Article

Reinforcement-Learning-Based Tracking Control with Fixed-Time Prescribed Performance for Reusable Launch Vehicle under Input Constraints

Shihao Xu [1], Yingzi Guan [1], Changzhu Wei [1,*] , Yulong Li [2] and Lei Xu [1]

[1] School of Astronautics, Harbin Institute of Technology, Harbin 150001, China; xsh_hit@163.com (S.X.); guanyz@hit.edu.cn (Y.G.); xullei517@126.com (L.X.)
[2] Beijing Institute of Space Launch Technology, China Academy of Launch Vehicle Technology, Beijing 100076, China; 556long@163.com
[*] Correspondence: weichangzhu@hit.edu.cn

Abstract: This paper proposes a novel reinforcement learning (RL)-based tracking control scheme with fixed-time prescribed performance for a reusable launch vehicle subject to parametric uncertainties, external disturbances, and input constraints. First, a fixed-time prescribed performance function is employed to restrain attitude tracking errors, and an equivalent unconstrained system is derived via an error transformation technique. Then, a hyperbolic tangent function is incorporated into the optimal performance index of the unconstrained system to tackle the input constraints. Subsequently, an actor-critic RL framework with super-twisting-like sliding mode control is constructed to establish a practical solution for the optimal control problem. Benefiting from the proposed scheme, the robustness of the RL-based controller against unknown dynamics is enhanced, and the control performance can be qualitatively prearranged by users. Theoretical analysis shows that the attitude tracking errors converge to a preset region within a preassigned fixed time, and the weight estimation errors of the actor-critic networks are uniformly ultimately bounded. Finally, comparative numerical simulation results are provided to illustrate the effectiveness and improved performance of the proposed control scheme.

Keywords: reinforcement learning-based control; prescribed performance control; fixed-time control; input constraints; reusable launch vehicle

1. Introduction

Recent years have witnessed an increasing demand for reliable and economical access to space. Reusable launch vehicles (RLV), as a cost-effective means of undertaking space missions, are attracting more and more attention from researchers [1]. A dynamic model of RLV provides strong non-linear and coupling characteristics due to the complex flight environment of the re-entry phase. External disturbances, uncertain structural and aerodynamic parameters, and input constraints inevitably exist during real flight, having a significant impact on the attitude control system. In this context, attitude control for RLV is a challenging topic and has elicited widespread interest. Various control methodologies, such as adaptive control [2], dynamic inversion control [3], robust control [4], sliding mode control [5,6], and neural network (NN) control [7,8], have been applied over the past decades. Nevertheless, there is still scope to develop an optimal control approach for RLV suffering from complicated non-linear dynamics, parametric uncertainties and limited inputs.

From a mathematical point of view, the Hamilton–Jacobi–Bellman (HJB) function and its solution are required to be established to solve the optimal control problems. However, it is difficult to derive an analytical solution from the HJB function for non-linear continuous-time systems. Given this, a reinforcement learning (RL) scheme with an actor-critic (AC)

structure was initially created by Werbos [9], whereby a critic network was exploited to approximate the value function, and an actor network was deployed to obtain the optimal control policy. Informed by Werbos' contribution, Vamvoudakis et al. developed an online AC algorithm to solve the continuous-time infinite horizon optimal control problem [10]. He et al. proposed a novel online learning and optimization structure by incorporating a reference network into the AC structure [11]. Ma et al. devised a learning-based adaptive sliding mode control scheme for a tethered space robot with limited inputs [12]. Although the above control strategies have provided excellent results in terms of optimal control, the existing problem concerns the need for accurate dynamic modeling [13]. Given the parametric uncertainties and unknown disturbances, it is, in practice, difficult to exactly determine the system dynamics for RLV, limiting the methods' applicability. Therefore, a further problem exists in that the aforementioned methods' robustness must be enhanced for practical systems with unknown dynamics. Fan et al. combined (ISMC) with a reinforcement learning control scheme for non-linear systems with partially unknown dynamics [14]. However, the input constraints were not considered, and the ISMC used may lead to unexpected oscillation [15]. Zhang et al. developed a learning-based H_∞ tracking control scheme for which the uncertainties and input constraints were considered [16]. Nevertheless, the control design is conservative, and the iterative algorithm is rather complicated.

It is of note that previous RL-based control methods only establish asymptotic or finite-time convergence [17]. Therefore, the upper bound of the convergence time is uncontrollable, and the transient and steady-state performance, namely, the maximum overshoot and the steady accuracy, cannot be quantitatively prearranged by users. As a promising solution to this problem, the prescribed performance control (PPC) method created by Bechlioulis et al. has attracted widespread attention [18]. The salient feature of PPC is that users can quantitatively pre-arrange both the transient performance and the steady tracking error. In [19], a novel PPC scheme combined with a command filter was proposed for a quadrotor unmanned aerial vehicle subject to error constraints. In [20], an NN-based adaptive non-affine tracking controller was devised for an air-breathing hypersonic vehicle with guaranteed prescribed performance. In [21], a data-driven PPC scheme was developed for an unmanned surface vehicle with unknown dynamics. However, the conventional PPC approach only ensures that the system states converge to the preset region as time tends to infinity [22], leading to an unsatisfactory solution for time-limited problems, such as the re-entry mission. Moreover, the input constraints and the optimality are not comprehensively considered in the conventional PPC paradigm. Therefore, the expected performance index cannot be consistently guaranteed and optimized for RLV suffering from poor aerodynamic maneuverability and limited control torques.

Motivated by the foregoing considerations, a novel RL-based tracking controller, with fixed-time prescribed performance for RLV subject to parameter uncertainties, external disturbances and input constraints, is investigated. The main contributions and characteristics of the proposed method can be summarized as follows.

- An online RL-based, nearly optimal, controller with limited inputs is developed by synthesizing the AC structure and the hyperbolic tangent performance index. In addition, the robustness of the learning-based controller is strengthened by incorporating a super-twisting-like sliding mode control.
- Compared with a previous learning-based controller described in [10–12], in which the system dynamics are required to be known exactly, the proposed control scheme only requires the input-output data pairs of RLV, such that the system dynamics can be completely unknown.
- In contrast to existing RL-based control schemes with asymptotic or finite-time convergence [10–12,17,21], the proposed control scheme can ensure that the tracking errors converge to a preset region within a preassigned fixed time. Moreover, the prescribed transient and steady-state performance can be guaranteed.

- Comparative numerical simulation investigations show that the proposed method can provide improved performance in terms of the transient response, and steady accuracy with less control effort.

2. Problem Statement and Preliminaries

2.1. Problem Statement

Following [2], the control-oriented model of the rigid-body RLV is given as follows:

$$\begin{cases} \dot{\Theta} = R\omega \\ I\dot{\omega} = -\Omega I\omega + M + \Delta D \end{cases} \tag{1}$$

where $\Theta = [\alpha, \beta, \sigma]^\top$ represents the attitude angle vector, $\omega = [p, q, r]^\top$ denotes the angular rate vector, $M = [M_x, M_y, M_z]^\top$ is the control input vector, and M_x, M_y, M_z are limited to the interval $[-\bar{M}, \bar{M}]$, $\Delta D = \Delta D_a + \Delta D_e$ is the unknown disturbance vector, ΔD_a is the aerodynamic torque vector, and ΔD_e is the external disturbance vector. ΔD_a can be formulated as:

$$\Delta D_a = \begin{bmatrix} \left(\frac{m_x^p L_r p}{V} + m_x^\sigma \sigma \right) q_V S_r L_r \\ \left(\frac{m_y^r L_r r}{V} + m_y^\beta \beta \right) q_V S_r L_r \\ \left(\frac{m_z^q L_r q}{V} + m_z^\alpha \alpha \right) q_V S_r L_r \end{bmatrix} \tag{2}$$

where m_x^p, m_y^r and m_z^q represent the damping moment coefficients, m_x^σ, m_y^β and m_z^α are the static stability moment coefficients, V is the velocity, q_V is the dynamic pressure, and S_r and L_r are the cross-sectional area and the reference length of RLV, respectively.

The skew-symmetric matrix Ω, the inertia matrix I, and the coordinate transformation matrix R are defined by

$$\Omega = \begin{bmatrix} 0 & -r & q \\ r & 0 & -p \\ -q & p & 0 \end{bmatrix}, I = \begin{bmatrix} I_{xx} & 0 & -I_{xz} \\ 0 & I_{yy} & 0 \\ -I_{zx} & 0 & I_{zz} \end{bmatrix},$$

$$R = \begin{bmatrix} -\cos\alpha\tan\beta & 1 & -\sin\alpha\tan\beta \\ \sin\alpha & 0 & -\cos\alpha \\ -\cos\alpha\cos\beta & -\sin\beta & -\sin\alpha\cos\beta \end{bmatrix}. \tag{3}$$

Defining the guidance command vector $\Theta_d = [\alpha_c, \beta_c, \sigma_c]^\top$, the attitude tracking error vector $e_1 = \Theta - \Theta_d = [e_{1\alpha}, e_{1\beta}, e_{1\sigma}]^\top$, and the angular rate tracking error vector $e_2 = R\omega - \dot{\Theta}_d$, the tracking error dynamics are given as:

$$\begin{cases} \dot{e}_1 = e_2 \\ \dot{e}_2 = B_1 M + \Delta D_1 \end{cases} \tag{4}$$

where $B_1 = RI^{-1}$ is the control matrix, $\Delta D_1 = -RI^{-1}\Omega I\omega - \ddot{\Theta}_c + \dot{R}\omega + RI^{-1}\Delta D$ denotes the lumped disturbance vector.

Assumption 1 ([2]). ΔD_1 *is bounded by* $||\Delta D_1|| \leq D_m$.

Assumption 2 ([2]). *During the re-entry phase, $\beta \neq \pm 90$ deg, thus R is always invertible.*

Control Objective: According to the tracking error dynamics (4), the control objective of this paper can be summarized as developing an RL-based optimal control scheme with guaranteed fixed-time prescribed performance such that the attitude tracking errors $e_{1i}(i = \alpha, \beta, \sigma)$ can converge to a preset region within a preassigned fixed time.

2.2. Preliminaries

Lemma 1 ([23]). *Considering a non-linear function $\alpha(x) = \ln[1 - \tanh^2(x)]$, the following equation*

$$\alpha(x) = \ln(4) - 2x\,\mathrm{sign}(x) + \kappa_\alpha \tag{5}$$

always holds, where κ_α is bounded by a real positive constant.

Lemma 2 ([24]). *Considering the following fixed-time prescribed performance function (FTPPF)*

$$\rho(t) = \begin{cases} (\rho_0 - \rho_\infty)\left[\dfrac{\sin(2\pi t/T)}{2\pi} - \dfrac{t}{T}\right] + \rho_0, & 0 \le t \le T \\ \rho_\infty, & t > T \end{cases} \tag{6}$$

where $\rho_0 > \rho_\infty > 0$ represent the initial and terminal values of FTPPF, respectively. $T > 0$ is the preassigned convergence time. It can be concluded that $\rho(t)$ is a positive, non-increasing and $\mathcal{C}^2$ continuous function with $\rho(0) = \rho_0$, $\rho(T) = \rho_\infty$ and $\dot{\rho}(T) = \ddot{\rho}(T) = 0$.

Lemma 3 ([25]). *For any $\mu > 0$, the following inequality holds*

$$0 \le |x| - x\tanh\left(\frac{x}{\mu}\right) \le k_p\mu \tag{7}$$

where k_p satisfies $k_p = e^{-(k_p+1)}$ (i.e., $k_p = 0.2785$).

3. Controller Design

3.1. Prescribed Performance Constraint

In this subsection, the following constraint is formulated to restrain the attitude tracking errors $e_{1i}(i = \alpha, \beta, \sigma)$ within the FTPPF

$$-\rho_i(t) < e_{1i}(t) < \rho_i(t) \tag{8}$$

where $\rho_i(t)$ is defined in Lemma 2. Subsequently, the equivalent unconstrained error variables $\eta_{1i}(i = \alpha, \beta, \sigma)$ can be derived via the error transformation method [18]

$$\eta_{1i} = \frac{1}{2}\ln\frac{1 + z_i}{1 - z_i} \tag{9}$$

with $z_i = e_{1i}/\rho_i$. Taking the first and second-order time derivatives of η_{1i} yields

$$\begin{cases} \dot{\eta}_{1i} = \xi_i\eta_{2i} \\ \dot{\eta}_{2i} = \ddot{e}_{2i} - \Lambda_i \end{cases} \tag{10}$$

with $\eta_{2i} = e_{2i} - \frac{\dot{\rho}_i}{\rho_i}e_{1i}$, $\xi_i = \frac{1}{2\rho_i}\left(\frac{1}{z_i+1} - \frac{1}{z_i-1}\right) \ge \frac{1}{\rho_{i\infty}}$ and $\Lambda_i = \frac{\rho_i\ddot{\rho}_i e_{1i} + \rho_i\dot{\rho}_i e_{2i} - \dot{\rho}_i^2 e_{1i}}{\rho_i^2}$. Substituting (4) into (10), one obtains

$$\begin{cases} \dot{\boldsymbol{\eta}}_1 = \boldsymbol{\xi}\boldsymbol{\eta}_2 \\ \dot{\boldsymbol{\eta}}_2 = \boldsymbol{B}_1\boldsymbol{M} + \boldsymbol{D}_1 - \boldsymbol{\Lambda} \end{cases} \tag{11}$$

where $\boldsymbol{\xi} = \mathbb{D}(\xi_i)$, $\boldsymbol{\eta}_2 = \mathbb{C}(\eta_{2i})$ and $\boldsymbol{\Lambda} = \mathbb{C}(\Lambda_i)$, and the symbols $\mathbb{C}(\cdot)$ and $\mathbb{D}(\cdot)$ represent the diagonal matrix and the column vector, respectively.

3.2. Reinforcement Learning-Based Control Design

Firstly, the following sliding variable s is defined as:

$$\boldsymbol{s} = \boldsymbol{\eta}_2 + c\boldsymbol{\eta}_1 \tag{12}$$

where c is a positive diagonal matrix. Taking the first time derivative of s along (11) yields

$$\dot{s} = \Xi + B_1 M + D_1 \tag{13}$$

where $\Xi = c\dot{\eta}_1 - \Lambda$. In order to achieve a satisfactory tracking performance, a traditional super-twisting controller based on (12) and (13) is developed as follows:

$$M = -B_1^{-1}\Gamma, \Gamma = \Xi + \mathbb{C}\left[\lambda_{1i}\text{sig}^{1/2}(s_i)\right] - M_d, \tag{14}$$

where $\text{sig}^{1/2}(s_i) = |s_i|^{1/2}\text{sign}(s_i)$, $\dot{M}_d = -\mathbb{D}(\lambda_{2i})\text{sign}(s_i)$, and λ_{1i} and λ_{2i} are positive constants. Nevertheless, the feasibility of (14) may not be guaranteed with limited control inputs, and the time-fuel performance index cannot be approximately optimized. To this end, an online RL-based, nearly optimal, controller is proposed by integrating the AC structure into the super-twisting controller for a comprehensive solution to the issue mentioned above.

Before elaborating the detailed design procedure, it is assumed that there exists a group of admissible control strategies [10]

$$M = M_o - B_1^{-1}\left\{\mathbb{C}\left[\lambda_{1i}\text{sig}^{1/2}(s_i)\right] - M_d\right\} \in \Omega_u. \tag{15}$$

Moreover, M can achieve the control objective with the following time-fuel performance index being satisfied

$$V(s) = \int_t^\infty \left\{V_s(\tau) + J[M_o(\tau)]\right\}\mathrm{d}\tau \tag{16}$$

where $V_s = s^\top Q s$, Q is a positive diagonal matrix, and $J(M_o)$ is chosen as [26]:

$$J(M_o) = 2\int_0^{M_o}\left[\lambda\tanh^{-1}\left(\frac{v}{\lambda}\right)^\top \varpi\right]\mathrm{d}v \tag{17}$$

where ϖ is selected as a positive diagonal matrix, v represents the variable of integration, and the upper and lower bounds of v are M_o and $\mathbf{0}$, respectively. The Lyapunov equation of (16) can be calculated as:

$$V_s(t) + J[M_o(t)] + \nabla V_s^\top \dot{s} = 0 \tag{18}$$

where $\nabla V_s = \partial V(s)/\partial s$. The optimal value function is defined as:

$$V^*(s) = \min_{M_o}\int_t^\infty \left\{V_s(\tau) + J[M_o(\tau)]\right\}\mathrm{d}\tau, \tag{19}$$

and the corresponding HJB equation can be formulated as:

$$\min_{M_o}\left\{V_s + J(M_o) + \nabla V_s^{*\top}\dot{s}\right\} = 0, \tag{20}$$

where $\nabla V_s^* = \partial V^*(s)/\partial s$.

Equation (20) is equivalent to

$$\frac{\partial}{\partial M_o^*}\left[V_s + J(M_o^*) + \nabla V_s^{*\top}\dot{s}\right] = 0. \tag{21}$$

Solving the partial derivative of (21) yields the following optimal control strategy

$$M_o^* = -\lambda\tanh(\bar{M}_o^*), \bar{M}_o^* = \frac{1}{2\lambda}\varpi^{-1}B_1^\top \nabla V_s^*. \tag{22}$$

Substituting M_o^* into (17), one obtains

$$J(M_o^*) = \lambda \nabla V_s^{*\top} B_1 \tanh(\bar{M}_o^*) + \lambda^2 \bar{\varpi} \ln\left[\mathbf{1}_c - \tanh^2(\bar{M}_o^*)\right]$$
$$= -\nabla V_s^{*\top} B_1 M_o^* + \lambda^2 \bar{\varpi} \ln\left[\mathbf{1}_c - \tanh^2(\bar{M}_o^*)\right], \tag{23}$$

where $\mathbf{1}_c$ is a column vector with all elements being one, and $\bar{\varpi}$ is a row vector generated by the elements on the main diagonal of ϖ. Combining (20) and (23), it can be derived that

$$V_s + \nabla V_s^{*\top} \dot{s} + \lambda^2 \bar{\varpi} \ln[\mathbf{1}_c - \tanh^2(\bar{M}_o^*)] = 0. \tag{24}$$

The nearly optimal control M_o^* can be obtained by solving (24) if ∇V_s^* is available. Inspired by the NN-based control scheme, the following approximation of $V^*(s)$ can be established

$$V^*(s) = W^{*\top} \sigma[s(t)] + \epsilon[s(t)] \tag{25}$$

where W^* is the optimal weight vector, $\sigma[s(t)]$ is the base vector, and $\epsilon[s(t)]$ is the approximation error of NN. Subsequently, the gradient of $V^*(s)$ is

$$\nabla V_s^* = W^{*\top} \nabla \sigma_s[s(t)] + \nabla \epsilon_s[s(t)] \tag{26}$$

where $\nabla \sigma_s = \partial \sigma / \partial s$, $\nabla \epsilon_s = \partial \epsilon / \partial s$. In light of the universal approximation property of NN for smooth functions on prescribed compact sets, the approximation errors $\epsilon[s(t)]$ and $\nabla \epsilon_s[s(t)]$ are bounded with a finite dimension of $\sigma[s(t)]$ [27]. Moreover, it is assumed that $||W^*||$, $\sigma[s(t)]$ and $\nabla \sigma_s[s(t)]$ are bounded [13,14,16,17,21].

Recalling (22), the NN-based nearly optimal control law can be formulated as:

$$\begin{cases} M = \hat{M}_o^* - B_1^{-1}\left\{\mathbb{C}\left[\lambda_{1i}\mathrm{sig}^{1/2}(s_i)\right] - M_d\right\} \\ \hat{M}_o^* = -\lambda \tanh\left[\dfrac{1}{2\lambda}\varpi^{-1}B_1^\top\left(\nabla\sigma_s^\top W^* + \nabla\epsilon_s\right)\right] \end{cases}. \tag{27}$$

Defining the Bellman error as [14]:

$$B_\epsilon = \lambda^2 \bar{\varpi}\left\{\ln\left[\mathbf{1}_c - \tanh^2(\bar{M}_o^*)\right] - \ln\left[\mathbf{1}_c - \tanh^2\left(\hat{M}_o^*\right)\right]\right\} \tag{28}$$

with $\hat{M}_o^* = \frac{1}{2\lambda}\varpi^{-1}B_1^\top\nabla\sigma_s^\top W^*$, (20) can be rewritten as:

$$V_s + W^{*\top}\nabla\sigma_s\Gamma + \lambda^2\bar{\varpi}\ln\left[\mathbf{1}_c - \tanh^2\left(\hat{M}_o^*\right)\right] + \epsilon_H = 0 \tag{29}$$

where $\epsilon_H = \nabla\epsilon_s^\top\Gamma + B_\epsilon + W^*\nabla\sigma_s D_2$ is the bounded HJB error [10].

In this paper, the optimal weight W^* is generated by the online RL scheme with the AC structure. In this context, the nearly optimal control policy is formulated as:

$$\begin{cases} M = \hat{M}_a^* - B_1^{-1}\left\{\mathbb{C}\left[\lambda_{1i}\mathrm{sig}^{1/2}(s_i)\right] - M_d\right\} \\ \hat{M}_a^* = -\lambda\tanh\left(\hat{M}_a^*\right), \hat{M}_a^* = \dfrac{1}{2\lambda}\varpi^{-1}B_1^\top\nabla\sigma_s^\top\hat{W}_a \end{cases} \tag{30}$$

where $\hat{W}_a$ is the weight of the actor network. Moreover, the performance index (16) can be estimated as:

$$\hat{V}_s = \hat{W}_c^\top\sigma_s[s(t)] \tag{31}$$

where $\hat{W}_c$ is the weight of the critic network. The adaptation laws for $\hat{W}_c$ and $\hat{W}_a$ are designed as:

$$\dot{\hat{W}}_c = -A_1\left\{\bar{\theta}\left[\theta^\top \hat{W}_c + V_s + J(\hat{M}_a^*)\right] + p_c(\hat{W}_c - \hat{W}_a)\right\}, \tag{32}$$

$$\dot{\hat{W}}_a = -A_2\left[p_a(\hat{W}_a - \hat{W}_c) - \frac{\lambda p_a}{p_c}\nabla\sigma_s B_1 \Psi \bar{\theta}^\top \hat{W}_c\right], \tag{33}$$

where $\theta = \nabla\sigma_s(\Gamma + B_1\hat{M}_a^*)$, $\bar{\theta} = \theta/(\theta^\top\theta + 1)^2$; A_1 and A_2 are positive diagonal matrices; $\Psi = \tanh\left(\hat{M}_a^*\right) - \tanh\left(\hat{M}_a^*/\kappa\right)$; p_c, p_a and κ are positive real constants. The projection operator $\mathrm{Proj}(\cdot)$ is imposed on (33) to guarantee that $\hat{W}_a$ is bounded [28]. It is assumed that $\theta_1 = \theta/(\theta^\top\theta + 1)$ is persistently exciting (PE) [29].

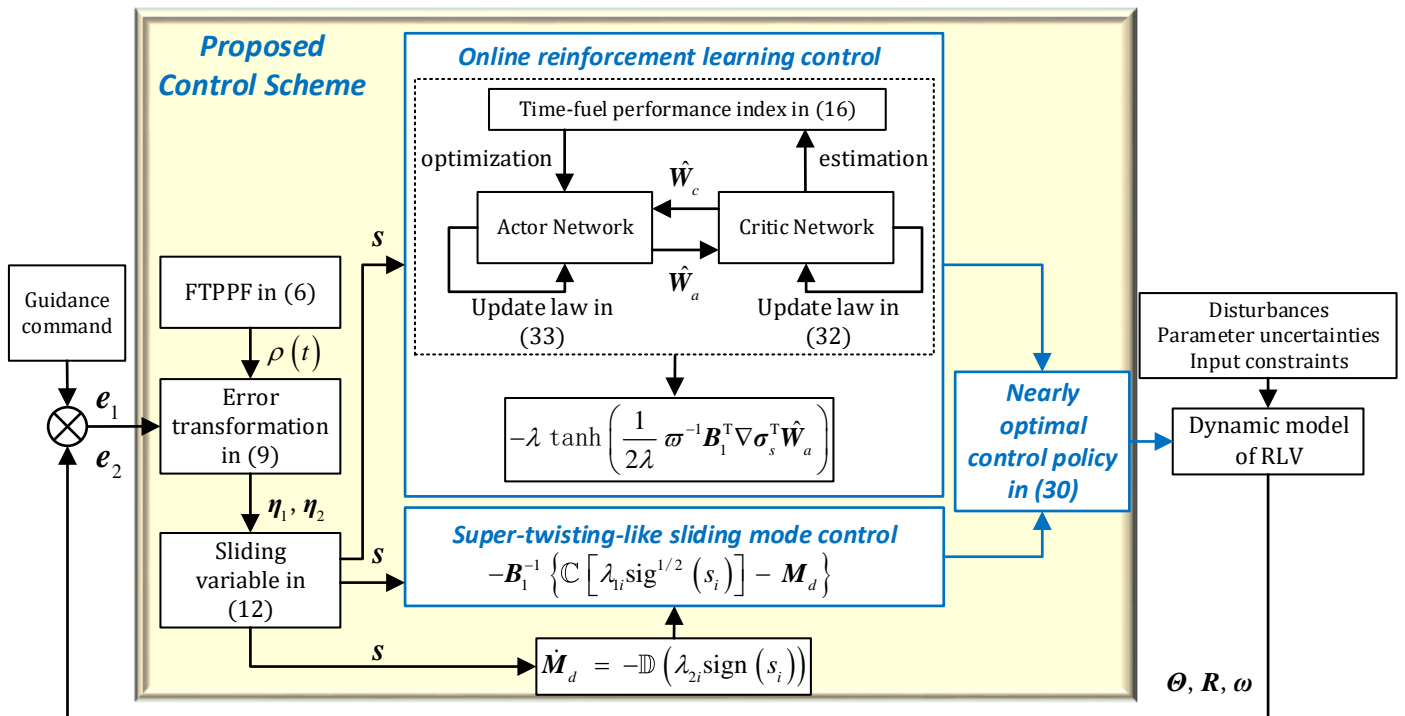The proposed control scheme is illustrated by a block diagram in Figure 1.

Figure 1. The block diagram of the proposed control scheme.

3.3. Stability Analysis

Theorem 1. *Considering the control-oriented RLV model (1) with input constraints, if the initial conditions satisfies $-\rho_{i0} < e_{1i}(0) < \rho_{i0}$, the nearly optimal control policy is chosen as (30) with the weight update laws (32) and (33), then the following results can be obtained:*

- *the sliding variable s, the weight estimation errors $\tilde{W}_a$ and $\tilde{W}_c$ are uniformly ultimately bounded (UUB);*
- *the attitude tracking errors e_{1i} uniformly obey the fixed-time performance envelops in (8).*

Proof of Theorem 1. Consider the Lyapunov function candidate as follows:

$$V = V(s) + \frac{1}{2}\tilde{W}_c^\top A_1^{-1}\tilde{W}_c + \frac{p_c}{2p_a}\tilde{W}_a^\top A_2^{-1}\tilde{W}_a \tag{34}$$

where $\tilde{W}_c = W^* - \hat{W}_c$ and $\tilde{W}_a = W^* - \hat{W}_a$. Taking the first time derivative of V yields

$$\dot{V} = \dot{V}(s) + \dot{V}_c + \dot{V}_a \tag{35}$$

with $\dot{V}_c = \tilde{W}_c^\top A_1^{-1} \dot{\hat{W}}_c$, $\dot{V}_a = \frac{p_c}{p_a} \tilde{W}_a^\top A_2^{-1} \dot{\hat{W}}_a$, and

$$\dot{V}(s) = \nabla V_s^\top \left(\mathbf{\Gamma} + D_1 + B_1 \hat{M}_a^* \right). \tag{36}$$

By invoking the HJB equation (29), $\nabla V_s^\top \mathbf{\Xi}$ can be rewritten as:

$$\nabla V_s^\top \mathbf{\Xi} = -V_s - \lambda^2 \bar{\omega} \ln\left[\mathbf{1}_c - \tanh^2\left(\hat{M}_o^* \right) \right] + \nabla V_s^\top \left\{ \mathbb{C} \left[\lambda_{1i} \mathrm{sig}^{1/2}(s_i) \right] - M_d \right\} - \varepsilon_e \tag{37}$$

where $\varepsilon_e = B_\epsilon + \nabla V_s^\top D_1$. Substituting (37) into (36) yields

$$\dot{V}(s) = -V_s - \lambda^2 \bar{\omega} \ln\left[\mathbf{1}_c - \tanh^2\left(\hat{M}_o^* \right) \right] - \lambda \nabla V_s^{*\top} B_1 \tanh\left(\hat{M}_a^* \right) - B_\epsilon. \tag{38}$$

Furthermore, according to (17), it can be deduced that

$$\lambda^2 \bar{\omega} \ln\left[\mathbf{1}_c - \tanh^2\left(\hat{M}_o^* \right) \right] = J(\underline{M}_o^*) + \nabla V_s^{*\top} B_1 \underline{M}_o^* \tag{39}$$

with $\underline{M}_o^* = -\lambda \tanh\left(\hat{M}_o^* \right)$. Equation (38) can be rewritten as:

$$\begin{aligned}
\dot{V}(s) &= -V_s - J(\underline{M}_o^*) - \nabla V_s^{*\top} B_1 \underline{M}_o^* - B_\epsilon - \lambda \left(W^{*\top} \nabla \sigma_s + \nabla \epsilon_s^\top \right) B_1 \tanh\left(\hat{M}_a^* \right) \\
&= -V_s - J(\underline{M}_o^*) + \varepsilon_v - \lambda \left(\hat{W}_a^\top + \tilde{W}_a^\top \right) \nabla \sigma_s B_1 \tanh\left(\hat{M}_a^* \right)
\end{aligned} \tag{40}$$

where $\varepsilon_v = -B_\epsilon + \nabla V_s^{*\top} B_1 \underline{M}_o^* - \nabla \epsilon_s^\top B_1 \hat{M}_a^*$. Noting that $\tanh(\cdot)$ is an odd function, it can be concluded that $\lambda \hat{W}_a^\top \nabla \sigma_s B_1 \tanh\left(\hat{M}_a^* \right) > 0$. Moreover, it is indicated that $J(\underline{M}_o^*) > 0$ from the definition in (17). Based on the above discussions, (40) can be simplified as:

$$\dot{V}_s(s) \leq -V_s - \lambda \tilde{W}_a^\top \nabla \sigma_s B_1 \tanh\left(\hat{M}_a^* \right) + \varepsilon_v. \tag{41}$$

Subsequently, incorporating (32) with $\dot{V}_c$ yields

$$\dot{V}_c = -\tilde{W}_c^\top \mathbf{\Gamma}_1 \dot{\hat{W}}_c = \tilde{W}_c^\top \left\{ \bar{\theta} \left[\theta^\top \hat{W}_c + V_s + J(\hat{M}_a^*) \right] + p_c(\hat{W}_c - \hat{W}_a) \right\}. \tag{42}$$

Recalling (29) and the definition of θ, (42) can be rearranged as:

$$\begin{aligned}
\dot{V}_c &= \tilde{W}_c^\top \left\{ \bar{\theta}[\theta^\top \hat{W}_c - \theta^\top W^* + \theta^\top W^* - \epsilon_H - \lambda^2 \bar{\omega} \ln[\mathbf{1}_c - \tanh^2(\hat{M}_o^*)] \right. \\
&\quad \left. - W^{*\top} \nabla \sigma_s \mathbf{\Gamma} + J(\hat{M}_a^*)] \right\} + p_c \tilde{W}_c^\top (\hat{W}_c - \hat{W}_a) \\
&= \tilde{W}_c^\top \bar{\theta} \left\{ -\theta^\top \tilde{W}_c + \left(\nabla \sigma_s B_1 \hat{M}_a^* \right)^\top W^* - \lambda^2 \bar{\omega} \ln\left[\mathbf{1}_c - \tanh^2\left(\hat{M}_o^* \right) \right] - \epsilon_H \right. \\
&\quad \left. + \lambda \hat{W}_a^\top \nabla \sigma_s B_1 \tanh\left(\hat{M}_a^* \right) + \lambda^2 \bar{\omega} \ln\left[\mathbf{1}_c - \tanh^2\left(\hat{M}_a^* \right) \right] \right\} + p_c \tilde{W}_c^\top (\hat{W}_c - \hat{W}_a) \\
&= \tilde{W}_c^\top \bar{\theta} \left\{ -\theta^\top \tilde{W}_c - \lambda \tilde{W}_a^\top \nabla \sigma_s B_1 \tanh\left(\hat{M}_a^* \right) - \lambda^2 \bar{\omega} \ln\left[\mathbf{1}_c - \tanh^2\left(\hat{M}_o^* \right) \right] - \epsilon_H \right. \\
&\quad \left. + \lambda^2 \bar{\omega} \ln\left[\mathbf{1}_c - \tanh^2\left(\hat{M}_a^* \right) \right] \right\} + p_c \tilde{W}_c^\top (\hat{W}_c - \hat{W}_a).
\end{aligned} \tag{43}$$

With the aid of Lemma 1 and Lemma 3, (43) can be further simplified as:

$$
\begin{aligned}
\dot{V}_c &= \tilde{W}_c^\top \bar{\theta} \Big[-\theta^\top \tilde{W}_c - \lambda \tilde{W}_a^\top \nabla \sigma_s B_1 \tanh\left(\hat{\tilde{M}}_a^*\right) - \epsilon_H - \lambda B_1^\top \nabla \sigma_s^\top \hat{W}_a \text{sign}\left(\hat{\tilde{M}}_a^*\right) \\
&\quad + \lambda B_1^\top \nabla \sigma_s^\top W^* \text{sign}\left(\hat{\tilde{M}}_o^*\right) + \lambda^2 \bar{\omega}(\varepsilon_a - \varepsilon_{a^*}) \Big] + p_c \tilde{W}_c^\top \left(\hat{W}_c - \hat{W}_a\right) \\
&= \tilde{W}_c^\top \bar{\theta} \Big[-\epsilon_H + \lambda^2 \bar{\omega}(\varepsilon_a - \varepsilon_{a^*}) + \lambda B_1^\top \nabla \sigma_s^\top W^* \text{sign}\left(\hat{\tilde{M}}_o^*\right) \Big] \\
&\quad - \tilde{W}_c^\top \bar{\theta} \left\{ \lambda \tilde{W}_a^\top \nabla \sigma_s B_1 \tanh\left(\hat{\tilde{M}}_a^*\right) + \lambda \hat{W}_a^\top \nabla \sigma_s B_1 \left[\tanh\left(\frac{\hat{\tilde{M}}_a^*}{\kappa}\right) + \kappa_1 \right] \right\} \\
&\quad - \tilde{W}_c^\top \bar{\theta} \theta^\top \tilde{W}_c + p_c \tilde{W}_c^\top \left(\hat{W}_c - \hat{W}_a\right)
\end{aligned}
\tag{44}
$$

where $||\varepsilon_a||$, $||\varepsilon_{a^*}||$ and $|\kappa_1|$ are bounded. For ease of notation, two bounded vectors $\aleph_1$ and $\aleph_2$ are defined as:

$$
\begin{aligned}
\aleph_1 &= \left\{ \lambda \nabla \sigma_s B_1 \tanh\left(\frac{\hat{\tilde{M}}_a^*}{\kappa}\right) - \lambda \nabla \sigma_s B_1 \tanh\left(\hat{\tilde{M}}_a^*\right) \right\} \bar{\theta}^\top W^*, \\
\aleph_2 &= \lambda W^{*\top} \nabla \sigma_s B_1 \text{sign}\left(\hat{\tilde{W}}_o^*\right) - \lambda W^{*\top} \nabla \sigma_s B_1 \left[\tanh\left(\frac{\hat{\tilde{M}}_a^*}{\kappa}\right) + \kappa_1 \right] \\
&\quad - \epsilon_H + \lambda^2 \bar{\omega}(\varepsilon_a - \varepsilon_{a^*}),
\end{aligned}
\tag{45}
$$

then (44) can be further sorted into the following structure

$$
\begin{aligned}
\dot{V}_c &= -\tilde{W}_c^\top \bar{\theta} \theta^\top \tilde{W}_c + p_c \tilde{W}_c^\top \left(\hat{W}_c - \hat{W}_a\right) + \tilde{W}_c \bar{\theta} \aleph_2 + \tilde{W}_a^\top \aleph_1 + \lambda \kappa_1 \tilde{W}_c^\top \bar{\theta} \tilde{W}_a^\top \nabla \sigma_s B_1 \\
&\quad + \lambda \hat{W}_c^\top \bar{\theta} \tilde{W}_a^\top \nabla \sigma_s B_1 \Psi.
\end{aligned}
\tag{46}
$$

Substituting (33), (41), and (46) into (35) yields

$$
\begin{aligned}
\dot{V} &\leq -V_s - \lambda \tilde{W}_a^\top \nabla \sigma_s B_1 \tanh\left(\hat{\tilde{M}}_a^*\right) + \varepsilon_- \tilde{W}_c^\top \bar{\theta} \theta^\top \tilde{W}_c + p_c \tilde{W}_c^\top \left(\hat{W}_c - \hat{W}_a\right) \\
&\quad + \tilde{W}_c \bar{\theta} \aleph_2 + \tilde{W}_a^\top \aleph_1 + \lambda \kappa_1 \tilde{W}_c^\top \bar{\theta} \tilde{W}_a^\top \nabla \sigma_s B_1 - \lambda \hat{W}_c^\top \bar{\theta} \tilde{W}_a^\top \nabla \sigma_s B_1 \\
&\quad + \lambda \hat{W}_c^\top \bar{\theta} \tilde{W}_a^\top \nabla \sigma_s B_1 \Psi + \tilde{W}_a^\top \left[p_c(\hat{W}_a - \hat{W}_c) - \lambda \nabla \sigma_s B_1 \Psi \bar{\theta}^\top \hat{W}_c \right] \\
&\leq -V_s - \lambda \tilde{W}_a^\top \nabla \sigma_s B_1 \tanh\left(\hat{\tilde{M}}_a^*\right) + \varepsilon_v - \tilde{W}_c^\top \bar{\theta} \theta^\top \tilde{W}_c + p_c(\tilde{W}_a - \tilde{W}_c)^\top (\tilde{W}_a - \tilde{W}_c) \\
&\quad + \tilde{W}_c \bar{\theta} \aleph_2 + \tilde{W}_a^\top \aleph_1 + \lambda \kappa_1 \tilde{W}_c^\top \bar{\theta} \tilde{W}_a^\top \nabla \sigma_s B_1 \\
&\leq -V_s - \tilde{W}_c^\top \bar{\theta} \theta^\top \tilde{W}_c + \varepsilon_{v1} + \tilde{W}_c^\top \bar{\theta} \aleph_3
\end{aligned}
\tag{47}
$$

where $\varepsilon_{v1} = \varepsilon_v - \lambda \tilde{W}_a^\top \nabla \sigma_s B_1 \tanh\left(\hat{\tilde{M}}_a^*\right) + \tilde{W}_a^\top \aleph_1$ and $\aleph_3 = \aleph_2 + \lambda \kappa_1 \tilde{W}_a^\top \nabla \sigma_s B_1$.

With the definition of the generalized vector $G = \left[s, \tilde{W}_c^\top \theta_1\right]^\top$, the inequality (47) can be rewritten as:

$$
\dot{V} \leq -G^\top \Pi G + G^\top \Pi_\theta + \varepsilon_{v1}
\tag{48}
$$

where $\Pi = \text{diag}(Q, I)$, $\Pi_\theta = \left[0, \aleph_3 / \left(\theta^\top \theta + 1\right)\right]^\top$. Given that Π_θ, $\hat{W}_a$ are bounded vectors, it can be concluded that the $\tilde{W}_a$ is a bounded vector and $||\Pi_\theta|| \leq \bar{\Pi}_\theta$ and $|\varepsilon_{v1}| \leq \bar{\varepsilon}_{v1}$. Therefore, $\dot{V}$ is negative if

$$
||G|| > \frac{\bar{\Pi}_\theta \sqrt{\bar{\Pi}_\theta^2 + 4\lambda_{\min}(\Pi)\bar{\varepsilon}_{v1}}}{2\lambda_{\min}(\Pi)},
\tag{49}
$$

which implies that the sliding variable s and $\tilde{W}_c^\top \theta_1$ are UUB. Furthermore, in view of the assumption that θ_1 is PE, the weight estimation error $\tilde{W}_c$ is also UUB [14]. Once the sliding

variable converges to the vicinity of origin, the gradient vector $\nabla\sigma[s(t)]$ together with $\hat{M}_a^*$ will also be small enough. Defining $d = \Xi + D_1 + B_1\hat{M}_a^*$, and assuming that there exists a small positive constant $\bar{d}$ satisfying $||d|| \leq \bar{d}$, the dynamic of the sliding variable can be represented as:

$$\begin{bmatrix} \dot{s} \\ \dot{M}_d \end{bmatrix} = \begin{bmatrix} d - \mathbb{C}\left[\lambda_{1i}\text{sig}^{1/2}(s_i)\right] - M_d \\ -\mathbb{D}(\lambda_{2i})\text{sign}(s_i) \end{bmatrix}. \tag{50}$$

According to the deduction lines in [30], the sliding variable can reach a small neighborhood of origin; thus, the equivalent unconstrained errors η_{1i} are bounded.

Recalling the error transformation method in (9), it can be derived that

$$z_i = T(\eta_{1i}) = \frac{\exp(\eta_{1i}) - \exp(-\eta_{1i})}{\exp(\eta_{1i}) + \exp(-\eta_{1i})}. \tag{51}$$

Note that $T(\eta_{1i})$ is a smooth, strictly increasing, function w.r.t. η_{1i} and satisfies $\lim_{\eta_{1i}\to-\infty} T(\eta_{1i}) = -1$, $\lim_{\eta_{1i}\to\infty} T(\eta_{1i}) = 1$; thus, it can be derived that $-1 < z_i < 1$ and $-\rho_i < e_{1i} < \rho_i$ hold true if η_{1i} is bounded. Following the above analysis, it can be deduced that e_{1i} never violates the fixed-time prescribed performance constraints in (8) for all $t > 0$, and $|e_{1i}| < \rho_{i\infty}$ for $t > T$.

This completes the proof of Theorem 1. $\square$

4. Numerical Simulations

In this section, numerical simulations for the re-entry phase of RLV are carried out to illustrate the effectiveness and improved performance of the proposed control scheme. The parameters of the RLV are based on [31]. The initial states of the re-entry phase are set as $\alpha_0 = 47$ deg, $\beta_0 = -2$ deg, $\sigma_0 = 2$ deg, $p_0 = 0$ deg/s, $q_0 = 0$ deg/s and $r_0 = 0$ deg/s. The guidance commands are designed as $\alpha_c = (45 - 0.5t)$ deg, $\beta_c = 0$ deg and $\sigma_c = 0$ deg. The uncertainties of the inertia parameters, the aerodynamic parameters and the air density are each set to +20% bias. The external disturbance ΔD_e is given as [32]:

$$\Delta D_e = \begin{bmatrix} (0.5 + \cos(\pi t/2) + \sin(\pi t/3)) \times p \\ (0.5 + \cos(\pi t/2) + \sin(\pi t/4)) \times r \\ (0.5 + \cos(\pi t/3) + \sin(\pi t/2)) \times q \end{bmatrix} \times 10^4 (\text{N} \cdot \text{m}). \tag{52}$$

The saturation bound of the control torques is set as $\bar{M} = 1 \times 10^5$ (N · m). The attitude tracking errors are required to be less than 0.5 deg for $t \geq 2$ s. Therefore, the parameters of the FTPPF are selected as $\rho_{0i} = \pi/36$, $\rho_{\infty i} = \pi/360$ and $T = 2$. Comprehensively considering the satisfactory transient and steady-state performance, the parameters of the proposed control scheme are rigorously chosen as $c = 2I_{3\times3}$, $\omega = I_{3\times3}$, $Q = I_{3\times3}$, $\lambda_{1i} = 2$, $\lambda_{2i} = 0.1$. Inspired by the work of [12,14], the suitable basis vectors can be selected as polynomial combinations of the concerned state variables in the performance index (16). The base vectors of the actor and critic networks are adjusted in repeated trials to balance the approximation error and the computational burden. They are identically selected as:

$$\sigma[s(t)] = \left[\frac{s_\alpha^4}{4}, \frac{s_\alpha^3}{3}, \frac{s_\alpha^2}{2}, \frac{s_\beta^4}{4}, \frac{s_\beta^3}{3}, \frac{s_\beta^2}{2}, \frac{s_\sigma^4}{4}, \frac{s_\sigma^3}{3}, \frac{s_\sigma^2}{2}\right]^\top. \tag{53}$$

The initial values of $\hat{W}_a$ and $\hat{W}_c$ are randomly generated in $(0, 1)$. The user-defined parameters of the weight adaptation laws are designed as $A_1 = I_{9\times9}$, $A_2 = 0.1I_{9\times9}$, $p_c = 0.001$, $p_a = 1$, $\kappa = 0.001$.

Furthermore, in order to demonstrate the superiority of the proposed control scheme, the RL-based finite-time control (RLFTC) with input constraints in [17], and the robust adaptive backstepping control (RABC) method in [33], are implemented. To provide a fair

comparison, the time-fuel performance index of the RLFTC is identical to the proposed method, and the generation method for the initial values of the actor network and the critic network remains the same. The other user-defined parameters of the RLFTC are intensively selected as $T_1 = 0.1, \lambda = 2/3, \Gamma_c = 0.2, \Gamma_a = 0.2, k_a = 0.1$. The parameters of the RABC remain the same as that in [33]. The simulation results are given in Figures 2–10.

The tracking performances of the attitude angles under three controllers are shown in Figures 2–5. It can clearly be seen that the proposed control scheme can provide faster convergence and smaller overshoot in the presence of parametric uncertainties and external disturbances. The angular rates of RLV and the sliding manifolds are demonstrated in Figures 6 and 7, which provide further evidence for the improved performance of the proposed control scheme. Moreover, by comparing the tracking performance with the control inputs illustrated in Figure 8, it can be observed that the proposed control scheme exhibits better transient and steady-state performance with limited control inputs. The evolution trajectories of $\hat{W}_a$ and $\hat{W}_c$ are depicted in Figures 9 and 10, respectively. It can be readily found that $\hat{W}_a$ and $\hat{W}_c$ are convergent to the same values, which indicates that the ideal weight vector W^* can be effectively estimated via the proposed adaptation law.
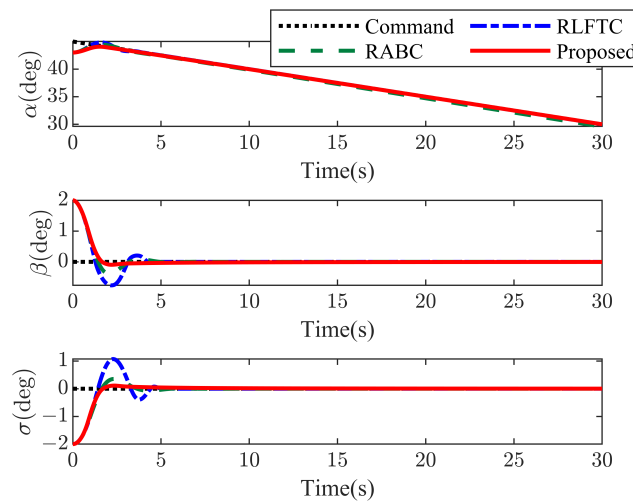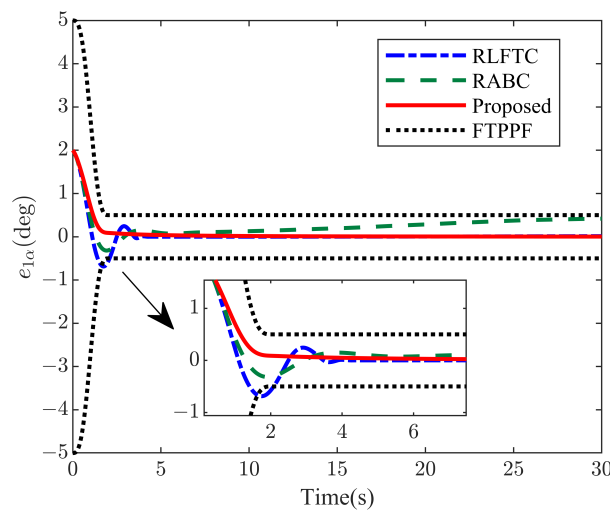

Figure 2. Time histories of the attitude angles.

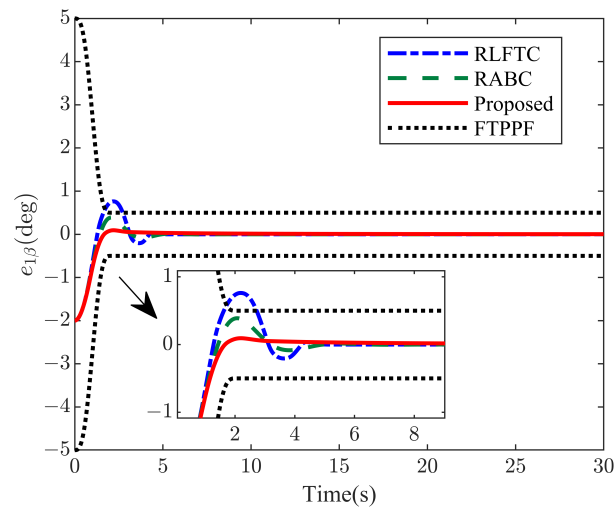**Figure 3.** Time histories of the attitude tracking error $e_{1\alpha}$.

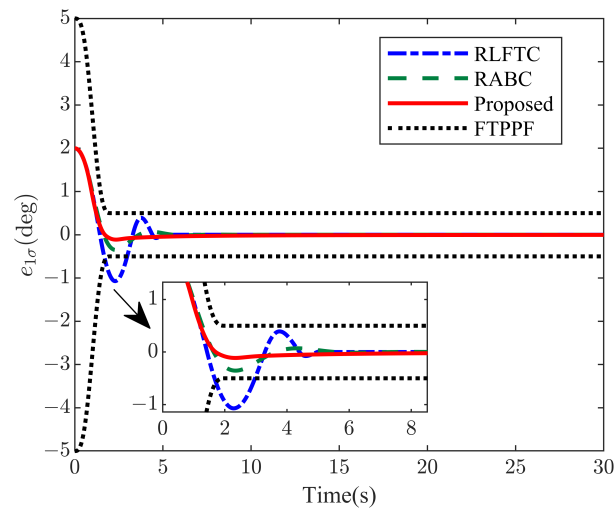**Figure 4.** Time histories of the attitude tracking error $e_{1\beta}$.

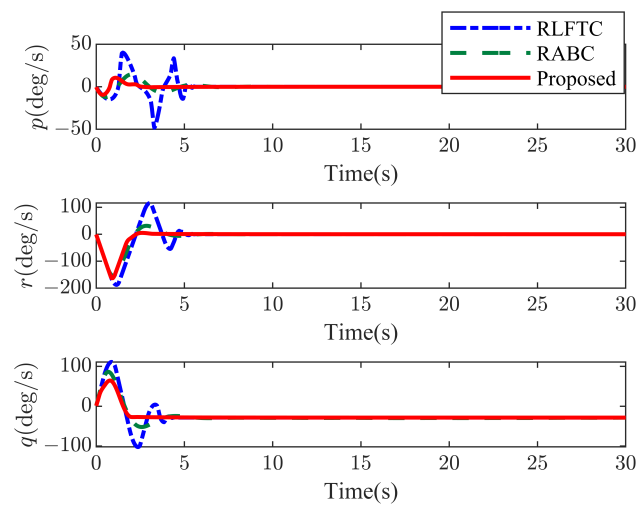**Figure 5.** Time histories of the attitude tracking error $e_{1\sigma}$.
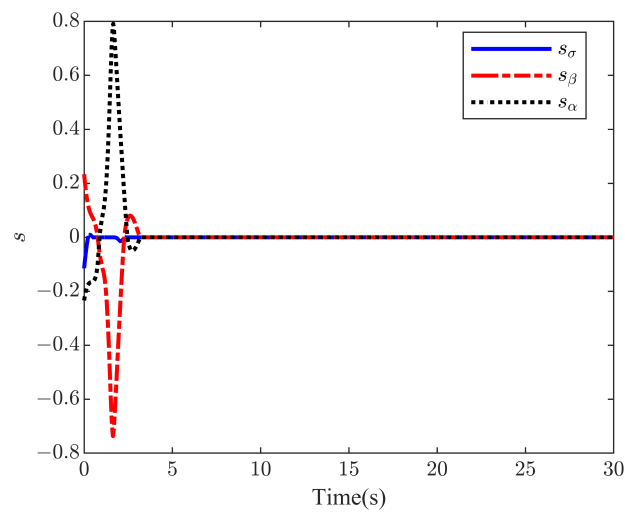

Figure 6. Time histories of the angular rates.
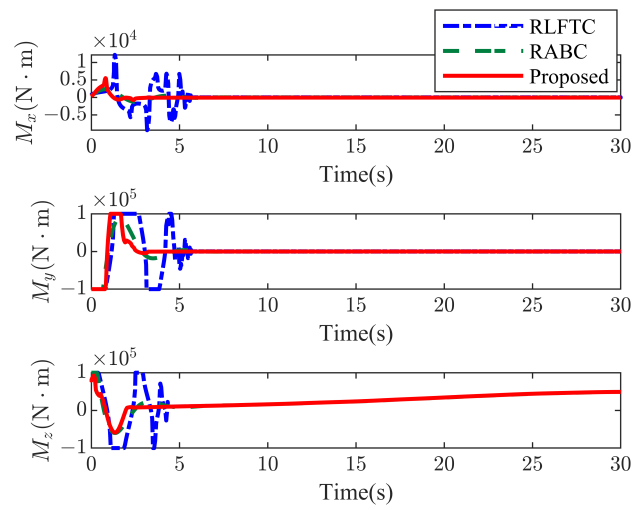
Figure 7. Time histories of the sliding manifolds.

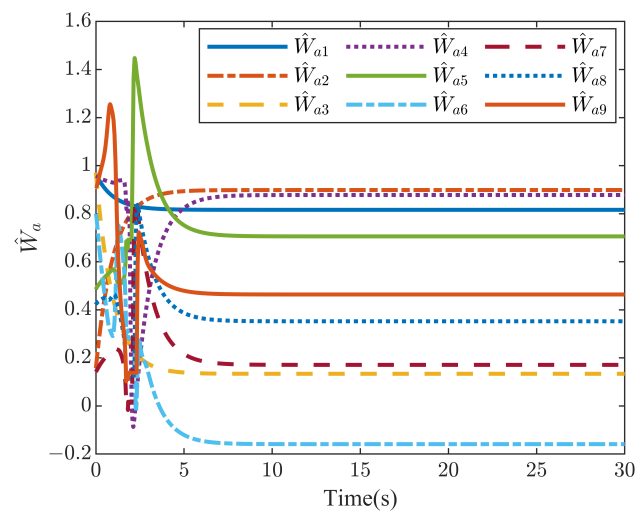**Figure 8.** Time histories of the control torques.

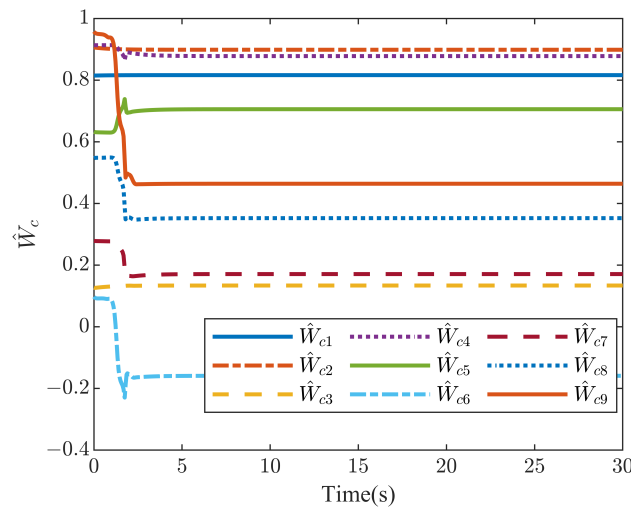**Figure 9.** Weights of the actor network.

Figure 10. Weights of the critic network.

To make a clear comparison, the maximum overshoot and the adjustment time are introduced to evaluate the transient performance of these controllers. Moreover, the integral absolute control effort (IACE) index $\int_0^{30}(|M_i|)d\tau, (i = x, y, z)$ and the integral of the time and absolute error (ITAE) $\int_0^{30} \tau(|e_{1i}|)d\tau, (i = \alpha, \beta, \sigma)$ index of three control schemes are calculated to evaluate the tracking accuracy and control effort. The performance indices of the three channels are summarized in Tables 1–3.

Table 1. Performance indexes of the α-channel.

Performance Index	Proposed Method	RABC	RLFTC
Maximum Overshoot	/	20.5%	34.0%
Adjustment Time [1] (s)	3.9	5.6	4.3
ITAE index (deg $\cdot$s^2)	1.6407	43.2600	1.6962
IACE index ($N \cdot m \cdot s$)	8.507×10^5	8.893×10^5	1.022×10^6

[1] The time when the attitude tracking error reaches to ± 0.05 deg.

Table 2. Performance indexes of the β-channel.

Performance Index	Proposed Method	RABC	RLFTC
Maximum Overshoot	4.5%	19.0%	38.2%
Adjustment Time (s)	3.1	4.1	4.6
ITAE index (deg $\cdot$s^2)	0.7347	1.5192	1.1507
IACE index ($N \cdot m \cdot s$)	1.8365×10^5	2.0571×10^5	4.155×10^5

Table 3. Performance indexes of the σ-channel.

Performance Index	Proposed Method	RABC	RLFTC
Maximum Overshoot	5.5%	17.7%	53.5%
Adjustment Time (s)	4.1	4.8	4.7
ITAE index (deg $\cdot$s^2)	0.8783	1.7853	1.7242
IACE index ($N \cdot m \cdot s$)	4.6577×10^3	5.0277×10^3	1.8364×10^4

From the foregoing simulation results, it can be concluded that the proposed control scheme outperforms RLFTC and RABC in terms of transient performance, tracking accuracy and control effort. Furthermore, by synthesizing the AC structure and the fixed-time PPC paradigm, the proposed control scheme offers an online RL-based model-free solution for controlling RLV and other complex industrial systems.

5. Conclusions

In this paper, an innovative RL-based tracking control scheme with fixed-time prescribed performance has been proposed for RLV under parametric uncertainties, external disturbances, and input constraints. By resorting to the FTPPF, fixed-time performance envelopes have been imposed on the attitude tracking errors. Combined with the AC-based online RL structure and the super-twisting-like sliding mode control, the optimal control policy and the performance index have been learned recursively, and the robustness of the learning process has been further enhanced. Moreover, theoretical analysis has demonstrated that the attitude tracking error can converge to a preset region within a preassigned fixed time, and that the sliding variable, the weight estimation errors of the actor and critic networks are UUB. Comparative simulation results have verified the effectiveness and improved performance of the proposed control scheme. The angular rate constraints will be addressed in our future work, and the optimal control problem for the underactuated RLV will be specifically addressed. Experimental investigations, such as hardware-in-the-loop simulations, will be undertaken.

Author Contributions: Conceptualization, S.X. and Y.G.; methodology, S.X.; software, S.X.; validation, Y.G., C.W. and Y.L.; formal analysis, S.X.; investigation, S.X.; resources, C.W.; data curation, Y.G.; writing—original draft preparation, S.X.; writing—review and editing, S.X., Y.G., C.W., Y.L. and L.X.; visualization, S.X., Y.G. and L.X.; supervision, C.W.; project administration, C.W.; funding acquisition, C.W. All authors have read and agreed to the published version of the manuscript.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

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

Abbreviations

The following abbreviations are used in this manuscript:

RLV	Reusable Launch Vehicles
NN	Neural Network
RL	Reinforcement Learning
AC	Actor-Critic
UUB	Uniformly Ultimately Bounded
HJB	Hamilton–Jacobi–Bellman
PPC	Prescribed Performance Control
FTPPF	Fixed-time Prescribed Performance Function
RLFTC	Reinforcement-Learning-Based Finite-Time Control
RABC	Robust Adaptive Backstepping Control
IACE	Integral Absolute Control Effort
ITAE	Integral of Time and Absolute Error

References

1. Stott, J.E.; Shtessel, Y.B. Launch vehicle attitude control using sliding mode control and observation techniques. *J. Frankl. Inst. B* **2012**, *349*, 397–412.
2. Tian, B.L.; Li, Z.Y.; Zhao, X.P.; Zong, Q. Adaptive Multivariable Reentry Attitude Control of RLV With Prescribed Performance. *IEEE Trans. Syst. Man Cybern. Syst.* **2022**, 1–5. doi:10.1109/TSMC.2022.3145508
3. Acquatella, P.; Briese, L.E.; Schnepper, K. Guidance command generation and nonlinear dynamic inversion control for reusable launch vehicles. *Acta Astronaut.* **2020**, *174*, 334–346.
4. Xu, B.; Wang, X.; Shi, Z.K. Robust adaptive neural control of nonminimum phase hypersonic vehicle model. *IEEE Trans. Syst. Man Cybern. Syst.* **2019**, *51*, 1107–1115.

5. Zhang, L.; Wei, C.Z.; Wu, R.; Cui, N.G. Fixed-time extended state observer based non-singular fast terminal sliding mode control for a VTVL reusable launch vehicle. *Aerosp. Sci. Technol.* **2018**, *82*, 70–79.

6. Ju, X.Z.; Wei, C.Z.; Xu, H.C.; Wang, F. Fractional-order sliding mode control with a predefined-time observer for VTVL reusable launch vehicles under actuator faults and saturation constraints. *ISA Trans.* **2022**, *in press.*

7. Cheng, L.; Wang, Z.B.; Gong, S.P. Adaptive control of hypersonic vehicles with unknown dynamics based on dual network architecture. *Acta Astronaut.* **2022**, *193*, 197–208.

8. Xu, B.; Shou, Y.X.; Shi, Z.K.; Yan, T. Predefined-Time Hierarchical Coordinated Neural Control for Hypersonic Reentry Vehicle. *IEEE Trans. Neural Netw. Learn. Syst.* **2022**. doi:10.1109/TNNLS.2022.3151198.

9. Werbos, P. Approximate dynamic programming for realtime control and neural modelling. In *Handbook of Intelligent Control: Neural, Fuzzy and Adaptive Approaches*; Van Nostrand Reinhold: New York, NY, USA, 1992; pp. 493–525.

10. Vamvoudakis, K.G.; Lewis, F.L. Online actor–critic algorithm to solve the continuous-time infinite horizon optimal control problem. *Automatica* **2010**, *46*, 878–888.

11. He, H.B.; Ni, Z.; Fu, J. A three-network architecture for on-line learning and optimization based on adaptive dynamic programming. *Neurocomputing* **2012**, *78*, 3–13.

12. Ma, Z.Q.; Huang, P.F.; Lin, Y.X. Learning-based Sliding Mode Control for Underactuated Deployment of Tethered Space Robot with Limited Input. *IEEE Trans. Aerosp. Electron. Syst.* **2021**, *58*, 2026–2038.

13. Wang, N.; Gao, Y.; Zhao, H.; Ahn, C.K. Reinforcement learning-based optimal tracking control of an unknown unmanned surface vehicle. *IEEE Trans. Neural Netw. Learn. Syst.* **2020**, *32*, 3034–3045.

14. Fan, Q.Y.; Yang, G.H. Adaptive actor–critic design-based integral sliding-mode control for partially unknown nonlinear systems with input disturbances. *IEEE Trans. Neural Netw. Learn. Syst.* **2015**, *27*, 165–177.

15. Kuang, Z.; Gao, H.; Tomizuka, M. Precise linear-motor synchronization control via cross-coupled second-order discrete-time fractional-order sliding mode. *IEEE/ASME Trans. Mechatronics* **2020**, *26*, 358–368.

16. Zhang, H.; Cui, X.; Luo, Y.; Jiang, H. Finite-horizon H_∞ tracking control for unknown nonlinear systems with saturating actuators. *IEEE Trans. Neural Netw. Learn. Syst.* **2017**, *29*, 1200–1212.

17. Wang, N.; Gao, Y.; Yang, C.; Zhang, X.F. Reinforcement learning-based finite-time tracking control of an unknown unmanned surface vehicle with input constraints. *Neurocomputing* **2022**, *484*, 26–37.

18. Bechlioulis, C.P.; Rovithakis, G.A. Robust adaptive control of feedback linearizable MIMO nonlinear systems with prescribed performance. *IEEE Trans. Autom. Control* **2008**, *53*, 2090–2099.

19. Cui, G.Z.; Yang, W.; Yu, J.P.; Li, Z.; Tao, C.B. Fixed-time prescribed performance adaptive trajectory tracking control for a QUAV. *IEEE Trans. Circuits Syst. II* **2021**, *69*, 494–498.

20. Bu, X.W. Guaranteeing prescribed performance for air-breathing hypersonic vehicles via an adaptive non-affine tracking controller. *Acta Astronaut.* **2018**, *151*, 368–379.

21. Wang, N.; Gao, Y.; Zhang, X.F. Data-driven performance-prescribed reinforcement learning control of an unmanned surface vehicle. *IEEE Trans. Neural Netw. Learn. Syst.* **2021**, *32*, 5456–5467.

22. Luo, S.B.; Wu, X.; Wei, C.S.; Zhang, Y.L.; Yang, Z. Adaptive finite-time prescribed performance attitude tracking control for reusable launch vehicle during reentry phase: An event-triggered case. *Adv. Space Res.* **2022**, *69*, 3814–3827.

23. Modares, H.; Sistani, M.B.N.; Lewis, F.L. A policy iteration approach to online optimal control of continuous-time constrained-input systems. *ISA Trans.* **2013**, *52*, 611–621.

24. Tan, J.; Guo, S.J. Backstepping control with fixed-time prescribed performance for fixed wing UAV under model uncertainties and external disturbances. *Int. J. Control* **2022**, *95*, 934–951.

25. Yuan, Y.; Wang, Z.; Guo, L.; Liu, H.P. Barrier Lyapunov functions-based adaptive fault tolerant control for flexible hypersonic flight vehicles with full state constraints. *IEEE Trans. Syst. Man Cybern. Syst.* **2018**, *50*, 3391–3400.

26. Lyshevski, S.E. Optimal control of nonlinear continuous-time systems: Design of bounded controllers via generalized non-quadratic functionals. In Proceedings of the 1998 American Control Conference. ACC (IEEE Cat. No. 98CH36207), Philadelphia, PA, USA, 26–26 June 1998; Volume 1, pp. 205–209.

27. Hornik, K.; Stinchcombe, M.; White, H. Universal approximation of an unknown mapping and its derivatives using multilayer feedforward networks. *Neural Netw.* **1990**, *3*, 551–560.

28. Kamalapurkar, R.; Walters, P.; Dixon, W.E. Model-based reinforcement learning for approximate optimal regulation. *Automatica* **2016**, *64*, 94–104.

29. Bhasin, S.; Kamalapurkar, R.; Johnson, M.; Vamvoudakis, K.G.; Lewis, F.L.; Dixon, W.E. A novel actor–critic–identifier architecture for approximate optimal control of uncertain nonlinear systems. *Automatica* **2013**, *49*, 82–92.

30. Moreno, J.A.; Osorio, M. Strict Lyapunov functions for the super-twisting algorithm. *IEEE Trans. Autom. Control* **2012**, *57*, 1035–1040.

31. Wei, C.Z.; Wang, M.Z.; Lu, B.G.; Pu, J.L. Accelerated Landweber iteration based control allocation for fault tolerant control of reusable launch vehicle. *Chin. J. Aeronaut.* **2022**, *35*, 175–184.

32. Zhang, C.F.; Zhang, G.S.; Dong, Q. Fixed-time disturbance observer-based nearly optimal control for reusable launch vehicle with input constraints. *ISA Trans.* **2022**, *122*, 182–197.

33. Wang, Z.; Wu, Z.; Du, Y. Robust adaptive backstepping control for reentry reusable launch vehicles. *Acta Astronaut.* **2016**, *126*, 258–264.