can be further extended to application verification in complex industrial scenarios in the future.

References

[1] R. Zhou, L. Meng, X. Yuan, and Z. Qiao, "Research and experimental analysis of hydraulic cylinder position control mechanism based on pressure detection," *Machines*, vol. 10, no. 1, p. 1, 2021.

[2] Y. Zhang, H. Zhang, K. Gao, Q. Zeng, F. Meng, and J. Cheng, "Research on intelligent control system of hydraulic support based on position and posture detection," *Machines*, vol. 11, no. 1, p. 33, 2022.

[3] D. Li, X. Zhao, Z. Zhao, C. Su, and J. Meng, "Stability analysis of the floating multi-robot coordinated towing system based on ship stability.," *Engineering Letters*, vol. 32, no. 6, pp. 1191–1200, 2024.

[4] X. Yang, D. Liu, and Q. Wei, "Online approximate optimal control for affine non-linear systems with unknown internal dynamics using adaptive dynamic programming," *IET Control Theory & Applications*, vol. 8, no. 16, pp. 1676–1688, 2014.

[5] Y. Zhu, D. Zhao, and Z. Zhong, "Adaptive optimal control of heterogeneous cacc system with uncertain dynamics," *IEEE Transactions on Control Systems Technology*, vol. 27, no. 4, pp. 1772–1779, 2018.

[6] S. Bhasin, R. Kamalapurkar, M. Johnson, K. G. Vamvoudakis, F. L. Lewis, and W. E. Dixon, "A novel actor–critic–identifier architecture for approximate optimal control of uncertain nonlinear systems," *Automatica*, vol. 49, no. 1, pp. 82–92, 2013.

[7] P. Werbos, "Approximate dynamic programming for real-time control and neural modeling," *Handbook of intelligent control*, 1992.

[8] D. Liu, D. Wang, F.-Y. Wang, H. Li, and X. Yang, "Neural-network-based online hjb solution for optimal robust guaranteed cost control of continuous-time uncertain nonlinear systems," *IEEE transactions on cybernetics*, vol. 44, no. 12, pp. 2834–2847, 2014.

[9] D. Wang, D. Liu, and H. Li, "Policy iteration algorithm for online design of robust control for a class of continuous-time nonlinear systems," *IEEE Transactions on Automation Science and Engineering*, vol. 11, no. 2, pp. 627–632, 2014.

[10] G. Wen, S. S. Ge, and F. Tu, "Optimized backstepping for tracking control of strict-feedback systems," *IEEE transactions on neural networks and learning systems*, vol. 29, no. 8, pp. 3850–3862, 2018.

[11] G. Wen, C. P. Chen, and S. S. Ge, "Simplified optimized backstepping control for a class of nonlinear strict-feedback systems with unknown dynamic functions," *IEEE Transactions on Cybernetics*, vol. 51, no. 9, pp. 4567–4580, 2020.

[12] G. Wen, B. Li, and B. Niu, "Optimized backstepping control using reinforcement learning of observer-critic-actor architecture based on fuzzy system for a class of nonlinear strict-feedback systems," *IEEE Transactions on Fuzzy Systems*, vol. 30, no. 10, pp. 4322–4335, 2022.

[13] S. Liu, H. Yan, L. Zhao, and D. Gao, "Fault estimate and reinforcement learning based optimal output feedback control for single-link robot arm model.," *Engineering Letters*, vol. 33, no. 1, pp. 21–28, 2025.

[14] Q. Yu, J. Ding, L. Wu, and X. He, "Event-triggered prescribed time adaptive fuzzy fault-tolerant control for nonlinear systems with full-state constraints.," *Engineering Letters*, vol. 32, no. 8, pp. 1577–1584, 2024.

[15] S. P. Bhat and D. S. Bernstein, "Finite-time stability of continuous autonomous systems," *SIAM Journal on Control and optimization*, vol. 38, no. 3, pp. 751–766, 2000.

[16] Y. Zhou, X. Wan, C. Huang, and X. Yang, "Finite-time stochastic synchronization of dynamic networks with nonlinear coupling strength via quantized intermittent control," *Applied Mathematics and Computation*, vol. 376, p. 125157, 2020.

[17] K. Xu, H. Wang, and P. X. Liu, "Adaptive fuzzy finite-time tracking control of nonlinear systems with unmodeled dynamics," *Applied Mathematics and Computation*, vol. 450, p. 127992, 2023.

[18] A. Polyakov, "Nonlinear feedback design for fixed-time stabilization of linear control systems," *IEEE transactions on Automatic Control*, vol. 57, no. 8, pp. 2106–2110, 2011.

[19] H. Wang and Z. Meng, "Fixed-time adaptive neural tracking control for high-order nonlinear switched systems with input saturation and dead-zone," *Applied Mathematics and Computation*, vol. 480, p. 128904, 2024.

[20] F. Wei, L. Zhang, B. Niu, and G. Zong, "Adaptive decentralized fixed-time neural control for constrained strong interconnected nonlinear systems with input quantization," *International Journal of Robust and Nonlinear Control*, vol. 34, no. 14, pp. 9899–9927, 2024.

[21] Y. Wu, H. Ma, M. Chen, and H. Li, "Observer-based fixed-time adaptive fuzzy bipartite containment control for multiagent systems with unknown hysteresis," *IEEE Transactions on Fuzzy Systems*, vol. 30, no. 5, pp. 1302–1312, 2021.

[22] Y. Zhang and F. Wang, "Observer-based fixed-time neural control for a class of nonlinear systems," *IEEE transactions on neural networks and learning systems*, vol. 33, no. 7, pp. 2892–2902, 2021.

[23] Y. Li, T. Yang, and S. Tong, "Adaptive neural networks finite-time optimal control for a class of nonlinear systems," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 31, no. 11, pp. 4451–4460, 2019.

[24] H. Wang, J. Ma, X. Zhao, B. Niu, M. Chen, and W. Wang, "Adaptive fuzzy fixed-time control for high-order nonlinear systems with sensor and actuator faults," *IEEE Transactions on Fuzzy Systems*, vol. 31, no. 8, pp. 2658–2668, 2023.