Fault Estimate and Reinforcement Learning Based Optimal Output Feedback Control for Single-Link Robot Arm Model

Sihan Liu*, Hailong Yan, Lixia Zhao, Dongxiang Gao

***Abstract*—This paper presents an optimal output feedback tracking control scheme for a single-link robot arm (SLRA) model, utilizing reinforcement learning (RL). Initially, a neural network (NN) state observer is developed to accurately estimate the system state. Then, an optimal output feedback controller is constructed based on the RL algorithm and the optimal backstepping control (OBC) method to effectively control the SLRA system. Subsequently, employing Lyapunov stability theory, it is demonstrated that all signals in the closed-loop system are guaranteed to remain bounded. In conclusion, a simulation case is presented to demonstrate the efficacy of the proposed approach.**

***Index Terms*—Single-link robot arm (SLRA), Neural network state observer, Optimal backstepping control, Reinforcement learning**

I. Introduction

With the rapid advancement of the robotics industry, the standalone robotic system featuring a Single-Link Robot Arm (SLRA) has gained widespread utilization across various industrial sectors [1, 2]. Renowned for its robustness and resilience in diverse industrial settings, SLRA is extensively employed in service robots, space exploration robots, and as operational units in other manufacturing domains. These robotic arms are particularly well-suited for high-precision tasks such as welding, processing, and quality control, playing an indispensable role in ensuring operational accuracy and reliability [3–5]. Furthermore, their exceptional adaptability and scalability enable customization to meet specific requirements across diverse industrial applications.

In recent years, there have been numerous successful adaptive control outcomes for intricate nonlinear systems such as SLRA [6–8]. It is widely recognized that unidentified system faults in industrial operations may lead to a deterioration in control system performance and result in unstable or even catastrophic incidents [9, 10]. Consequently, fault-tolerant control has emerged as a pivotal area of investigation within the realm of control theory [11]. A primary challenge lies in the accurate estimation of unknown faults within the controlled system when it encounters unforeseen time-varying system faults, with the aim of achieving optimal control performance. To tackle this challenge, an effective approach is to design a fault estimator and reconstruct the fault signal through the fault estimator, thereby designing a fault-tolerant control scheme [12–14]. This scheme can effectively eliminate the influence of fault signals on system state and system output. The accuracy of the fault estimator will directly affect the rationality of the controller. Therefore, the design of the fault estimator is particularly important in fault estimation research.

At the same time, it is difficult to maximize or minimize the performance index of the control system in practice due to the influence of the environment and the limitations of the engineering system [15]. Therefore, the optimization problem is a difficult problem in the current control field and has gradually become the focus of attention [16]. Finally, the optimal control problem of the nonlinear system was transformed into the solution of Hamilton-Jacobi-Bellman (HJB) partial differential equations. In [17], the identifier-Actor-critic architecture was proposed, where the actor neural network performs control actions, the critic neural network is used to estimate these actions, and then the evaluation is returned to the actor, and the dynamic process of the uncertain system can be approximated by the neural network identifier [18]. According to [19], for a class of nonlinear large-scale systems with strict feedback structure. In [20], a fuzzy distributed adaptive optimal control method is proposed and FLS is used to identify the uncertain nonlinear function of the system. In [21], the output feedback adaptive neural network optimization control problem is solved for a quarter of the vehicle active electric suspension system. However, none of the above studies have discussed how to handle unknown system failures, so it will be a meaningful topic to design an optimal control strategy for the SLRA system that can handle unknown system failures.

Given this, drawing on the aforementioned research, this paper aims to develop an optimal output feedback controller based on a fault estimator for the SLRA model. This will ensure that the dynamic process of the system can be reconstructed even in the presence of an unknown fault, while simultaneously optimizing the performance index of the controlled system.

The remaining sections of the paper are structured as follows: Section II provides a problem description and background knowledge, Section III presents the principal findings, Section IV offers an in-depth stability analysis, Section V demonstrates a simulation example, and finally, the conclusion section summarizes the key outcomes.

Manuscript received July 21, 2024; revised November 20, 2024.

Sihan Liu is a teacher of School of Information Engineering, Liaoning Institute of Science and Engineering, Jinzhou 121010, China. (Corresponding author, e-mail: 1350907346@qq.com).

Hailong Yan is an Associate Professor of School of Information Engineering, Liaoning Institute of Science and Engineering, Jinzhou 121010, China. (e-mail: 77904555@qq.com).

Lixia Zhao is a Professor of School of Information Engineering, Liaoning Institute of Science and Engineering, Jinzhou 121010, China. (e-mail: 396476831@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: 1074501593@qq.com).

II. PROBLEM FORMULATION AND PRELIMINARIES

For single-link robot arm, as shown in Fig. 1, its dynamic behavior can be described by the following equations:

$$M\ddot{q}+\frac{1}{2}mgl\sin(q)=\tau+\zeta(t) \tag{1}$$

where q is the angle position, $\dot{q}$ is the angular velocity, $\ddot{q}$ is the angular acceleration, $g=9.8m/s^2$ is the acceleration due to gravity, M is the inertia, l is the length of the link, m is the mass of the link, and τ is the control force.

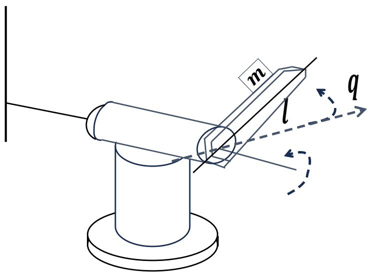

Fig. 1: Model diagram of a single-link robot arm.

According to Equation (1), its dynamic equation can be transformed into a state space of the following form

$$\begin{cases}\dot{x}_1=x_2\\ \dot{x}_2=bu+f(x)+\zeta(t)\\ y=x_1\end{cases} \tag{2}$$

where $x=[x_1,x_2]^T\in\mathbb{R}^2$ denotes the state variables, and x_1 is a measurable state, x_2 is an unmeasured state. u and y are the control input and output, respectively. $f(x)=\frac{1}{2M}mgl\sin(q)$ is a known continuous function, and b is a known control gain constant. In order to be controllable (2), the requirement is that $b\neq 0$. $\zeta(t)$ represents the time-varying fault of the unknown system.

Definition 1 For system (2), if the control protocol u is continuous and satisfies $u(0)=0$, then $u\in\Omega_u$ is admissible control and u makes system (2) stable while ensuring that the performance cost function is finite. Ω_u denotes an admissible control set.

Control Objective: For the machine arm system (2), a fault estimator-based output feedback optimal control strategy is proposed to ensure that: 1) the output signal y can track the reference signal y_r; 2) while saving communication resources and minimizing the cost function, all signals in system (2) are bounded.

To achieve the above control objectives, the following lemmas and assumptions need to be introduced.

Lemma 1 [22] (Young's inequality) For $\forall\gamma_1,\gamma_2\in\mathbb{R}$, one has

$$\gamma_1\gamma_2\leq\frac{\varpi^\iota}{\iota}|\gamma_1|^\iota+\frac{1}{o\varpi^o}|\gamma_2|^o \tag{3}$$

where $\varpi>0,\iota>1,o>1$ and $(\iota-1)(o-1)=1$.

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

$$\sup_{x\in\Omega_x}|f(x)-W^T\Psi(x)|\leq\varepsilon \tag{4}$$

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

$$W^*=\arg\min_{W\in\mathbb{R}^m}\{\sup_{x\in\Omega_x}|f(x)-W^T\Psi(x)|\} \tag{5}$$

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

$$f(x)=W^{*T}\Psi(x)+\varepsilon(x) \tag{6}$$

where $\varepsilon(x)$ is the NN approximation error, which can be bounded by $|\varepsilon(x)|\leq\overline{\varepsilon}$, where $\overline{\varepsilon}$ 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] Let $V(t)\in R$ be a positive continuous function. If it satisfies the inequality $\dot{V}(t)\leq-\beta V(t)+c$, where $\beta,c\in R^+$, then the following inequality holds:

$$V(t)\leq e^{-\beta t}V(0)+\frac{c}{\beta}(1-e^{-\beta t}) \tag{7}$$

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

Assumption 2. The Gaussian function $\Psi(x)$ satisfies the following global Lipschitz continuity condition:

$$\|\Psi(x)-\Psi(\overline{x})\|\leq L_\Psi\|x-\overline{x}\| \tag{8}$$

where L_Ψ is the Lipschitz constant.

III. MAIN RESULT

In this section, we assume that the state x_2 in system (2) is unmeasurable, and then we establish a neural network observer with an intermediate variable fault estimator. Then, an optimal controller is constructed based on a reinforcement learning algorithm.

A. Design of NN state observer and fault estimator

With the help of Lemma 2, it follows that the nonlinear terms in system (2) can be approximated by $\hat{f}(\hat{x}|\hat{W})=\hat{W}_f^T\Psi_f(\hat{x})$ and $\hat{f}(\hat{x}|W^*)=W_f^{*T}\Psi_f(\hat{x})$, where $\hat{x}$ and $\hat{W}$ represent the estimations of x and W^*, respectively. Define the variables errors $\delta_f=f(\overline{x})-\hat{f}(\hat{\overline{x}}|\hat{W})$ and $\varepsilon_f=f(x)-\hat{f}(\hat{x}|W^*)$, and there exist positive constants $\overline{\delta}_f$ and $\overline{\varepsilon}_f$ such that $|\delta_f|\leq\overline{\delta}_f$ and $|\varepsilon_f|\leq\overline{\varepsilon}_f$.

Define the state observer as follows:

$$\begin{cases}\dot{\hat{x}}_1=\hat{x}_2+l_1(y-\hat{x}_1)\\ \dot{\hat{x}}_2=bu+\hat{f}(\hat{x}|\hat{W})+\hat{\zeta}(t)+l_2(y-\hat{x}_1)\\ \hat{y}=\hat{x}_1\end{cases} \tag{9}$$

where $\hat{\zeta}(t)$ is estimate of $\zeta(t)$ and $l_i,i=1,2$ represent the design parameters. Define the observer error $\tilde{x}=x-\hat{x}$, the NN approximation error $\tilde{W}_f=W_f^*-\hat{W}_f$, and the fault estimation error $\tilde{\zeta}(t)=\zeta(t)-\hat{\zeta}(t)$. It follows from (2) and (12) that

$$\dot{\tilde{x}}=A\tilde{x}+\sum_{i=1}^{n}B(\delta_f+\tilde{\zeta}(t)) \tag{10}$$

where $A = \begin{bmatrix} -l_1 & 1 \\ -l_2 & 0 \end{bmatrix}$, $B = [0,1]^T$. A suitable choice of parameter l_i ensures that the matrix A is a strict Hurwitz matrix and there exist any positive definite matrices P and Q such that $A^TP + PA = -Q$.

The intermediate variable $\Xi(t)$ with tuning parameters η is defined as follows:

$$\Xi(t) = \zeta(t) - \eta x_2 \tag{11}$$

where $\eta > 0$. In addition, $\Xi(t)$ and $\zeta(t)$ are replaced by Ξ and ζ next for ease of writing.

The derivative of the intermediate variable Ξ yields

$$\begin{aligned} \dot{\Xi} &= \dot{\zeta} - \eta \dot{x}_2 \\ &= \dot{\zeta} - \eta\big(bu + f(x) + \zeta(t)\big) \end{aligned} \tag{12}$$

The adaptation law $\hat{\Xi}$ for the intermediate variable is designed as follows:

$$\dot{\hat{\Xi}} = -\eta\big(bu + \hat{f}(\hat{x}|\hat{W}) + \hat{\zeta}\big) \tag{13}$$

It can be obtained from (12) and (13) that the error between the intermediate variable and its estimate are denoted by:

$$\begin{aligned} \dot{\tilde{\Xi}} &= \dot{\zeta} - \eta(\delta_f + \tilde{\zeta}) \\ &= \dot{\zeta} - \eta(\delta_f + \tilde{\Xi} + \eta\tilde{x}_2) \end{aligned} \tag{14}$$

Define the intermediate variable error as $\tilde{\Xi} = \Xi - \hat{\Xi}$. Subsequently, consider the following Lyapunov function:

$$V_0 = \tilde{x}^TP\tilde{x} + \frac{1}{2}\tilde{\Xi}^2 \tag{15}$$

Then, the derivative of V_0 is equal to

$$\begin{aligned} \dot{V}_0 =& \dot{\tilde{x}}^TP\tilde{x} + \tilde{x}^TP\dot{\tilde{x}} + \tilde{\Xi}\dot{\tilde{\Xi}} \\ =& \tilde{x}^T(A^TP + PA)\tilde{x} + 2\tilde{x}^TPB(\delta_f + \tilde{\zeta}) \\ &+ \tilde{\Xi}\big(\dot{\zeta} - \eta(\delta_f + \tilde{\Xi} + \eta\tilde{x}_2)\big) \\ =& -\tilde{x}^TQ\tilde{x} + 2\tilde{x}^TPB(\delta_f + \tilde{\zeta}) + \tilde{\Xi}\dot{\zeta} \\ & - \eta\delta_f\tilde{\Xi} - \eta\tilde{\Xi}^2 - \eta^2\tilde{x}_2\tilde{\Xi} \end{aligned} \tag{16}$$

According to Young's inequality, it follows that

$$\begin{aligned} 2\tilde{x}^TPB\delta_f &\le \|\tilde{x}\|^2 + \|P\|^2\overline{\delta}_f^2 \\ 2\tilde{x}^TPB\tilde{\zeta} &\le (\|P\|^2 + 2\eta^2)\|\tilde{x}\|^2 + 2\tilde{\Xi}^2 \\ \tilde{\Xi}\dot{\zeta} &\le \frac{1}{2}\tilde{\Xi}^2 + \frac{1}{2}\overline{\zeta}^2 \\ -\eta\delta_f\tilde{\Xi} &\le \frac{1}{2}\eta^2\overline{\delta}_f^2 + \frac{1}{2}\tilde{\Xi}^2 \\ -\eta^2\tilde{x}_2\tilde{\Xi} &\le \frac{1}{2}\eta^4\|\tilde{x}\|^2 + \frac{1}{2}\tilde{\Xi}^2 \end{aligned} \tag{17}$$

Substituting (17) into (16) can be obtained

$$\begin{aligned} \dot{V}_0 \le& -\big(\lambda_{\min}(Q) - 1 - \|P\|^2 - 2\eta^2 - \frac{1}{2}\eta^4\big)\|\tilde{x}\|^2 \\ & - \frac{7}{2}\tilde{\Xi}^2 + (\|P\|^2 + \frac{1}{2}\eta^2)\overline{\delta}_f^2 + \sum_{i=1}^{n}\frac{1}{2}\overline{\zeta}^2 \\ \le& -\tau\|\tilde{x}\|^2 - \frac{7}{2}\tilde{\Xi}^2 + \sigma_0 \end{aligned} \tag{18}$$

where

$$\begin{aligned} \tau =& \lambda_{\min}(Q) - 1 - \|P\|^2 - 2\eta^2 - \frac{1}{2}\eta^4, \\ \sigma_0 =& (\|P\|^2 + \frac{1}{2}\eta^2)\overline{\delta}_f^2 + \sum_{i=1}^{n}\frac{1}{2}\overline{\zeta}^2. \end{aligned} \tag{19}$$

B. Optimized backstepping controller design

In the following, a optimized output feedback algorithm is developed based on the NN state observer (9) and the intermediate variable (11). The optimized backstepping control is designed by the following two steps. Consider the tracking error coordinate transformation as follows:

$$\begin{aligned} z_1 &= x_1 - y_r \\ z_2 &= x_2 - \hat{\alpha}_1^* \end{aligned} \tag{20}$$

Among them, y_r serves as the reference signal, while α_1 and α_1^* represent the virtual control and optimal virtual control correspondingly.

Step 1: From (2) and (20), the derivative of z_1 can be calculated

$$\dot{z}_1 = \dot{x}_1 - \dot{y}_r = x_2 - \dot{y}_r \tag{21}$$

The optimal performance index function is chosen as

$$J_1(z_1) = \int_t^\infty h_1\Big(z_1(v), \alpha_1\big(z_1(v)\big)\Big)dv \tag{22}$$

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 (22), the optimal performance index function can be obtained

$$\begin{aligned} J_1^*(z_1) &= \int_t^\infty h_1(z_1(v), \alpha_1^*(z_1(v)))dv \\ &= \min_{\alpha_1 \in \Omega}\{\int_t^\infty h_1\Big(z_1(v), \alpha_1\big(z_1(v)\big)\Big)dv\} \end{aligned} \tag{23}$$

Replace x_2 in (21) with the optimal virtual control α_1^*, and subsequently define the HJB equation associated with (21) and (23) as

$$H_1(z_1, \alpha_1^*, \frac{\mathrm{d}J_1^*}{\mathrm{d}z_1}) = z_1^2 + {\alpha_1^*}^2 + \frac{\mathrm{d}J_1^*}{\mathrm{d}z_1}(\alpha_1^* - \dot{y}_r) = 0 \tag{24}$$

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

$$\alpha_1^* = -\frac{1}{2}\frac{\mathrm{d}J_1^*(z_1)}{\mathrm{d}z_1} \tag{25}$$

According to (9) and (20), $z_1 = x_1 - y_r$ can be estimated as $\hat{z}_1 = \hat{x}_1 - y_r$, so replacing z_1 with $\hat{z}_1$ yields $\frac{\mathrm{d}J_1^*(\hat{z}_1)}{\mathrm{d}\hat{z}_1}$. To achieve the fixed-time optimal control objective, $\frac{\mathrm{d}J_1^*(\hat{z}_1)}{\mathrm{d}\hat{z}_1}$ is estimated and decomposed into

$$\frac{\mathrm{d}J_1^*(\hat{z}_1)}{\mathrm{d}\hat{z}_1} = 2\varrho_1\hat{z}_1 + J_1^o(\hat{z}_1) \tag{26}$$

where $\varrho_1 > \frac{7}{4}$ is design parameter. $J_1^o(\hat{z}_1) = -2\varrho_1\hat{z}_1 + \frac{\mathrm{d}J_1^*(\hat{z}_1)}{\mathrm{d}\hat{z}_1} \in \mathbb{R}$ is a continuous function, and substituting (26) into (25) has

$$\alpha_1^* = -\varrho_1\hat{z}_1 - \frac{1}{2}J_1^o(\hat{z}_1) \tag{27}$$

Since $J_1^o(\hat{z}_1)$ is continuous unknown function, it can be approximated by NN as follows:

$$J_1^o(\hat{z}_1) = W_{J1}^{*T}\Psi_{J1}(\hat{z}_1) + \varepsilon_{J1}(\hat{z}_1) \tag{28}$$

where $W_{J1}^* \in \mathbb{R}^{m1}$ represents the ideal weight vector, $\Psi_{J1}(\hat{z}_1) \in \mathbb{R}^{m1}$ is the basis function vector, and $\varepsilon_{J1}(\hat{z}_1) \in \mathbb{R}$ represents the approximation error bounded by $\|\varepsilon_{J1}(\hat{z}_1)\| \le \overline{\varepsilon}_{J1}$ as arbitrarily small. Then, (26) and (27) can be reorganized as

$$\frac{\mathrm{d}J_1^*(\hat{z}_1)}{\mathrm{d}\hat{z}_1} = 2\varrho_1\hat{z}_1 + W_{J1}^{*T}\Psi_{J1}(\hat{z}_1) + \varepsilon_{J1} \tag{29}$$

$$\alpha_1^* = -\varrho_1 \hat{z}_1 - \frac{1}{2} W_{J1}^{*T} \Psi_{J1}(\hat{z}_1) - \frac{1}{2} \varepsilon_{J1} \tag{30}$$

Since W_{J1}^* is unknown constant vector, the optimal virtual control (30) is not available for the controlled system. To derive the effective optimized virtual control, the following RL algorithm with critic and actor is performed.

$$\frac{\mathrm{d}\hat{J}_1^*(\hat{z}_1)}{\mathrm{d}\hat{z}_1} = 2\varrho_1 \hat{z}_1 + \hat{W}_{c1}^T \Psi_{J1}(\hat{z}_1) \tag{31}$$

$$\hat{\alpha}_1^* = -\varrho_1 \hat{z}_1 - \frac{1}{2} \hat{W}_{a1}^T \Psi_{J1}(\hat{z}_1) \tag{32}$$

where $\frac{\mathrm{d}\hat{J}_1^*(\hat{z}_1)}{\mathrm{d}\hat{z}_1}$ and $\hat{\alpha}_1^*$ are the estimates of $\frac{\mathrm{d}J_1^*(\hat{z}_1)}{\mathrm{d}\hat{z}_1}$ and α_1^*, respectively. $\hat{W}_{c1}^T \Psi_{J1}(\hat{z}_1) \in \mathbb{R}^{m1}$ and $\hat{W}_{a1}^T \Psi_{J1}(\hat{z}_1) \in \mathbb{R}^{m1}$ 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.

$$\dot{\hat{W}}_{c1} = -\kappa_{c1} \Psi_{J1}(\hat{z}_1) \Psi_{J1}^T(\hat{z}_1) \hat{W}_{c1} \tag{33}$$

$$\begin{aligned}\dot{\hat{W}}_{a1} = &- \Psi_{J1}(\hat{z}_1) \Psi_{J1}^T(\hat{z}_1) \big(\kappa_{a1}(\hat{W}_{a1} \\ &- \hat{W}_{c1}) + \kappa_{c1} \hat{W}_{c1} \big)\end{aligned} \tag{34}$$

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 $x_2 = \tilde{x}_2 + \hat{x}_2$, (21) can be rewritten as

$$\dot{\hat{z}}_1 = \tilde{x}_2 + \hat{x}_2 - \dot{y}_r = \hat{z}_2 + \hat{\alpha}_1^* + \tilde{x}_2 - \dot{y}_r \tag{35}$$

Substituting (32) into (35) to get

$$\dot{\hat{z}}_1 = - \varrho_1 \hat{z}_1 + \hat{z}_2 - \frac{1}{2} \hat{W}_{a1}^T \Psi_{J1}(\hat{z}_1) + \tilde{x}_2 - \dot{y}_r \tag{36}$$

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

$$V_1 = \frac{1}{2} \hat{z}_1^2 + \frac{1}{2} \tilde{W}_{c1}^T \tilde{W}_{c1} + \frac{1}{2} \tilde{W}_{a1}^T \tilde{W}_{a1} \tag{37}$$

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

$$\begin{aligned}\dot{V}_1 =& \hat{z}_1 \big(- \varrho_1 \hat{z}_1 + \hat{z}_2 - \frac{1}{2} \hat{W}_{a1}^T \Psi_{J1}(\hat{z}_1) + \tilde{x}_2 - \dot{y}_r \big) \\ &+ \kappa_{c1} \tilde{W}_{c1}^T \Psi_{J1}(\hat{z}_1) \Psi_{J1}^T(\hat{z}_1) \hat{W}_{c1} + \tilde{W}_{a1}^T \Psi_{J1}(\hat{z}_1) \\ &\times \Psi_{J1}^T(\hat{z}_1) \big(\kappa_{a1}(\hat{W}_{a1} - \hat{W}_{c1}) + \kappa_{c1} \hat{W}_{c1} \big)\end{aligned} \tag{38}$$

The Young's inequality yields the following results:

$$\begin{aligned}\hat{z}_1 \hat{z}_2 &\leq \frac{1}{2} \hat{z}_1^2 + \frac{1}{2} \hat{z}_2^2 \\ \hat{z}_1 \tilde{x}_2 &\leq \frac{1}{2} \hat{z}_1^2 + \frac{1}{2} \|\tilde{x}\|^2 \\ -\hat{z}_1 \dot{y}_r &\leq \frac{1}{2} \hat{z}_1^2 + \frac{1}{2} \dot{y}_r^2 \\ -\frac{1}{2} \hat{z}_1 \hat{W}_{a1}^T \Psi_{J1}(\hat{z}_1) &\leq \frac{1}{4} \hat{z}_1^2 + \frac{1}{4} \hat{W}_{a1}^T \Psi_{J1}(\hat{z}_1) \Psi_{J1}^T(\hat{z}_1) \hat{W}_{a1}\end{aligned} \tag{39}$$

Along with (38) and (39), we can calculate:

$$\begin{aligned}\dot{V}_1 \leq& - (\varrho_1 - \frac{7}{4}) \hat{z}_1^2 + \kappa_{c1} \tilde{W}_{c1}^T \Psi_{J1}(\hat{z}_1) \Psi_{J1}^T(\hat{z}_1) \hat{W}_{c1} \\ &+ \kappa_{a1} \tilde{W}_{a1}^T \Psi_{J1}(\hat{z}_1) \Psi_{J1}^T(\hat{z}_1) \hat{W}_{a1} + \frac{1}{2} \|\tilde{x}\|^2 \\ &+ (\kappa_{c1} - \kappa_{a1}) \tilde{W}_{a1}^T \Psi_{J1}(\hat{z}_1) \Psi_{J1}^T(\hat{z}_1) \hat{W}_{c1} \\ &+ \frac{1}{4} \hat{W}_{a1}^T \Psi_{J1}(\hat{z}_1) \Psi_{J1}^T(\hat{z}_1) \hat{W}_{a1} + \frac{1}{2} \hat{z}_2^2 + \frac{1}{2} \dot{y}_r^2\end{aligned} \tag{40}$$

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

$$\begin{aligned}\tilde{W}_{c1}^T \Psi_{J1}(\hat{z}_1) \Psi_{J1}^T(\hat{z}_1) \hat{W}_{c1} =& \frac{1}{2} W_{J1}^{*T} \Psi_{J1}(\hat{z}_1) \Psi_{J1}^T(\hat{z}_1) W_{J1}^* \\ &- \frac{1}{2} \tilde{W}_{c1}^T \Psi_{J1}(\hat{z}_1) \Psi_{J1}^T(\hat{z}_1) \tilde{W}_{c1} \\ &- \frac{1}{2} \hat{W}_{c1}^T \Psi_{J1}(\hat{z}_1) \Psi_{J1}^T(\hat{z}_1) \hat{W}_{c1} \\ \tilde{W}_{a1}^T \Psi_{J1}(\hat{z}_1) \Psi_{J1}^T(\hat{z}_1) \hat{W}_{a1} =& \frac{1}{2} W_{J1}^{*T} \Psi_{J1}(\hat{z}_1) \Psi_{J1}^T(\hat{z}_1) W_{J1}^* \\ &- \frac{1}{2} \tilde{W}_{a1}^T \Psi_{J1}(\hat{z}_1) \Psi_{J1}^T(\hat{z}_1) \tilde{W}_{a1} \\ &- \frac{1}{2} \hat{W}_{a1}^T \Psi_{J1}(\hat{z}_1) \Psi_{J1}^T(\hat{z}_1) \hat{W}_{a1} \\ \tilde{W}_{a1}^T \Psi_{J1}(\hat{z}_1) \Psi_{J1}^T(\hat{z}_1) \hat{W}_{c1} \leq& - \frac{1}{2} \tilde{W}_{a1}^T \Psi_{J1}(\hat{z}_1) \Psi_{J1}^T(\hat{z}_1) \tilde{W}_{a1} \\ &- \frac{1}{2} \hat{W}_{c1}^T \Psi_{J1}(\hat{z}_1) \Psi_{J1}^T(\hat{z}_1) \hat{W}_{c1}\end{aligned} \tag{41}$$

Subsequently, we can acquire

$$\begin{aligned}\dot{V}_1 \leq& - (\varrho_1 - \frac{7}{4}) \hat{z}_1^2 - \frac{\kappa_{c1}}{2} \tilde{W}_{c1}^T \Psi_{J1}(\hat{z}_1) \Psi_{J1}^T(\hat{z}_1) \tilde{W}_{c1} \\ &- (\kappa_{a1} - \frac{\kappa_{c1}}{2}) \tilde{W}_{a1}^T \Psi_{J1}(\hat{z}_1) \Psi_{J1}^T(\hat{z}_1) \tilde{W}_{a1} \\ &- \frac{\kappa_{a1}}{2} \hat{W}_{c1}^T \Psi_{J1}(\hat{z}_1) \Psi_{J1}^T(\hat{z}_1) \hat{W}_{c1} - (\frac{\kappa_{a1}}{2} - \frac{1}{4}) \\ &\times \hat{W}_{a1}^T \Psi_{J1}(\hat{z}_1) \Psi_{J1}^T(\hat{z}_1) \hat{W}_{a1} + \frac{1}{2} \hat{z}_2^2 + \frac{1}{2} \|\tilde{x}\|^2 \\ &+ \frac{1}{2} \dot{y}_r^2 + \frac{\kappa_{c1} + \kappa_{a1}}{2} W_{J1}^{*T} \Psi_{J1}(\hat{z}_1) \Psi_{J1}^T(\hat{z}_1) W_{J1}^*\end{aligned} \tag{42}$$

The following inequality holds when $\lambda_{\Psi_{J1}}^{\min}$ is the minimum eigenvalue of $\Psi_{J1}(\hat{z}_1) \Psi_{J1}^T(\hat{z}_1)$.

$$\begin{aligned}&- \tilde{W}_{c1}^T \Psi_{J1}(\hat{z}_1) \Psi_{J1}^T(\hat{z}_1) \tilde{W}_{c1} \leq -\lambda_{\Psi_{J1}}^{\min} \tilde{W}_{c1}^T \tilde{W}_{c1} \\ &- \tilde{W}_{a1}^T \Psi_{J1}(\hat{z}_1) \Psi_{J1}^T(\hat{z}_1) \tilde{W}_{a1} \leq -\lambda_{\Psi_{J1}}^{\min} \tilde{W}_{a1}^T \tilde{W}_{a1}\end{aligned} \tag{43}$$

According to the design parameters $\kappa_{a1} > \frac{\kappa_{c1}}{2}$ and $\kappa_{a1} > \frac{1}{2}$, as well as (43), it can yield

$$\begin{aligned}\dot{V}_1 \leq& - (\varrho_1 - \frac{7}{4}) \hat{z}_1^2 - \frac{\kappa_{c1}}{2} \lambda_{\Psi_{J1}}^{\min} \tilde{W}_{c1}^T \tilde{W}_{c1} \\ &- (\kappa_{a1} - \frac{\kappa_{c1}}{2}) \lambda_{\Psi_{J1}}^{\min} \tilde{W}_{a1}^T \tilde{W}_{a1} + \frac{1}{2} \|\tilde{x}\|^2 \\ &+ \frac{1}{2} \hat{z}_2^2 + \sigma_1\end{aligned} \tag{44}$$

where $\sigma_1 = \frac{1}{2} \dot{y}_r^2 + \frac{\kappa_{c1} + \kappa_{a1}}{2} W_{J1}^{*T} \Psi_{J1}(\hat{z}_1) \Psi_{J1}^T(\hat{z}_1) W_{J1}^*$. 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.

$$\begin{aligned}\dot{z}_2 &= \dot{x}_2 - \dot{\hat{\alpha}}_1^* \\ &= bu + f(x) + \zeta - \dot{\hat{\alpha}}_1^*\end{aligned} \tag{45}$$

The selection of the most suitable integral cost function is detailed as follows:

$$\begin{aligned}J_2^*(z_2) &= \int_t^\infty h_2 \Big(z_2(v), u^* \big(z_2(v) \big) \Big) dv \\ &= \min_{u \in \Omega} \{ \int_t^\infty h_2 \Big(z_2(v), u \big(z_2(v) \big) \Big) dv \}\end{aligned} \tag{46}$$

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

Based on (46), the HJB equation is constructed as

$$H_2(z_2,u^*,\frac{\mathrm{d}J_2^*}{\mathrm{d}z_2}) = z_2^2+u^{*2}+\frac{\mathrm{d}J_2^*}{\mathrm{d}z_2}\big(u^*+f(x)+\zeta-\dot{\hat{\alpha}}_1^*\big)=0 \tag{47}$$

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

$$u^*=-\frac{1}{2}\frac{\mathrm{d}J_2^*(z_2)}{\mathrm{d}z_2} \tag{48}$$

The tracking error z_2 can be estimated as $\hat{z}_2=\hat{x}_2-\dot{\hat{\alpha}}_1^*$, where z_2 is substituted with $\hat{z}_2$, resulting in

$$u^*=-\frac{1}{2}\frac{\mathrm{d}J_2^*(\hat{z}_2)}{\mathrm{d}\hat{z}_2} \tag{49}$$

Then, $\frac{\mathrm{d}J_2^*(\hat{z}_2)}{\mathrm{d}\hat{z}_2}$ can be factored as

$$\frac{\mathrm{d}J_2^*(\hat{z}_2)}{\mathrm{d}\hat{z}_2}=2\varrho_2\hat{z}_2+2f(\hat{x})+2\hat{\zeta}+J_2^o(\hat{z}_2) \tag{50}$$

where $\varrho_2>\frac{9}{4}$ is design parameter. $J_2^o(\hat{z}_2)=-2\varrho_2\hat{z}_2-2f(\hat{x})-2\hat{\zeta}+\frac{\mathrm{d}J_2^*(\hat{z}_2)}{\mathrm{d}\hat{z}_2}\in\mathbb{R}$. Using NN, we can approximate $f(\hat{x})$ as $W_f^{*T}\Psi_f(\hat{x})+\varepsilon_f(\hat{x})$, the optimal controller u^* can be expressed as

$$u^*=-\varrho_2\hat{z}_2-W_f^{*T}\Psi_f(\hat{x})-\varepsilon_f-\hat{\zeta}-\frac{1}{2}J_2^o(\hat{z}_2) \tag{51}$$

Since $J_2^o(\hat{z}_2)$ is unknown continuous term, it can also be approximated using NN as follows:

$$J_2^o(\hat{z}_2)=W_{J2}^{*T}\Psi_{J2}(\hat{z}_2)+\varepsilon_{J2}(\hat{z}_2) \tag{52}$$

where $W_{J2}^*\in\mathbb{R}^{m2}$ is the ideal weight vector, $\Psi_{J2}(\hat{z}_2)\in\mathbb{R}^{m2}$ is the NN basis function vector, and the NN approximation error $\varepsilon_{J2}(\hat{z}_2)\in\mathbb{R}$ is bounded.

Similarly, we can derive the following conclusion

$$\frac{\mathrm{d}J_2^*(\hat{z}_2)}{\mathrm{d}\hat{z}_2}=2\varrho_2\hat{z}_2+2\hat{\zeta}+2W_f^{*T}\Psi_f(\hat{x})+W_{J2}^{*T}\Psi_{J2}(\hat{z}_2)+\varepsilon_2 \tag{53}$$

$$u^*=-\varrho_2\hat{z}_2-\hat{\zeta}-W_f^{*T}\Psi_f(\hat{x})-\frac{1}{2}W_{J2}^{*T}\Psi_{J2}(\hat{z}_2)-\frac{1}{2}\varepsilon_2 \tag{54}$$

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

The optimal control (54), however, remains unattainable, necessitating the execution of an RL algorithm featuring both a critic and an actor to acquire viable control signal.

$$\frac{\mathrm{d}\hat{J}_2^*(\hat{z}_2)}{\mathrm{d}\hat{z}_2}=2\varrho_2\hat{z}_2+2\hat{\zeta}+2\hat{W}_f^T\Psi_{J2}(\hat{x})+\hat{W}_{c2}^T\Psi_{J2}(\hat{z}_2) \tag{55}$$

$$\hat{u}^*=\frac{1}{b}\big(-\varrho_2\hat{z}_2-\hat{\zeta}-\hat{W}_f^T\Psi_{J2}(\hat{x})-\frac{1}{2}\hat{W}_{a2}^T\Psi_{J2}(\hat{z}_2)\big) \tag{56}$$

where $\frac{\mathrm{d}\hat{J}_2^*(\hat{z}_2)}{\mathrm{d}\hat{z}_2}$ is the estimate of $\frac{\mathrm{d}J_2^*(\hat{z}_2)}{\mathrm{d}\hat{z}_2}$, and $\hat{u}^*$ is the final actual optimal controller. $\hat{W}_{c2}^T\Psi_{J2}(\hat{z}_2)\in\mathbb{R}^{m2}$ and $\hat{W}_{a2}^T\Psi_{J2}(\hat{z}_2)\in\mathbb{R}^{m2}$ are the NN weight vectors of critic and actor, respectively.

As with the previous steps, the corresponding three adaptive update laws are designed as follows:

$$\dot{\hat{W}}_f=\Gamma_f\Psi_{Jn}(\hat{x})\hat{z}_2-\kappa_f\hat{W}_f \tag{57}$$

$$\dot{\hat{W}}_{c2}=-\kappa_{c2}\Psi_{J2}(\hat{z}_2)\Psi_{J2}^T(\hat{z}_2)\hat{W}_{c2} \tag{58}$$

$$\dot{\hat{W}}_{a2}=-\Psi_{J2}(\hat{z}_2)\Psi_{J2}^T(\hat{z}_2)\big(\kappa_{a2}(\hat{W}_{a2}-\hat{W}_{c2})+\kappa_{c2}\hat{W}_{c2}\big) \tag{59}$$

where $\Gamma_f>0$, $\kappa_f>0$, $\kappa_{c2}>0$ 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 the actual optimal control (56), the $\dot{\hat{z}}_2$ can be expressed as follows

$$\begin{aligned}\dot{\hat{z}}_2=&b\hat{u}^*+\tilde{W}_f^T\Psi_f(\hat{x})+\hat{W}_f^T\Psi_f(\hat{x})+\varepsilon_f+\zeta-\dot{\hat{\alpha}}_1^*\\=&-\varrho_2\hat{z}_2+\tilde{\zeta}-\frac{1}{2}\hat{W}_{a2}^T\Psi_{J2}(\hat{z}_2)+\tilde{W}_f^T\Psi_f(\hat{x})+\varepsilon_f-\dot{\hat{\alpha}}_1^*\end{aligned} \tag{60}$$

Subsequently, the nth step Lyapunov function should be selected as

$$V_2=\frac{1}{2}\hat{z}_2^2+\frac{1}{2\Gamma_f}\tilde{W}_f^T\tilde{W}_f+\frac{1}{2}\tilde{W}_{c2}^T\tilde{W}_{c2}+\frac{1}{2}\tilde{W}_{a2}^T\tilde{W}_{a2} \tag{61}$$

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

Combining (56)-(60), the derivative of (61) can be calculated as

$$\begin{aligned}\dot{V}_2=&\hat{z}_2\big(-\varrho_2\hat{z}_2+\tilde{\zeta}-\frac{1}{2}\hat{W}_{a2}^T\Psi_{J2}(\hat{z}_2)+\tilde{W}_f^T\Psi_f(\hat{x})\\&+\varepsilon_f-\dot{\hat{\alpha}}_1^*\big)+\kappa_f\tilde{W}_f^T\hat{W}_f+\kappa_{c2}\tilde{W}_{c2}^T\Psi_{J2}(\hat{z}_2)\\&\times\Psi_{J2}^T(\hat{z}_2)\hat{W}_{c2}+\tilde{W}_{a2}^T\Psi_{J2}(\hat{z}_2)\Psi_{J2}^T(\hat{z}_2)\\&\times\big(\kappa_{a2}(\hat{W}_{a2}-\hat{W}_{c2})+\kappa_{c2}\hat{W}_{c2}\big)\end{aligned} \tag{62}$$

Using the Young's inequality, we have

$$\begin{aligned}\hat{z}_2\varepsilon_f\le&\frac{1}{2}\hat{z}_2^2+\frac{1}{2}\bar{\varepsilon}_f^2\\\hat{z}_2\tilde{\zeta}\le&\frac{1}{2}\hat{z}_2^2+2\tilde{\Xi}^2+2\eta^2\|\tilde{x}\|^2\\-\hat{z}_2\dot{\hat{\alpha}}_1^*\le&\frac{1}{2}\hat{z}_2^2+\frac{1}{2}\dot{\hat{\alpha}}_1^{*2}\\-\frac{1}{2}\hat{z}_2\hat{W}_{a2}^T\Psi_{J2}(\hat{z}_2)\le&\frac{1}{4}\hat{z}_2^2+\frac{1}{4}\hat{W}_{a2}^T\Psi_{J2}(\hat{z}_2)\Psi_{J2}^T(\hat{z}_2)\hat{W}_{a2}\end{aligned} \tag{63}$$

Meanwhile, according to the previous steps, the following results hold:

$$\begin{aligned}\tilde{W}_f\hat{W}_f\le&-\frac{1}{2}\tilde{W}_f^T\tilde{W}_f+\frac{1}{2}W_f^{*T}W_f^*\\\tilde{W}_{c2}^T\Psi_{J2}(\hat{z}_2)\Psi_{J2}^T(\hat{z}_2)\hat{W}_{c2}=&\frac{1}{2}W_{J2}^{*T}\Psi_{J2}(\hat{z}_2)\Psi_{J2}^T(\hat{z}_2)W_{J2}^*\\&-\frac{1}{2}\tilde{W}_{c2}^T\Psi_{J2}(\hat{z}_2)\Psi_{J2}^T(\hat{z}_2)\tilde{W}_{c2}\\&-\frac{1}{2}\hat{W}_{c2}^T\Psi_{J2}(\hat{z}_2)\Psi_{J2}^T(\hat{z}_2)\hat{W}_{c2}\\\tilde{W}_{a2}^T\Psi_{J2}(\hat{z}_2)\Psi_{J2}^T(\hat{z}_2)\hat{W}_{a2}=&\frac{1}{2}W_{J2}^{*T}\Psi_{J2}(\hat{z}_2)\Psi_{J2}^T(\hat{z}_2)W_{J2}^*\\&-\frac{1}{2}\tilde{W}_{a2}^T\Psi_{J2}(\hat{z}_2)\Psi_{J2}^T(\hat{z}_2)\tilde{W}_{a2}\\&-\frac{1}{2}\hat{W}_{a2}^T\Psi_{J2}(\hat{z}_2)\Psi_{J2}^T(\hat{z}_2)\hat{W}_{a2}\\\tilde{W}_{a2}^T\Psi_{J2}(\hat{z}_2)\Psi_{J2}^T(\hat{z}_2)\hat{W}_{c2}\le&-\frac{1}{2}\tilde{W}_{a2}^T\Psi_{J2}(\hat{z}_2)\Psi_{J2}^T(\hat{z}_2)\tilde{W}_{a2}\\&-\frac{1}{2}\hat{W}_{c2}^T\Psi_{J2}(\hat{z}_2)\Psi_{J2}^T(\hat{z}_2)\hat{W}_{c2}\end{aligned} \tag{64}$$

Substituting (63) and (64) into (62) yields

$$
\begin{aligned}
\dot{V}_2 \leq & -(\varrho_2-\frac{7}{4})\hat{z}_2^2-\frac{\kappa_f}{2}\tilde{W}_f^T\tilde{W}_f-\frac{\kappa_{c2}}{2}\tilde{W}_{c2}^T\Psi_{J2}(\hat{z}_2)\\
&\times\Psi_{J2}^T(\hat{z}_2)\tilde{W}_{c2}-(\kappa_{a2}-\frac{\kappa_{c2}}{2})\tilde{W}_{a2}^T\Psi_{J2}(\hat{z}_2)\\
&\times\Psi_{J2}^T(\hat{z}_2)\tilde{W}_{a2}-\frac{\kappa_{a2}}{2}\hat{W}_{c2}^T\Psi_{J2}(\hat{z}_2)\Psi_{J2}^T(\hat{z}_2)\hat{W}_{c2}\\
&-(\frac{\kappa_{a2}}{2}-\frac{1}{4})\hat{W}_{a2}^T\Psi_{J2}(\hat{z}_2)\Psi_{J2}^T(\hat{z}_2)\hat{W}_{a2}+2\eta^2\|\tilde{x}\|^2\\
&+2\tilde{\Xi}^2+\frac{1}{2}\bar{\varepsilon}_f^2+\frac{1}{2}\dot{\hat{\alpha}}_1^{*2}+\frac{\kappa_f}{2}W_f^{*T}W_f^*+\frac{\kappa_{c2}+\kappa_{a2}}{2}\\
&\times W_{J2}^{*T}\Psi_{J2}(\hat{z}_2)\Psi_{J2}^T(\hat{z}_2)W_{J2}^*\\
\leq & -(\varrho_2-\frac{7}{4})\hat{z}_2^2-\frac{\kappa_f}{2}\tilde{W}_f^T\tilde{W}_f-\frac{\kappa_{c2}}{2}\lambda_{\Psi_{J2}}^{\min}\tilde{W}_{c2}^T\tilde{W}_{c2}\\
&-(\kappa_{a2}-\frac{\kappa_{c2}}{2})\lambda_{\Psi_{J2}}^{\min}\tilde{W}_{a2}^T\tilde{W}_{a2}+2\eta^2\|\tilde{x}\|^2+2\tilde{\Xi}^2+\sigma_2
\end{aligned} \tag{65}
$$

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

IV. STABILITY ANALYSIS

Theorem 1 If the optimal control strategy proposed in this paper is applied to the SLRA system with unknown fault (2), where the state observer is (9), the intermediate variable disturbance observer and its adaptive law are (13), the observer, critic, and actuator parameters are adaptive laws (57) and (33), (58) and (34), the optimal virtual controller is (32), and the fixed-time optimal actual controller is (56), then the control strategy can ensure that the control signal in the closed-loop system is always consistently ultimately bounded, and the reference signal can track the expected target.

Proof: Construct a Lyapunov function $V = \sum_{i=0}^{2} V_i$, and by integrating the preceding steps, we can compute

$$
\begin{aligned}
\dot{V} \leq & -(\tau-\frac{1}{2}+2\eta^2)\|\tilde{x}\|^2-\frac{3}{2}\tilde{\Xi}^2-(\varrho_1-\frac{7}{4})\hat{z}_1^2\\
&-(\varrho_2-\frac{9}{4})\hat{z}_2^2-\frac{\kappa_f}{2}\tilde{W}_f^T\tilde{W}_f-\sum_{i=1}^{2}\frac{\kappa_{ci}}{2}\lambda_{\Psi_{Ji}}^{\min}\tilde{W}_{ci}^T\tilde{W}_{ci}\\
&-\sum_{i=1}^{2}(\kappa_{ai}-\frac{\kappa_{ci}}{2})\lambda_{\Psi_{Ji}}^{\min}\tilde{W}_{ai}^T\tilde{W}_{ai}+\sigma_0+\sigma_1+\sigma_2\\
\leq & -\breve{\tau}\|\tilde{x}\|^2-\frac{3}{2}\tilde{\Xi}^2-(\varrho_1-\frac{7}{4})\hat{z}_1^2-(\varrho_2-\frac{9}{4})\hat{z}_2^2\\
&-\frac{\kappa_f}{2}\tilde{W}_f^T\tilde{W}_f-\sum_{i=1}^{2}\frac{\kappa_{ci}}{2}\lambda_{\Psi_{Ji}}^{\min}\tilde{W}_{ci}^T\tilde{W}_{ci}\\
&-\sum_{i=1}^{2}(\kappa_{ai}-\frac{\kappa_{ci}}{2})\lambda_{\Psi_{Ji}}^{\min}\tilde{W}_{ai}^T\tilde{W}_{ai}+\sigma
\end{aligned} \tag{66}
$$

where $\breve{\tau} = \tau - \frac{1}{2} + 2\eta^2$ and $\sigma = \sigma_0 + \bar{\sigma}_1 + \bar{\sigma}_2$. As all the terms in a are bounded, there exists a constant $c > 0$ such that $|\sigma| < c$.

In accordance with Lemma 3, the subsequent inequality can be deduced via algebraic manipulation.

$$
\dot{V} \leq -\beta V + c \tag{67}
$$

where $\beta = \min\{\frac{\tau}{\lambda_{\max}(P)}, \frac{3}{2}, \varrho_1 - \frac{7}{4}, \varrho_2 - \frac{9}{4}, \frac{\kappa_f}{2}, \frac{\kappa_{ci}}{2}\lambda_{\Psi_{Ji}}^{\min}, (\kappa_{ai} - \frac{\kappa_{ci}}{2})\lambda_{\Psi_{Ji}}^{\min}, i = 1, 2\}$.

Then, there was the following:

$$
V \leq e^{-\beta t}V(0) + \frac{c}{\beta}(1 - e^{-\beta t}) \tag{68}
$$

■

The proof of Theorem 1 is completed.

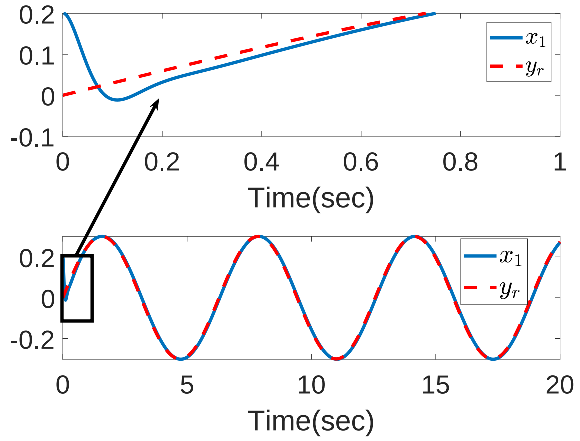

Fig. 2: The trajectories of x_1 and y_r.

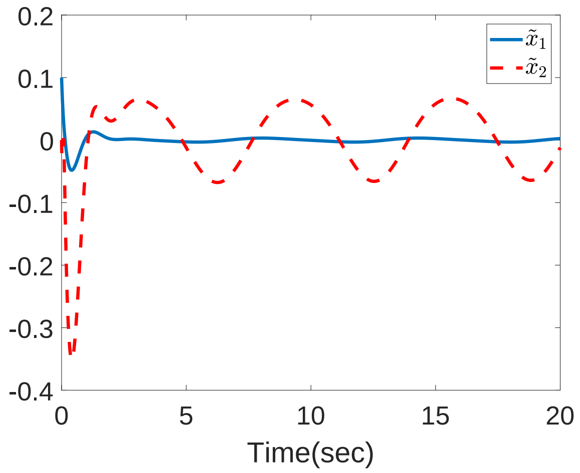

Fig. 3: The trajectories of $\tilde{x}_1$ and $\tilde{x}_2$.

V. SIMULATION EXAMPLE

Subsequently, a numerical simulation example will be used to confirm the effectiveness of the suggested approach. The corresponding system parameters are selected as $m = 10\,kg$, $M = 0.5\,kg/m^2$, $l = 1\,m$, and the reference signal is selected as $y_r = 0.3\sin t$, and time-varying fault $\zeta(t) = 0.5\sin(0.6t) + 0.4\cos(0.4t) + 0.2$.

In addition, the control parameters are designed as $\rho_1 = 12$, $\rho_2 = 15$, $l_1 = 16$, $l_2 = 18$, $\kappa_f = 0.5$, $\kappa_{c1} = 10$, $\kappa_{c2} = 10$, $\kappa_{a1} = 12$, $\kappa_{a2} = 12$ and $\eta = 4$.

The initial values of the system are set as $x_1(0) = x_2(0) = 0.2$, $\hat{x}_1(0) = 0.1$, $\hat{x}_2(0) = 0.2$, $\hat{\Xi}(0) = 0$, $\hat{W}_f(0) = [0.2, \ldots, 0.2]^T \in \mathbb{R}^{6\times1}$, $\hat{W}_{c1}(0) = \hat{W}_{a1}(0) = [0.5, \ldots, 0.5]^T \in \mathbb{R}^{6\times1}$ and $\hat{W}_{c2}(0) = \hat{W}_{a2}(0) = [0.4, \ldots, 0.4]^T \in \mathbb{R}^{6\times1}$.

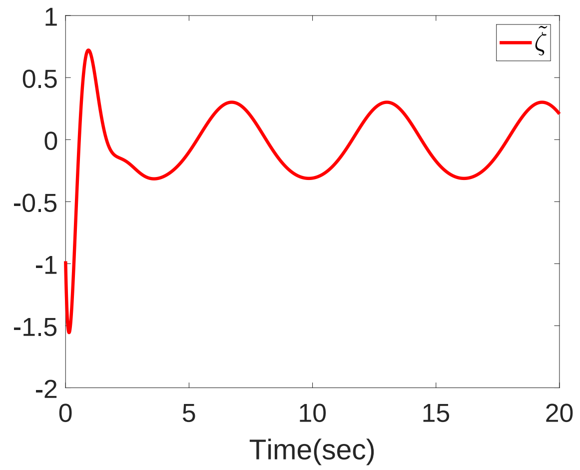

Fig. 4: The trajectory of the fault estimation error $\tilde{\zeta}$.

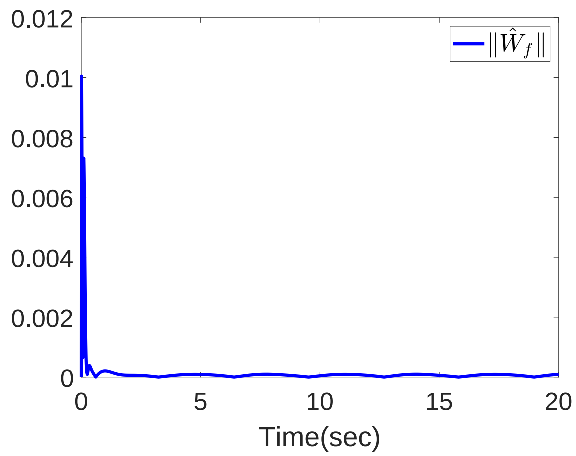

Fig. 5: The norm of the $\hat{\Theta}_f$.

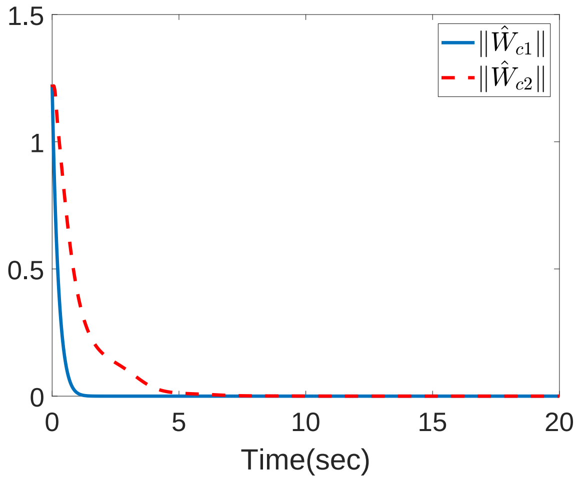

Fig. 6: The norms of the $\hat{W}_{c1}$ and $\hat{W}_{c2}$.

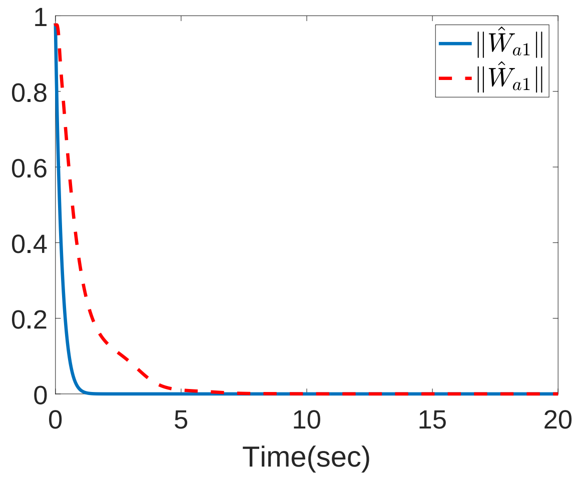

Fig. 7: The norms of the $\hat{W}_{a1}$ and $\hat{W}_{a2}$.

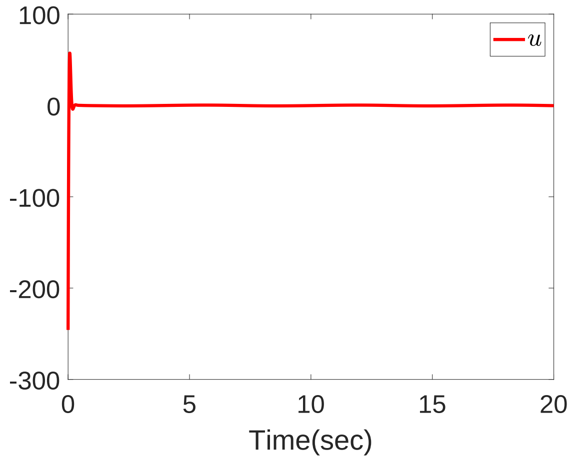

Fig. 8: Control input u.

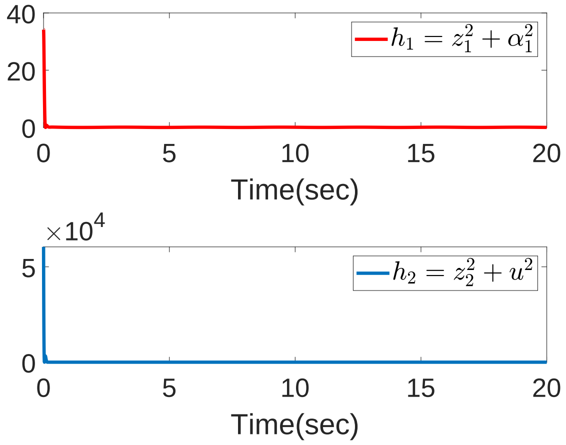

Fig. 9: Cost functions h_1 and h_2.

The simulation result graph is shown in Figs. 2-9. This indicates that the proposed control scheme has excellent tracking performance and is able to estimate and reconstruct unknown faults while ensuring that the optimal performance index is minimized.

Fig. 4 illustrates that the error between the actual and estimated fault values can converge to a region near zero. Figs. 5-7 demonstrate that the adaptive law derived from the reinforcement learning algorithm achieves asymptotic convergence. Figs. 8 and 9 present the response curves of both the optimal controller and the minimum performance function.

VI. Conclusion

For the single-link robot arm model, an adaptive output feedback optimal control strategy based on neural networks for fault estimation is proposed. This scheme can effectively estimate and suppress unknown system faults when they occur. At the same time, by combining optimal inverse dynamics techniques and reinforcement learning algorithms, the optimal virtual iscontrol signal and controller for the system were found, greatly saving control resources. As a result, this approach plays a crucial role in enhancing the performance of the single-link robotic arm system.

References

[1] M. E. Moran, "Evolution of robotic arms," *Journal of Robotic Surgery*, vol. 1, no. 2, pp. 103–111, 2007.

[2] J. de Jesus Rubio, E. Garcia, C. A. Ibanez, and C. Torres, "Stabilization of the robotic arms," *IEEE Latin America Transactions*, vol. 13, no. 8, pp. 2567–2573, 2015.

[3] R. Gautam, A. Gedam, A. Zade, and A. Mahawadiwar, "Review on development of industrial robotic arm," *International Research Journal of Engineering and Technology*, vol. 4, no. 03, p. 429, 2017.

[4] W. Wallach and C. Allen, "Framing robot arms control," *Ethics and Information Technology*, vol. 15, pp. 125–135, 2013.

[5] 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.

[6] X. Li, M. Liu, W. Wang, Y. Li, and M. Li, "Fault-tolerant control method of robotic arm based on machine vision," in *2018 Chinese Control And Decision Conference*, pp. 484–489, IEEE, 2018.

[7] F. Jia, F. Cao, G. Lyu, and X. He, "A novel framework of cooperative design: Bringing active fault diagnosis into fault-tolerant control," *IEEE Transactions on Cybernetics*, vol. 53, no. 5, pp. 3301–3310, 2022.

[8] X. Guo, C. Wang, and L. Liu, "Adaptive fault-tolerant control for a class of nonlinear multi-agent systems with multiple unknown time-varying control directions," *Automatica*, vol. 167, p. 111802, 2024.

[9] L. Wu, J. Park, and N. Zhao, "Robust adaptive fault-tolerant tracking control for nonaffine stochastic nonlinear systems with full-state constraints," *IEEE Transactions on Cybernetics*, vol. 50, no. 8, pp. 3793–3805, 2019.

[10] 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.

[11] A. A. Amin, M. S. Iqbal, and M. H. Shahbaz, "Development of intelligent fault-tolerant control systems with machine leaprning, deep learning, and transfer learning algorithms: A review," *Expert Systems with Applications*, p. 121956, 2023.

[12] G. Wang and C. Yi, "Fault estimation for nonlinear systems by an intermediate estimator with stochastic failure," *Nonlinear Dynamics*, vol. 89, no. 2, pp. 1195–1204, 2017.

[13] X. Wang, Z. Fei, T. Wang, and L. Yang, "Dynamic event-triggered actuator fault estimation and accommodation for dynamical systems," *Information Sciences*, vol. 525, pp. 119–133, 2020.

[14] T. Zhang, F. Deng, Y. Sun, and P. Shi, "Fault estimation and fault-tolerant control for linear discrete time-varying stochastic systems," *Science China Information Sciences*, vol. 64, pp. 1–16, 2021.

[15] R. S. Sutton, A. G. Barto, and R. J. Williams, "Reinforcement learning is direct adaptive optimal control," *IEEE Control Systems Magazine*, vol. 12, no. 2, pp. 19–22, 1992.

[16] 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.

[17] 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.

[18] 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.

[19] I. Michailidis, S. Baldi, E. B. Kosmatopoulos, and P. A. Ioannou, "Adaptive optimal control for large-scale nonlinear systems," *IEEE Transactions on Automatic Control*, vol. 62, no. 11, pp. 5567–5577, 2017.

[20] S. Tong, K. Sun, and S. Sui, "Observer-based adaptive fuzzy decentralized optimal control design for strict-feedback nonlinear large-scale systems," *IEEE Transactions on Fuzzy Systems*, vol. 26, no. 2, pp. 569–584, 2017.

[21] Y. Li, Y. Liu, and S. Tong, "Observer-based neuro-adaptive optimized control of strict-feedback nonlinear systems with state constraints," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 33, no. 7, pp. 3131–3145, 2021.

[22] 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.

[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] 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.