Adaptive Optimal Control Based on Critic-Actor Architecture for Hydraulic Support Cylinder System with Asymmetric Output Error Constraints

Wei Teng*, Guangming Wang

Abstract—**This paper takes the hydraulic support cylinder system (HSCS) as the research object. Based on its working principle and physical characteristics, a nonlinear mathematical model describing the electro-hydraulic control power system of the hydraulic support is constructed. Firstly, by combining reinforcement learning technology and asymmetric output constraint theory, an adaptive asymmetric output constraint optimal control strategy is proposed, aiming to enhance the robustness of the system and meet industrial constraint requirements. Secondly, by adopting a simplified optimal backstepping design method, an adaptive optimal controller is designed to ensure that the performance of each subsystem reaches the optimum and the output error of the system is always limited within the preset asymmetric constraint range. Finally, the simulation verification shows that the proposed method has good effectiveness and engineering feasibility.**

Index Terms—**Hydraulic support cylinder, Adaptive optimal control, Reinforcement learning, Asymmetric output constraints**

I. Introduction

As the tide of industrial intelligence sweeps across the landscape, the intelligent control system governing the "trinity" of fully mechanized mining operations has emerged as a cornerstone technology for safeguarding safety and amplifying efficiency. Among its critical elements, the electro-hydraulic control cylinder of the hydraulic support stands out, yet it grapples with formidable challenges owing to its pronounced nonlinear hysteresis and parameters that vary with time. These hurdles include the intricacies of dynamic modeling and the shortfall in servo tracking precision [1, 2]. Fortunately, the evolution of adaptive nonlinear control theory heralds an ingenious resolution, paving the way for the intelligent advancement of such sophisticated industrial systems [3, 4].

Optimal control for nonlinear systems is one of the core aspects of modern control theory, aiming to optimize the performance indicators of control systems [5]. It integrates the fundamental conditions and methods distilled from practical problems, with the research object being controlled dynamic systems or motion processes. It seeks the best control scheme among the allowable ones to ensure the system achieves the optimal performance when transitioning from the initial state to the target state [6]. With the rapid development of digital technology and electronic computers, optimal control has been widely applied in production, military, and economic activities, playing a significant role in the national economy and national defense. The optimal problem is theoretically equivalent to solving the Hamilton-Jacobi-Bellman (HJB) equation [7], but due to its strong nonlinearity and dynamic uncertainty, it is difficult to solve directly through analysis. To overcome this challenge, reinforcement learning (RL) and adaptive dynamic programming (ADP) have been proven to be effective solutions. RL and ADP were initially proposed by Werbos for discrete systems [8] and later extended to continuous systems [9, 10], but they are only applicable to affine nonlinear systems. For the control problem of nonlinear mismatched systems, in [11] proposed an optimal control method based on the backstepping framework, ensuring the optimization of each subsystem. To simplify complexity and alleviate the continuous excitation condition, in [12–14] further simplified the optimal backstepping control strategy.

Constraints are a frequent occurrence in practical control systems. Considering that control systems often have to strike a balance between performance requirements and physical limitations, the constraint problem holds significant importance in practical control systems [15, 16]. To date, the barrier Lyapunov function (BLF) has become the mainstream method for dealing with constraints due to its ability to constrain state variables within a predefined compact set [17–21]. In [17], the BLF method was used to control a multi-input multi-output (MIMO) nonlinear system to achieve practical stability under output constraint conditions. In [18], a neural network adaptive fault-tolerant controller based on integral type BLF was adopted, and it was proposed that the full state constraint of the uncertain nonlinear system could be satisfied even if the initial value violated the predefined compact set limit. However, it should be noted that the BLF technique in [17, 18] is only used to handle symmetric constraints. Recently, in [19–21], further research on the tracking control of nonlinear systems under time-varying asymmetric constraint conditions has been conducted using different types of BLF. That is, the constraints are allowed to be asymmetric and time-varying, and the boundary conditions have been greatly relaxed in these works, which is more in line with the application of actual systems.

Based on the above research, this paper takes the working principle of the hydraulic support electro-hydraulic control cylinder as the control object, and combines the reinforcement learning algorithm and adaptive control theory, aiming to optimize the overall stability and robustness of the system operation. Through the modeling analysis, controller

Manuscript received May 6, 2025; revised July 18, 2025.

Wei Teng is a teacher of School of Computer and Software Engineering, University of Science and Technology Liaoning, Anshan 114051, China (Corresponding author, e-mail: tengjiawei1978@163.com).

Guangming Wang is an electrical engineer of Ansteel Engineering Technology Corporation Limited, Liaoning, Anshan, 114051, China (e-mail: wgm0515@163.com).

design and simulation verification of the working process of the hydraulic support electro-hydraulic control cylinder system, an optimal control strategy combining reinforcement learning and asymmetric output constraints is proposed. The specific contributions are summarized as follows: 1) Mathematical analysis and modeling of the working principle and physical process of Hydraulic support cylinder; 2) Keeping all signals in the Hydraulic support cylinder system bounded while achieving performance optimization of each subsystem; 3) The system output error strictly conforms to the pre-designed form and avoids any violation of constraints throughout the process.

II. MODEL DESCRIPTION AND PRELIMINARIES

Assuming that the electro-hydraulic control cylinder system of the hydraulic support operates in the direction shown in Figure 1 during operation, according to the force balance equation and the flow balance equation, its dynamic equation can be expressed as:

$$M\ddot{X}_p = p_1A_1 - p_2A_2 - B\dot{X}_p + F_T \tag{1}$$

where $p_1 = (p_s + p_r)/2$, $p_2 = (p_s - p_r)/2$. 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.

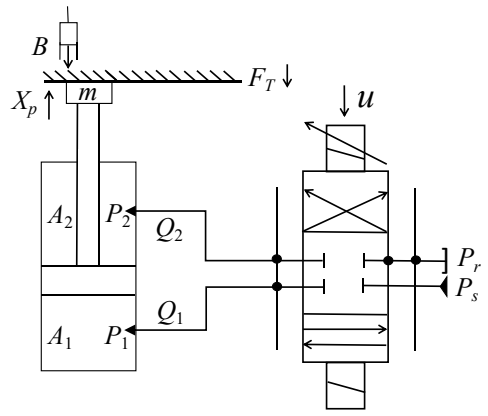

Fig. 1: Model diagram of hydraulic support cylinder.

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:

$$\begin{cases} \dot{\xi}_1 = m\xi_2 \\ \dot{\xi}_2 = \xi_3 - \dfrac{B}{m}\xi_2 - F_T - mg \\ \dot{\xi}_3 = \gamma_1 u - \gamma_2\xi_2 - \gamma_3(p_1 - p_2) \\ y = \xi_1 \end{cases} \tag{2}$$

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 paper, x_v denotes the displacement of the spool in the proportional valve, and k_v represents a positive constant.

Then, for $\forall t \geq t_0$, the system output is constrained by a asymmetric time-varying boundary condition, defined as follows:

$$-\underline{\varrho}_{a1}(t) \leq \xi_1 \leq \overline{\varrho}_{a1}(t) \tag{3}$$

where $\underline{\varrho}_{a1}(t)$ and $\overline{\varrho}_{a1}(t)$ represent time-varying boundary functions, with the constraint being asymmetric when $\underline{\varrho}_{a1}(t) \neq \overline{\varrho}_{a1}(t)$.

Additionally, the bounded 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.

Lemma 1 [22] Let $f(x)$ be a continuous function defined on a compact set Ω_x. Then for $\forall\varepsilon > 0$, there exist the NN $\zeta^T\Psi(x)$ such that

$$\sup_{x\in\Omega_x} |f(x) - \zeta^T\Psi(x)| \leq \varepsilon \tag{5}$$

where $\zeta = [\zeta_1, \zeta_2, \ldots, \zeta_m]^T \in \mathbb{R}^m$ 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 ζ^* of NN is defined as

$$\zeta^* = \arg\min_{\zeta\in\mathbb{R}^m}\{\sup_{x\in\Omega_x} |f(x) - \zeta^T\Psi(x)|\} \tag{6}$$

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

$$f(x) = \zeta^{*T}\Psi(x) + \varepsilon(x) \tag{7}$$

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

Assumption 1. There exist time-varying functions $\underline{\varrho}_{b1}(t)$ and $\overline{\varrho}_{b1}(t)$ such that the reference signal y_r satisfies $-\underline{\varrho}_{a1}(t) < \underline{\varrho}_{b1}(t) \leq y_r \leq \overline{\varrho}_{b1}(t) < \overline{\varrho}_{a1}(t)$. Additionally, the derivatives of $-\underline{\varrho}_{a1}(t)$, $\underline{\varrho}_{b1}(t)$, y_r, $\overline{\varrho}_{b1}(t)$, and $\overline{\varrho}_{a1}(t)$ are known and bounded.

III. MAIN RESULT

In this section, we will combine the reinforcement learning algorithm and the asymmetric output constraint control method to design an optimal backstepping control strategy under the critic-actor architecture, thereby constructing an optimal controller.

A. Optimized backstepping design

First, consider the following tracking error coordinate transformation:

$$\begin{aligned} z_1 &= \xi_1 - y_r \\ z_2 &= m\xi_2 - \hat{\alpha}_1^* \\ z_3 &= \xi_3 - \hat{\alpha}_2^* \end{aligned} \tag{8}$$

where y_r is selected as the reference signal and set to $0.2\sin(t)$. $\alpha_{i-1}(i = 2, 3)$ and $\hat{\alpha}_{i-1}^*$ represent the virtual control and actual optimal virtual control correspondingly.

Step 1: In order to impose the asymmetric time-varying constraints on the output, an asymmetric Lyapunov function as given in [19] is introduced:

$$\breve{V} = \frac{z_1^2}{(\overline{\phi}_1(t) - z_1)(\underline{\phi}_1(t) + z_1)} \tag{9}$$

where $\underline{\phi}_1(t)$ and $\overline{\phi}_1(t)$ are positive barrier functions. Define set $\phi_{z_1} = \{-\underline{\phi}_1(t) < z_1 < \overline{\phi}_1(t)\}$, ensuring that $\breve{V}$ is well-defined within this compact set. Simultaneously, the boundary functions designated as $\underline{\phi}_1(t) = \underline{\varrho}_{a_1}(t) - \underline{\varrho}_{b_1}(t)$ and $\overline{\phi}_1(t) = \overline{\varrho}_{a_1}(t) - \overline{\varrho}_{b_1}(t)$ are chosen to satisfy condition (3).

Then, the derivative of $\breve{V}$ is as follows:

$$\dot{\breve{V}} = Pz_1(\dot{z}_1 + Q) \tag{10}$$

where

$$P = \frac{2\underline{\phi}_1\overline{\phi}_1 - \underline{\phi}_1 z_1 + \overline{\phi}_1 z_1}{[(\overline{\phi}_1 - z_1)(\underline{\phi}_1 + z_1)]^2} \tag{11}$$

$$Q = \frac{-\dot{\underline{\phi}}_1\overline{\phi}_1 - \underline{\phi}_1\dot{\overline{\phi}}_1 + z_1(\dot{\underline{\phi}}_1 - \dot{\overline{\phi}}_1)}{2\underline{\phi}_1\overline{\phi}_1 - \underline{\phi}_1 z_1 + \overline{\phi}_1 z_1} \tag{12}$$

From (4) and (8), the derivative of z_1 can be calculated

$$\dot{z}_1 = m\xi_2 - \dot{y}_r \tag{13}$$

The optimal performance index function is chosen as

$$J_1(z_1) = \int_t^\infty h_1\Big(z_1(v), \alpha_1\big(z_1(v)\big)\Big)dv \tag{14}$$

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

$$\begin{aligned} J_1^*(z_1) &= \int_t^\infty h_1(z_1(v), \alpha_1^*(z_1(v)))dv \\ &= \min_{\alpha_1 \in \Omega_{z_1}} \{\int_t^\infty h_1\Big(z_1(v), \alpha_1\big(z_1(v)\big)\Big)dv\} \end{aligned} \tag{15}$$

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

$$H_1(z_1, \alpha_1^*, \frac{\mathrm{d}J_1^*}{\mathrm{d}z_1}) = z_1^2 + \alpha_1^{*2} + \frac{\mathrm{d}J_1^*}{\mathrm{d}z_1}(\alpha_1^* - \dot{y}_r) = 0 \tag{16}$$

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

$$\alpha_1^* = -\frac{1}{2}\frac{\mathrm{d}J_1^*(z_1)}{\mathrm{d}z_1} \tag{17}$$

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

$$\frac{\mathrm{d}J_1^*(z_1)}{\mathrm{d}z_1} = (\frac{2\chi_1}{P} + \frac{7P}{2})z_1 + J_1^o(\xi_1, z_1) \tag{18}$$

where $\chi_1 > 0$ is design parameter. $J_1^o(\xi_1, z_1) = -(\frac{2\chi_1}{P} + \frac{7P}{2})z_1 + \frac{\mathrm{d}J_1^*(z_1)}{\mathrm{d}z_1} \in \mathbb{R}$ is a continuous function, and substituting (18) into (17) has

$$\alpha_1^* = -(\frac{\chi_1}{P} + \frac{7P}{4})z_1 - \frac{1}{2}J_1^o(\xi_1, z_1) \tag{19}$$

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

$$J_1^o(\xi_1, z_1) = \zeta_{J1}^{*T}\Psi_{J1}(\xi_1, z_1) + \varepsilon_{J1}(\xi_1, z_1) \tag{20}$$

where ζ_{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 $|\varepsilon_{J1}(\xi_1, z_1)| \leq \overline{\varepsilon}_{J1}$ as arbitrarily small. Then, (18) and (19) can be reorganized as

$$\frac{\mathrm{d}J_1^*(z_1)}{\mathrm{d}z_1} = (\frac{2\chi_1}{P} + \frac{7P}{2})z_1 + \zeta_{J1}^{*T}\Psi_{J1} + \varepsilon_{J1} \tag{21}$$

$$\alpha_1^* = -(\frac{\chi_1}{P} + \frac{7P}{4})z_1 - \frac{1}{2}\zeta_{J1}^{*T}\Psi_{J1} - \frac{1}{2}\varepsilon_{J1} \tag{22}$$

Since ζ_{J1}^* is unknown constant vector, the optimal virtual control (22) is not available for the controlled system. To derive the effective optimized virtual control, the following RL algorithm with critic and actor is performed.

$$\frac{\mathrm{d}\hat{J}_1^*(z_1)}{\mathrm{d}z_1} = (\frac{2\chi_1}{P} + \frac{7P}{2})z_1 + \hat{\zeta}_{c1}^T\Psi_{J1} \tag{23}$$

$$\hat{\alpha}_1^* = -(\frac{\chi_1}{P} + \frac{7P}{4})z_1 - \frac{1}{2}\hat{\zeta}_{a1}^T\Psi_{J1} \tag{24}$$

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. $\hat{\zeta}_{c1}^T\Psi_{J1}$ and $\hat{\zeta}_{a1}^T\Psi_{J1}$ 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.

$$\dot{\hat{\zeta}}_{c1} = -\kappa_{c1}\Psi_{J1}\Psi_{J1}^T\hat{\zeta}_{c1} \tag{25}$$

$$\dot{\hat{\zeta}}_{a1} = -\Psi_{J1}\Psi_{J1}^T\big(\kappa_{a1}(\hat{\zeta}_{a1} - \hat{\zeta}_{c1}) + \kappa_{c1}\hat{\zeta}_{c1}\big) \tag{26}$$

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 (24), (13) can be rewritten as

$$\dot{z}_1 = z_2 - (\frac{\chi_1}{P} + \frac{7P}{4})z_1 - \frac{1}{2}\hat{\zeta}_{a1}^T\Psi_{J1} - \dot{y}_r \tag{27}$$

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

$$V_1 = \breve{V} + \frac{1}{2}\tilde{\zeta}_{c1}^T\tilde{\zeta}_{c1} + \frac{1}{2}\tilde{\zeta}_{a1}^T\tilde{\zeta}_{a1} \tag{28}$$

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

Then, the derivative of V_1 is

$$\begin{aligned} \dot{V}_1 =& Pz_1\big(z_2 - (\frac{\chi_1}{P} + \frac{7P}{4})z_1 - \frac{1}{2}\hat{\zeta}_{a1}^T\Psi_{J1} - \dot{y}_r + Q\big) \\ &+ \kappa_{c1}\tilde{\zeta}_{c1}^T\Psi_{J1}\Psi_{J1}^T\hat{\zeta}_{c1} + \tilde{\zeta}_{a1}^T\Psi_{J1}\Psi_{J1}^T\big(\kappa_{a1}(\hat{\zeta}_{a1} \\ &- \hat{\zeta}_{c1}) + \kappa_{c1}\hat{\zeta}_{c1}\big) \end{aligned} \tag{29}$$

The Young's inequality yields the following results

$$\begin{aligned} Pz_1z_2 &\le \frac{1}{2}P^2z_1^2+\frac{1}{2}z_2^2 \\ PQz_1 &\le \frac{1}{2}P^2z_1^2+\frac{1}{2}Q^2 \\ -Pz_1\dot{y}_r &\le \frac{1}{2}P^2z_1^2+\frac{1}{2}\dot{y}_r^2 \\ -\frac{1}{2}Pz_1\hat{\zeta}_{a1}^T\Psi_{J1} &\le \frac{1}{4}P^2z_1^2+\frac{1}{4}\hat{\zeta}_{a1}^T\Psi_{J1}\Psi_{J1}^T\hat{\zeta}_{a1} \end{aligned} \tag{30}$$

Along with (29) and (30), we can calculate:

$$\begin{aligned} \dot{V}_1 \le & -\chi_1 z_1+\kappa_{c1}\tilde{\zeta}_{c1}^T\Psi_{J1}\Psi_{J1}^T\hat{\zeta}_{c1} \\ & +\kappa_{a1}\tilde{\zeta}_{a1}^T\Psi_{J1}\Psi_{J1}^T\hat{\zeta}_{a1}+\frac{1}{2}z_2^2+\frac{1}{2}\dot{y}_r^2 \\ & +(\kappa_{c1}-\kappa_{a1})\tilde{\zeta}_{a1}^T\Psi_{J1}\Psi_{J1}^T\hat{\zeta}_{c1} \\ & +\frac{1}{4}\hat{\zeta}_{a1}^T\Psi_{J1}\Psi_{J1}^T\hat{\zeta}_{a1}+\frac{1}{2}Q^2 \end{aligned} \tag{31}$$

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

$$\begin{aligned} \tilde{\zeta}_{c1}^T\Psi_{J1}\Psi_{J1}^T\hat{\zeta}_{c1} =& \frac{1}{2}\zeta_{J1}^{*T}\Psi_{J1}\Psi_{J1}^T\zeta_{J1}^*-\frac{1}{2}\tilde{\zeta}_{c1}^T\Psi_{J1} \\ & \times\Psi_{J1}^T\tilde{\zeta}_{c1}-\frac{1}{2}\hat{\zeta}_{c1}^T\Psi_{J1}\Psi_{J1}^T\hat{\zeta}_{c1} \\ \tilde{\zeta}_{a1}^T\Psi_{J1}\Psi_{J1}^T\hat{\zeta}_{a1} =& \frac{1}{2}\zeta_{J1}^{*T}\Psi_{J1}\Psi_{J1}^T\zeta_{J1}^*-\frac{1}{2}\tilde{\zeta}_{a1}^T\Psi_{J1} \\ & \times\Psi_{J1}^T\tilde{\zeta}_{a1}-\frac{1}{2}\hat{\zeta}_{a1}^T\Psi_{J1}\Psi_{J1}^T\hat{\zeta}_{a1} \\ \tilde{\zeta}_{a1}^T\Psi_{J1}\Psi_{J1}^T\hat{\zeta}_{c1} \le & -\frac{1}{2}\tilde{\zeta}_{a1}^T\Psi_{J1}\Psi_{J1}^T\tilde{\zeta}_{a1} \\ & -\frac{1}{2}\hat{\zeta}_{c1}^T\Psi_{J1}\Psi_{J1}^T\hat{\zeta}_{c1} \end{aligned} \tag{32}$$

Subsequently, we can acquire

$$\begin{aligned} \dot{V}_1 \le & -\chi_1 z_1-\frac{\kappa_{c1}}{2}\tilde{\zeta}_{c1}^T\Psi_{J1}\Psi_{J1}^T\tilde{\zeta}_{c1} \\ & -(\kappa_{a1}-\frac{\kappa_{c1}}{2})\tilde{\zeta}_{a1}^T\Psi_{J1}\Psi_{J1}^T\tilde{\zeta}_{a1} \\ & -\frac{\kappa_{a1}}{2}\hat{\zeta}_{c1}^T\Psi_{J1}\Psi_{J1}^T\hat{\zeta}_{c1}-(\frac{\kappa_{a1}}{2}-\frac{1}{4}) \\ & \times\hat{\zeta}_{a1}^T\Psi_{J1}\Psi_{J1}^T\hat{\zeta}_{a1}+\frac{1}{2}z_2^2+\frac{1}{2}\dot{y}_r^2 \\ & +\frac{\kappa_{c1}+\kappa_{a1}}{2}\zeta_{J1}^{*T}\Psi_{J1}\Psi_{J1}^T\zeta_{J1}^*+\frac{1}{2}Q^2 \end{aligned} \tag{33}$$

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

$$\begin{aligned} -\tilde{\zeta}_{c1}^T\Psi_{J1}\Psi_{J1}^T\tilde{\zeta}_{c1} &\le -\lambda_{\Psi_{J1}}^{\min}\tilde{\zeta}_{c1}^T\tilde{\zeta}_{c1} \\ -\tilde{\zeta}_{a1}^T\Psi_{J1}\Psi_{J1}^T\tilde{\zeta}_{a1} &\le -\lambda_{\Psi_{J1}}^{\min}\tilde{\zeta}_{a1}^T\tilde{\zeta}_{a1} \end{aligned} \tag{34}$$

According to the design parameters $\kappa_{a1}>\frac{\kappa_{c1}}{2}$ and $\kappa_{a1}>\frac{1}{2}$, as well as (34), it can yield

$$\begin{aligned} \dot{V}_1 \le & -\chi_1 z_1-\frac{\kappa_{c1}}{2}\lambda_{\Psi_{J1}}^{\min}\tilde{\zeta}_{c1}^T\tilde{\zeta}_{c1} \\ & -(\kappa_{a1}-\frac{\kappa_{c1}}{2})\lambda_{\Psi_{J1}}^{\min}\tilde{\zeta}_{a1}^T\tilde{\zeta}_{a1}+\frac{1}{2}z_2^2+\sigma_1 \end{aligned} \tag{35}$$

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

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

$$\dot{z}_2=\xi_3-\frac{B}{m}\xi_2-F_T-mg-\dot{\hat{\alpha}}_1^* \tag{36}$$

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

$$\begin{aligned} J_2^*(z_2) &= \int_t^\infty h_2\Big(z_2(v),\alpha_2^*\big(z_2(v)\big)\Big)dv \\ &= \min_{\alpha_2\in\Omega_{z_2}}\{\int_t^\infty h_2\Big(z_2(v),\alpha_2\big(z_2(v)\big)\Big)dv\} \end{aligned} \tag{37}$$

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 (37), the HJB equation is constructed as

$$\begin{aligned} H_2(z_2,\alpha_2^*,\frac{\mathrm{d}J_2^*}{\mathrm{d}z_2}) =& z_2^2+\alpha_2^{*2}+\frac{\mathrm{d}J_2^*}{\mathrm{d}z_2}\big(\alpha_2^*+\zeta_{f2}^{*T}\Psi_{f2}(\xi) \\ & +\varepsilon_f(\xi)-\dot{\hat{\alpha}}_1^*\big)=0 \end{aligned} \tag{38}$$

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

$$\alpha_2^*=-\frac{1}{2}\frac{\mathrm{d}J_2^*(z_2)}{\mathrm{d}z_2} \tag{39}$$

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

$$\begin{aligned} \frac{\mathrm{d}J_2^*(z_2)}{\mathrm{d}z_2} =& (2\chi_2+\frac{9}{2})z_2+2\zeta_{f2}^{*T}\Psi_{f2}+2\varepsilon_{f2} \\ & +J_2^o(\xi_2,z_2) \end{aligned} \tag{40}$$

where $\chi_2>0$ is design parameter. $J_2^o(\xi_2,z_2)=-(2\chi_2+\frac{9}{2})z_2-2\zeta_{f2}^{*T}\Psi_{f2}-2\varepsilon_{f2}+\frac{\mathrm{d}J_2^*(z_2)}{\mathrm{d}z_2}$ is a continuous function, and the α_2^* can be expressed as

$$\alpha_2^*=-(\chi_2+\frac{9}{4})z_2-\zeta_{f2}^{*T}\Psi_{f2}-\varepsilon_{f2}-\frac{1}{2}J_2^o(\xi_2,z_2) \tag{41}$$

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

$$J_2^o(\xi_2,z_2)=\zeta_{J2}^{*T}\Psi_{J2}+\varepsilon_{J2} \tag{42}$$

where ζ_{J2}^* is the ideal weight vector, Ψ_{J2} is the NN basis function vector, and the NN approximation error ε_{J2} is bounded.

Similarly, we can derive the following conclusion

$$\frac{\mathrm{d}J_2^*(z_2)}{\mathrm{d}z_2}=(2\chi_2+\frac{9}{2})z_2+2\zeta_{f2}^{*T}\Psi_{f2}+\zeta_{J2}^{*T}\Psi_{J2}+\varepsilon_2 \tag{43}$$

$$\alpha_2^*=-(\chi_2+\frac{9}{4})z_2-\zeta_{f2}^{*T}\Psi_{f2}-\frac{1}{2}\zeta_{J2}^{*T}\Psi_{J2}-\frac{1}{2}\varepsilon_2 \tag{44}$$

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

The optimal control (44), however, remains unattainable, necessitating the execution of an RL algorithm featuring both a critic and an actor to acquire viable control signal.

$$\frac{\mathrm{d}\hat{J}_2^*(z_2)}{\mathrm{d}z_2}=(2\chi_2+\frac{9}{2})z_2+2\hat{\zeta}_{f2}^T\Psi_{f2}+\hat{\zeta}_{c2}^T\Psi_{J2} \tag{45}$$

$$\hat{\alpha_2}^*=-(\chi_2+\frac{9}{4})z_2-\hat{\zeta}_{f2}^T\Psi_{J2}-\frac{1}{2}\hat{\zeta}_{a2}^T\Psi_{J2} \tag{46}$$

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. $\hat{\zeta}_{c2}^T\Psi_{J2}$ and $\hat{\zeta}_{a2}^T\Psi_{J2}$ 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:

$$\dot{\hat{\zeta}}_{f2}=\Gamma_{f2}\Psi_{J2}z_2-\kappa_{f2}\hat{\zeta}_{f2} \tag{47}$$

$$\dot{\hat{\zeta}}_{c2} = -\kappa_{c2}\Psi_{J2}\Psi_{J2}^T\hat{\zeta}_{c2} \tag{48}$$

$$\dot{\hat{\zeta}}_{a2} = -\Psi_{J2}\Psi_{J2}^T\big(\kappa_{a2}(\hat{\zeta}_{a2}-\hat{\zeta}_{c2})+\kappa_{c2}\hat{\zeta}_{c2}\big) \tag{49}$$

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

$$\begin{aligned}\dot{z}_2 =& -(\chi_2+\frac{9}{4})z_2+z_3-\frac{1}{2}\hat{\zeta}_{a2}^T\Psi_{J2}\\ &+\tilde{\zeta}_{f2}^T\Psi_{f2}+\varepsilon_{f2}-\dot{\hat{\alpha}}_1^*\end{aligned} \tag{50}$$

Subsequently, the Lyapunov function V_2 is established as

$$V_2=\frac{1}{2}z_2^2+\frac{1}{2\Gamma_{f2}}\tilde{\zeta}_{f2}^T\tilde{\zeta}_{f2}+\frac{1}{2}\tilde{\zeta}_{c2}^T\tilde{\zeta}_{c2}+\frac{1}{2}\tilde{\zeta}_{a2}^T\tilde{\zeta}_{a2} \tag{51}$$

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

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

$$\begin{aligned}\dot{V}_2 =& z_2\big(-(\chi_2+\frac{9}{4})z_2+z_3-\frac{1}{2}\hat{\zeta}_{a2}^T\Psi_{J2}+\tilde{\zeta}_{f2}^T\Psi_{f2}\\ &+\varepsilon_{f2}-\dot{\hat{\alpha}}_1^*\big)+\frac{\kappa_{f2}}{\Gamma_{f2}}\tilde{\zeta}_{f2}^T\hat{\zeta}_{f2}+\kappa_{c2}\tilde{\zeta}_{c2}^T\Psi_{J2}\Psi_{J2}^T\hat{\zeta}_{c2}\\ &+\tilde{\zeta}_{a2}^T\Psi_{J2}\Psi_{J2}^T\big(\kappa_{a2}(\hat{\zeta}_{a2}-\hat{\zeta}_{c2})+\kappa_{c2}\hat{\zeta}_{c2}\big)\end{aligned} \tag{52}$$

Using the Young's inequality, we have

$$\begin{aligned}z_2z_3 &\le \frac{1}{2}z_2^2+\frac{1}{2}z_3^2\\ z_2\varepsilon_{f2} &\le \frac{1}{2}z_2^2+\frac{1}{2}\bar{\varepsilon}_{f2}^2\\ -z_2\dot{\hat{\alpha}}_1^* &\le \frac{1}{2}z_2^2+\frac{1}{2}\dot{\hat{\alpha}}_1^{*2}\\ -\frac{1}{2}z_2\hat{\zeta}_{a2}^T\Psi_{J2} &\le \frac{1}{4}z_2^2+\frac{1}{4}\hat{\zeta}_{a2}^T\Psi_{J2}\Psi_{J2}^T\hat{\zeta}_{a2}\end{aligned} \tag{53}$$

Substituting (53) into (52) yields

$$\begin{aligned}\dot{V}_2 \le& -\chi_2z_2^2-\frac{\kappa_{f2}}{2\Gamma_{f2}}\tilde{\zeta}_{f2}^T\tilde{\zeta}_{f2}-\frac{\kappa_{c2}}{2}\tilde{\zeta}_{c2}^T\Psi_{J2}\Psi_{J2}^T\tilde{\zeta}_{c2}\\ &-(\kappa_{a2}-\frac{\kappa_{c2}}{2})\tilde{\zeta}_{a2}^T\Psi_{J2}\Psi_{J2}^T\tilde{\zeta}_{a2}-\frac{\kappa_{a2}}{2}\hat{\zeta}_{c2}^T\Psi_{J2}\Psi_{J2}^T\hat{\zeta}_{c2}\\ &-(\frac{\kappa_{a2}}{2}-\frac{1}{4})\hat{\zeta}_{a2}^T\Psi_{J2}\Psi_{J2}^T\hat{\zeta}_{a2}+\frac{\kappa_{c2}+\kappa_{a2}}{2}\zeta_{J2}^{*T}\Psi_{J2}\\ &\times\Psi_{J2}^T\zeta_{J2}^*+\frac{1}{2}\bar{\varepsilon}_{f2}^2+\frac{1}{2}\dot{\hat{\alpha}}_1^{*2}+\frac{\kappa_{f2}}{2}\zeta_{f2}^{*T}\zeta_{f2}^*+\frac{1}{2}z_3^2\\ \le& -\chi_2z_2^2-\frac{\kappa_{f2}}{2\Gamma_{f2}}\tilde{\zeta}_{f2}^T\tilde{\zeta}_{f2}-\frac{\kappa_{c2}}{2}\lambda_{\Psi_{J2}}^{\min}\tilde{\zeta}_{c2}^T\tilde{\zeta}_{c2}\\ &-(\kappa_{a2}-\frac{\kappa_{c2}}{2})\lambda_{\Psi_{J2}}^{\min}\tilde{\zeta}_{a2}^T\tilde{\zeta}_{a2}-\frac{1}{2}z_2^2+\frac{1}{2}z_3^2+\sigma_2\end{aligned} \tag{54}$$

where $\sigma_2=\frac{1}{2}\bar{\varepsilon}_{f2}^2+\frac{\kappa_{c2}+\kappa_{a2}}{2}\zeta_{J2}^{*T}\Psi_{J2}\Psi_{J2}^T\zeta_{J2}^*+\frac{\kappa_{f2}}{2}\zeta_{f2}^{*T}\zeta_{f2}^*+\frac{1}{2}\dot{\hat{\alpha}}_1^{*2}$ 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

$$\begin{aligned}\dot{z}_3 &= \dot{\xi}_3-\dot{\hat{\alpha}}_2^*\\ &= \gamma_1u-\gamma_2\xi_2-\gamma_3(p_1-p_2)-\dot{\hat{\alpha}}_2^*\end{aligned} \tag{55}$$

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

$$\begin{aligned}J_3^*(z_3) &= \int_t^\infty h_3\Big(z_3(v),u^*\big(z_3(v)\big)\Big)dv\\ &= \min_{u\in\Omega_{z_3}}\{\int_t^\infty h_3\Big(z_3(v),u\big(z_3(v)\big)\Big)dv\}\end{aligned} \tag{56}$$

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

Based on (56), the HJB equation is constructed as

$$\begin{aligned}H_3(z_3,u^*,\frac{\mathrm{d}J_3^*}{\mathrm{d}z_3}) =& z_3^2+u^{*2}+\frac{\mathrm{d}J_3^*}{\mathrm{d}z_3}\big(u^*+\zeta_{f3}^{*T}\Psi_{f3}\\ &+\varepsilon_{f3}-\dot{\hat{\alpha}}_2^*\big)=0\end{aligned} \tag{57}$$

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

$$u^*=-\frac{1}{2}\frac{\mathrm{d}J_3^*(z_3)}{\mathrm{d}z_3} \tag{58}$$

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

$$\frac{\mathrm{d}J_3^*(z_3)}{\mathrm{d}z_3}=(2\chi_3+\frac{7}{2})z_3+2\zeta_{f3}^{*T}\Psi_{f3}+2\varepsilon_{f3}+J_3^o(\xi_3,z_3) \tag{59}$$

where $\chi_3>0$ is design parameter. $J_3^o(\xi_3,z_3)=-(2\chi_3+\frac{7}{2})z_3-2\zeta_{f3}^{*T}\Psi_{f3}-2\varepsilon_{f3}+\frac{\mathrm{d}J_3^*(z_3)}{\mathrm{d}z_3}$ is a continuous function, and the u^* can be expressed as

$$u^*=-(\chi_3+\frac{7}{4})z_3-\zeta_{f3}^{*T}\Psi_{f3}-\varepsilon_{f3}-\frac{1}{2}J_3^o(\xi_3,z_3) \tag{60}$$

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

$$J_3^o(\xi_3,z_3)=\zeta_{J3}^{*T}\Psi_{J3}+\varepsilon_{J3} \tag{61}$$

where ζ_{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

$$\frac{\mathrm{d}J_3^*(z_3)}{\mathrm{d}z_3}=(2\chi_3+\frac{7}{2})z_3+2\zeta_{f3}^{*T}\Psi_{f3}+\zeta_{J3}^{*T}\Psi_{J3}+\varepsilon_3 \tag{62}$$

$$u^*=-(\chi_3+\frac{7}{4})z_3-\zeta_{f3}^{*T}\Psi_{f3}-\frac{1}{2}\zeta_{J3}^{*T}\Psi_{J3}-\frac{1}{2}\varepsilon_3 \tag{63}$$

where $\varepsilon_3=2\varepsilon_{f3}+\varepsilon_{J3}$. For (63), however, remains unattainable, necessitating the execution of an RL algorithm featuring both a critic and an actor to acquire viable control signal.

$$\frac{\mathrm{d}\hat{J}_3^*(z_3)}{\mathrm{d}z_3}=(2\chi_3+\frac{7}{2})z_3+2\hat{\zeta}_{f3}^T\Psi_{f3}+\hat{\zeta}_{c3}^T\Psi_{J3} \tag{64}$$

$$\hat{u}^*=-(\chi_3+\frac{7}{4})z_3-\hat{\zeta}_{f3}^T\Psi_{J3}-\frac{1}{2}\hat{\zeta}_{a3}^T\Psi_{J3} \tag{65}$$

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. $\hat{\zeta}_{c3}^T\Psi_{J3}$ and $\hat{\zeta}_{a3}^T\Psi_{J3}$ are the NN weight vectors of critic and actor, respectively.

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

$$\dot{\hat{\zeta}}_{f3}=\Gamma_{f3}\Psi_{J3}z_3-\kappa_{f3}\hat{\zeta}_{f3} \tag{66}$$

$$\dot{\hat{\zeta}}_{c3}=-\kappa_{c3}\Psi_{J3}\Psi_{J3}^T\hat{\zeta}_{c3} \tag{67}$$

$$\dot{\hat{\zeta}}_{a3}=-\Psi_{J3}\Psi_{J3}^T\big(\kappa_{a3}(\hat{\zeta}_{a3}-\hat{\zeta}_{c3})+\kappa_{c3}\hat{\zeta}_{c3}\big) \tag{68}$$

where $\Gamma_{f3}>0$, $\kappa_{f3}>0$, $\kappa_{c3}>0$ and $\kappa_{a3}>0$ are design parameters, while κ_{c3} and κ_{a3} satisfy $\kappa_{a3}>\frac{\gamma_1^2}{2}$, $\kappa_{a3}>\frac{\kappa_{c3}}{3}$.

Following (55) and (65), we obtain $\dot{z}_3$

$$\dot{z}_3=\gamma_1\big(-(\chi_3+\frac{7}{4})z_3-\frac{1}{2}\hat{\zeta}_{a3}^T\Psi_{J3}+\tilde{\zeta}_{f3}^T\Psi_{f3}+\varepsilon_{f3}-\dot{\hat{\alpha}}_1^*\big) \tag{69}$$

Subsequently, the Lyapunov function V_3 is established as

$$V_3 = \frac{1}{2}z_3^2 + \frac{1}{2\Gamma_{f3}}\tilde{\zeta}_{f3}^T\tilde{\zeta}_{f3} + \frac{1}{2}\tilde{\zeta}_{c3}^T\tilde{\zeta}_{c3} + \frac{1}{2}\tilde{\zeta}_{a3}^T\tilde{\zeta}_{a3} \tag{70}$$

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

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

$$\begin{aligned}\dot{V}_3 =& \gamma_1 z_3\big(-(\chi_3 + \frac{7}{4})z_3 - \frac{1}{2}\hat{\zeta}_{a3}^T\Psi_{J3} + \tilde{\zeta}_{f3}^T\Psi_{f3} + \varepsilon_{f3} \\ & - \dot{\hat{\alpha}}_2^*\big) + \frac{\kappa_{f3}}{\Gamma_{f3}}\tilde{\zeta}_{f3}^T\hat{\zeta}_{f3} + \kappa_{c3}\tilde{\zeta}_{c3}^T\Psi_{J3}\Psi_{J3}^T\hat{\zeta}_{c3} \\ & + \tilde{\zeta}_{a3}^T\Psi_{J3}\Psi_{J3}^T\big(\kappa_{a3}(\hat{\zeta}_{a3} - \hat{\zeta}_{c3}) + \kappa_{c3}\hat{\zeta}_{c3}\big)\end{aligned} \tag{71}$$

Using the Young's inequality, we have

$$\begin{aligned}\gamma_1 z_3\varepsilon_{f3} &\le \frac{1}{2}z_3^2 + \frac{\gamma_1^2}{2}\bar{\varepsilon}_{f3}^2 \\ -\gamma_1 z_3\dot{\hat{\alpha}}_2^* &\le \frac{1}{2}z_3^2 + \frac{\gamma_1^2}{2}\dot{\hat{\alpha}}_2^{*2} \\ -\frac{1}{2}\gamma_1 z_3\hat{\zeta}_{a3}^T\Psi_{J3} &\le \frac{1}{4}z_3^2 + \frac{\gamma_1^2}{4}\hat{\zeta}_{a3}^T\Psi_{J3}\Psi_{J3}^T\hat{\zeta}_{a3}\end{aligned} \tag{72}$$

Substituting (72) into (71) yields

$$\begin{aligned}\dot{V}_3 \le& -\chi_3 z_3^2 - \frac{\kappa_{f3}}{2\Gamma_{f3}}\tilde{\zeta}_{f3}^T\tilde{\zeta}_{f3} - \frac{\kappa_{c3}}{2}\tilde{\zeta}_{c3}^T\Psi_{J3}\Psi_{J3}^T\tilde{\zeta}_{c3} \\ & -(\kappa_{a3} - \frac{\kappa_{c3}}{2})\tilde{\zeta}_{a3}^T\Psi_{J3}\Psi_{J3}^T\tilde{\zeta}_{a3} - \frac{\kappa_{a3}}{2}\hat{\zeta}_{c3}^T\Psi_{J3}\Psi_{J3}^T\hat{\zeta}_{c3} \\ & -(\frac{\kappa_{a3}}{2} - \frac{\gamma_1^2}{4})\hat{\zeta}_{a3}^T\Psi_{J3}\Psi_{J3}^T\hat{\zeta}_{a3} + \frac{\kappa_{c3}+\kappa_{a3}}{2}\zeta_{J3}^{*T}\Psi_{J3} \\ & \times\Psi_{J3}^T\zeta_{J3}^* + \frac{1}{2}\bar{\varepsilon}_{f3}^2 + \frac{1}{2}\dot{\hat{\alpha}}_2^{*2} + \frac{\kappa_{f3}}{2}\zeta_{f3}^{*T}\zeta_{f3}^* - \frac{1}{2}z_3^2 + \sigma_3 \\ \le& -\chi_3 z_3^2 - \frac{\kappa_{f3}}{2\Gamma_{f3}}\zeta_{f3}^T\tilde{\zeta}_{f3} - \frac{\kappa_{c3}}{2}\lambda_{\Psi_{J3}}^{\min}\tilde{\zeta}_{c3}^T\tilde{\zeta}_{c3} \\ & -(\kappa_{a3} - \frac{\kappa_{c3}}{2})\lambda_{\Psi_{J3}}^{\min}\tilde{\zeta}_{a3}^T\tilde{\zeta}_{a3} - \frac{1}{2}z_3^2 + \sigma_3\end{aligned} \tag{73}$$

where $\sigma_3 = \frac{\gamma_1^2}{2}\bar{\varepsilon}_{f3}^2 + \frac{\kappa_{c3}+\kappa_{a3}}{2}\zeta_{J3}^{*T}\Psi_{J3}\Psi_{J3}^T\zeta_{J3}^* + \frac{\gamma_1^2}{2}\dot{\hat{\alpha}}_2^{*2} + \frac{\kappa_{f3}}{2}\zeta_{f3}^{*T}\zeta_{f3}^*$ 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 Apply the control strategy proposed in this paper to (2), where the adaptive laws for NN parameters, critics and actors are (47), (66) and (25), (48), (67) and (26), (49), (68), respectively, and the optimal virtual control and control input are (24), (46), and (65). Based on this, the following conclusions can be drawn: 1) This control strategy can ensure that all signals in the closed-loop system remain bounded and achieve the performance optimization of each subsystem; 2) The output error strictly follows the design form defined in (3) and no violation of constraints occurs.

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

$$\begin{aligned}\dot{V} \le& -\sum_{i=1}^{3}\chi_i z_i^2 - \sum_{i=2}^{3}\frac{\kappa_{f3}}{2\Gamma_{f3}}\tilde{\zeta}_{fi}^T\tilde{\zeta}_{fi} - \sum_{i=1}^{3}\frac{\kappa_{ci}}{2}\lambda_{\Psi_{Ji}}^{\min}\tilde{\zeta}_{ci}^T\tilde{\zeta}_{ci} \\ & -\sum_{i=1}^{3}(\kappa_{ai} - \frac{\kappa_{ci}}{2})\lambda_{\Psi_{Ji}}^{\min}\tilde{\zeta}_{ai}^T\tilde{\zeta}_{ai} + \sum_{i=1}^{3}\bar{\sigma}_i \\ \le& -\Theta V + \Delta\end{aligned} \tag{74}$$

where $\Theta = \min\{2\chi_i, \frac{\kappa_{fj}}{\Gamma_{fj}}, \kappa_{ci}, (\kappa_{ai} - \frac{\kappa_{ci}}{2})\lambda_{\Psi_{Ji}}^{\min}, i = 1, 2, 3, j = 2, 3\}$, $\Delta = \sum_{i=0}^{3}\bar{\sigma}_i$. ■

The proof of Theorem 1 is completed.

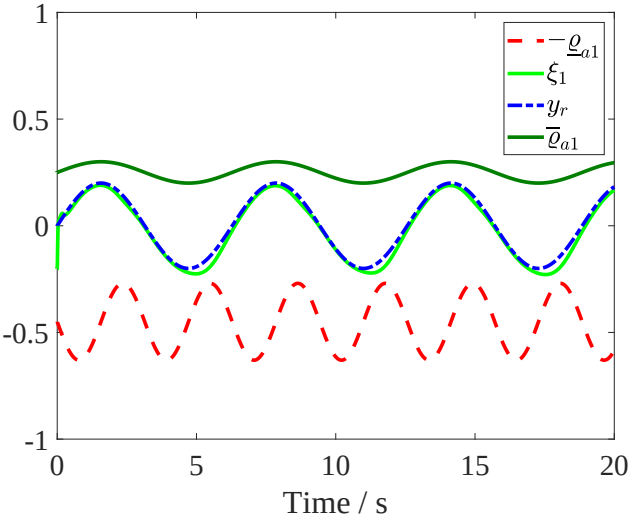

Fig. 2: ξ_1, y_r and the constraints $-\underline{\varrho}_{a1}$, $\overline{\varrho}_{a1}$.

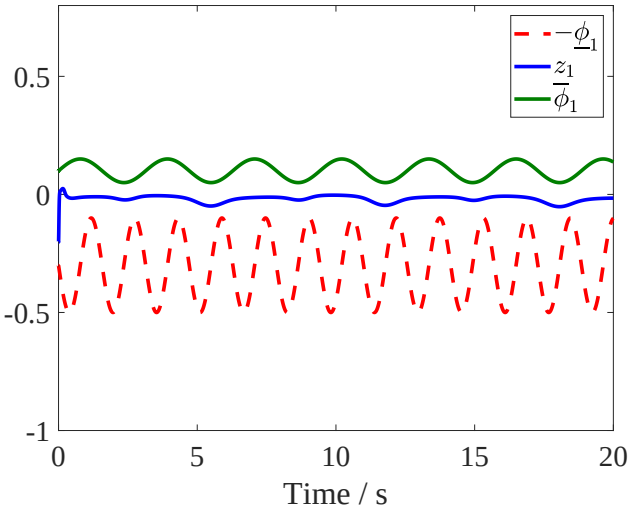

Fig. 3: z_1 and the constraints $\underline{\phi}_1$, $\overline{\phi}_1$.

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:

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 output ξ_1 is constrained within the range of $-0.18\sin(2t) - 0.45 < \xi_1 < 0.05\sin(t) + 0.25$. The reference signal is selected as $y_r = 0.2\sin(t)$.

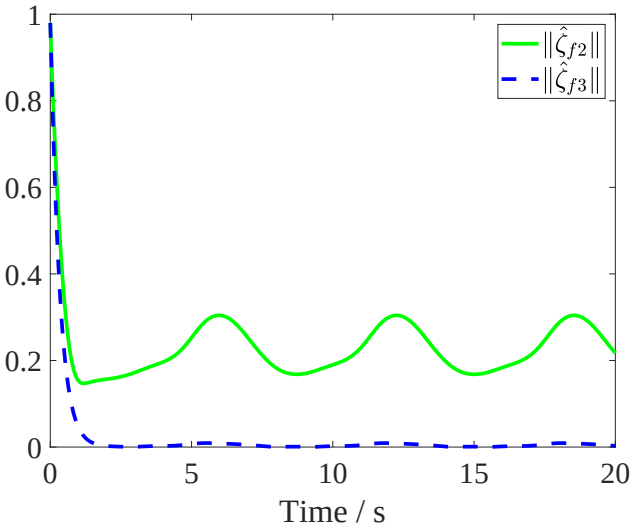

Fig. 4: The norms of the $\hat{\zeta}_{f2}$ and $\hat{\zeta}_{f3}$.

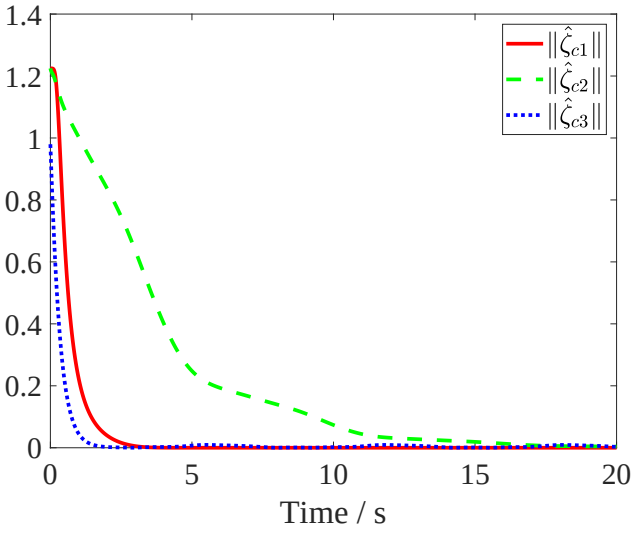

Fig. 5: The norms of the $\hat{\zeta}_{c1}$, $\hat{\zeta}_{c2}$ and $\hat{\zeta}_{c3}$.

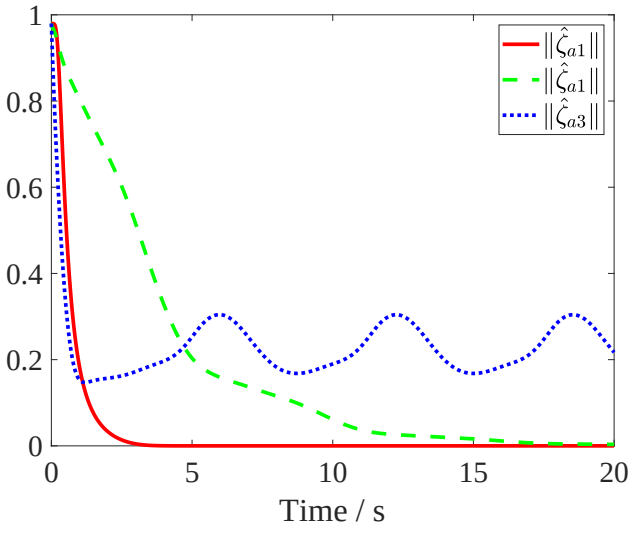

Fig. 6: The norms of the $\hat{\zeta}_{a1}$, $\hat{\zeta}_{a2}$ and $\hat{\zeta}_{a3}$.

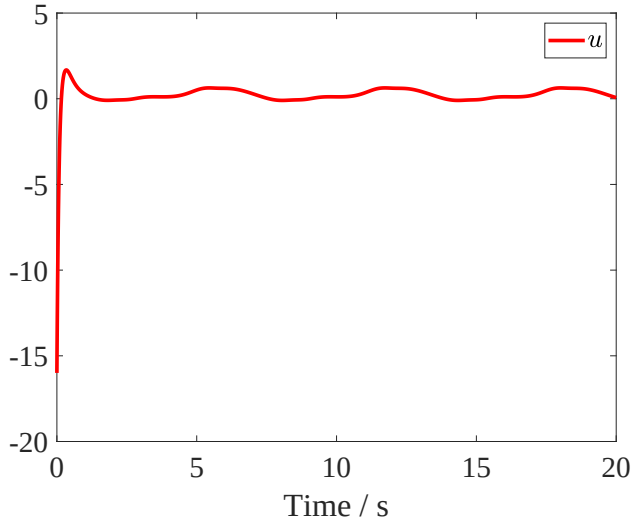

Fig. 7: Control input u.

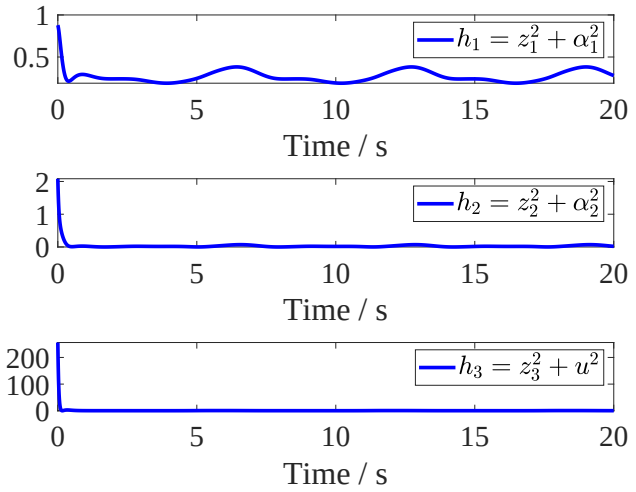

Fig. 8: Cost functions h_1, h_2 and h_3.

The control parameters are designed as $\chi_1 = 8$, $\chi_2 = 12$, $\chi_3 = 18$, $\kappa_{f2} = 15$, $\kappa_{f3} = 20$, $\kappa_{c1} = \kappa_{c2} = \kappa_{c3} = 10$, $\kappa_{a1} = \kappa_{a2} = \kappa_{a3} = 12$. Furthermore, the initial values are set as $\xi_1(0) = -0.2$, $\xi_2(0) = \xi_3(0) = 0.2$, $\hat{\zeta}_{f2}(0) = \hat{\zeta}_{f3}(0) = [0.2, \ldots, 0.2]^T \in \mathbb{R}^{6\times1}$, $\hat{\zeta}_{c1}(0) = \hat{\zeta}_{a1}(0) = [0.5, \ldots, 0.5]^T \in \mathbb{R}^{6\times1}$, $\hat{\zeta}_{c2}(0) = \hat{\zeta}_{a2}(0) = [0.4, \ldots, 0.4]^T \in \mathbb{R}^{6\times1}$, $\hat{\zeta}_{c3}(0) = \hat{\zeta}_{a3}(0) = [0.4, \ldots, 0.4]^T \in \mathbb{R}^{6\times1}$.

The simulation results show that the dual neural network structure based on the critic-actor framework proposed in this paper can efficiently evaluate the value function of the current control strategy, generate adaptive compensation terms to optimize the control law, and ensure that the output error of the hydraulic support cylinder system is always within the preset range by dynamically adjusting the system control gain online. Figures 2 and 3 verify the excellent tracking performance of the system under this strategy. Figures 4 to 8 further demonstrate that the critic adaptive law, the actor adaptive law, and the optimal controller designed in this paper all exhibit fast convergence characteristics and maintain stability, thereby effectively achieving the optimal control state of the system.

V. Conclusion

Based on the working principle of the hydraulic support electro-hydraulic control system, a more precise control mathematical model was constructed. Combined with reinforcement learning technology, an optimal backstepping controller with a critic-actor mechanism was designed. By introducing the asymmetric constraint theory method, this controller not only achieved performance optimization of each subsystem but also effectively ensured that the output error always met the preset asymmetric constraint conditions. Simulation results show that the proposed method significantly improves the robustness and convergence efficiency of the system, fully verifying the effectiveness and engineering feasibility of the control strategy. This research provides a solid theoretical support and feasible practical solution for the performance optimization of the HSCS, and has the potential for further application verification in complex industrial scenarios.

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] E. Guan, H. Miao, P. Li, J. Liu, and Y. Zhao, "Dynamic model analysis of hydraulic support," *Advances in Mechanical Engineering*, vol. 11, no. 1, p. 1687814018820143, 2019.

[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] Y. Wang, J. Zhao, H. Wang, and H. Ding, "Output feedback control for the driving cylinder of hydraulic support with error constraint," *Journal of Vibration and Control*, vol. 29, no. 13-14, pp. 3126–3136, 2023.

[5] 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.

[6] 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.

[7] 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.

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

[9] D. Liu, D. Wang, F. 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.

[10] 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.

[11] 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.

[12] 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.

[13] D. Gao, W. Cui, L. Wu, Y. Zhang, and Y. Tao, "Fixed-time optimized control for nonlinear strict-feedback systems based on reinforcement learning and disturbance observer," *Applied Mathematics and Computation*, vol. 508, p. 129628, 2026.

[14] 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.

[15] K. P. Tee, S. S. Ge, and E. H. Tay, "Barrier lyapunov functions for the control of output-constrained nonlinear systems," *Automatica*, vol. 45, no. 4, pp. 918–927, 2009.

[16] X. Jin, "Adaptive fault-tolerant control for a class of output-constrained nonlinear systems," *International Journal of Robust and Nonlinear Control*, vol. 25, no. 18, pp. 3732–3745, 2015.

[17] Y. Li and S. Tong, "Adaptive fuzzy output constrained control design for multi-input multioutput stochastic nonstrict-feedback nonlinear systems," *IEEE Transactions on Cybernetics*, vol. 47, no. 12, pp. 4086–4095, 2016.

[18] L. Wu, X. He, L. Guo, S. Huang, and Y. Hu, "Neural network adaptive switched fault-tolerant control of uncertain nonlinear systems with full state constraints," *Neurocomputing*, vol. 598, p. 128034, 2024.

[19] X. Jin, "Adaptive fixed-time control for mimo nonlinear systems with asymmetric output constraints using universal barrier functions," *IEEE Transactions on Automatic Control*, vol. 64, no. 7, pp. 3046–3053, 2018.

[20] F. Wang, Y. Gao, C. Zhou, and Q. Zong, "Disturbance observer-based backstepping formation control of multiple quadrotors with asymmetric output error constraints," *Applied Mathematics and Computation*, vol. 415, p. 126693, 2022.

[21] Z. Sun, J. Li, C. Wen, and C. Chen, "Adaptive event-triggered prescribed-time stabilization of uncertain nonlinear systems with asymmetric time-varying output constraint," *IEEE Transactions on Automatic Control*, vol. 69, no. 8, pp. 5454–5461, 2024.

[22] 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.