Preprints.org

Article

Not peer-reviewed version

Adaptive Actor-Critic Optimal Tracking Control for a Class of High-Order Nonlinear Systems with Partially Unknown Dynamics

Dengguo Xu , Xinsuo Li , Fapeng Li , Jingbei Tian *

Posted Date: 5 February 2026

doi: 10.20944/preprints202602.0353.v1

Keywords: optimal tracking control; reinforcement learning; actor-critic architecture; partially unknown nonlinear system

 doi:10.20944/preprints202602.0353.v1

Article

Adaptive Actor-Critic Optimal Tracking Control for a Class of High-Order Nonlinear Systems with Partially Unknown Dynamics

Dengguo Xu [1,2], Xinsuo Li [1,2], Fapeng Li [1] and Jingbei Tian [1,*]

1 School of Automation, Guangxi University of Science and Technology, Liuzhou, 545616, China
2 Guangxi Key Laboratory of Logistics Unmanned Aircraft Technology for Transportation Industry, Liuzhou, 545616, China
* Correspondence: 2005090@gxust.edu.cn

Abstract

This study develops an adaptive optimal tracking control law using neural network (NN)-based reinforcement learning (RL) for high-order partially unknown nonlinear systems. By designing a cost function associated with the sliding mode surface (SMS), the original tracking control problem is equivalently transformed into solving the optimal control problem related to the tracking Hamilton-Jacobi-Bellman (HJB) equation. Since the analytical solution of the HJB equation is generally intractable, we employ a policy iteration algorithm derived from the HJB equation, where both the partial derivative of the optimal tracking cost function and the optimal control law are approximated by NNs. The proposed RL framework achieves simplification through actor-critic training laws derived under the condition that a simple function is zero. Finally, two simulative examples are provided to demonstrate the effectiveness and advantages of the proposed adaptive optimal tracking control method.

Keywords: optimal tracking control; reinforcement learning; actor-critic architecture; partially unknown nonlinear system

---

1. Introduction

In recent decades, optimal tracking control for nonlinear systems has remained one of the most significant research topics in control theory, with extensive applications in industrial manufacturing [1], aerospace [2], and robotics [3,4]. However, the majority of existing controller designs fail to incorporate formal optimization frameworks for performance cost functions, despite energy/time efficiency being critical in practical engineering. This study addresses this gap by developing an optimal tracking control law for partially unknown continuous-time nonlinear systems. Using an actor-critic reinforcement learning strategy derived from the HJB equation, the proposed method enables system states to achieve asymptotic tracking of desired reference trajectory with guaranteed convergence.

Nonlinear tracking control problems can be effectively addressed through nonlinear control techniques including feedback linearization[5,6], backstepping approach [7–10], and sliding mode control(SMC) [3,11–13], among others. Feedback linearization, a mature methodology in nonlinear control theory, transforms nonlinear systems into partially or fully linear equivalents through coordinated nonlinear state feedback and coordinate transformations (including dynamic compensation). This enables the application of established linear control design methods. Unlike feedback linearization, the backstepping approach provides an alternative framework for nonlinear systems unbounded by linear constraints. This methodology synthesizes feedback control laws through an iterative recursive procedure. Its principal advantage lies in preserving beneficial nonlinearities while guaranteeing asymptotic stability for regulation and tracking tasks. However, these methods are mainly applicable to strictly feedback nonlinear systems. To enhance transient response performance, SMC has garnered significant research attention. As a robust control strategy, SMC's distinctive feature lies in its adaptive

 doi:10.20944/preprints202602.0353.v1

control structure switching based on state-dependent sliding surfaces. Its multiple variants have been developed, including terminal SMC [11], integral SMC [3,12], and hierarchical SMC [13]. Recent years have witnessed increasingly in-depth research on adaptive optimal control integrated with sliding mode surface (SMS). Zhang et al. addressed the problem of SMS-based adaptive optimal control for a class of switched continuous-time nonlinear systems with average dwell time by employing an actor-critic RL strategy [14]. Zhao et al. investigated the SMS-based approximate optimal control problem in the context of nonlinear multi-player Stackelberg-Nash games [15]. Furthermore, Zhang et al. studied the adaptive optimal control scheme based on a hierarchical SMS for a class of switched continuous-time nonlinear systems subject to unknown disturbances, utilizing an actor-critic NNs architecture [16]. However, all of these are aimed at the problem of optimal regulation.

Designing optimal control laws for nonlinear systems typically requires solving HJB equation. However, the inherent nonlinearity of HJB equation makes analytical solutions challenging to derive via conventional methods. To address this, RL algorithms have been integrated into the optimal control framework, enabling feasible solutions. As a machine learning paradigm, RL allows agents to learn optimal policies through environmental interactions [17,20]. This RL-based function approximation approach has successfully enabled adaptive optimal control, emerging as a prominent methodology for complex nonlinear control in recent decades.

Current research extensively explores optimal tracking for nonlinear systems using adaptive dynamic programming (ADP) [18,19]. This method combines dynamic programming (DP) and RL. Notably, Modares et al. proposed a novel integral RL formulation for continuous-time nonlinear systems [20]. This approach adapts methodologies originally developed for optimal regulation problems to address optimal tracking control. Recent advancements in RL-based tracking control have demonstrated significant progress across diverse nonlinear systems. Wen et al. developed an optimized tracking control framework employing neural network-based actor-critic reinforcement learning for a class of nonlinear dynamic systems [21]. Wang et al. subsequently proposed a novel actor-critic RL scheme to achieve optimal tracking control for unmanned surface vehicles, explicitly addressing complex unknowns including dead-zone input nonlinearities, uncertain system dynamics, and external disturbances [22]. Further innovations include fuzzy integral RL-based fault-tolerant control algorithm, which integrates RL techniques with fuzzy-augmented models to handle partially unknown systems with actuator faults [23]. Specialized applications have been demonstrated, including robust online tracking control for space manipulators and nearly optimal trajectory tracking of autonomous surface vehicles [24,25]. While the backstepping technique has been widely applied in optimal tracking control of strict-feedback nonlinear systems [26–29], the same problem for high-order canonical nonlinear systems remains largely unexplored.

Motivated by the above discussions, this paper addresses a class of canonical form high-order nonlinear system by proposing a novel adaptive optimal tracking control scheme. The main contributions are in three aspects. First, an adaptive optimal tracking control scheme is proposed via a SMS for high-order canonical nonlinear systems. By constructing a cost function specifically related to the SMS, the original control problem is equivalently transformed into one that seeks an optimal control strategy. The designed SMS constrains the states of the error dynamic system, thereby forcing the tracking error to converge to zero with predefined dynamics. Second, in most of the existing literature, such as [19–22], the requirement of persistent excitation is necessary for training the RL optimal control with adaptive parameters. The proposed optimization method can avoid this requirement. Third, an adaptive optimal tracking law is developed using an actor-critic NNs RL framework. Compared with the [30] method, our method achieves higher tracking accuracy and computational efficiency with fewer conditional constraint.

The rest of this paper is organized as follows: Section 2 describes the optimal tracking control problem of a class of canonical form high-order nonlinear system. In Section 3, we present a SMS-based adaptive optimal tracking control using actor-critic RL. Based on Lyapunov theory, it is proved that the error signals are semi-globally uniformly ultimately bounded (SGUUB), and the tracking error can

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 5 February 2026 doi:10.20944/preprints202602.0353.v1

be steered into a small neighborhood of zero in Section 4. Simulation are given to demonstrate the effectiveness of the proposed method in Section 5. Finally, the conclusion summarized in Section 6.

2. Problem Description

A canonical form nonlinear system with relative degree n is given by [31]

$$\begin{aligned} \dot{x}_i(t) &= x_{i+1}(t), \quad i = 1, \cdots, n-1 \\ \dot{x}_n(t) &= f(\bar{x}) + gu, \end{aligned} \tag{1}$$

where $\bar{x}(t) = [x_1(t), x_2(t), \cdots, x_n(t)]^T \in \mathbb{R}^n$ is the system state vector, $u \in \mathbb{R}$ is the control input, $f(\bar{x}) \in \mathbb{R}$ is the unknown and bounded nonlinear dynamic function, and $g \in \mathbb{R}$ is the known control gain constant.

Assumption 1. *The system (1) is stabilizable, meaning there exists an admissible control policy $u \in \Psi(\Omega)$ that ensures global stability of the closed-loop system.*

Assumption 2. *The reference trajectory $y_d(t)$ and its i-th order derivatives $y_d^{(i)}(t)$ $(i = 1, ..., n)$ are assumed to be bounded.*

Definition 1 ([32])**.** *For a nonlinear system $\dot{\zeta} = f(\zeta, t)$, the solution $\zeta(t) \in \mathbb{R}^n$ is called SGUUB if for any initial state $\zeta(t_0) \in \Omega$, where Ω is a compact set, there exist constants $\sigma > 0$ and $T(\sigma, \zeta(t_0)) > 0$ such that $\|\zeta(t)\| \leqslant \sigma$ for all $t \geq t_0 + T(\sigma, \zeta(t_0))$.*

The control objective of this paper is to design an optimized tracking controller for partially unknown nonlinear dynamic system (1) using a NN-based actor-critic RL strategy. The proposed control law not only ensures all closed-loop error signals are SGUUB, but also guarantees that the output state $x_1(t)$ and its derivative signals rapidly track the reference trajectory $y_d(t)$ and its corresponding derivatives, respectively.

3. Optimal Tracking Control Design

3.1. Tracking HJB Equation

The tracking errors are defined as $z_1(t) = x_1(t) - y_d(t), z_2(t) = x_2(t) - \dot{y}_d(t), \cdots, z_n(t) = x_n(t) - y_d^{(n-1)}(t)$. Based on the dynamic model (1) and the above tracking error definitions, the error dynamical system is derived as follows:

$$\begin{aligned} \dot{z}_i(t) &= z_{i+1}(t), \quad i = 1, \cdots, n-1 \\ \dot{z}_n(t) &= f(\bar{z}, \bar{y}_d) + gu - y_d^{(n)}(t) \end{aligned} \tag{2}$$

where $\bar{z} = [z_1(t), z_2(t), \cdots, z_n(t)]^T \in \mathbb{R}^n$, and $\bar{y}_d = [y_d(t), \dot{y}_d(t), \cdots, y_d^{(n-1)}(t)]^T \in \mathbb{R}^n$.

Based on the tracking errors, the SMS [33] is constructed as

$$s(t) = K_0^T \bar{z}(t) = k_1 z_1(t) + \cdots + k_{n-1} z_{n-1}(t) + z_n(t), \tag{3}$$

where $s(t) \in \mathbb{R}$, $K_0 = [k_1, \cdots, k_{n-1}, 1]^T \in \mathbb{R}^n$, and $k_j > 0$ $(j = 1, \cdots, n-1)$ are design parameters.

Taking the time derivative of $s(t)$ along the error dynamical system (2) yields

$$\begin{aligned} \dot{s}(t) = K_1^T \bar{z}(t) + f(\bar{z}, \bar{y}_d) + gu - y_d^{(n)}(t) = k_1 z_2(t) + \cdots + k_{n-1} z_n(t) \\ + f(\bar{z}, \bar{y}_d) + gu - y_d^{(n)}(t), \end{aligned} \tag{4}$$

where $K_1 = [0, k_1, \cdots, k_{n-1}]^T \in \mathbb{R}^n$.

The design parameters k_j are chosen such that the polynomial $\rho^{n-1} + k_{n-1}\rho^{n-2} + \cdots + k_2\rho + k_1$ is Hurwitz, i.e., all its roots lie in the open left-half plane. Note that when the tracking errors $z_1(t), \cdots, z_n(t)$ are constrained to the SMS $s(t) = 0$, they will converge to a small neighborhood of zero.

Define an infinite-horizon cost function with respect to the SMS s as follow

$$V(s) = \int_t^\infty r(s(\tau), u(s))d\tau, \tag{5}$$

where $r(s,u) = s^2 + u^2$ denotes the utility function.

Definition 2. *For the error dynamical system (2), the control policy $u(s)$ is defined as admissible with respect to $V(s)$ on the compact set Ω, denoted by $u(s) \in \Psi(\Omega)$, if $u(s)$ is continuous, $u(0) = 0$, $u(s)$ can stabilize (2) on Ω and $V(s)$ is finite for $\forall s \in \Omega$.*

The optimal control objective for the error dynamical system (2) is to find an admissible control law that minimizes the cost function (5) over an infinite horizon, thus achieving the desired control task. Accordingly, the optimal cost function is defined as

$$V^*(s) = \min_{u\in\Psi(\Omega)} \left(\int_t^\infty r(s(\tau), u(s))d\tau \right) = \int_t^\infty r(s(\tau), u^*(s))d\tau, \tag{6}$$

where $V^*(s)$ and $u^*(s)$ denote the optimal cost function and optimal control law, respectively, and $\Psi(\Omega)$ represents the set of admissible controls over the domain Ω.

By calculating the time derivative of the cost function (6) along the sliding mode dynamic (4) and applying the optimality condition, the following HJB equation is derived:

$$\begin{aligned} H(s,u^*,\nabla V^*(s)) &= r(s,u^*) + \nabla V^*(s)\dot{s} \\ &= s^2 + u^{*2} + \nabla V^*(s)(K_1^T\bar{z}(t) + f(\bar{z},\bar{y}_d) + gu^* - y_d^{(n)}(t)) \\ &= 0 \end{aligned} \tag{7}$$

where $\nabla V^*(s) = \partial V^*(s)/\partial s$ denotes the partial derivative of $V^*(s)$ with respect to s.

Assuming the solution of (7) exists and is unique, the optimal control u^* can be obtained by applying the stationarity condition $\frac{\partial H(s,u^*,\nabla V^*(s))}{\partial u^*} = 0$ as

$$u^* = -\frac{g}{2}\nabla V^*(s). \tag{8}$$

Substituting (8) into (7) yields

$$s^2 - \frac{g^2}{4}(\nabla V^*(s))^2 + \nabla V^*(s)\left(K_1^T\bar{z}(t) + f(\bar{z},\bar{y}_d) - {y_d}^{(n)}(t)\right) = 0. \tag{9}$$

The partial derivative $\nabla V^*(s)$ can be obtained by solving the HJB equation (9). The optimal control protocol can then be derived by combining the result with (8). However, solving (9) analytically is difficult due to the inherent system nonlinearity. Furthermore, the lack of complete knowledge about the system dynamics $f(\bar{z},\bar{y}_d)$ further complicates the solution of (9). To overcome these challenges, this paper employs a reinforcement learning method with an actor-critic architecture.

3.2. NNs Approximation in Actor-Critic RL

To achieve the optimal tracking control, the term $\nabla V^*(s)$ in (9) is decomposed as

$$\nabla V^*(s) = 2\gamma_u s(t) + \frac{2}{g^2}K_1^T\bar{z}(t) + V^0(s,\bar{z}), \tag{10}$$

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 5 February 2026 doi:10.20944/preprints202602.0353.v1

where $\gamma_u > 0$ is the designed parameter and $V^0(s, \bar{z}) = -2\gamma_u s(t) - \frac{2}{g^2} K_1^T \bar{z}(t) + \nabla V^*(s)$ is the unknown function in this dual architecture.

Remark 1. *In (10), the terms* $2\gamma_u s(t)$ *and* $\frac{2}{g^2} K_1^T \bar{z}(t)$ *are designed to ensure the boundedness and stability of the system signals.*

According to the Weierstrass approximation theorem [34], the continuous function $V^0(s, \bar{z})$ can be approximated by weighted sum of basis functions. Given the relationship $s(t) = K^T \bar{z}(t)$ between s and $\bar{z}$, the approximation takes the form as

$$V^0(s, \bar{z}) = W^{*T} \Phi(\bar{z}) + \varepsilon(\bar{z}), \tag{11}$$

where $W^* = [w_1, w_2, \cdots, w_N] \in \mathbb{R}^N$ is the ideal weight vector and bounded by a positive constant $\bar{W}$ such that $\|W^*\| \leq \bar{W}$. Moreover, $\Phi = [\phi_1, \phi_2, \cdots, \phi_N] \in \mathbb{R}^N$ and $\varepsilon(\bar{z}) \in \mathbb{R}$ are, respectively, activation function vector and the approximation error. With the number of neurons in the hidden layer $N \to \infty$, the NN approximation error $\varepsilon(\bar{z}) \to 0$.

Putting (11) into (10) yields

$$\nabla V^*(s) = 2\gamma_u s(t) + \frac{2}{g^2} K_1^T \bar{z}(t) + W^{*T} \Phi(\bar{z}) + \varepsilon(\bar{z}), \tag{12}$$

since the ideal weight vector W^* is unknown, (12) is approximated by critic NN as

$$\nabla \hat{V}^*(s) = 2\gamma_u s(t) + \frac{2}{g^2} K_1^T \bar{z}(t) + \hat{W}_c^T \Phi(\bar{z}), \tag{13}$$

where $\hat{W}_c$ denotes the current estimate of W^*. The weight estimation error is then defined as $\tilde{W}_c = \hat{W}_c - W^*$.

According to (8) and (13), the optimal control to the system is approximated by an additional actor NN as

$$\hat{u}^* = -g\gamma_u s(t) - \frac{1}{g} K_1^T \bar{z}(t) - \frac{g}{2} \hat{W}_a^T \Phi(\bar{z}), \tag{14}$$

where $\hat{W}_a$ is the current estimate of W^*. The actor NN weight estimation error is defined as $\tilde{W}_a = \hat{W}_a - W^*$.

The critic and actor adaptive training laws are designed as

$$\dot{\hat{W}}_c(t) = -\gamma_c \Big(\Phi(\bar{z}) \Phi^T(\bar{z}) + \sigma I_m \Big) \hat{W}_c(t), \tag{15}$$

$$\dot{\hat{W}}_a(t) = -\Big(\Phi(\bar{z}) \Phi^T(\bar{z}) + \sigma I_m \Big) \big[\gamma_a \big(\hat{W}_a(t) - \hat{W}_c(t) \big) - \gamma_c \hat{W}_c(t) \big], \tag{16}$$

where $\gamma_c > 0$ and $\gamma_a > 0$ are the learning rates for the critic and actor networks, respectively. Here, $\sigma > 0$ is a regularization parameter, and I_m denotes the $m \times m$ identity matrix.

Inserting both (13) and (14) into (7), the approximated HJB equation yields

$$\begin{aligned} \hat{H}(s, \hat{u}^*, \nabla \hat{V}^*(s)) = s^2 &+ g^2 \Big(-\gamma_u s(t) - \frac{1}{g^2} K_1^T \bar{z}(t) - \frac{1}{2} \hat{W}_a^T \Phi(\bar{z}) \Big)^2 \\ &+ \Big(2\gamma_u s(t) + \frac{2}{g^2} K_1^T \bar{z}(t) + \hat{W}_c^T(t) \Phi(\bar{z}) \Big) \\ &\times \Big(-\gamma_u g^2 s(t) - \frac{g^2}{2} \hat{W}_a^T \Phi(\bar{z}) + f(\bar{z}, \bar{y}_d) - y_d^{(n)}(t) \Big). \end{aligned} \tag{17}$$

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 5 February 2026 doi:10.20944/preprints202602.0353.v1

The Bellman residual error E is introduced, with its expression as

$$\begin{aligned} E &= \hat{H}(s, \hat{u}^*, \nabla \hat{V}^*(s)) - H(s, u^*, \nabla V^*(s)) \\ &= \hat{H}(s, \hat{u}^*, \nabla \hat{V}^*(s)) \end{aligned} \tag{18}$$

According to the previous analysis and the optimal control theory, the approximate optimized control $\hat{u}^*$ should ensure $E \to 0$. If the condition $E = 0$ is met and has a unique solution, it is equivalent to

$$\frac{\partial \hat{H}(s, \hat{u}^*, \nabla \hat{V}^*(s))}{\partial \hat{W}_a} = \frac{g^2}{2} \Phi(\bar{z}) \Phi^T(\bar{z}) \big(\hat{W}_a(t) - \hat{W}_c(t) \big) = 0. \tag{19}$$

To make (19) hold, construct the following function $\Gamma(t)$ such that

$$\Gamma(t) = \big(\hat{W}_a(t) - \hat{W}_c(t) \big)^T \big(\hat{W}_a(t) - \hat{W}_c(t) \big). \tag{20}$$

Obviously, $\Gamma(t) = 0$ ensures that (19) holds. When the weight vectors of both networks are trained to satisfy $\hat{W}_a(t) = \hat{W}_c(t)$, equations (13) and (14) fulfill the relation (8), and the control action approaches the optimal solution.

Theorem 1. *Under the adaptive laws given by (15) and (16), the function $\Gamma(t)$ converges to zero asymptotically.*

Proof. The partial derivatives of $\Gamma(t)$ with respect to the weight estimates are

$$\frac{\partial \Gamma(t)}{\partial \hat{W}_a(t)} = -\frac{\partial \Gamma(t)}{\partial \hat{W}_c(t)} = 2\big(\hat{W}_a(t) - \hat{W}_c(t) \big). \tag{21}$$

The time derivative of $\Gamma(t)$ along the trajectories of the system is

$$\dot{\Gamma}(t) = \frac{\partial \Gamma(t)}{\partial \hat{W}_a(t)} \dot{\hat{W}}_a(t) + \frac{\partial \Gamma(t)}{\partial \hat{W}_c(t)} \dot{\hat{W}}_c(t) = \frac{\partial \Gamma(t)}{\partial \hat{W}_a(t)} \Big(\dot{\hat{W}}_a(t) - \dot{\hat{W}}_c(t) \Big). \tag{22}$$

Substituting the adaptive laws (15) and (16) into (22) yields

$$\begin{aligned} \dot{\Gamma}(t) &= 2\big(\hat{W}_a - \hat{W}_c \big)^T \Big[-\gamma_a \Big(\Phi(\bar{z}) \Phi^T(\bar{z}) + \sigma I_m \Big) \big(\hat{W}_a - \hat{W}_c \big) \Big] \\ &= -2\gamma_a \big(\hat{W}_a - \hat{W}_c \big)^T \Big(\Phi(\bar{z}) \Phi^T(\bar{z}) + \sigma I_m \Big) \big(\hat{W}_a - \hat{W}_c \big). \end{aligned} \tag{23}$$

Noting that $\Gamma(t) = \|\hat{W}_a - \hat{W}_c\|^2$, the expression in (23) can be rewritten as a function of $\Gamma(t)$ itself. Since the matrix $\big(\Phi(\bar{z}) \Phi^T(\bar{z}) + \sigma I_m \big)$ is positive definite, it follows that

$$\begin{aligned} \dot{\Gamma}(t) &= -2\gamma_a \big(\hat{W}_a - \hat{W}_c \big)^T \Big(\Phi(\bar{z}) \Phi^T(\bar{z}) + \sigma I_m \Big) \big(\hat{W}_a - \hat{W}_c \big) \\ &\leq -2\gamma_a \lambda_{\min} \|\hat{W}_a - \hat{W}_c\|^2 = -2\gamma_a \lambda_{\min} \Gamma(t), \end{aligned} \tag{24}$$

where $\lambda_{\min} > 0$ is the minimum eigenvalue of the positive definite matrix $\big(\Phi(\bar{z}) \Phi^T(\bar{z}) + \sigma I_m \big)$.

This inequality, $\dot{\Gamma}(t) \leq -2\gamma_a \lambda_{\min} \Gamma(t)$, implies that $\Gamma(t)$ converges to zero exponentially. □

Remark 2. *The critic NN in (14) may be used to determine the actor without using another NN for the actor. However, to obtain the form of (19) and to facilitate the derivation of weight update laws and subsequent stability analysis, separate NNs are necessary for the actor and critic.*

Remark 3. *As demonstrated in this subsection, the proposed RL approach not only features a simple structure but also eliminates the need for explicit knowledge of the system dynamic $f(\bar{z}, \bar{y}_d)$ when solving the optimal control problem.*

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 5 February 2026 doi:10.20944/preprints202602.0353.v1

4. Main Results

Lemma 1 ([35]). *$X(t) \in \mathbb{R}$ is a positive continuous function and has the bounded initial value $X(0)$. If the inequality $\dot{X}(t) \leqslant -\alpha X(t) + \beta$ holds, where $\alpha > 0$ and $\beta > 0$ are constants, then*

$$X(t) \leqslant e^{-\alpha t} X(0) + \frac{\beta}{\alpha}(1 - e^{-\alpha t}). \tag{25}$$

Theorem 2. *Consider the partially unknown nonlinear system (1) under Assumptions 1 and 2, with a bounded initial state. The optimal control law is derived from the actor-critic RL framework, where the critic and actor are updated according to (13) and (14), using the learning rules given in (15) and (16), respectively. If the parameters satisfy $\gamma_u > \frac{1}{2} + \frac{1}{g^2}$, $\gamma_c > \frac{\gamma_a}{2}$ and $\gamma_a > 1$, then all estimation errors are SGUUB. Moreover, the tracking errors converge to an arbitrarily small neighborhood of zero.*

Proof. Choose the Lyapunov candidate function as

$$L(t) = \frac{1}{2}s^2 + \frac{1}{2}\tilde{W}_c^T(t)\tilde{W}_c(t) + \frac{1}{2}\tilde{W}_a^T(t)\tilde{W}_a(t). \tag{26}$$

Calculating the time derivative (26) along (4), (15) and (16) yields

$$\begin{aligned}\dot{L}(t) = {} & s\left(-g^2\gamma_u s - \frac{g^2}{2}\hat{W}_a^T\Phi(\bar{z}) + f(\bar{z}, \bar{y}_d) - y_d^{(n)}(t)\right) - \gamma_c \tilde{W}_c^T(t)\Big(\Phi(\bar{z})\Phi^T(\bar{z}) \\ & + \sigma I_m\Big)\hat{W}_c(t) - \gamma_a \tilde{W}_a^T(t)\left(\Phi(\bar{z})\Phi^T(\bar{z}) + \sigma I_m\right)\left(\hat{W}_a(t) - \hat{W}_c(t)\right) \\ & + \gamma_c \tilde{W}_a^T(t)\left(\Phi(\bar{z})\Phi^T(\bar{z}) + \sigma I_m\right)\hat{W}_c(t).\end{aligned} \tag{27}$$

It follows from $\tilde{W}_c = \hat{W}_c - W^*$ and $\tilde{W}_a = \hat{W}_a - W^*$ that

$$\hat{W}_a^T\Phi(\bar{z}) = \tilde{W}_a^T\Phi(\bar{z}) + W^{*T}\Phi(\bar{z}) \tag{28}$$

and

$$\hat{W}_a(t) - \hat{W}_c(t) = \hat{W}_a(t) - W^* + W^* - \hat{W}_c(t) = \tilde{W}_a - \tilde{W}_c. \tag{29}$$

Using (28) and (29), (27) can be rewritten as

$$\begin{aligned}\dot{L}(t) = {} & -g^2\gamma_u s^2 - \frac{g^2}{2}s\tilde{W}_a^T\Phi(\bar{z}) - \frac{g^2}{2}sW^{*T}\Phi(\bar{z}) + sf(\bar{z}, \bar{y}_d) - sy_d^{(n)}(t) \\ & - \gamma_c \tilde{W}_c^T(t)\left(\Phi(\bar{z})\Phi^T(\bar{z}) + \sigma I_m\right)\hat{W}_c(t) - \gamma_a \tilde{W}_a^T(t)\left(\Phi(\bar{z})\Phi^T(\bar{z}) + \sigma I_m\right) \\ & \times \left(\tilde{W}_a(t) - \tilde{W}_c(t)\right) + \gamma_c \tilde{W}_a^T(t)\left(\Phi(\bar{z})\Phi^T(\bar{z}) + \sigma I_m\right)\hat{W}_c(t).\end{aligned} \tag{30}$$

From Cauchy-Schwartz and Young's inequalities, we obtain

$$\begin{aligned} & sf(\bar{z}, \bar{y}_d) \leq \frac{1}{2}s^2 + \frac{1}{2}f^2(\bar{z}, \bar{y}_d), \\ & -sy_d^{(n)}(t) \leq \frac{1}{2}s^2 + \frac{1}{2}\left(y_d^{(n)}(t)\right)^2, \\ & -s\tilde{W}_a^T\Phi(\bar{z}) \leq \frac{1}{2}s^2 + \frac{1}{2}\tilde{W}_a^T(t)\left(\Phi(\bar{z})\Phi^T(\bar{z}) + \sigma I_m\right)\tilde{W}_a(t), \\ & -sW^{*T}\Phi(\bar{z}) \leq \frac{1}{2}s^2 + \frac{1}{2}W^{*T}\left(\Phi(\bar{z})\Phi^T(\bar{z}) + \sigma I_m\right)W^*.\end{aligned} \tag{31}$$

 doi:10.20944/preprints202602.0353.v1

Under these inequalities, (30) reduces to

$$\begin{aligned}\dot{L}(t) \leq &-\left[g^2\left(\gamma_u - \frac{1}{2}\right) - 1\right]s^2 + \frac{1}{2}f^2(\bar{z}, \bar{y}_d) + \frac{1}{2}\left(y_d^{(n)}(t)\right)^2 - \gamma_c \tilde{W}_c^T(t)\left(\Phi(\bar{z})\Phi^T(\bar{z}) + \sigma I_m\right) \\ &\times \hat{W}_c(t) - \left(\gamma_a - \frac{g^2}{4}\right)\tilde{W}_a^T(t)\left(\Phi(\bar{z})\Phi^T(\bar{z}) + \sigma I_m\right)\tilde{W}_a(t) + \gamma_a \tilde{W}_a^T(t)\left(\Phi(\bar{z})\Phi^T(\bar{z})\right. \\ &\left. + \sigma I_m\right)\tilde{W}_c(t) + \gamma_c \tilde{W}_a^T(t)\left(\Phi(\bar{z})\Phi^T(\bar{z}) + \sigma I_m\right)\hat{W}_c(t) + \frac{g^2}{4}W^{*T}\left(\Phi(\bar{z})\Phi^T(\bar{z}) + \sigma I_m\right)W^*.\end{aligned} \tag{32}$$

From the definition $\tilde{W}_c = \hat{W}_c - W^*$, the following equation can be derived

$$\begin{aligned}&\tilde{W}_c^T(t)\left(\Phi(\bar{z})\Phi^T(\bar{z}) + \sigma I_m\right)\hat{W}_c(t) \\ &= \tilde{W}_c^T(t)\left(\Phi(\bar{z})\Phi^T(\bar{z}) + \sigma I_m\right)\tilde{W}_c(t) + \hat{W}_c^T(t)\left(\Phi(\bar{z})\right. \\ &\left.\times \Phi^T(\bar{z}) + \sigma I_m\right)\hat{W}_c(t) - W^{*T}(t)\left(\Phi(\bar{z})\Phi^T(\bar{z}) + \sigma I_m\right)W^*.\end{aligned} \tag{33}$$

Moreover, the following inequality can be derived according to Young's inequality,

$$\begin{aligned}\tilde{W}_a^T(t)\left(\Phi(\bar{z})\Phi^T(\bar{z}) + \sigma I_m\right)\tilde{W}_c \leq &\frac{1}{2}\tilde{W}_a^T(t)\left(\Phi(\bar{z})\Phi^T(\bar{z}) + \sigma I_m\right)\tilde{W}_a(t) \\ &+ \frac{1}{2}\tilde{W}_c^T(t)\left(\Phi(\bar{z})\Phi^T(\bar{z}) + \sigma I_m\right)\tilde{W}_c(t), \\ \tilde{W}_a^T(t)\left(\Phi(\bar{z})\Phi^T(\bar{z}) + \sigma I_m\right)\hat{W}_c \leq &\frac{1}{2}\tilde{W}_a^T(t)\left(\Phi(\bar{z})\Phi^T(\bar{z}) + \sigma I_m\right)\tilde{W}_a(t) \\ &+ \frac{1}{2}\hat{W}_c^T(t)\left(\Phi(\bar{z})\Phi^T(\bar{z}) + \sigma I_m\right)\hat{W}_c(t).\end{aligned} \tag{34}$$

According to the above equation (33) and inequality (34), the following inequality can be derived from (32)

$$\begin{aligned}\dot{L}(t) \leq &-\left[g^2\left(\gamma_u - \frac{1}{2}\right) - 1\right]s^2 - \left(\gamma_c - \frac{\gamma_a}{2}\right)\tilde{W}_c^T(t)\left(\Phi(\bar{z})\Phi^T(\bar{z}) + \sigma I_m\right)\tilde{W}_c(t) \\ &- \frac{1}{2}\left(\gamma_a - \gamma_c - \frac{g^2}{2}\right)\tilde{W}_a^T(t)\left(\Phi(\bar{z})\Phi^T(\bar{z}) + \sigma I_m\right)\tilde{W}_a(t) - \frac{\gamma_c}{2}\hat{W}_c^T(t)\left(\Phi(\bar{z})\Phi^T(\bar{z})\right. \\ &\left. + \sigma I_m\right)\hat{W}_c(t) + \left(\gamma_c + \frac{g^2}{4}\right)W^{*T}\left(\Phi(\bar{z})\Phi^T(\bar{z}) + \sigma I_m\right)W^* + \frac{1}{2}f^2(\bar{z}, \bar{y}_d) + \frac{1}{2}\left(y_d^{(n)}(t)\right)^2.\end{aligned} \tag{35}$$

This inequality (35) can be simplified to

$$\begin{aligned}\dot{L}(t) \leq &-\left[g^2\left(\gamma_u - \frac{1}{2}\right) - 1\right]s^2 - \left(\gamma_c - \frac{\gamma_a}{2}\right)\tilde{W}_c^T(t)\left(\Phi(\bar{z})\Phi^T(\bar{z}) + \sigma I_m\right)\tilde{W}_c(t) \\ &- \frac{1}{2}\left(\gamma_a - \gamma_c - \frac{g^2}{2}\right)\tilde{W}_a^T(t)\left(\Phi(\bar{z})\Phi^T(\bar{z}) + \sigma I_m\right)\tilde{W}_a(t) + C(t),\end{aligned} \tag{36}$$

where $C(t) = (\gamma_c + \frac{g^2}{4})W^{*T}(\Phi(\bar{z})\Phi^T(\bar{z}) + \sigma I_m)W^* + \frac{1}{2}f^2(\bar{z}, \bar{y}_d) + \frac{1}{2}\left(y_d^{(n)}(t)\right)^2$ can be bounded by a constant β, i.e., $|C(t)| \leq \beta$.

Let $\alpha = \min\left\{g^2\left(\gamma_u - \frac{1}{2}\right) - 1, (\gamma_c - \frac{\gamma_a}{2})\sigma, \frac{1}{2}(\gamma_a - \gamma_c - \frac{g^2}{2})\sigma\right\}$, then (36) can be rewritten as

$$\dot{L}(t) \leq -\alpha L(t) + \beta \tag{37}$$

Applying Lemma 1, we rigorously obtain the following inequality:

$$L(t) \leq e^{-\alpha t}L(0) + \frac{\beta}{\alpha}(1 - e^{-\alpha t}). \tag{38}$$

 doi:10.20944/preprints202602.0353.v1

The above inequality implies that the signals $\bar{z}(t)$, $\tilde{W}_c(t)$, and $\tilde{W}_a(t)$ are SGUUB. Moreover, by selecting the design parameters to make α sufficiently large, the tracking errors $z_1(t), \cdots, z_n(t)$ can be made to converge to an arbitrarily small neighborhood of zero. This completes the proof. □

5. Simulation Experiment

In this section, the effectiveness and advantages of the proposed adaptive optimal tracking control method are demonstrated through two simulation examples.

Example 1: Consider a second-order nonlinear dynamic model as presented in [30]:

$$\begin{aligned} \dot{x}_1(t) &= x_2(t), \\ \dot{x}_2(t) &= (1 - x_1(t))^2 - x_2(t) + gu, \end{aligned} \tag{39}$$

where $x_1(t)$ and $x_2(t)$ are the system states, u is the control input, and $g = 1$ is the control gain, with initial conditions $x_1(0) = 2$ and $x_2(0) = 3$.

Based on the reference trajectory $y_d = 5sin(t)$, the tracking errors can be obtained

$$\begin{aligned} z_1(t) &= x_1(t) - 5\sin(t), \\ z_2(t) &= x_2(t) - 5\cos(t). \end{aligned} \tag{40}$$

The error dynamics derived in (39) are

$$\begin{aligned} \dot{z}_1(t) &= z_2(t), \\ \dot{z}_2(t) &= (1 - z_1(t) - 5sin(t))^2 - z_2(t) - 5cos(t) + u + 5\sin(t). \end{aligned} \tag{41}$$

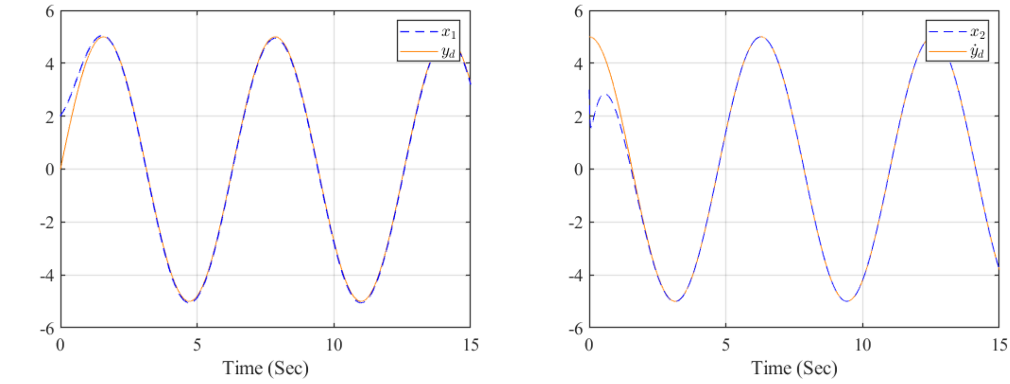
Following from (3), the SMS is designed as

$$s(t) = 2z_1(t) + z_2(t). \tag{42}$$

The critic and actor network weight vectors are defined as $\hat{W}_c = [\hat{W}_{c,1}, \hat{W}_{c,2}, \hat{W}_{c,3}]^T$ and $\hat{W}_a = [\hat{W}_{a,1}, \hat{W}_{a,2}, \hat{W}_{a,3}]^T$, respectively. The activation function vector is given by $\Phi(\bar{z}) = [z_1^2, z_1 z_2, z_2^2]^T$. The learning rates and regularization parameters for the training laws of both the critic and actor (15) and (16) are set to $\gamma_c = 3$, $\gamma_a = 5$, and $\sigma = 1$. The weight vectors are initialized to $\hat{W}_c(0) = [-1, 0, 1]^T$ and $\hat{W}_a(0) = [0, 1, 0]^T$.

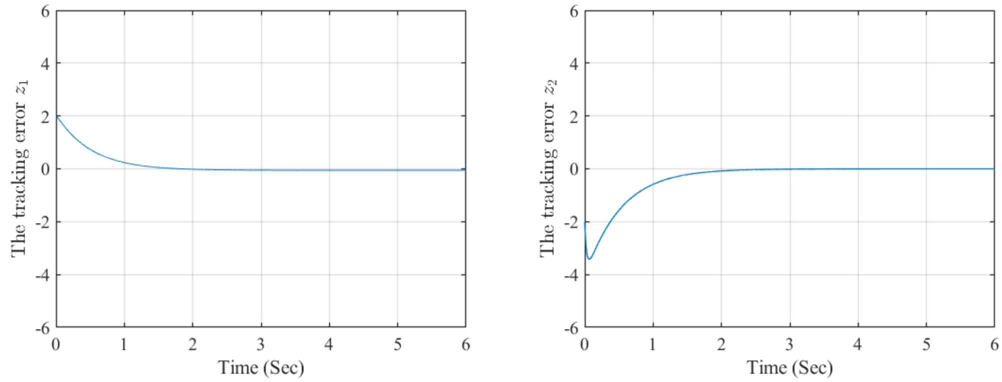
The simulation results are presented in Figures 1–fg4. Specifically, Figure 1 shows that the states $x_1(t)$ and $x_2(t)$ closely follow the reference trajectories y_d and its derivative $\dot{y}_d$, respectively. The corresponding tracking errors z_1 and z_2, shown in Figure 2, converge to zero. Furthermore, the neural network weights for the actor and critic remain bounded, as illustrated in Figure 3. Finally, the evolution of the utility function is depicted in Figure 4. Compared with the simulation results in [30], the method proposed in this paper achieves superior tracking performance.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 5 February 2026 doi:10.20944/preprints202602.0353.v1



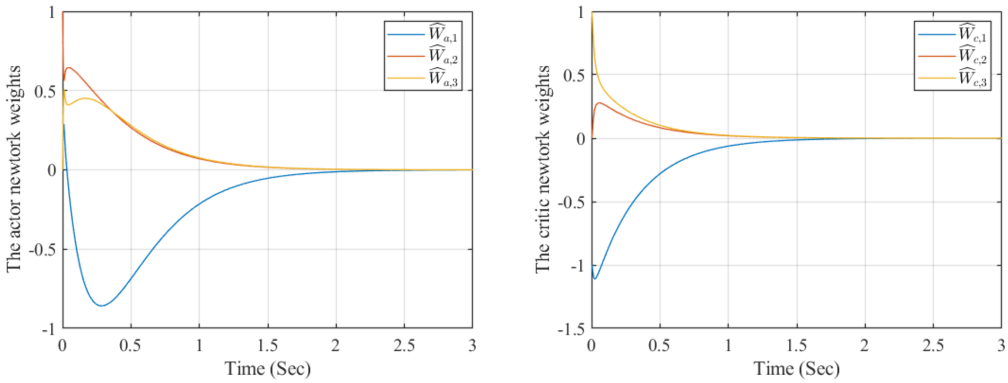
(**a**) (**b**)

Figure 1. The tracking performances.



(**a**) (**b**)

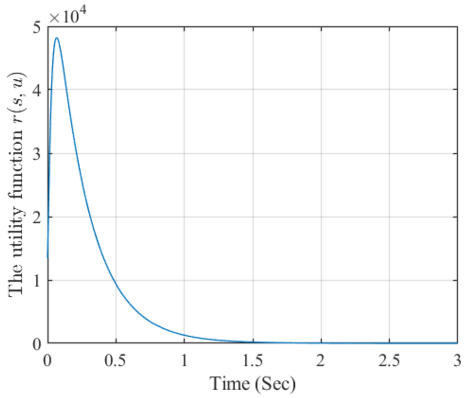
Figure 2. The tracking errors.



(**a**) (**b**)

Figure 3. The actor and critic network weights.

 doi:10.20944/preprints202602.0353.v1



(**a**)

Figure 4. The utility function.

Remark 4. *The purpose of presenting Example 1 is to compare the performance of our proposed algorithm with that of Algorithm in [30]. To further verify its adaptability, we select a higher-order problem as Example 2.*

Example 2: This example considers a third-order system. The system dynamics are described as follows:

$$\begin{aligned} \dot{x}_1(t) &= x_2(t), \\ \dot{x}_2(t) &= x_3(t), \\ \dot{x}_3(t) &= x_1(t)x_2(t) + e^{-2x_3(t)} + gu. \end{aligned} \tag{43}$$

The control gain is $g = 2$, and the initial states are $x_1(0) = 4$, $x_2(0) = 0$, $x_3(0) = 1.5$.
Given the reference trajectory $y_d = 2(sin(t) + cos(0.5t))$, the tracking errors can be calculated

$$\begin{aligned} z_1(t) &= x_1(t) - 2(sin(t) + cos(0.5t)), \\ z_2(t) &= x_2(t) - 2(cos(t) - 0.5sin(0.5t)), \\ z_3(t) &= x_3(t) + 2(sin(t) + 0.25cos(0.5t)) \end{aligned} \tag{44}$$

The error dynamics derived in (43) are

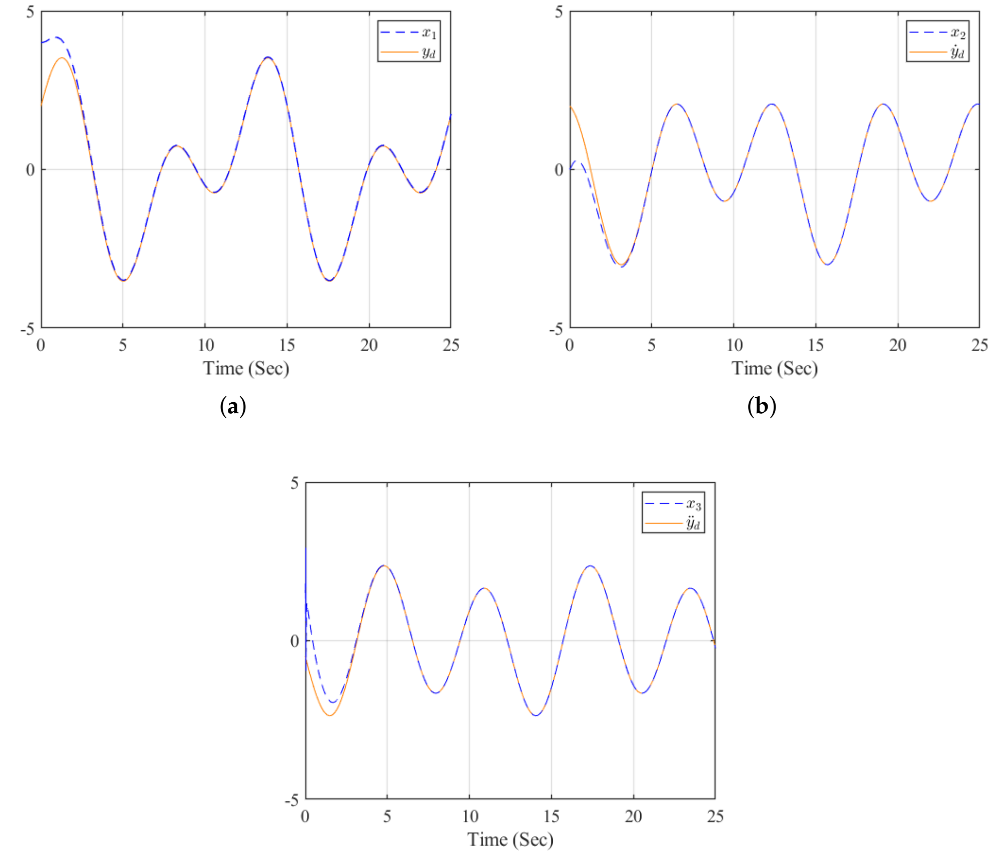
$$\begin{aligned} \dot{z}_1(t) &= z_2(t), \\ \dot{z}_2(t) &= z_3(t), \\ \dot{z}_3(t) &= [z_1(t) + 2(\sin(t) + \cos(0.5t))][z_2(t) + 2(\cos(t) - 0.5\sin(0.5t)] \\ &\quad + \exp[-2(z_3(t) - 2(\sin(t) + 0.25\cos(0.5t)))] + gu + 2(\cos(t) - 0.125\sin(0.5t)). \end{aligned} \tag{45}$$

Based on (3), the SMS is designed as

$$s(t) = z_1(t) + 2z_2(t) + z_3(t). \tag{46}$$

The activation function vector is given by $\Phi(\bar{z}) = \left[z_1^2, z_1z_2, z_1z_3, z_2^2, z_2z_3, z_3^2\right]^T$. The weight vectors are initialized to $\hat{W}_c(0) = [-1, 0, 1, 2, 0, 1]^T$ and $\hat{W}_a(0) = [0, 40, 0, 20, 0, 10]^T$. The parameters are set to $\gamma_u = 80$, $\gamma_c = 3$, $\gamma_a = 5$, and $\sigma = 1$.

The simulation results of Example 2 are presented in Figures 5–8. As shown in Figure 5, the states $x_1(t)$, $x_2(t)$ and $x_3(t)$ closely follow the reference trajectories y_d and its first- and second-order derivatives $\dot{y}_d$ and $\ddot{y}_d$, respectively. The corresponding tracking errors z_1, z_2, and z_3 depicted in Figure 6, converge to a small neighborhood of zero. Furthermore, Figure 7 demonstrates that the neural network weights for both the actor and critic remain bounded. Finally, the evolution of the utility function is shown in Figure 8. These results demonstrate that the proposed high-order tracking control scheme for canonical nonlinear systems maintains the boundedness of the error signal while ensuring accurate tracking of each system state to its corresponding reference signal.



(a) (b) (c)

Figure 5. The tracking performances.

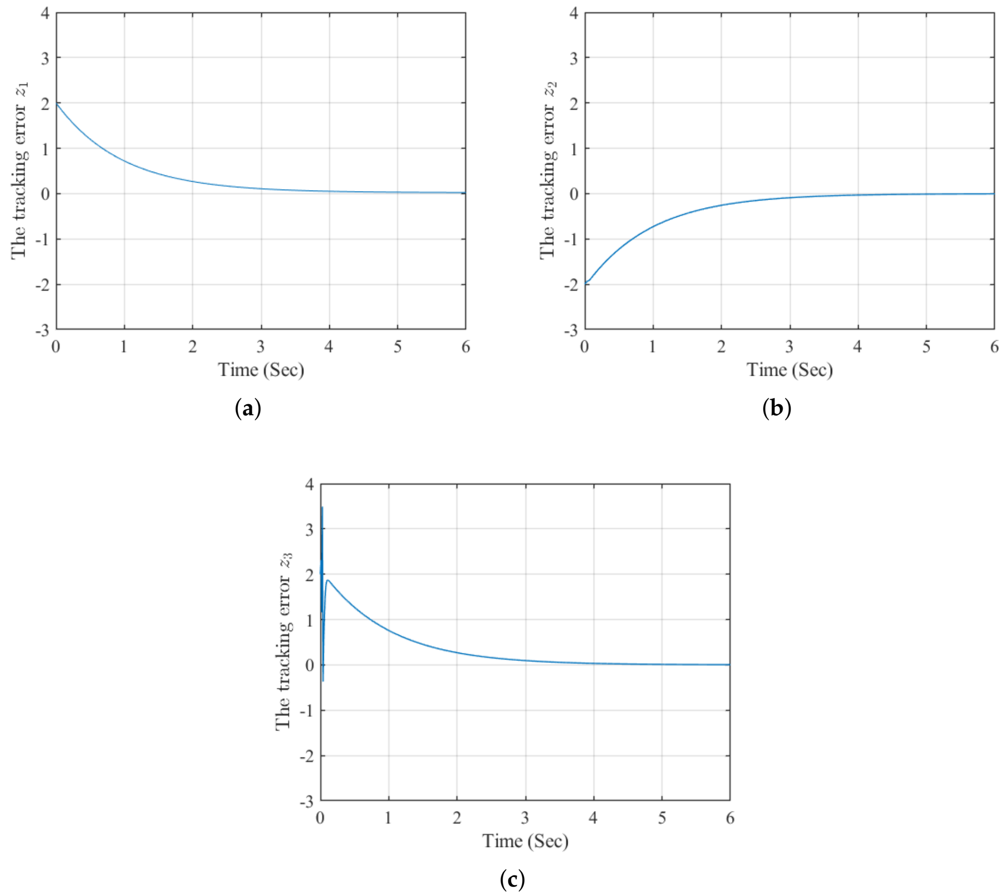


Figure 6. The tracking errors.

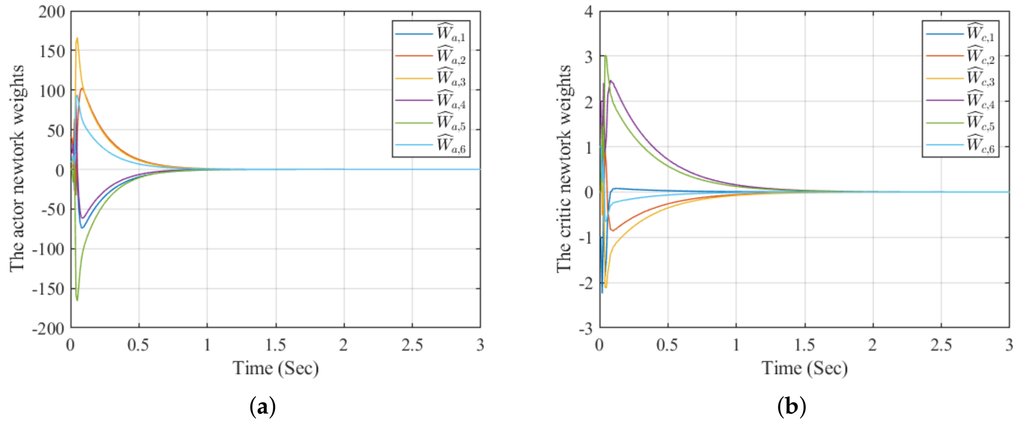
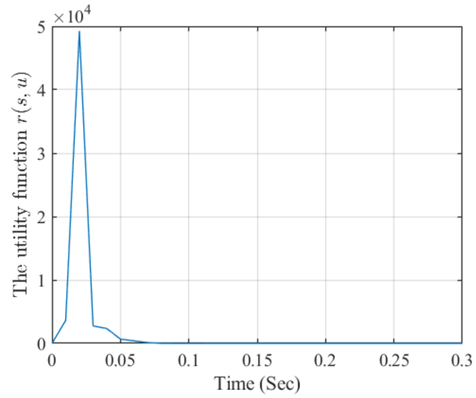


Figure 7. The actor and critic network weights.

 doi:10.20944/preprints202602.0353.v1



(**a**)

Figure 8. The utility function.

6. Conclusions

This paper proposes a SMS-based adaptive optimal tracking control scheme, utilizing an NN-based RL approach, for a class of high-order nonlinear systems with partial unknowns. A cost function defined in terms of the SMS is constructed to derive the optimal control law. Accordingly, the HJB equation is addressed within an actor-critic NN framework. Leveraging the inherent relationship between the SMS and the tracking errors significantly simplifies the computational process. Under the premise of ensuring the stability of the closed-loop system, the update law of the actor-critic network is indirectly derived under the condition that a simple function is zero, without the need for persistence excitation condition. Furthermore, the proposed design requires no prior knowledge of the internal system dynamics. Meanwhile, the proposed method offers a significant advantage over [30] by eliminating the need for a Riccati-like equation. Comparative simulation results demonstrate the superior tracking performance of the proposed scheme over existing methods. Finally, future work will extend this framework to tackle the inverse optimal tracking problem, addressing the challenge of manual cost function specification.

Author Contributions: D.X.: investigation, methodology, writing-review; X.L.: investigation, software; F.L.: formal analysis; J.T.: conceptualization, writing-original draft preparation; All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by the National Natural Science Foundation of China under Grant No. 61463002 and the Doctoral Foundation of Guangxi University of Science and Technology Grant No. Xiaokebo 22Z04.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data presented in this study are already in the article. For further consultation, please contact the corresponding author.

Acknowledgments: The authors thank the journal editors and reviewers for their valuable suggestions and opinions.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Li P, Duan G, Zhang B, et al. High-Order Fully Actuated Approach for Output Tracking Control of Flexible Servo Systems Subject to Uncertainties and Disturbances[J]. IEEE Transactions on Industrial Electronics, 2025.
2. Bu X, Qi Q. Fuzzy optimal tracking control of hypersonic flight vehicles via single-network adaptive critic design[J]. IEEE Transactions on Fuzzy Systems, 2020, 30(1): 270-278.

3. Lee J, Chang P H, Jin M. Adaptive integral sliding mode control with time-delay estimation for robot manipulators[J]. IEEE Transactions on Industrial Electronics, 2017, 64(8): 6796-6804.
4. Wang C, Liu Z, Sun S, et al. Trajectory tracking of a mobile robot in underground roadways based on hierarchical model predictive control. Actuators, 2026, 15(1):47.
5. Voos H. Nonlinear control of a quadrotor micro-UAV using feedback-linearization[C]//2009 IEEE International Conference on Mechatronics. IEEE, 2009: 1-6.
6. Bonna R, Camino J F. Trajectory tracking control of a quadrotor using feedback linearization[C]//International Symposium on Dynamic Problems of Mechanics. Sao Paulo, Brazil: University of Campinas, 2015, 1(9).
7. Yousefizadeh S, Bendtsen J D, Vafamand N, et al. Tracking control for a DC microgrid feeding uncertain loads in more electric aircraft: Adaptive backstepping approach[J]. IEEE Transactions on Industrial Electronics, 2018, 66(7): 5644-5652.
8. Zhao K, Song Y. Removing the feasibility conditions imposed on tracking control designs for state-constrained strict-feedback systems[J]. IEEE Transactions on Automatic Control, 2018, 64(3): 1265-1272.
9. Yu J, Ma Y, Yu H, et al. Adaptive fuzzy dynamic surface control for induction motors with iron losses in electric vehicle drive systems via backstepping[J]. Information Sciences, 2017, 376: 172-189.
10. Song Y D, Zhou S. Tracking control of uncertain nonlinear systems with deferred asymmetric time-varying full state constraints[J]. Automatica, 2018, 98: 314-322.
11. Wang Y, Gu L, Xu Y, et al. Practical tracking control of robot manipulators with continuous fractional-order nonsingular terminal sliding mode[J]. IEEE Transactions on industrial electronics, 2016, 63(10): 6194-6204.
12. Qiao L, Zhang W. Adaptive non-singular integral terminal sliding mode tracking control for autonomous underwater vehicles[J]. IET Control Theory & Applications, 2017, 11(8): 1293-1306.
13. Hwang C L, Chiang C C, Yeh Y W. Adaptive fuzzy hierarchical sliding-mode control for the trajectory tracking of uncertain underactuated nonlinear dynamic systems[J]. IEEE Transactions on Fuzzy Systems, 2013, 22(2): 286-299.
14. Zhang H, Wang H, Niu B, et al. Sliding-mode surface-based adaptive actor-critic optimal control for switched nonlinear systems with average dwell time[J]. Information Sciences, 2021, 580: 756-774.
15. Zhao H, Zhao N, Zong G, et al. Sliding-mode surface-based approximate optimal control for nonlinear multiplayer Stackelberg-Nash games via adaptive dynamic programming[J]. Communications in Nonlinear Science and Numerical Simulation, 2024, 132: 107928.
16. Zhang H, Zhao X, Wang H, et al. Hierarchical sliding-mode surface-based adaptive actor–critic optimal control for switched nonlinear systems with unknown perturbation[J]. IEEE Transactions on Neural Networks and Learning Systems, 2022, 35(2): 1559-1571.
17. Huang J, Xu D, Li Y, et al. Inverse reinforcement learning for discrete-time linear systems based on inverse optimal control[J]. ISA transactions, 2025, 163: 108-119.
18. Xu D, Wang Q, Li Y. Optimal guaranteed cost tracking of uncertain nonlinear systems using adaptive dynamic programming with concurrent learning[J]. International Journal of Control, Automation and Systems, 2020, 18(5): 1116-1127.
19. Na J, Lv Y, Zhang K, et al. Adaptive identifier-critic-based optimal tracking control for nonlinear systems with experimental validation[J]. IEEE Transactions on Systems, Man, and Cybernetics: Systems, 2020, 52(1): 459-472.
20. Modares H, Lewis F L. Optimal tracking control of nonlinear partially-unknown constrained-input systems using integral reinforcement learning[J]. Automatica, 2014, 50(7): 1780-1792.
21. Wen G, Chen C L P, Ge S S, et al. Optimized adaptive nonlinear tracking control using actor–critic reinforcement learning strategy[J]. IEEE transactions on industrial informatics, 2019, 15(9): 4969-4977.
22. Wang N, Gao Y, Zhao H, et al. Reinforcement learning-based optimal tracking control of an unknown unmanned surface vehicle[J]. IEEE Transactions on Neural Networks and Learning Systems, 2020, 32(7): 3034-3045.
23. Zhang H, Zhang K, Cai Y, et al. Adaptive fuzzy fault-tolerant tracking control for partially unknown systems with actuator faults via integral reinforcement learning method[J]. IEEE Transactions on Fuzzy Systems, 2019, 27(10): 1986-1998.
24. Zhuang H, Zhou H, Shen Q, et al. Optimal robust online tracking control for space manipulator in task space using off-policy reinforcement learning[J]. Aerospace Science and Technology, 2024, 153: 109446.
25. Wu T, Zhang Y, Yang X, et al. Predefined-time nearly optimal trajectory tracking control for autonomous surface vehicles with unknown dynamics[J]. Ocean Engineering, 2025, 327: 121021.

doi:10.20944/preprints202602.0353.v1

26. Wen G, Ge S S, Tu F. Optimized backstepping for tracking control of strict-feedback systems[J]. IEEE transactions on neural networks and learning systems, 2018, 29(8): 3850-3862.
27. Liu Y, Zhu Q, Wen G. Adaptive tracking control for perturbed strict-feedback nonlinear systems based on optimized backstepping technique[J]. IEEE Transactions on Neural Networks and Learning Systems, 2020, 33(2): 853-865.
28. Huang Z, Bai W, Li T, et al. Adaptive reinforcement learning optimal tracking control for strict-feedback nonlinear systems with prescribed performance[J]. Information Sciences, 2023, 621: 407-423.
29. Yuan H, Cao L, Lin W, et al. Learning-observer-based fixed-time optimal tracking control for robotic manipulators with full-state constraints[J]. Neurocomputing, 2025: 130186.
30. Wen G, Niu B. Optimized tracking control based on reinforcement learning for a class of high-order unknown nonlinear dynamic systems[J]. Information Sciences, 2022, 606: 368-379.
31. Song Y, Wang Y, Holloway J, et al. Time-varying feedback for regulation of normal-form nonlinear systems in prescribed finite time[J]. Automatica, 2017, 83: 243-251.
32. Zhu J, Wen G, Veluvolu K C. Optimized backstepping consensus control using adaptive observer-critic–actor reinforcement learning for strict-feedback multi-agent systems[J]. Journal of the Franklin Institute, 2024, 361(6): 106693.
33. Hušek P. Adaptive sliding mode control with moving sliding surface[J]. Applied Soft Computing, 2016, 42: 178-183.
34. Dao P N, Phung M H. Nonlinear robust integral based actor–critic reinforcement learning control for a perturbed three-wheeled mobile robot with mecanum wheels[J]. Computers and Electrical Engineering, 2025, 121: 109870.
35. Lin J, Wang M, Yan H, et al. Prescribed-Time Optimal Tracking Control for a Class of Stochastic Systems Using Reinforcement Learning[J]. Journal of the Franklin Institute, 2025: 107881.