AIMS Mathematics, 11(3): 6297–6328.
DOI: 10.3934/math.2026260
Received: 18 December 2025
Revised: 03 March 2026
Accepted: 05 March 2026
Published: 12 March 2026

https://www.aimspress.com/journal/Math

Research article

Attitude tracking control of quadrotor unmanned aerial vehicle based on an adaptive predefined time value iteration approach

Zhixin Feng and Renming Yang*

School of Information Science and Electrical Engineering, Shandong Jiaotong University, Jinan, 250357, Shandong, China

* **Correspondence:** Email: renmingyang0222@163.com.

Abstract: This paper addresses the attitude tracking problem of quadrotor unmanned aerial vehicles (UAVs) by introducing a novel predefined-time stable adaptive value iteration (PTS-AVI) control scheme. Unlike conventional value iteration-based adaptive dynamic programming (VI-ADP) methods—where the running cost is independent of the value function—the proposed cost function explicitly incorporates the value function V, thereby ensuring predefined-time stability (PTS) throughout both the training and deployment phases. This functional dependency was systematically addressed via an auxiliary time-scale partial differential equation (PDE) formulated in the s-domain. To circumvent the need to solve the complex Hamilton–Jacobi–Bellman (HJB) equation directly, a parameter update method was employed to approximate its solution, yielding an optimal control policy that satisfies predefined-time stability criteria. In addition, a high-precision predefined-time disturbance observer was designed to estimate and compensate for unknown disturbances. Both theoretical analysis and simulation results confirmed that the proposed control scheme guarantees state convergence to equilibrium within a user-specified time, regardless of initial conditions. This work differs from existing studies by integrating predefined-time stability requirements into the VI-ADP formulation and coupling it with a predefined-time disturbance observer for attitude tracking.

Keywords: predefined-time stable; adaptive value iteration; unmanned aerial vehicle; adaptive dynamic programming; disturbance observer
Mathematics Subject Classification: 93C10, 93C40, 93D05, 93D21

1. Introduction

Unmanned aerial vehicles (UAVs) have been widely studied due to their broad applications and the stringent requirements on safety and maneuverability. For example, [1–3] focused on disturbance-related issues in UAV systems and employed techniques such as event-triggered control,

fuzzy control, and disturbance observers to address underactuation and strong coupling characteristics. [4–7] achieved faster convergence and more accurate disturbance compensation by developing finite-time observers and sliding mode control strategies. Furthermore, [8–10] applied fixed-time stability theory to investigate UAV stabilization problems with guaranteed fixed convergence time. Among various modules, attitude tracking is a fundamental inner-loop task that directly determines the flight stability and tracking accuracy. Existing attitude controllers can be broadly categorized by their convergence properties, including asymptotic, finite-time, fixed-time, and predefined-time schemes. This paper focuses on predefined-time attitude tracking control, where a user-specified settling-time upper bound is explicitly enforced.

It should be noted that infinite-time stability control, while straightforward to analyze and implement, does not guarantee a bounded convergence time. Finite-time stability control [11–13] improves upon this by ensuring convergence within a finite duration, yet the settling time remains highly dependent on initial conditions, thereby limiting its applicability. To overcome this limitation, fixed-time stability control has been introduced [14–17], which eliminates dependence on initial conditions. However, the convergence time in such schemes is governed by multiple control parameters whose relationships are often implicit and non-intuitive. To achieve more predictable and tunable convergence behavior, predefined-time stability theory has recently been developed [18–21], allowing the user to directly specify the convergence time as a control parameter. Predefined-time control explicitly prescribes the settling-time upper bound through design parameters, which is attractive for time-critical UAV attitude maneuvers. This enables the UAV to achieve rapid attitude regulation as much as possible, thereby improving operational efficiency and enhancing safety during flight. Prescribed performance control enforces error evolution within predefined envelopes (transient/steady-state bounds) and can be viewed as complementary, focusing on constraint satisfaction over time, whereas predefined-time focuses on a hard convergence deadline.

In addition to the above time-domain stability formulations, it is common to integrate them with other control methodologies to further enhance system performance. Optimal control represents one such approach, which designs control strategies by optimizing performance metrics—such as minimal time or energy consumption—to achieve theoretically optimal system behavior. The adaptation of optimal control for nonlinear systems has long been a fundamental topic in control theory. In recent years, several theoretical results on reinforcement learning with finite-time stability have been reported [22, 23]. Studies [24, 25] have investigated fixed-time optimal control problems, while [26, 27] focused on predefined-time optimal control for nonlinear systems. In contrast, this paper analyzes nonlinear systems subject to disturbances by integrating an improved predefined-time stability theory with reinforcement learning techniques. In this context, adaptive dynamic programming (ADP) and value iteration (VI) methods have been extensively studied to approximately solve the Hamilton-Jacobi-Bellman (HJB) equation for continuous-time nonlinear systems. Classical ADP methods generally rely on infinite-horizon cost functions to ensure asymptotic stability and optimality. Recently, [28] revisited the adaptive optimal control problem and proposed a class of continuous-time nonlinear VI methods, overcoming the limitation that VI could not be directly applied to continuous systems. Nevertheless, their work does not impose any bound on the convergence time, nor does it account for the influence of external disturbances. While predefined-time control and optimal control have been studied separately, their integration with value iteration-based optimal control under disturbances remains relatively unexplored, especially for

quadrotor UAV systems.

To enhance the transient response performance of quadrotor UAVs under uncertain conditions, high-performance disturbance observers have become an active research topic. For instance, existing studies have combined fast nonsingular sliding mode control with adaptive predefined-time disturbance observers for quadrotor UAVs, and experimental results have demonstrated their advantages in rapid disturbance estimation and predefined-time transient performance [29]. Moreover, disturbance observers based on fast fixed-time distributed neural formation control have been developed for multirotor systems, where fast fixed-time dynamics are incorporated to improve convergence speed in the presence of unknown disturbances and networked environments [30]. [31–33] employed sliding mode control strategies to address disturbance rejection in UAV systems. In addition, [34–36] also presented a number of effective theoretical results for addressing disturbance effects in nonlinear systems as well as UAV systems. In optimal control problems, many effective disturbance-handling methods and performance guarantees have also been reported in [37–39]. Distinct from these observer-oriented sliding mode or formation control frameworks, this study embeds a predefined-time disturbance observer into a value-iteration-based optimal attitude tracking architecture and investigates the coupled learning-control behavior. In doing so, the proposed approach preserves optimal control objectives while enabling the system to achieve a user-specified upper bound on the settling time.

Although the aforementioned studies have integrated finite-time, fixed-time, and predefined-time stability concepts with optimal control, incorporating predefined-time stability criteria into the VI-ADP framework under unknown disturbances remains a significant research gap. Furthermore, ensuring optimal transient performance under uncertain conditions requires high-precision observers. Predefined-time and fast fixed-time disturbance observers have been developed to quickly estimate perturbations. However, these are predominantly coupled with sliding mode control or neural formation frameworks. While effective in disturbance rejection, sliding mode methods often introduce chattering, and formation frameworks do not inherently optimize the internal control cost. On the other hand, disturbance compensation mechanisms within optimal control have been reported. Distinct from these existing studies, embedding a predefined-time disturbance observer directly into a value-iteration-based optimal attitude tracking architecture remains uninvestigated.

Motivated by this gap, this paper extends the VI-ADP framework to a disturbed UAV model and integrates it with predefined-time stability criteria to address the attitude tracking problem of quadrotor UAVs. First, the dynamic model of the quadrotor UAV is established and a predefined-time disturbance observer is designed. Then, an optimal control formulation is developed, incorporating the predefined-time stability criterion into the cost function, which leads to the derivation of predefined-time stability theorems for both the training and software deployment phases. Moreover, radial basis function neural networks (RBF-NNs) are employed to approximate the value function and the Hamiltonian, thereby circumventing the need to solve the complex HJB equation directly while still attaining a near-optimal solution. Simulation results demonstrate that the proposed PTS-AVI-based quadrotor controller achieves significantly faster convergence and improved disturbance suppression capability, rendering it highly suitable for security-critical and time-sensitive UAV applications.

The main contributions of this paper are summarized as follows:

1) Unlike the traditional VI-ADP methods, where the cost function is independent of the value

function, this paper ingeniously embeds the value function V and the predefined-time stability criterion directly into the cost function, presenting a novel adaptive value iteration framework with predefined-time stability. This design simplifies the stability proof and fundamentally ensures that the optimal control strategy can make the system state converge within the user-specified time, and the convergence time is independent of the initial state.

2) We introduce a predefined-time term based on the traditional HJB equation, and the radial basis function neural network is used to approximate the value function and Hamiltonian function, avoiding the direct solution of complex HJB equations, thus expanding the application of optimal control theory in time-constrained systems.

3) Unlike fixed-time disturbance observers [17], for unknown disturbances in the system, a novel disturbance observer is designed, where its observation error can converge to zero within a predefined time. This observer can quickly and accurately estimate disturbances and effectively counteract their impact through feedforward compensation, significantly enhancing system robustness.

4) We provide complete theoretical guarantees for both the training and deployment phases, establishing a complete continuous-time value iteration theory for cost functions that depends on value functions, and proves the predefined-time stability of the system during both the training and software deployment phases.

The rest of this paper is organized as follows:

Section 2 introduces preliminary knowledge. In Section 3, we propose several main results. Section 4 gives the emulation outcome. Section 5 summarizes this paper and discusses future research directions.

Note: The real-valued function f belongs to the class $C^k(Q)$, where $k \in Z_+$, and f is continuous when $k = 0$, or continuously differentiable on Q when $k \geq 1$. $\mathcal{P} = \left\{V \in C^1(\mathbb{R}^n) \mid V \text{ is positive semi-definite}\right\}$ and $\mathcal{P}_+ = \{V \in \mathcal{P} \mid V \text{ is positive definite}\}$.

2. Materials and methods

The attitude kinematic equation of the UAV can be expressed as [10]

$$\ddot{\varphi} = \frac{-J_r\varpi_r\dot{\theta} + \left(I_{yy} - I_{zz}\right)\dot{\theta}\dot{\psi}}{I_{xx}} + \frac{l}{I_{xx}}u_\varphi + d_\varphi \tag{2.1}$$

$$\ddot{\theta} = \frac{-J_r\varpi_r\dot{\varphi} + \left(I_{zz} - I_{xx}\right)\dot{\varphi}\dot{\psi}}{I_{yy}} + \frac{l}{I_{yy}}u_\theta + d_\theta \tag{2.2}$$

$$\ddot{\psi} = \frac{\left(I_{xx} - I_{yy}\right)\dot{\varphi}\dot{\theta}}{I_{zz}} + \frac{1}{I_{zz}}u_\psi + d_\psi, \tag{2.3}$$

where $l \in \mathbb{R}$ represents the distance from each rotor to the center of mass of the UAV, I_{xx} , I_{yy} , $I_{zz} \in \mathbb{R}$ are the three-axis moments of inertia of the body, respectively, $u_\varphi, u_\theta, u_\psi$ are the control inputs for the roll angle φ, pitch angle θ, and yaw angle ψ, respectively, $d_\varphi, d_\theta, d_\psi$ are the disturbances acting on the roll angle φ, pitch angle θ, and yaw angle ψ, respectively, $J_r \in R$ is the total moment of inertia of the propeller and motor rotor, and $\varpi_r = \varpi_1 + \varpi_2 - \varpi_3 - \varpi_4$ represents the rotor angular velocity.

As illustrated in Figure 1, two coordinate frames are used in the quadrotor modeling: The earth-fixed inertial frame $\{E\} = (O_e, x_e, y_e, z_e)$, and the body-fixed frame $\{B\} = (O_b, x_b, y_b, z_b)$ are attached to

the UAV at its center of mass O_b. The attitude is described by the roll-pitch-yaw Euler angles (ϕ, θ, ψ), which parameterize the rotation from $\{B\}$ to $\{E\}$ through the rotation matrix $R_E^B(\phi, \theta, \psi)$. The total thrust is generated along the body z_b axis by the four rotors, with individual thrusts f_i and reaction moments M_i $(i = 1, \ldots, 4)$, while gravity G acts along the inertial z_e axis. Unless otherwise stated, translational quantities (position/velocity) are expressed in $\{E\}$, whereas angular velocity and control torques are expressed in $\{B\}$. Set $\eta = [\varphi, \theta, \psi]^T$, $u = [u_\varphi, u_\theta, u_\psi]^T$, $G = \left[\frac{l}{I_{xx}}, \frac{l}{I_{yy}}, \frac{1}{I_{zz}}\right]^T$, $d = [d_\varphi, d_\theta, d_\psi]^T$. Then the UAV system can be simplified as:

$$\ddot{\eta} = f(\dot{\eta}) + Gu + d. \tag{2.4}$$

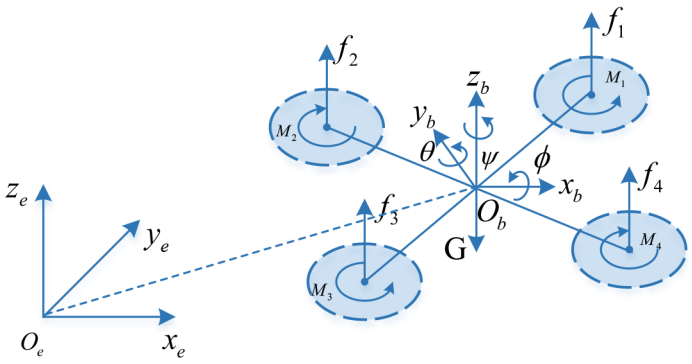

Figure 1. Schematic of the UAV coordinate system.

Assumption 2.1. *Assume that the external time-varying disturbance d of the quadrotor UAV system is bounded satisfying* $\|d\| < d_1$, *where* d_1 *is a positive constant.*

Remark 2.1. *This assumption primarily states that the disturbances in the UAV attitude control system are bounded, which is a general assumption [4].*

The general research process for optimal control problems can be described as follows. Considering the following nonlinear system:

$$\dot{x} = f(x) + gu + d =: f_x(x, u, d, t), \tag{2.5}$$

where $x \in \mathbb{R}^n$ is the state of the system, $f(x) \in \mathbb{R}^n$ is the nonlinear team of x, $u \in \mathbb{R}^n$ is the control input, and $d \in \mathbb{R}^n$ is the lump disturbance.

In the optimal control problem, the performance metrics function of optimal control decides the optimal performance, and the performance metrics function is given as

$$\mathcal{J}(x_0; u) = \int_0^\infty r(x, u)dt. \tag{2.6}$$

Taking advantage of the performance metrics function, we can find the optimal controller and its association function:

$$V^*(x_0) = \inf_u \mathcal{J}(x_0; u), \quad u^* = \arg\inf_u \mathcal{J}(x_0; u). \tag{2.7}$$

Noting that directly solving the above equation is quite difficult. To achieve this destination, we take advantage of the continuous-time VI method.

The continuous-time VI method relies on the HJB equation, and the HJB equation is described as

$$\inf_{u\in\mathbb{R}^m} H\left(x, \partial_x V(x), u\right) = 0, \quad V(0) = 0, \tag{2.8}$$

where the Hamiltonian function H is defined as:

$$H(x, p, u) = p f_x(x, u, d, t) + r(x, u). \tag{2.9}$$

Now, we present an assumption.

Assumption 2.2. *There exists an unique continuous function $\alpha_1 : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}^m$ such that*

$$\alpha_1(x, p) = \arg\inf_{u\in\mathbb{R}^m} H(x, p, u). \tag{2.10}$$

The HJB equation provides the necessary conditions for the optimal value function, and Assumption 2.2 ensures the existence of optimal control, which is a common assumption. The combination of the two allows optimal control to be expressed as a state-feedback policy:

$$u^*(t) = \mu^*(x(t)) \tag{2.11}$$

$$\mu^*(x) = \arg\inf H\left(x, \partial_x V^*(x), u\right), \quad \forall x \in \mathbb{R}^n. \tag{2.12}$$

Based on the above analysis, rather than solving the cost function directly to obtain the optimal policy, the optimal policy is derived via the HJB equation, and Assumption 2.2 guarantees the existence and uniqueness of the minimizer of the Hamiltonian with respect to the control input, which is required to define a well-posed optimal feedback policy (2.11) [28]. In light of these optimization principles and the challenges associated with directly solving the HJB equation, the following section analyzes the value iteration (VI) method as an alternative approach to solving the HJB equation. The corresponding results are presented below.

3. Results

In the section, to counteract the unknown disturbances in the system, a predefined-time observer is developed to accurately estimate these disturbances in the UAV model. This estimated disturbance is incorporated into the control law as a compensatory term, thereby reducing its negative impact on system stability and settling time.

In the following, we first give the definition of being predefined-time stable.

Definition 3.1. *(See [20]) The origin of system* (2.5) *is predefined-time stable if it is fixed-time stable and the settling time $T_c^{'}$ is independent of any system parameters and initial states or can be arbitrarily chosen in advance.*

Lemma 3.1. *For the system $\dot{x} = \phi(x, t)$, if there exists a radially unbounded Lyapunov function $V(x)$ such that*

$$\dot{V} \leq -\frac{2}{\alpha\beta T_c^{'}}(2\beta V + V^{1-\frac{\alpha}{2}} + \beta^2 V^{1+\frac{\alpha}{2}}) \tag{3.1}$$

holds for any solution $x(t, x_0)$ of the system $\dot{x} = \phi(x, t)$, then the trajectory of the system is predefined-time stable, where $T_c^{'} > 0$ is a predefined-time constant, $\alpha \in (0, 1)$, and $\beta > 0$ is an assignable parameter.

Now, inspired by the fixed-time disturbance observer constructed in [17], we develop a predefined-time disturbance observer result.

Theorem 3.1. *Consider the dynamical system* (2.4). *By constructing the following predefined-time disturbance observer, the observation error of the time-varying disturbance converges within the predefined time* $T_c^{'}$*:*

$$\hat{d} = \lambda \hat{x}_e + \dot{x}_e. \tag{3.2}$$

Here, $\hat{d}$ *denotes the disturbance estimate,* $x_e = \dot{\eta} - \dot{\eta}_a$*, where* $\dot{\eta}_a$ *is the state of the auxiliary function, the observer gain satisfies* $\lambda > 0$*, and* $\hat{x}_e$ *is the estimate of the state error* x_e*. Its derivative is defined as follows:*

$$\dot{\hat{x}}_e = \dot{x}_e + \frac{2}{\alpha\beta T_c^{'}}\Big(\beta sign(\tilde{x}_e) + 2^{-\frac{2-\alpha}{2}} sign(\tilde{x}_e)^{1-\alpha} + \beta^2 2^{-\frac{2+\alpha}{2}} sign(\tilde{x}_e)^{1+\alpha}\Big) + \tilde{x}_e. \tag{3.3}$$

The estimation error of x_e *is defined by* $\tilde{x}_e = x_e - \hat{x}_e$*.*

Proof. Construct the following auxiliary function:

$$\ddot{\eta}_a = f(\dot{\eta}) + Gu + \lambda x_e. \tag{3.4}$$

Choose the Lyapunov function as:

$$V = \frac{1}{2}\tilde{x}_e^T \tilde{x}_e. \tag{3.5}$$

Taking the time derivative of (3.5) and invoking (3.3), we obtain:

$$\begin{aligned}
\dot{V} &= -\tilde{x}_e^T \frac{2}{\alpha\beta T_c^{'}}\Big(\beta sgn(\tilde{x}_e) + 2^{-\frac{2-\alpha}{2}} sgn(\tilde{x}_e)^{1-\alpha} + \beta^2 2^{-\frac{2+\alpha}{2}} sgn(\tilde{x}_e)^{1+\alpha}\Big) - \tilde{x}_e^T \tilde{x}_e \\
&\leq -\frac{2}{\alpha\beta T_c^{'}}\Big(\beta\|\tilde{x}_e\|^2 + 2^{-\frac{2-\alpha}{2}}\|\tilde{x}_e\|^{2-\alpha} + \beta^2 2^{-\frac{2+\alpha}{2}}\|\tilde{x}_e\|^{2+\alpha}\Big) \\
&\leq -\frac{2}{\alpha\beta T_c^{'}}\Big(\beta\|\tilde{x}_e\|^2 + 2^{-\frac{2-\alpha}{2}}\|\tilde{x}_e\|^{2*\frac{2-\alpha}{2}} + \beta^2 2^{-\frac{2+\alpha}{2}}\|\tilde{x}_e\|^{2*\frac{2+\alpha}{2}}\Big) \\
&\leq -\frac{2}{\alpha\beta T_c^{'}}\Big(2\beta\frac{1}{2}\|\tilde{x}_e\|^2 + (\frac{1}{2}\|\tilde{x}_e\|^2)^{\frac{2-\alpha}{2}} + \beta^2(\frac{1}{2}\|\tilde{x}_e\|^2)^{\frac{2+\alpha}{2}}\Big) \\
&\leq -\frac{2}{\alpha\beta T_c^{'}}\Big(2\beta V + V^{1-\frac{\alpha}{2}} + \beta^2 V^{1+\frac{\alpha}{2}}\Big),
\end{aligned} \tag{3.6}$$

where $sgn(\tilde{x}_e) = sign(\tilde{x}_e) * \tilde{x}_e$.

Hence, the above inequality matches the predefined-time convergence form in Lemma 3.1, implying that the error between the true and estimated states decays to zero before the predefined time $T_c^{'}$. Consequently, we have:

$$\begin{aligned}
\tilde{d} &= d - \hat{d} \\
&= d - \lambda\hat{x}_e - \ddot{\eta} + \ddot{\eta}_a \\
&= d - \lambda\hat{x}_e - (f(\dot{\eta}) + Gu + d) + f(\dot{\eta}) + Gu + \lambda x_e \\
&= \lambda\tilde{x}_e.
\end{aligned} \tag{3.7}$$

When $\tilde{x}_e = 0$, it follows immediately that $\tilde{d} = d - \hat{d} = 0$. Therefore, the disturbance estimate $\hat{d}$ accurately tracks the true disturbance. The proof is complete. □

Let the control law be $u = (u_{01} - G^{-1}\hat{d})$, and then the UAV system can be rewritten as

$$\ddot{\eta} = f(\dot{\eta}) + Gu_{01} + e_d, \tag{3.8}$$

where the error of disturbance $e_d = d - \hat{d}$.

Consider the attitude tracking problem. Set the desired roll angle, pitch angle, and yaw angle as $\eta_d = [\varphi_d, \theta_d, \psi_d]^T$. Define the variables $e = \eta - \eta_d \in \mathbb{R}^3$, $x = [x_1, x_2, x_3]^T = \dot{e} = \dot{\eta} - \dot{\eta}_d \in \mathbb{R}^3$. Then set $z = \begin{bmatrix} e \\ x \end{bmatrix} \in \mathbb{R}^6$ and

$$\dot{z} = \begin{bmatrix} \dot{e} \\ \dot{x} \end{bmatrix} = \begin{bmatrix} x \\ f_f(x + \dot{\eta}_d) - \ddot{\eta}_d \end{bmatrix} + \begin{bmatrix} 0_{3\times3}, 0_{3\times3} \\ 0_{3\times3}, G \end{bmatrix} u_0 + \begin{bmatrix} 0_{3\times3} \\ I_3 \end{bmatrix} e_d, \tag{3.9}$$

where $f_f(x + \dot{\eta}_d) = [f_{f1}(x + \dot{\eta}_d), f_{f2}(x + \dot{\eta}_d), f_{f3}(x + \dot{\eta}_d)]^T$ is given as

$$\begin{aligned} f_{f1}(x + \dot{\eta}_d) &= \frac{1}{I_{xx}}\left[\left(I_{yy} - I_{zz}\right)\left(x_2 + \dot{\theta}_d\right)\left(x_3 + \dot{\psi}_d\right) - J_r\varpi_r\left(x_2 + \dot{\theta}_d\right)\right] \\ &= \frac{1}{I_{xx}}\left[\left(I_{yy} - I_{zz}\right)\left(x_2x_3 + x_2\dot{\psi}_d + \dot{\theta}_dx_3 + \dot{\theta}_d\dot{\psi}_d\right)\right. \\ &\quad \left. - J_r\Omega_rx_2 - J_r\varpi_r\dot{\theta}_d\right]. \end{aligned} \tag{3.10}$$

$$\begin{aligned} f_{f2}(x + \dot{\eta}_d) &= \frac{1}{I_{yy}}\left[(I_{zz} - I_{xx})\left(x_1 + \dot{\phi}_d\right)\left(x_3 + \dot{\psi}_d\right) - J_r\varpi_r\left(x_1 + \dot{\phi}_d\right)\right] \\ &= \frac{1}{I_{yy}}\left[(I_{zz} - I_{xx})\left(x_1x_3 + x_1\dot{\psi}_d + \dot{\phi}_dx_3 + \dot{\phi}_d\dot{\psi}_d\right)\right. \\ &\quad \left. - J_r\varpi_rx_1 - J_r\varpi_r\dot{\phi}_d\right]. \end{aligned} \tag{3.11}$$

$$\begin{aligned} f_{f3}(x + \dot{\eta}_d) &= \frac{1}{I_{zz}}\left[\left(I_{xx} - I_{yy}\right)\left(x_1 + \dot{\phi}_d\right)\left(x_2 + \dot{\theta}_d\right)\right] \\ &= \frac{1}{I_{zz}}\left[\left(I_{xx} - I_{yy}\right)\left(x_1x_2 + x_1\dot{\theta}_d + \dot{\phi}_dx_2 + \dot{\phi}_d\dot{\theta}_d\right)\right]. \end{aligned} \tag{3.12}$$

Based on the above Eq (3.9), the attitude dynamics equation for tracking can be expressed as

$$\dot{z} = F(z, t) + G_zu_0 + D_ze_d, \tag{3.13}$$

where $F(z, t) = \left[x, f_f(x + \dot{\eta}_d) - \ddot{\eta}_d\right]^T$, $u_0 = [0_{3\times1}, u_{01}]^T$, $G_z = \begin{bmatrix} 0_{3\times3}, 0_{3\times3} \\ 0_{3\times3}, G \end{bmatrix}$, $D_z = [0_{3\times3}, I_3]^T$.

As is well known, the Hamilton-Jacobi-Bellman (HJB) equation is widely applied in the analysis of nonlinear systems for both asymptotic and finite-time stability. It is formulated within the framework of optimal control theory, and solving the HJB equation yields optimal feedback policies. The core idea is to transform the original control problem into the problem of finding a value function (also referred to as a cost-to-go or performance index function), which represents the optimal cost or cumulative reward from the current state until certain terminal conditions are met. The HJB equation ensures that this value function satisfies appropriate dynamic programming principles. In essence, the HJB equation states that, for any time t and state z, the time derivative of the value function along the system trajectory equals the minimum of the instantaneous cost rate over admissible control inputs. Solving this partial differential equation yields the optimal value function $V^*(z, t)$, and then the optimal control policies $u_0^*(z, t)$ can be obtained by solving the corresponding Hamiltonian function minimization problem.

Definition 3.2. *For system* (3.13)*, the control policy* $\mu : \mathbb{R}^n \to \mathbb{R}^m$ *is admissible for performance metrics function* (2.6) *if, for all* $z_0 \in \mathbb{R}^n$*, system* (3.13) *is predefined-time stable at the origin under* $u_0 = \mu(z)$ *and* $J(z_0; u_0) < \infty$*.*

Among them, the design of the cost function is the most critical part for the HJB equation to satisfy predefined-time stability. Based on this, the following cost function about UAV module (3.13) is set.

$$r(z, u_0) = \frac{2}{\alpha\beta T_c'}\left(2\beta V + V^{1-\frac{\alpha}{2}} + \beta^2 V^{1+\frac{\alpha}{2}}\right) + u_0^T R u_0, \tag{3.14}$$

where $u_0^T R u_0$ is the energy consumption optimization term. To facilitate the process analysis, we define

$$\Phi(V(z)) = \frac{2}{\alpha\beta T_c'}\left(2\beta V + V^{1-\frac{\alpha}{2}} + \beta^2 V^{1+\frac{\alpha}{2}}\right). \tag{3.15}$$

R and $P = \begin{bmatrix} Q_e & S \\ S^\top & Q_x \end{bmatrix} > 0$ are positive definite matrices, and V is defined as

$$V = z^T P z. \tag{3.16}$$

Let A denote the set of allowed control policies. The following lemma shows that $\mu^* \in A$, and system (3.13) is stable [28].

Next, we propose a lemma to prove that, under the designed r, the convergence time of optimal control is within the predefined time.

Remark 3.1. *We understand [28] continuous-time VI as the evolution with respect to s since the running cost contains V. In the paper, we establish equivalent DP mappings and Lyapunov comparisons in the s-network domain, thereby inheriting its convergence and approximation conclusions.*

Remark 3.2. *In classical ADP/VI frameworks (e.g., [28]), the running cost* $r(z, u_0)$ *is assumed to be independent of the value function. In the paper, however,* $r(z, u_0; V) = \Phi(V(z)) + u_0^T R u_0$ *explicitly contains the value function V. To ensure that the dynamic programming operator still defines a monotone mapping, we reinterpret the VI process as the evolution along an auxiliary time variable s, governed by the PDE*

$$\partial_s V(z, s) + \inf_{u_0}\left\{\nabla V(z, s)^T \dot{z} + r\left(z, u_0; V(z, s)\right)\right\} = 0. \tag{3.17}$$

This formulation guarantees that the update of V is monotonic in s, since $\Phi(\cdot)$ *is continuous, strictly increasing, and radially unbounded. Consequently, the same monotonicity and convergence arguments used in [28] can be applied to show that* $V(z, s)$ *converges uniformly on compact sets to the unique solution of the PTS-HJB equation as* $s \to -\infty$*. Therefore, the proposed PTS-AVI algorithm inherits the convergence and optimality properties of the standard VI method, despite the presence of V in the cost.*

Lemma 3.2. *(PTStab Lyapunov condition) Consider the nonlinear system* (3.13)*, let* $V(z) \in C^1$ *be positive definite and radially unbounded, and let* $R \in \mathbb{R}^{m\times m}$ *be symmetric positive definite.*

Define the running cost by

$$r(z,u_0)=\frac{2}{\alpha\beta T_c'}\Big(2\beta V(z)+V^{1-\frac{\alpha}{2}}(z)+\beta^2V^{1+\frac{\alpha}{2}}(z)\Big)+u_0^TRu_0, \tag{3.18}$$

where parameters $\alpha\in(0,1)$, $\beta>0$, $T_c'>0$.

If the optimal value function $V^*(\cdot)$ *solves the PTS-HJB equation*

$$0=\inf_{u_0\in\mathbb{R}^m}\Big\{\nabla V^*(z)^T\dot{z}+r(z,u_0)\Big\}, \tag{3.19}$$

and $u_0^*(z)$ *attains this minimum, then along the closed-loop system* $\dot{z}=F(z,t)+G_zu_0^*+D_ze_d$, *the Lyapunov derivative satisfies*

$$\dot{V}(z)\ \le\ -\frac{2}{\alpha\beta T_c'}\Big(2\beta V(z)+V^{1-\frac{\alpha}{2}}(z)+\beta^2V^{1+\frac{\alpha}{2}}(z)\Big), \tag{3.20}$$

which means that system (3.13) *converges within a predefined time* T_c'.

Proof. Fix z and define the Hamiltonian

$$\mathcal{H}(z,u_0)\ :=\ \nabla V^*(z)^T\dot{z}\ +\ r(z,u_0). \tag{3.21}$$

Substitute (3.18) into (3.21), and list the terms related to u_0 separately:

$$\begin{aligned}\mathcal{H}(z,u_0)=\nabla V^*(z)^T(F(z,t)+D_ze_d)+\frac{2}{\alpha\beta T_c'}\Big(2\beta V+V^{1-\frac{\alpha}{2}}+\beta^2V^{1+\frac{\alpha}{2}}\Big)\\ +\ \underbrace{\Big(\nabla V^*(z)^TG_z\,u_0+u_0^TRu_0\Big)}_{=:\ \Theta(z,u_0)}.\end{aligned} \tag{3.22}$$

Denote

$$a(z)\ :=\ G_z^T\nabla V^*(z). \tag{3.23}$$

Then the quadratic-form term can be written as

$$\Theta(z,u_0)\ =\ a(z)^Tu_0\ +\ u_0^TRu_0. \tag{3.24}$$

Complete the square for (3.24) (for specific details, please refer to the Appendix):

$$\Theta(z,u_0)=\big(u_0+\tfrac{1}{2}R^{-1}a(z)\big)^TR\big(u_0+\tfrac{1}{2}R^{-1}a(z)\big)-\tfrac{1}{4}\,a(z)^TR^{-1}a(z). \tag{3.25}$$

From (3.25), it can be seen that $\Theta(z,u_0)$ attains its minimum value at

$$u_0^*(z)\ =\ -\tfrac{1}{2}R^{-1}a(z)\ =\ -\tfrac{1}{2}R^{-1}G_z^T\nabla V^*(z) \tag{3.26}$$

and the minimum value is

$$\min_{u_0}\Theta(z,u_0)\ =\ -\tfrac{1}{4}\,a(z)^TR^{-1}a(z)\ \le\ 0. \tag{3.27}$$

Therefore, the PTS-HJB Eq (3.19) is equivalent to

$$0\ =\ \nabla V^*(z)^T(F(z,t)+D_ze_d)+\frac{2}{\alpha\beta T_c'}\Big(2\beta V+V^{1-\frac{\alpha}{2}}+\beta^2V^{1+\frac{\alpha}{2}}\Big)-\tfrac{1}{4}\,a(z)^TR^{-1}a(z). \tag{3.28}$$

From (3.28), it can be obtained that:

$$\nabla V^*(z)^T(F(z,t)+D_z e_d) = -\frac{2}{\alpha\beta T_c'}\left(2\beta V + V^{1-\frac{\alpha}{2}} + \beta^2 V^{1+\frac{\alpha}{2}}\right) + \tfrac{1}{4}\, a(z)^T R^{-1} a(z). \tag{3.29}$$

The derivative of V along the closed-loop $u_0 = u_0^*(z)$ is

$$\begin{aligned}
\dot{V}(z) &= \nabla V^*(z)^T(F(z,t)+D_z e_d) \\
&= \nabla V^*(z)^T(F(z,t)+D_z e_d) \;+\; \nabla V^*(z)^T G_z u_0^*(z) \\
&= \left[-\frac{2}{\alpha\beta T_c'}\left(2\beta V + V^{1-\frac{\alpha}{2}} + \beta^2 V^{1+\frac{\alpha}{2}}\right) + \tfrac{1}{4}\, a^T(z) R^{-1} a(z)\right] \;+\; a^T(z) u_0^* \\
&= -\frac{2}{\alpha\beta T_c'}\left(2\beta V + V^{1-\frac{\alpha}{2}} + \beta^2 V^{1+\frac{\alpha}{2}}\right) + \tfrac{1}{4}\, a^T(z) R^{-1} a(z) \;-\; \tfrac{1}{2}\, a^T(z) R^{-1} a(z) \\
&= -\frac{2}{\alpha\beta T_c'}\left(2\beta V + V^{1-\frac{\alpha}{2}} + \beta^2 V^{1+\frac{\alpha}{2}}\right) \;-\; \tfrac{1}{4}\, a(z)^T R^{-1} a(z).
\end{aligned} \tag{3.30}$$

Since $R > 0$, we have $a^T(z)R^{-1}a(z) > 0$, thus from (3.30), we can obtain

$$\dot{V}(z) \;\le\; -\frac{2}{\alpha\beta T_c'}\left(2\beta V(z) + V^{1-\frac{\alpha}{2}}(z) + \beta^2 V^{1+\frac{\alpha}{2}}(z)\right). \tag{3.31}$$

This is the required PTS Lyapunov inequality (3.20). According to the predefined-time stability theory (from which a uniformly bounded convergence time can be obtained by integrating this type of differential inequality), it is known that $V(z(t)) \to 0$ within T_c', and thus $z(t) \to 0$. □

Lemma 3.1 provides a predefined-time Lyapunov criterion. Lemma 2 shows that if the PTS-HJB equation is satisfied with the running cost (3.18), then the closed-loop Lyapunov derivative automatically matches the predefined-time Lyapunov inequality in Lemma 3.1. Based on these lemmas, we establish the predefined-time stability of the closed-loop error system in the subsequent theorems.

Lemma 3.3. *[28] Assume that $V^* \in \mathcal{P}$ is intrinsic, r is positive semi-definite and continuous, $r(z,\cdot)$ is positive definite for all $z \in \mathbb{R}^n$, system* (3.13) *with $y = r(z,0)$ is zero state observable, and Assumption 2.1 is satisfied. Then, we have $V^* \in \mathcal{P}_+$ and $\mu^* \in A$, where A is the set of allowable control policies.*

Remark 3.3. *The definition of zero-state observability can be found in [40, Definition 6.5]. The zero-state observability in Lemma 3.3 ensures that the optimal control law μ^* is not only optimal but also a control law that stabilizes system* (3.13).

The above destination is to find the optimal controller and its association function. Since it is not easy to directly solve the HJB equation, literature [28] proposes a continuous-time nonlinear VI method to take advantage of numerical impersonation to simulate the following equation and obtain the optimal function value and policies.

$$0 = \partial_s V(z,s) + \inf_{u_0 \in \mathbb{R}^m} H\left(z, \partial_z V(z,s), u_0\right), \quad s \le 0. \tag{3.32}$$

Assumption 3.1. *$V_0 \in \mathcal{P}$ is intrinsic, and when $V(\cdot, 0) = V_0(\cdot)$, there exists a unique solution to the HJB Eq* (3.32) *on $\mathbb{R}^n \times \mathbb{R}_-$.*

If both Assumption 2.2 and Assumption 3.1 hold, then there exists $\mu_v \in C^0(\mathbb{R}^n \times \mathbb{R}_-)$ such that the following relationship holds for all $(z, s) \in \mathbb{R}^n \times \mathbb{R}_-$:

$$\mu_v(z, s) = \arg \inf_{u_0 \in \mathbb{R}^m} H\left(z, \partial_z V(z, s), u_0\right). \tag{3.33}$$

Theorem 3.2. *[28] The conditions in Lemma 3.3 hold. For any V_0 satisfying Assumption 2, $\lim_{s\to-\infty} V(\cdot, s) = V^*(\cdot)$ is uniform on any compact subset in $\mathbb{R}^n$. Furthermore, for all $s < 0$, $V(\cdot, s) \in P$. Additionally, if r is positive definite, then for all $s < 0$, $V(\cdot, s) \in \mathcal{P}_+$.*

Theorem 3.2 indicates that by solving the generalized HJB equation using the backward time integral method, a general VI method for nonlinear systems is obtained. Here, it is obvious that r is positive definite; therefore, the VI method proposed in the literature can be applied. In addition, Theorem 3.2 solves the difficulty of directly solving the HJB equation through backward time integration, enabling the optimal policies with respect to the time variable s to converge to the optimal policies derived from Eq (2.10).

The quadrotor UAV attitude information is η, the tracking error is $e = \eta - \eta_d$, and the tracking error system (3.13) is considered. Training/software deployment uses the same form of just-in-time cost (substituting $V(z, s)$ during training and using $\hat{V}(z)$ during software deployment):

$$r(z, u_0; V) = \Phi(V) + u_0^T R u_0, \tag{3.34}$$

with $0 < \alpha < 1$, $\beta > 0$, $T_c > 0$, $R > 0$ and $\tilde{\mu} = 1 - \frac{\alpha}{2} \in (0, 1)$, $\tilde{\nu} = 1 + \frac{\alpha}{2} > 1$.

Assumption 3.2. *In the HJB/VI PDE of training-form $s-$, there exists $\varepsilon_1 \in [0, 1)$ such that for any (z, s),*

$$\inf_{u_0} \nabla_z V(z, s)^T \left(F(z, t) + G_z u_0 + D_z e_d\right) \geq -\varepsilon_1 \Phi\left(V(z, s)\right) \tag{3.35}$$

holds.

Remark 3.4. *Assumption 3.2 and the variable s in this paper represent an auxiliary time variable, which are used to implement the VI algorithm within the continuous-time framework and do not represent real time. $s-$ denotes the auxiliary variable $s \to -\infty$, which can also be represented as $s \leq 0$ in some formulas. The evolution of $s-$ simulates the iterative process of the VI algorithm. In reference [28], precisely this auxiliary variable s was used to break through the limitation that VI could only be applied to discrete systems, thereby enabling the VI process for continuous systems.*

Assumption 3.2 indicates that within the admissible control range, the worst-case instantaneous growth rate of the value function V along any trajectory will not counteract the designed dissipation term $\Phi(V)$ beyond a proportion ε_1. When $\varepsilon_1 < 1$, the cost term can still provide net dissipation, thereby ensuring the validity of the key inequality for predefined-time decay.

Next, we present the main results. By constructing a predefined-time form of r, we prove the predefined-time stability theorem for both the training and deployment phases of UAVs under the HJB equation framework within reinforcement learning. This represents a novel form and approach, distinct from conventional asymptotic stability, finite-time stability, and fixed-time stability. Furthermore, its application in UAVs is also a new attempt.

3.1. Training phase (HJB/VI with s−)

In the training phase, the following result holds.

Theorem 3.3. *Let $V(z, s) \in C^1$ satisfy the following HJB/VI PDE with $s \leq 0$:*

$$\frac{\partial V}{\partial s}(z, s) + \inf_{u_0}\{\nabla_z V(z, s)^T (F(z, t) + G_z u_0 + D_z e_d) + r(z, u_0; V)\} = 0, \tag{3.36}$$

and assume that Assumption 3.2 holds, with $V(z, s)$ being positive definite and radially unbounded with respect to z. Here, r meets (3.34)*, $\tilde{\mu} = 1 - \frac{\alpha}{2} \in (0, 1)$, $\tilde{\nu} = 1 + \frac{\alpha}{2} > 1$, and we set*

$$\Phi(V) = \frac{2}{\alpha\beta T_c}\Big(2\beta V + V^{\tilde{\mu}} + \beta^2 V^{\tilde{\nu}}\Big). \tag{3.37}$$

Where $\Phi(V)$ is continuous and monotonically increasing and $u_0^T R u_0 \geq 0$, then $V(z, s)$ evolves along the auxiliary time variable s, yielding

$$\frac{\partial V}{\partial s}(z, s) \leq -(1 - \varepsilon_1)\, \Phi(V(z, s)). \tag{3.38}$$

Furthermore, for any initial condition $V(0)$, there exists a prescribed upper bound on the settling time $T_c^{'}/(1 - \varepsilon_1)$. Thus, $V(z, s)$ achieves predefined-time convergence.

Proof. At the minimization point $u_0^*(z, s)$ of the PDE for $s \leq 0$, we have

$$\partial_s V(z, s) + \nabla_z V(z, s)^T (F(z, t) + G_z u_0^* + D_z e_d) + r(z, u_0^*; V) = 0. \tag{3.39}$$

Because $r(z, u_0^*; V) = \Phi(V(z, s)) + u_0^{*T} R u_0^* \geq \Phi(V)$, we have

$$\partial_s V \leq -(\nabla_z V^T (F(z, t) + G_z u_0^* + D_z e_d) + \Phi(V)). \tag{3.40}$$

From Assumption 3.2, we have $\nabla_z V^T (F(z, t) + G_z u_0^* + D_z e_d) \geq -\varepsilon_1 \Phi(V)$, and substituting this inequality into (3.40), we obtain

$$\partial_s V \leq -(1 - \varepsilon_1)\Phi(V). \tag{3.41}$$

As a result, (3.41) can be written as

$$\partial_s V \leq -(1 - \varepsilon_1)\frac{2}{\alpha\beta T_c}\Big(2\beta V + V^{\tilde{\mu}} + \beta^2 V^{\tilde{\nu}}\Big). \tag{3.42}$$

For $s \in [s_0, s_f]$ and $V(s_f) = 0$, according to Lemma 3.2, for any initial value $V(s_0) \geq 0$, there exists a uniform upper bound of $T_c^{'}/(1 - \varepsilon_1)$. □

The work above uses the continuous nonlinear VI method to conduct the transformation to solve for the optimal value function and optimal policies, changing from solving the cost price function (2.6) to solving the HJB equation. Next, adapting theory will be used to update the parameters, and approximate updates will be made to the optimal value function and Hamiltonian.

Value function and HJB equation approximation

Take advantage of function approximation theory. For all $(z, u_0) \in K_z \times K_{u_0}$, ingest two sets of continuously differentiable linearly independent basis functions $\{\phi_{Ri}\}_{i=1}^{\infty}$ and $\{\psi_{Ri}\}_{i=1}^{\infty}$ to approximate V and H at each auxiliary time variable s.

$$\begin{aligned} \lim_{N\to\infty} \hat{V}_N(z, w(s)) &= V(z, s) \\ \lim_{N\to\infty} \hat{H}_N(z, u_0, c(s)) &= H\left(z, \partial_z V(z, s), u_0\right) \end{aligned} \tag{3.43}$$

where $w = [w_1, w_2, \ldots, w_N]^T, c = [c_1, c_2, \ldots, c_N]^T$ are the weights of the basis functions, N is the quantity of basis functions, and $K_z \subset \mathbb{R}^n$ and $K_{u_0} \subset \mathbb{R}^m$ are two compact sets, where $z(0)$ is in the internal of K_z. Select basis functions and a radial basis function neural network (RBF-NN) to approximate $\hat{V}_N$ and $\hat{H}_N$.

$$\hat{V}_N(z, w) = \sum_{i=1}^{N} w_i \phi_{Ri}(z), \tag{3.44}$$

$$\hat{H}_N(z, u_0, c) = \sum_{i=1}^{N} c_i \psi_{Ri}(z, u_0), \tag{3.45}$$

where each basis function is

$$\phi_{Ri}(z) = \exp\left(-\frac{\left\|z - c_{\phi_R,i}\right\|^2}{\sigma_{\phi_R,i}^2}\right), \tag{3.46}$$

$$\psi_{Ri}(z, u_0) = \exp\left(-\frac{\left\|(z, u_0) - c_{\psi_R,i}\right\|^2}{\sigma_{\psi_R,i}^2}\right), \tag{3.47}$$

where $c_{(\cdot),i}$ is the i-th RBF factory, $\sigma_{(\cdot),i}^2$ is the width parameter, w_i, c_i are the corresponding weights, and N is the RBF quantity. Define vectors $\Phi(z) = [\phi_{R1}(z), \ldots, \phi_{RN}(z)]^T$ and $\Psi(x) = [\psi_{R1}(z), \ldots, \psi_{RN}(z)]^T$.

Define the following feature covariance matrix/gram information matrix function:

$$\begin{aligned} K_{\phi_R}\left(t_f\right) &= \int_0^{t_f} \Phi(z)\Phi^T(z)dt, \\ K_{\psi_R}\left(t_f\right) &= \int_0^{t_f} \Psi(z, u_0)\Psi^T(z, u_0)dt, \end{aligned} \tag{3.48}$$

and the weight function can be expressed as

$$\begin{aligned} W_{\phi_R} &= K_{\phi_R}^{-1} \int_0^{t_f} \Phi(z) y_{\phi_R}(t)dt, \\ W_{\psi_R} &= K_{\psi_R}^{-1} \int_0^{t_f} \Psi(z, u_0) y_{\psi_R}(t)dt, \end{aligned} \tag{3.49}$$

where $y = r(z, 0)$.

To achieve the persistent excitation condition in adapting theory, the following assumption is made:

Assumption 3.3. *There exist $\gamma > 0$ and $t_0 > 0$, such that for all $t_f > t_0$, $(z(\cdot), u_0(\cdot))$ holds on $K_z \times K_{u_0}$ within the interval $\left[0, t_f\right]$, and*

$$\frac{1}{t_f} K_{\phi_R}\left(t_f\right) > \gamma I_N, \quad \frac{1}{t_f} K_{\psi_R}\left(t_f\right) > \gamma I_N. \tag{3.50}$$

Assumption 3.3 adds the persistent excitation (PE) condition to adaptive control.

Remark 3.5. *In Assumption 3.3, conditions* (3.50) *should be understood in the sense of positive definite matrices, i.e., $K_{\varphi_R}\left(t_f\right) > \gamma t_f I_N$, $K_{\psi_R}\left(t_f\right) > \gamma t_f I_N$, for all $t_f \geq t_0$. This ensures that the Gramian matrices $K_{\varphi_R}(t_f)$ and $K_{\psi_R}(t_f)$ are uniformly positive definite, which is the standard PE condition in adaptive/learning control. Under this condition, the regression problems admit unique solutions for the weight vectors W_{φ_R} and W_{ψ_R}, and the function approximators $\hat{V}_N$ and $\hat{H}_N$ converge to their true counterparts.*

Assumption 3.4. *There exists $\bar{N} > 0$, such that for any $N > \bar{N}$ and $z \in K_z$, $\inf_{u_0 \in K_{u_0}} \hat{H}_N(z, u_0, c)$ exists and is continuously differentiable with respect to c on $\mathbb{R}^N$. Furthermore, there exists a unique continuous $\hat{\mu}_N$ such that*

$$\hat{\mu}_N(z, c) = \arg \inf_{u_0 \in K_{u_0}} \hat{H}_N(z, u_0, c), \quad \forall (z, c) \in K_z \times \mathbb{R}^N. \tag{3.51}$$

Theorem 3.4. *[28] Consider system* (3.13) *and cost function* (2.6). *The conditions in Lemma 3.3 hold, and r is positive definite. Under assumptions 2–4 and since $V_0(\cdot) = \hat{V}_N(\cdot, \hat{w}(0))$, for any $\varepsilon > 0$, there exist s_f, t_f, N, and K_z such that*

$$\begin{aligned} &\sup_{z \in K_z} \left| \hat{V}_N\left(z, \hat{w}\left(s_f\right)\right) - V^*(z) \right| < \varepsilon, \\ &\sup_{z \in K_z} \left| \hat{\mu}_N\left(z, \hat{c}\left(s_f\right)\right) - \mu^*(z) \right| < \varepsilon. \end{aligned} \tag{3.52}$$

Among them, $\hat{w}$ and $\hat{c}$ are determined by the following formulas, respectively.

$$\frac{d}{ds}\hat{w} = K_\phi^{-1}\left(t_f\right) \int_0^{t_f} \Phi(z) \hat{H}_N\left(z, \hat{\mu}_N(z, \hat{c}), \hat{c}\right) dt \tag{3.53}$$

$$\hat{c} = K_\psi^{-1}\left(t_f\right) \int_0^{t_f} \Psi(z, u_0) \left(d\hat{V}_N(z, \hat{w}) + r dt\right) \tag{3.54}$$

When s approaches s_f, $V_N \to V((\cdot), s_f)$, and the approximation of optimal control is obtained:

$$u_{0N}(z) \approx -\frac{1}{2} R^{-1} G_z^T \partial_z V_N\left(z, w\left(s_f\right)\right). \tag{3.55}$$

The above content constitutes the predefined-time Lyapunov convergence proof for the learning and training phase, as well as the logic and computational methods for parameter updates. In summary, Eqs (3.43) to (3.55) implement the continuous-time VI as a "two-step batch least squares" process: First, over the interval $[0, t_f]$, features are accumulated to construct the information matrices K_{ϕ_R}, K_{ψ_R} and their corresponding right-hand sides.

The linear parameters c for the Hamiltonian are then regressed using the normal equation, thereby obtaining the nominal policy $\hat{\mu}$.

Subsequently, using this policy, the Hamiltonian $\hat{H}_N(z, \hat{\mu}_N(z, \hat{c}), \hat{c})$ is evaluated, and a single VI update in the s-direction is performed on the value function weights w via (3.53).

Assumption 3.5. *According to the universal approximation property of neural networks, there exists an ideal constant weight vector w^* such that the optimal value function $V^*(z)$ and its gradient $\nabla V^*(z)$ on a compact set Ω can be represented as:*

$$\begin{aligned} V^*(z) &= (w^*)^T \Phi(z) + \epsilon(z), \\ \nabla V^*(z) &= (w^*)^T \nabla\Phi(z) + \nabla\epsilon(z), \end{aligned} \tag{3.56}$$

where $\Phi(z)$ is the basis function vector, and $\epsilon(z)$ is the bounded reconstruction error. It is assumed that there exist positive constants $W_M, \epsilon_M, \epsilon'_M$ such that $\|w^*\| \le W_M$, $|\epsilon(z)| \le \epsilon_M$, and $\|\nabla\epsilon(z)\| \le \epsilon'_M$ for all $z \in \Omega$.

Remark 3.6. *Assumption 3.5 is grounded in the well-established universal approximation property (UAP) of neural networks [41]. It states that for a sufficiently large number of neurons (or basis functions), a linearly parameterized neural network can approximate any continuous function (in this case, the optimal value function $V^*(z)$ and its gradient) to an arbitrary degree of accuracy on a compact set Ω. The terms $\epsilon(z)$ and $\nabla\epsilon(z)$ represent the inherent reconstruction errors arising from the finite network structure. Since the UAV's operational workspace Ω is a compact set, these reconstruction errors are naturally bounded by finite constants ϵ_M and ϵ'_M. This boundedness is a prerequisite for establishing the uniformly ultimately bounded (UUB) stability of the closed-loop system.*

Convergence and stability analysis of the neural approximation is as follows.

Theorem 3.5. *Consider the value function approximation error dynamics. Under the proposed update law Eq* (3.53)*, the weight estimation errors $\tilde{w} = \hat{w} - w^*$ are UUB.*

The proof of the theory can be found in the Appendix.

Finally, iterate until the residual converges, and then freeze the parameters for deployment.

Then the following is the stability analytics of the system convergence time in the deployment phase.

Allow the existence of approximation/implementation error $\Delta(z)$ during software deployment, and the closed-loop function relationship is

$$\dot{\hat{V}}(z) = -r(z, u_0(z); \hat{V}) + \Delta(z). \tag{3.57}$$

Lemma 3.4. *For the system $\dot{x} = \phi(x, t)$, if there exists a radially unbounded Lyapunov function $V(x)$ such that*

$$\dot{V} \le -\frac{2}{\alpha\beta T_c}(2\beta V + V^{1-\frac{\alpha}{2}} + \beta^2 V^{1+\frac{\alpha}{2}}) + \theta \tag{3.58}$$

holds for any solution $x(t, x_0)$ of the system $\dot{x} = \phi(x, t)$, then the trajectory of the system is practical predefined-time stable, and the residual set of the solution of system $\dot{x} = \phi(x, t)$ can be given by

$$\{\lim_{t\to T'_c} x \mid V(x) \le min\{\frac{\theta\alpha T_c}{4(1-\gamma)}, (\frac{\theta\alpha\beta T_c}{2(1-\gamma)})^{\frac{2}{2-\alpha}}, (\frac{\theta\alpha T_c}{2\beta(1-\gamma)})^{\frac{2}{2+\alpha}}\}\}, \tag{3.59}$$

where $T_c > 0$ is a predefined-time constant, $\theta > 0$ is a boundary constant number, $\alpha \in (0,1)$ and β are assignable parameters, $0 < \gamma < 1$ is a given constant number, and the settling time is given by $T_c^{'} = T_c/\gamma$.

The proof of the lemma is provided in the Appendix.

3.2. Deployment phase

In the deployment phase, the learned policy guarantees predefined-time stability as follows.

Theorem 3.6. *Assume that $\hat{V}(z) \in C^1$ is positive definite and radially unbounded with respect to z, and the software deployment closed-loop satisfies*

$$\dot{\hat{V}}(z) = -\Phi(\hat{V}(z)) - u_0^T R u_0 + \Delta(z), \tag{3.60}$$

where Δ is the lumped error between the parameter estimate and the optimal parameter. If Δ satisfies $|\Delta(z)| \le \bar{\Delta}$, $\bar{\Delta} \ge 0$ and there exist constants $\alpha, \beta, T_c > 0$, $\tilde{\mu} = 1 - \frac{\alpha}{2} \in (0,1)$, $\tilde{\nu} = 1 + \frac{\alpha}{2} > 1$, i.e.,

$$\Phi(\hat{V}) \ge \frac{2}{\alpha\beta T_c}\left(2\beta V + V^{\tilde{\mu}} + \beta^2 V^{\tilde{\nu}}\right), \tag{3.61}$$

then, the following conclusions can be drawn:

1) If $\Delta \equiv 0$, then $\hat{V}(t)$ converges within the preset time T_c (independent of the initial value).

2) If there is only an absolute error $\bar{\Delta}$, then $\hat{V}(t)$ converges within the preset time T_c/γ.

Proof. Directly write the closed-loop differential from the assumption:

$$\dot{\hat{V}} = -\Phi(\hat{V}) - u_0^T R u_0 + \Delta \le -\Phi(\hat{V}) + \Delta. \tag{3.62}$$

From Eq (3.62), we can obtain

$$\dot{\hat{V}} \le -\frac{2}{\alpha\beta T_c}(2\beta V + V^{\tilde{\mu}} + \beta^2 V^{\tilde{\nu}}) + \Delta. \tag{3.63}$$

1) Error-free case: When $\Delta \equiv 0$, (3.63) degenerates to

$$\dot{\hat{V}} \le -\frac{2}{\alpha\beta T_c}\left(2\beta V + V^{\tilde{\mu}} + \beta^2 V^{\tilde{\nu}}\right). \tag{3.64}$$

From Lemma 3.1, the predefined-time upper bound T_c is as shown in the theorem.

2) Absolute error case: If $|\Delta| \le \bar{\Delta}$, then

$$\dot{\hat{V}} \le -\Phi(\hat{V}) + \bar{\Delta}. \tag{3.65}$$

Thus, from (3.61), we obtain that

$$\dot{\hat{V}} \le -\Phi(\hat{V}) + \bar{\Delta} \le -\frac{2}{\alpha\beta T_c}\left(2\beta\hat{V} + \hat{V}^{\tilde{\mu}} + \beta^2\hat{V}^{\tilde{\nu}}\right) + \bar{\Delta}. \tag{3.66}$$

By Lemma 3.4, starting from any initial value $\hat{V}(0)$, $\dot{\hat{V}}$ will converges within predefined time T_c/γ. □

Figure 2 presents the overall algorithmic framework of this study, including both the training and deployment phases. First, the attitude error dynamics are constructed based on the measured attitude and the reference attitude. Then, a predefined-time disturbance observer is designed to estimate the lumped disturbances, thereby compensating for the attitude dynamics and enhancing robustness. Next, a predefined-time stability criterion is embedded into the running cost to facilitate the subsequent optimal policy derivation, and the corresponding HJB equation is formulated. To avoid directly solving the highly nonlinear HJB equation, an auxiliary time variable s is introduced and a neural-network-based approximation is employed to compute the Hamiltonian and the optimal value function, with the associated parameters updated iteratively. As a result, the optimal control policy is obtained during the training phase and then applied in the deployment phase to achieve predefined-time stabilization for the UAV attitude control system.

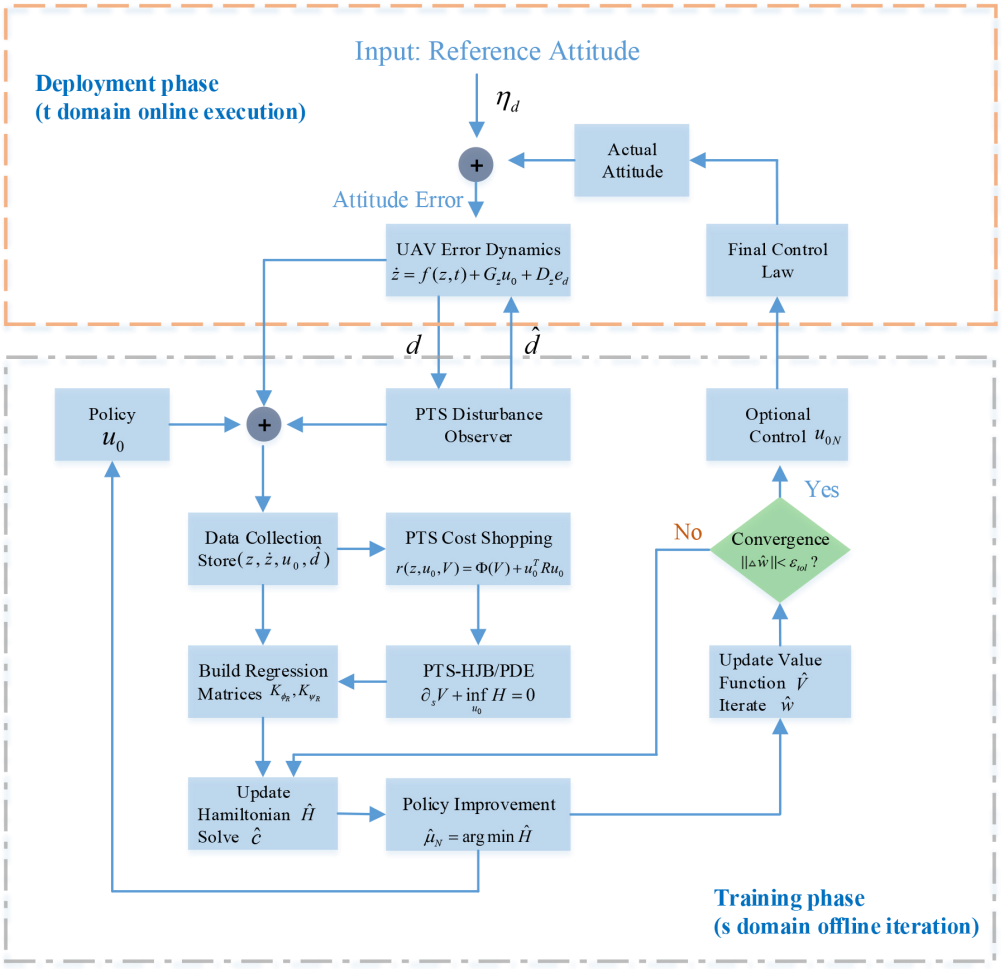

Figure 2. A summary of the proposed PTS-AVI control architecture.

4. Simulation results

To verify the effectiveness of the quadrotor UAV attitude tracking controller based on the PTS-AVI method proposed in this paper, the UAV attitude tracking model (3.13) incorporating external

disturbances is constructed for numerical simulation.

The inertial parameters and disturbances are set as $I_{xx} = 0.02, I_{yy} = 0.02, I_{zz} = 0.04, l = 0.1$, and $d = [0.05sin(0.5t), 0.05cos(0.7t), 0.02sin(1.1t)]^T$. The reference torque is $\eta_d = [\varphi_d, \theta_d, \psi_d]^T = [0.06cos(0.3t), 0.06cos(0.4t), 0.04cos(0.2t)]^T$ and $R = 0.1I_3$. The gyroscopic term is neglected, $\Omega_r = 0$, since its effect is small compared with the control torques under the considered operating conditions. The predefined-time parameters are $T_c' = 2s, \alpha = 0.5, \beta = 10$, and the exploring noise explore_std is the number of the level ranging from -0.35 to 0.35.

Simulations were implemented in MATLAB R2023b. The numerical solver was ode45 with step size $T_s = 0.001s$. The experiments were executed on a workstation with an Intel(R) Core(TM) i7-8750H CPU @ 2.20GHz (2.20 GHz), RAM: Hynix DDR4 32G and Windows 11.

Remark 4.1. *Parameter selection guidelines: The proposed method involves predefined-time design parameters α, β, T_c', the observer gain λ, and learning-related parameter N. These parameters play the following roles. First, the desired predefined upper bound of the settling time T_c' should be selected according to task requirements. For $\alpha \in (0, 1)$, a typical choice is $\alpha = 0.5$. Increasing β appropriately can further accelerate the convergence speed. For the observer gain λ, a larger λ improves the disturbance tracking performance, but may increase sensitivity to measurement noise. The number of RBF nodes N should be chosen to balance approximation accuracy and real-time computational capability. In practice, a moderate value of N can be adopted initially, and N should be increased only when the approximation error is noticeably large. For specific systems, the parameters can be tuned within reasonable ranges to achieve optimal overall performance.*

Remark 4.2. *It is noted that the observer dynamics in Eq* (3.3) *and the robust control terms contain discontinuous functions (e.g., $sign(\cdot)$ and $sign(\cdot)^\alpha$), which may induce high-frequency oscillation, known as chattering, in practical implementation. To mitigate this issue and ensure smooth control inputs for the UAV motors, a continuous approximation approach is adopted in the simulation and potential hardware deployment. Specifically, the $sign(s)$ function is replaced by the hyperbolic tangent function: $sign(s) \approx tanh(\frac{s}{\varpi})$, where $\varpi > 0$ is a small scalar determining the width of the boundary layer. Although this smoothing technique implies that the system theoretically achieves practical predefined-time stability (converging to a small residual set $\Omega = \{s : |s| \leq \epsilon\}$) rather than the origin, the tracking error can be maintained within an acceptable range by tuning ϖ. Furthermore, this modification prevents the excitation of unmodeled high-frequency dynamics, thereby improving the overall safety and feasibility of the engineering system. It should be pointed out that, when studying the problem of pose tracking, errors cannot be reduced to zero, and having a certain amount of error is reasonable.*

Set the training phase duration to 10 s, and the deployment verification phase duration is 10 s. We have the following simulation results:

Figure 3 shows the comparison between the external disturbances and their estimated values. It can be observed that, under the action of the predefined-time disturbance observer designed in this paper, the disturbances on all three axes can achieve accurate estimation of the true disturbances within the predefined time T_c', while maintaining good tracking performance. This indicates that the observer can effectively compensate for external disturbances during both the training and deployment phases, providing a reliable guarantee for the subsequent control law design, thereby enhancing the system's robustness and stability.

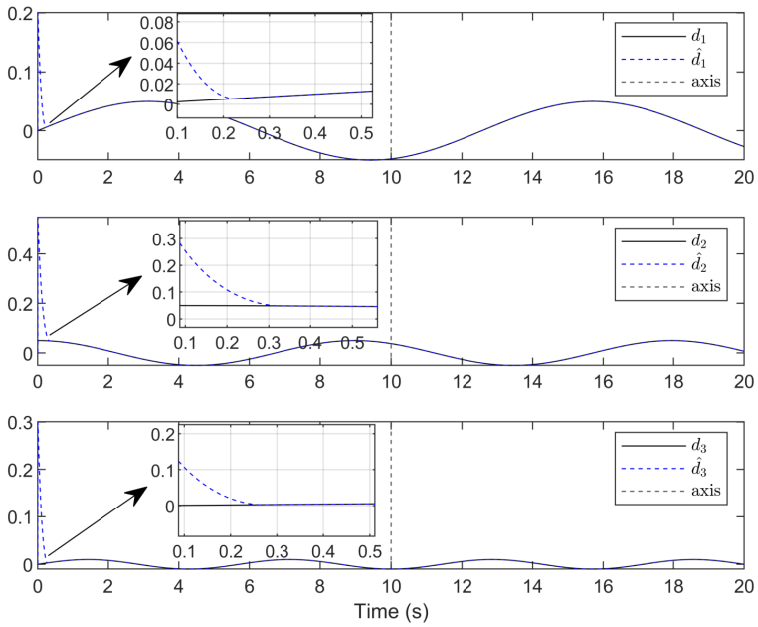

Figure 3. Disturbances acting on the UAV and their observed estimates.

Figures 4 and 5 report attitude tracking performance and tracking errors. During training (0–10 s), exploration noise is injected to excite the system for learning, while the attitude errors remain bounded and converge within the prescribed time. During deployment (10–20 s), the learned policy yields improved tracking accuracy with reduced error oscillations.

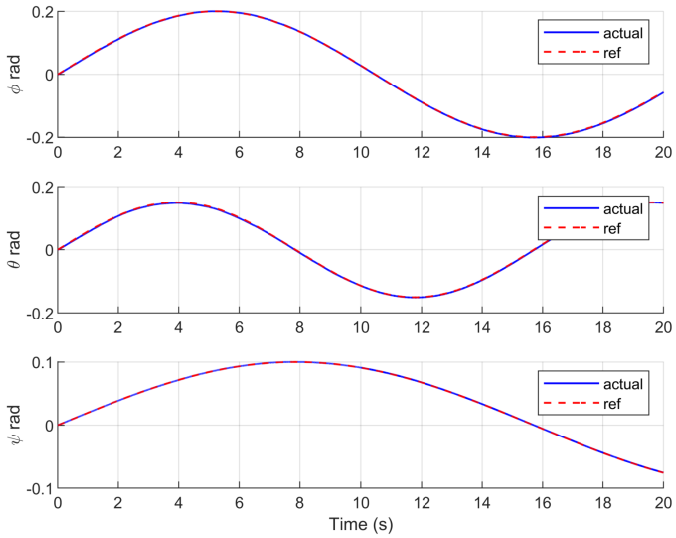

Figure 4. Attitude trajectory tracking of the UAV.

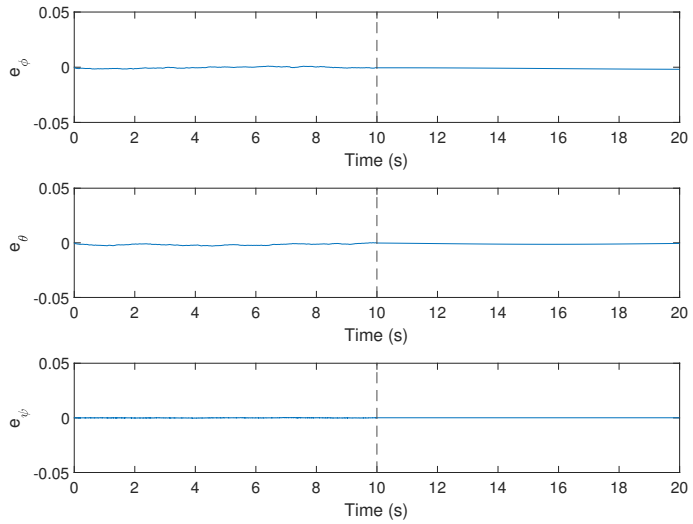

Figure 5. UAV attitude trajectory tracking error.

Figures 6 and 7 show the variation of the three-axis angular velocity trajectory tracking and errors. During the deployment phase, the angular velocity errors on each axis rapidly decay to near zero, verifying the effectiveness of the controller in rate control. However, in the training phase, affected by exploration noise and the initially imperfect policy, the rate errors exhibit significant fluctuations, which align with the expected characteristics of the early stages of reinforcement learning training.

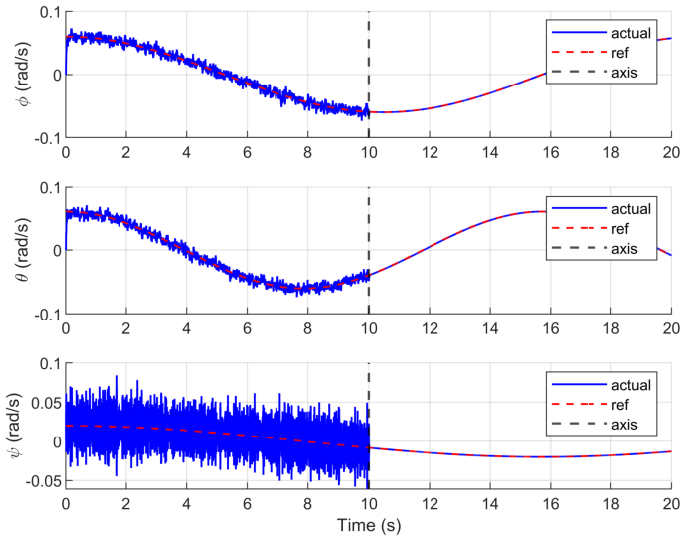

Figure 6. Angular velocity trajectory tracking of the UAV.

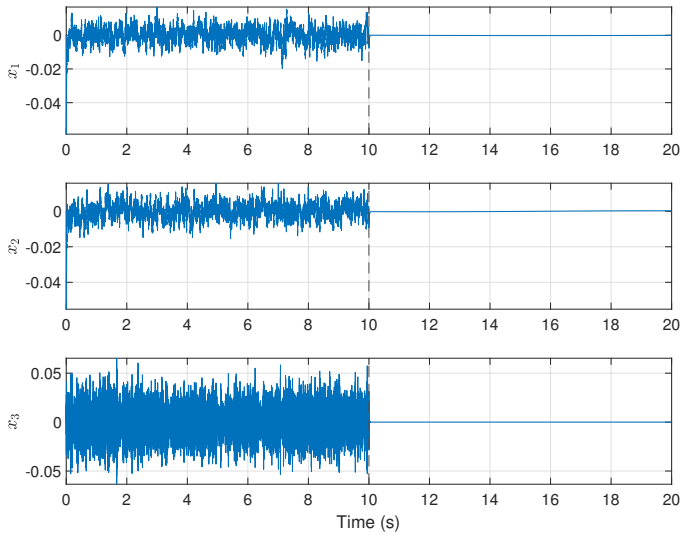

Figure 7. Angular velocity trajectory tracking error of the UAV.

Figure 8 illustrates the variation of the closed-form optimal control input u_0. During the training phase, the control input has a larger magnitude and contains high-frequency components, a result of the combined effects of exploration noise and iterative weight updates. In contrast, during the deployment phase, the control input becomes significantly smoother and remains within a reasonable range, indicating that the controller possesses good practical implementability.

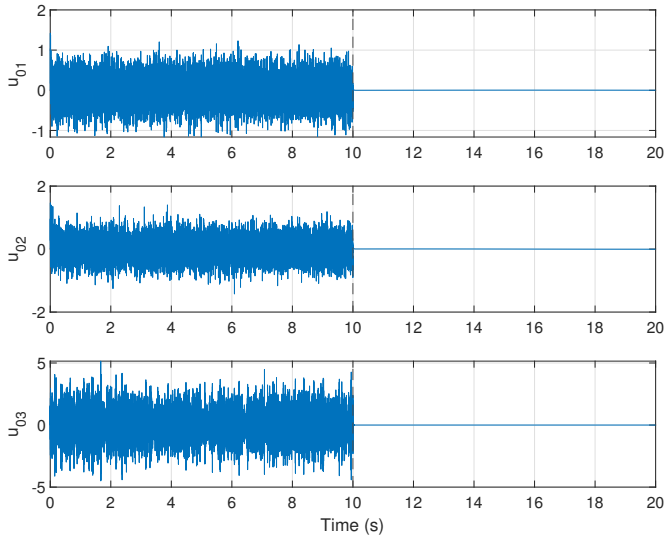

Figure 8. Control input for attitude control.

Figure 9 shows the variation of the approximate value function $\hat{V}$ and the HJB residual. It can be observed that during the training process, the residual $|H|$ decreases significantly over time and approaches zero in the later stages of training. The value function remains non-negative and exhibits a gradual convergence trend. This is consistent with the theoretical analysis, demonstrating that the PTS-

AVI method effectively approximates the PTS-HJB equation during the training phase and ultimately converges to the optimal value function.

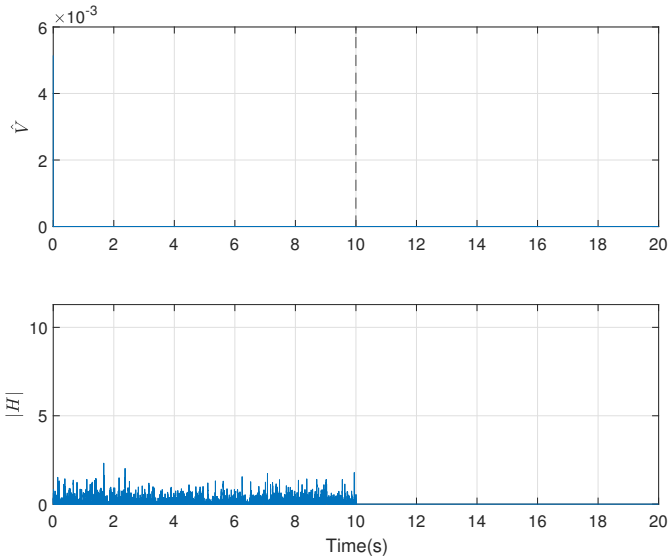

Figure 9. Variations of the approximate value function $\hat{V}$ and the HJB residual.

From Figure 10, it can be seen that in the three comparative experiments, the proposed approach in this paper was evaluated against three representative existing methods under identical simulation conditions: The method in [28] corresponds to the simulation results (a) and (d) in Figure 10, the fixed-time method in [25] corresponds to the simulation results of (b) and (e) in Figure 10, and the method in [27] corresponds to the simulation results (c) and (f) in Figure 10. From the attitude tracking errors e_ϕ, e_θ, e_ψ and angular velocity errors x_1, x_2, x_3, it can be observed that during the training phase (0–10 s), all methods exhibit random fluctuations induced by exploration noise and external disturbances. However, the method proposed in this paper yields significantly smaller error variations and achieves a faster and smoother error attenuation after the switching instant at 10 s, as show in Figures 4–7. In contrast, the PTS-AVI method consistently presents the smallest peak errors and the shortest settling time across all channels. The tracking errors converge more rapidly to the vicinity of zero and remain stable thereafter, demonstrating superior disturbance rejection capability and improved convergence performance.

To rigorously evaluate the robustness and sensitivity of the proposed predefined-time disturbance observer (PTDO), two sets of comparative simulations were conducted. (1) Varying convergence time T_c: The observer was tested under three different user-defined time parameters: $T_c = \{0.5\,\text{s}, 1\,\text{s}, 2\,\text{s}\}$, with a fixed disturbance Amplitude (Amp). (2) Varying disturbance intensity: The observer was tested under three distinct disturbance magnitudes: Small (Amp = 0.5), Medium (Amp = 2), and Large (Amp = 5.0), with a fixed $T_c = 1$ s. Crucially, to ensure a fair comparison, all simulation cases were initialized with the same estimation error state.

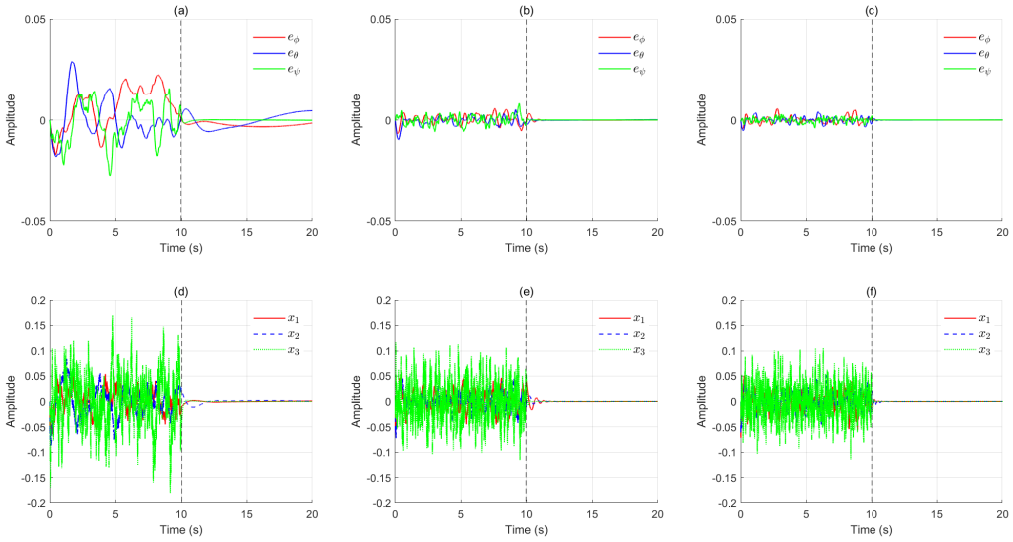

Figure 10. Attitude angle and attitude angular rate comparison experiment.

The results shown in Figure 11 clearly demonstrate the predefined-time property of the proposed observer. Although the convergence trajectories start from the same initial estimation error, they differ significantly due to the prescribed values of T_c. Specifically, the disturbance estimation errors converge within the predefined times of 0.5 s, 1 s, and 2 s, respectively, which confirms that the settling time is an independent and tunable design parameter.

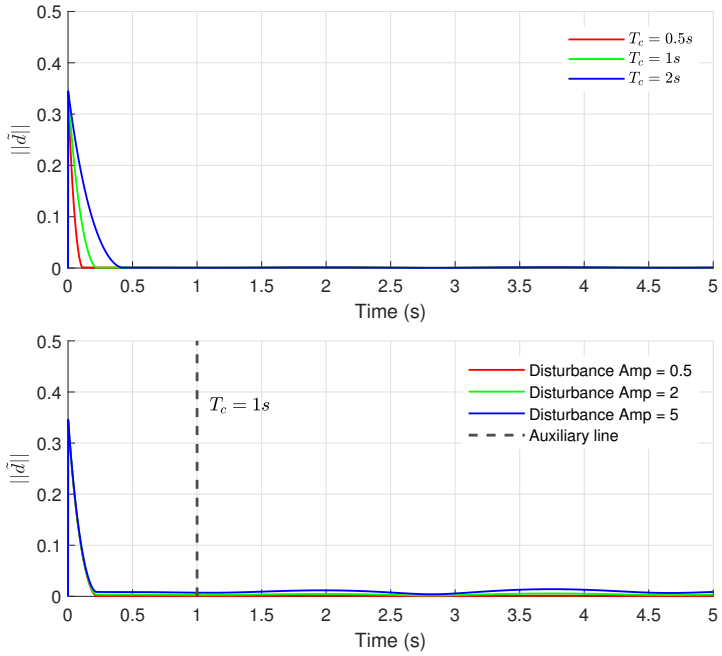

Figure 11. Sensitivity and robustness analysis of the disturbance observer.

Moreover, the disturbance amplitude simulations highlight the robustness of the proposed observer. Even when the disturbance magnitude is increased by ten times from (Amp = 0.5) to (Amp = 5), the observer is still able to successfully track the time-varying disturbance, although the estimation error shows an increasing tendency. Most importantly, the convergence time remains unchanged, i.e., the estimation error still converges at (T_c = 1 s).

In summary, the sensitivity analysis verifies that the proposed observer achieves guaranteed finite-time convergence, and the convergence process can be strictly regulated by the predefined parameter T_c. Moreover, as shown in Figure 11, the observer remains stable even under large-amplitude disturbances (Amp = 5), indicating strong robustness against severe wind gusts, provided that the disturbance remains within the physical capability of the UAV propulsion system. This ensures that reliable feedforward compensation can be provided for the subsequent VI-ADP controller.

Based on the aforementioned simulation results, the following conclusions can be drawn:

(1) The predefined-time disturbance observer can achieve rapid and accurate estimation of disturbances within a predefined time, thereby enhancing system robustness.

(2) By embedding a value function term into the cost function, the PTS-AVI control method ensures that the control law satisfies the predefined-time stability condition during both the training and deployment phases.

(3) The simulations validate the theoretical derivations presented in this paper: regardless of the initial state, both the system states and rates can converge to zero within a user-defined time T_c'. This result outperforms traditional VI-ADP methods, whose convergence time depends on the initial conditions.

In summary, the PTS-AVI control strategy proposed in this paper achieves high precision, disturbance rejection, and rapid convergence for UAV attitude control.

5. Summary and outlook

This paper studies the attitude tracking problem of UAVs. By combining the value iteration method of reinforcement learning, adapting optimal control and predefined time stability, the applicable UAV optimization procedure is proposed. This method only requires given learning parameters to obtain the optimal value function and optimal policies through training. Finally, the simulation results demonstrate that the proposed predefined-time optimal control scheme achieves faster convergence while maintaining accurate attitude tracking. This study focuses on the attitude (inner-loop) subsystem and adopts an RBF-based approximation approach, whose scalability mainly depends on the dimension of the basis functions and the system state dimension. When extending the proposed method to multi-UAV scenarios, additional challenges should be addressed, including communication constraints, coupled objectives (e.g., formation control), and higher-dimensional function approximation problems. These factors may increase computational burden and further complicate the performance evaluation conditions. In the future, the adapting optimal control of predefined-time value iteration for underactuated quadrotor UAVs will be studied. In addition, future work will further evaluate the performance limits and practical applicability of the proposed approach on higher-fidelity platforms, particularly under input constraints and actuator saturation. Moreover, hardware implementation and real-world experimental validation will be

conducted to demonstrate the feasibility of the proposed algorithm in practical applications. It should also be noted that like most reinforcement learning-based control methods, the convergence of the critic network weights relies on the persistence of excitation (PE) condition. In highly stable hovering scenarios with minimal state variation, the learning process may slow down due to a lack of diverse data samples. Future work will also explore adding probing noise intelligently to maintain learning efficiency without disturbing the flight stability.

Author contributions

The conceptualization and design of this study were carried out by Z. Feng and R. Yang. Z. Feng was responsible for data collection and surveys. Data visualization and drafting of the initial manuscript were completed by Z. Feng. R. Yang rigorously reviewed and edited the manuscript, providing critical theoretical guidance. Project supervision and funding acquisition were handled by Renming Yang. All authors approved the final version of the paper.

Use of Generative-AI tools declaration

The authors declare that they have not used Artificial Intelligence (AI) tools in the creation of this article.

Acknowledgments

This article received support from the National Natural Science Foundation of China (G61773015), the Natural Science Foundation of Shandong Province (ZR2023MF019), and the Research Leader Studio of Jinan (202228119).

Conflict of interest

All authors declare no conflicts of interest in this paper.

References

1. B. Tian, J. Cui, H. Lu, L. Liu, Q. Zong, Attitude control of UAVs based on event-triggered supertwisting algorithm, *IEEE T. Ind. Inform.*, **17** (2021), 1029–1038. https://doi.org/10.1109/TII.2020.2981367
2. X. Zhang, Y. Wang, G. Zhu, X. Chen, Z. Li, C. Wang, et al., Compound adaptive fuzzy quantized control for quadrotor and its experimental verification, *IEEE T. Cybernetics*, **51** (2021), 1121–1133. https://doi.org/10.1109/TCYB.2020.2987811
3. C. Li, Y. Wang, X. Yang, Adaptive fuzzy control of a quadrotor using disturbance observer, *Aerosp. Sci. Technol.*, **128** (2022), 107784. https://doi.org/10.1016/j.ast.2022.107784
4. Q. Chen, Y. Ye, Z. Hu, J. Na, S. Wang, Finite-time approximation-free attitude control of quadrotors: Theory and experiments, *IEEE T. Aero. Elec. Sys.*, **57** (2021), 1780–1792. https://doi.org/10.1109/TAES.2021.3050647

5. W. Zhu, L. Wang, B. Tian, Integral sliding mode control method for quadrotor unmanned aerial vehicle based on finite-time observer, *J. Chin. Inertial Technol.*, **31** (2023), 1244–1253. https://doi.org/10.13695/j.cnki.12-1222/o3.2023.12.011
6. K. Liu, P. Yang, R. Wang, L. Jiao, T. Li, J. Zhang, Observer-based adaptive fuzzy finite-time attitude control for quadrotor UAVs, *IEEE T. Aero. Elec. Sys.*, **59** (2023), 8637–8654. https://doi.org/10.1109/TAES.2023.3308552
7. K A. Alattas, M T. Vu, O. Mofid, F F. EI-Sousy, A. Fekih, S. Mobayen, Barrier function-based nonsingular finite-time tracker for quadrotor UAVs subject to uncertainties and input constraints, *Mathematics*, **10** (2022), 1659. https://doi.org/10.3390/math10101659
8. G. Cui, W. Yang, J. Yu, Z. Li, C. Tao, Fixed-time prescribed performance adaptive trajectory tracking control for a QUAV, *IEEE T. Circuits II*, **69** (2022), 494–498. https://doi.org/10.1109/TCSII.2021.3084240
9. B. Gao, Y. J. Liu, L. Liu, Fixed-time neural control of a quadrotor UAV with input and attitude constraints, *IEEE/CAA J. Automatic*, **10** (2023), 281–283. https://doi.org/10.1109/JAS.2023.123045
10. H. Jiang, Q. Ma, J. Guo, Fuzzy-based fixed-time attitude control of quadrotor unmanned aerial vehicle with full-state constraints: Theory and experiments, *IEEE T. Fuzzy Syst.*, **32** (2024), 1108–1115. https://doi.org/10.1109/TFUZZ.2023.3318572
11. H. Li, S. Zhao, W. He, R. Lu, Adaptive finite-time tracking control of full state constrained nonlinear systems with dead-zone, *Automatica*, **100** (2019), 99–107. https://doi.org/10.1016/j.automatica.2018.10.030
12. Y. Liu, X. Liu, Y. Jing, H. Wang, X. Li, Annular domain finite-time connective control for large-scale systems with expanding construction, *IEEE T. Syst. Man Cy. Syst.*, **51** (2021), 6159–6169. https://doi.org/10.1109/TSMC.2019.2960009
13. J. Yu, P. Shi, L. Zhao, Finite-time command filtered backstepping control for a class of nonlinear systems, *Automatica*, **92** (2018), 173–180. https://doi.org/10.1016/j.automatica.2018.03.033
14. X. Jin, Adaptive fixed-time control for MIMO nonlinear systems with asymmetric output constraints using universal barrier functions, *IEEE T. Automat. Contr.*, **64** (2019), 3046–3053. https://doi.org/10.1109/TAC.2018.2874877
15. D. Li, S S. Ge, T H. Lee, Fixed-time-synchronized consensus control of multiagent systems, *IEEE T. Control Netw.*, **8** (2021), 89–98. https://doi.org/10.1109/TCNS.2020.3034523
16. C. Hu, H. He, H. Jiang, Fixed/preassigned-time synchronization of complex networks via improving fixed-time stability, *IEEE T. Cybernetics*, **51** (2021), 2882–2892. https://doi.org/10.1109/TCYB.2020.2977934
17. K. Liu, R. Wang, S. Zheng, S. Dong, G. Sun, Fixed-time disturbance observer-based robust fault-tolerant tracking control for uncertain quadrotor UAV subject to input delay, *Nonlinear Dyn.*, **107** (2022), 2363–2390. https://doi.org/10.1007/s11071-021-07080-0
18. J. D. Sánchez-Torres, A. J. Muñoz-Vázquez, M. Defoort, E. Jiménez-Rodríguez, A. G. Loukianov, A class of predefined-time controllers for uncertain second-order systems, *Eur. J. Control*, **53** (2020), 52–58. https://doi.org/10.1016/j.ejcon.2019.10.003

19. E. Jiménez-Rodríguez, A. J. Muñoz-Vázquez, J. D. Sánchez-Torres, A. G. Loukianov, A note on predefined-time stability, *IFAC-PapersOnLine.*, **51** (2018), 520–525. https://doi.org/10.1016/j.ifacol.2018.07.332

20. A. K. Pal, S. Kamal, S. K. Nagar, B. Bandyopadhyay, L. Fridman, Design of controllers with arbitrary convergence time, *Automatica*, **112** (2020), 108710. https://doi.org/10.1016/j.automatica.2019.108710

21. J. Ni, L. Liu, Y. Tang, C. Liu, Predefined-time consensus tracking of second-order multiagent systems, *IEEE T. Syst. Man Cyb. Syst.*, **51** (2019), 2550–2560. https://doi.org/10.1109/TSMC.2019.2916257

22. P. Yang, S. Zhang, X. Yu, W. He, Reinforcement-Learning-Based Finite Time Fault Tolerant Control for a Manipulator With Actuator Faults, *IEEE T. Cybernetics*, **55** (2025), 2621–2632. https://doi.org/10.1109/TCYB.2025.3557681

23. Y. Gu, J. Zhao, Z Y. Sun, X. Xie, Reinforcement learning-based optimized multi-agent finite-time optimal synchronisation control and its application to the harmonic oscillator, *Nonlinear Dyn.*, **112** (2024), 13175–13188. https://doi.org/10.1007/s11071-024-09758-7

24. X. Bu, M. Lv, H. Lei, Discrete-time optimal control ensuring fixed-time prescribed performance for SSP, *IEEE T. Aero. Elec. Sys.*, **61** (2025), 3398–3407. https://doi.org/10.1109/TAES.2024.3485012

25. H. Yue, J. Xia, J. Zhang, J. H. Park, X. Xie, Event-based adaptive fixed-time optimal control for saturated fault-tolerant nonlinear multiagent systems via reinforcement learning algorithm, *Neural Netw.*, **183** (2025), 106952. https://doi.org/10.1016/j.neunet.2024.106952

26. L. Jin, L. Su, X. Fei, K. Wang, H. Shen, Adaptive neural inverse optimal control for uncertain switched nonlinear systems: an improved predefined-time method, *J. Control Decis.*, **2024** (2024), 1–15. https://doi.org/10.1080/23307706.2024.2400686

27. H. Shen, G. Cui, Z. Li, Q. Chang, Adaptive predefined-time optimal tracking control of flexible-joint robots, *IEEE Internet Things*, **12** (2025), 53132–53142. https://doi.org/10.1109/JIOT.2025.3616620

28. T. Bian, Z P. Jiang, Reinforcement learning and adaptive optimal control for continuous-time nonlinear systems: A value iteration approach, *IEEE T. Neur. Net. Lear.*, **33** (2022), 2781–2790. https://doi.org/10.1109/TNNLS.2020.3045087

29. K. Liu, Y. Yang, H. Lyu, W. Yang, Y. Zhang, Z. Tan, et al., Adaptive predefined-time disturbance observer-based fast nonsingular sliding mode control strategy for consumer quadrotor UAVs: Theory and experiments, *IEEE T. Consum. Electr.*, **71** (2025), 10629–10641. https://doi.org/10.1109/TCE.2025.3615655

30. K. Liu, W. Yang, L. Jiao, Z. Yuan, C Y. Wen, Fast fixed-time distributed neural formation control-based disturbance observer for multiple quadrotor UAVs under unknown disturbances, *IEEE T. Aero. Elec. Sys.*, **61** (2025), 13137–13155. https://doi.org/10.1109/TAES.2025.3578046

31. J. J. Xiong, Y. Chen, BFNN-based parameter adaptive sliding mode control for an uncertain TQUAV with time-varying mass, *Int. J. Robust Nonlin.*, **35** (2025), 4658–4668. https://doi.org/10.1002/rnc.7932

32. K A. Alattas, M T. Vu, O. Mofid, F F M. El-Sousy, A K. Alanazi, J. Awrejcewicz, et al., Adaptive nonsingular terminal sliding mode control for performance improvement of perturbed nonlinear systems, *Mathematics*, **10** (2022), 1064. https://doi.org/10.3390/math10071064

33. A. Najafi, M T. Vu, S. Mobayen, J. H. Asad, A. Fekih, Adaptive barrier fast terminal sliding mode actuator fault tolerant control approach for quadrotor UAVs, *Mathematics*, **10** (2022), 3009. https://doi.org/10.3390/math10163009

34. S. Li, S. Wang, Y. Zhang, X. Wang, Y. Zhang, W. Wu, et al., Distributed bearing-based fault-tolerant formation control of fixed-wing UAV swarm with prescribed performance, *Aerosp. Sci. Technol.*, **168** (2026), 110897. https://doi.org/10.1016/j.ast.2025.110897

35. Y. Fu, B. Wang, H. Zhao, M. Zhou, N. Li, Z. Gao, Adaptive safety attitude control of a hybrid VTOL UAV under transition flight subject to multiple faults and uncertainties, *Aerosp. Sci. Technol.*, **163** (2025), 110284. https://doi.org/10.1016/j.ast.2025.110284

36. J. Gu, Y. Wang, A constrained reinforcement learning based approach for cooperative control of multi-UAV in dense obstacle environments, *Sci. China Technol. Sci.*, **69** (2026), 1120601. https://doi.org/10.1007/s11431-025-3076-2

37. Y. Feng, Y. Zhou, H W. Ho, Performance-guaranteed quadrotor control with incremental adaptive dynamic programming and disturbance compensation, *Eng. Appl. Artif. Intel.*, **163** (2026), 113127. https://doi.org/10.1016/j.engappai.2025.113127

38. X. Song, P. Sun, S. Song, V. Stojanovic, Saturated-threshold event-triggered adaptive global prescribed performance control for nonlinear Markov jumping systems and application to a chemical reactor mode, *Expert Syst. Appl.*, **249** (2024), 123490. https://doi.org/10.1016/j.eswa.2024.123490

39. Z. Peng, X. Song, S. Song, I. Tejado, V. Stojanovic, Resilient consensus tracking control and vibration suppression of moving multiple helicopter flexible slung-load systems, *J. Vib. Control*, **2025** (2025), 10775463251321551. https://doi.org/10.1177/10775463251321551

40. H. K. Khalil, J. W. Grizzle, *Nonlinear systems*, Upper Saddle River, NJ: Prentice hall, 2002.

41. F L. Lewis, D. Vrabie, K. G. Vamvoudakis, Reinforcement learning and feedback control: Using natural decision methods to design optimal adaptive controllers, *IEEE Control Syst. Mag.*, **32** (2012), 76–105. https://doi.org/10.1109/MCS.2012.2214134

Appendix

To make the completion-of-square step (3.24)–(3.25) explicit, define $b(z) := \frac{1}{2}R^{-1}a(z)$. Then

$$\begin{aligned}(u_0 + b)^T R(u_0 + b) &= u_0^T Ru_0 + u_0^T Rb + b^T Ru_0 + b^T Rb \\ &= u_0^T Ru_0 + 2u_0^T Rb + b^T Rb,\end{aligned} \tag{5.1}$$

where we used $R = R^T$ and the scalar identity $b^T Ru_0 = u_0^T Rb$. Substituting $b = \frac{1}{2}R^{-1}a$ gives

$$2u_0^T Rb = 2u_0^T R\Big(\tfrac{1}{2}R^{-1}a\Big) = u_0^T a = a^T u_0, \quad b^T Rb = \tfrac{1}{4}a^T R^{-1}a. \tag{5.2}$$

Hence,

$$(u_0 + \tfrac{1}{2}R^{-1}a)^T R(u_0 + \tfrac{1}{2}R^{-1}a) = u_0^T R u_0 + a^T u_0 + \tfrac{1}{4}a^T R^{-1} a, \tag{5.3}$$

which implies

$$\Theta(z, u_0) = a^T u_0 + u_0^T R u_0 = (u_0 + \tfrac{1}{2}R^{-1}a)^T R(u_0 + \tfrac{1}{2}R^{-1}a) - \tfrac{1}{4}a^T R^{-1} a. \tag{5.4}$$

This proves that (3.24) and (3.25) are equivalent.

Proof of Lemma 3.4. To prove Lemma 3.4, one needs to show that the residual set of (3.59) holds. Therefore, we divide into three cases to obtain the residual subset separately. First, based on (3.58), we have

$$\dot{V} \leq -\frac{2}{\alpha\beta T_c}(2\gamma\beta V + V^{1-\frac{\alpha}{2}} + \beta^2 V^{1+\frac{\alpha}{2}}) + (\theta - \frac{4(1-\gamma)V}{\alpha T_c}). \tag{5.5}$$

From Eq (5.5), if

$$\theta - \frac{4(1-\gamma)V}{\alpha T_c} < 0, \tag{5.6}$$

which implies that $V(x) > \frac{\theta\alpha T_c}{4(1-\gamma)}$, then $\dot{V} \leq -\frac{2}{\alpha\beta T_c}(2\gamma\beta V + V^{1-\frac{\alpha}{2}} + \beta^2 V^{1+\frac{\alpha}{2}})$ holds.

Second, it is obvious that

$$\dot{V} \leq -\frac{2}{\alpha\beta T_c}(2\beta V + \gamma V^{1-\frac{\alpha}{2}} + \beta^2 V^{1+\frac{\alpha}{2}}) + (\theta - \frac{2(1-\gamma)V^{1-\frac{\alpha}{2}}}{\alpha\beta T_c}). \tag{5.7}$$

Equation (5.7) gives $\dot{V} \leq -\frac{2}{\alpha\beta T_c}(2\beta V + \gamma V^{1-\frac{\alpha}{2}} + \beta^2 V^{1+\frac{\alpha}{2}})$ if

$$\theta - \frac{2(1-\gamma)V^{1-\frac{\alpha}{2}}}{\alpha\beta T_c} < 0, \tag{5.8}$$

and then $V^{1-\frac{\alpha}{2}}(x) > \frac{\theta\alpha\beta T_c}{2(1-\gamma)}$.

Third, based on (3.58), we have

$$\dot{V} \leq -\frac{2}{\alpha\beta T_c}(2\beta V + V^{1-\frac{\alpha}{2}} + \gamma\beta^2 V^{1+\frac{\alpha}{2}}) + (\theta - \frac{2(1-\gamma)\beta V^{1+\frac{\alpha}{2}}}{\alpha T_c}), \tag{5.9}$$

from which, one gets $\dot{V} \leq -\frac{2}{\alpha\beta T_c}(2\beta V + V^{1-\frac{\alpha}{2}} + \gamma\beta^2 V^{1+\frac{\alpha}{2}})$, if

$$\theta - \frac{2(1-\gamma)\beta V^{1+\frac{\alpha}{2}}}{\alpha T_c} < 0. \tag{5.10}$$

Namely, $V^{1+\frac{\alpha}{2}}(x) > \frac{\theta\alpha T_c}{2\beta(1-\gamma)}$. From the above (5.6), (5.8), and (5.10), if at least one case is true, then $\dot{V} < 0$ holds, which implies that the system $\dot{x} = \phi(x, t)$ is strictly decreasing unless $\left\{x \mid V(x) \leq \frac{\theta\alpha T_c}{2(1-\gamma)}\right\}$, $\left\{x \mid V^{1-\frac{\alpha}{2}}(x) \leq \frac{\theta\alpha T_c}{2(1-\gamma)}\right\}$, and $\left\{x \mid V^{1+\frac{\alpha}{2}}(x) \leq \frac{\theta\alpha T_c}{2(1-\gamma)}\right\}$. Thus, the residual states set of the state trajectories can be represented as

$$\left\{\lim_{t\to T_c'} x \mid V(x) \leq min\left\{\frac{\theta\alpha T_c}{4(1-\gamma)}, (\frac{\theta\alpha\beta T_c}{2(1-\gamma)})^{\frac{2}{2-\alpha}}, (\frac{\theta\alpha T_c}{2\beta(1-\gamma)})^{\frac{2}{2+\alpha}}\right\}\right\}. \tag{5.11}$$

Now, we give the practically predefined-time stabilization time function.

Noting (5.5) and (5.6), it is easy to obtain

$$-\frac{2}{\alpha\beta T_c}(2\beta V + V^{1-\frac{\alpha}{2}} + \beta^2 V^{1+\frac{\alpha}{2}}) + \theta \leq -\frac{2}{\alpha\beta T_c}(2\gamma\beta V + V^{1-\frac{\alpha}{2}} + \beta^2 V^{1+\frac{\alpha}{2}}). \tag{5.12}$$

Then integrating both sides of (3.58) from $t_0 = 0$ to t_f, one gets:

$$\begin{aligned}
T_2(x_0) &\leq \int_0^{V_0} \frac{1}{\frac{2}{\alpha\beta T_c}(\,2\beta V + V^{1-\frac{\alpha}{2}} + \beta^2 V^{1+\frac{\alpha}{2}}\,) - \theta} dV \\
&\leq \frac{\alpha\beta T_c}{2} \int_0^{V_0} \frac{1}{(\,2\gamma\beta V + V^{1-\frac{\alpha}{2}} + \beta^2 V^{1+\frac{\alpha}{2}}\,)} dV \\
&\leq \frac{\alpha\beta T_c}{2\gamma} \int_0^{V_0} \frac{1}{(\,2\beta V + V^{1-\frac{\alpha}{2}} + \beta^2 V^{1+\frac{\alpha}{2}}\,)} dV \\
&\leq \int_0^{V_0} \frac{\frac{T_c}{\gamma}\frac{\alpha}{2}\beta V^{\frac{\alpha}{2}-1} dV}{2\beta V^{\frac{\alpha}{2}} + 1 + \beta^2 V^{\alpha}} \\
&\leq \frac{T_c}{\gamma} \int_0^{V_0} \frac{d\beta V^{\frac{\alpha}{2}}}{(\,1 + \beta V^{\frac{\alpha}{2}}\,)^2} \\
&\leq -\frac{T_c}{\gamma} \frac{1}{(\,1 + \beta V^{\frac{\alpha}{2}}\,)} \Big|_0^{V_0} \\
&\leq \frac{1}{\gamma} T_c (\,1 - \frac{1}{1 + \beta V_0^{\frac{\alpha}{2}}}\,)
\end{aligned} \tag{5.13}$$

Similarly, from (5.7)–(5.10), the following inequalities can be obtained:

$$-\frac{2}{\alpha\beta T_c}(2\beta V + V^{1-\frac{\alpha}{2}} + \beta^2 V^{1+\frac{\alpha}{2}}) + \theta \leq -\frac{2}{\alpha\beta T_c}(2\beta V + \gamma V^{1-\frac{\alpha}{2}} + \beta^2 V^{1+\frac{\alpha}{2}}), \tag{5.14}$$

and

$$-\frac{2}{\alpha\beta T_c}(2\beta V + V^{1-\frac{\alpha}{2}} + \beta^2 V^{1+\frac{\alpha}{2}}) + \theta \leq -\frac{2}{\alpha\beta T_c}(2\beta V + V^{1-\frac{\alpha}{2}} + \gamma\beta^2 V^{1+\frac{\alpha}{2}}). \tag{5.15}$$

According to (5.14) and (5.15), one can get the same outcome as (5.13), where $0 < \gamma < 1$, and the stabilization time function $T_2(x_0) \leq T_c^{'} = \frac{1}{\gamma} T_c$.

From (5.11), since $V(x)$ is continuous, we obtain that there exists some positive constant number δ such that $\|x\| \leq \delta$ holds, which implies that the system $\dot{x} = \phi(x, t)$ is practical predefined-time stable. This completes the proof. □

The proof of Theorem 3.5 is next.

Proof. Let w^* be the ideal bounded weights such that the optimal value function is $V^*(z) = (w^*)^T \Phi(z) + \epsilon(z)$, where $\epsilon(z)$ is the reconstruction error. Let the Lyapunov function candidate be:

$$L_{\hat{w}} = \frac{1}{2} \tilde{w}^T \alpha_2^{-1} \tilde{w}, \tag{5.16}$$

where $\tilde{w} = \hat{w} - w^*$ is the weight error and $\alpha_2 > 0$ is the learning rate parameter.

Substituting the gradient-based update law derived from minimizing the Bellman residual (Hamiltonian error) $E = \frac{1}{2}e_H^2$, we have $\dot{\hat{w}} = -\alpha_2 \frac{\partial E}{\partial \hat{w}}$. Utilizing the chain rule and the normalized gradient descent form:

$$\dot{L}_{\hat{w}} = -\tilde{w}^T \frac{\sigma\sigma^T}{m^2}\tilde{w} + \tilde{w}^T \Xi, \tag{5.17}$$

where $\sigma = \Phi(z)$, m is the normalization term, and Ξ represents the bounded terms related to the residual error $\epsilon(z)$ and the disturbance compensation error.

Under the PE condition, the matrix $\sigma\sigma^T$ is positive definite. Let matrix $Q_1 := \frac{\sigma\sigma^T}{m^2}$. For the quadratic form $-\tilde{w}^T Q\tilde{w}$, we have $-\tilde{w}^T Q\tilde{w} \leq -\lambda_{\min}(Q)\|\tilde{w}\|^2$, where $\lambda_{\min}(Q)$ is the smallest eigenvalue of matrix Q (strictly speaking, it is a positive number related to the intensity of the excitation). Thus, applying Young's inequality ($\tilde{w}^T\Xi \leq \|\tilde{w}\| \cdot \|\Xi\| \leq \frac{1}{2}\|\tilde{w}\|^2 + \frac{1}{2}\|\Xi\|^2$), it follows that:

$$\begin{aligned} \dot{L}_{\tilde{w}} &\leq -\lambda_{\min}(Q)\|\tilde{w}\|^2 + \left(\frac{1}{2}\|\tilde{w}\|^2 + \frac{1}{2}\|\Xi\|^2\right) \\ &\leq -\left(\lambda_{\min}(Q) - \frac{1}{2}\right)\|\tilde{w}\|^2 + \frac{1}{2}\|\Xi\|^2 \\ &\leq -\lambda_{\min}\|\tilde{w}\|^2 + C, \end{aligned} \tag{5.18}$$

where $\lambda_{\min} = \lambda_{\min}(Q) - \frac{1}{2}$ depends on the minimum eigenvalue of the regressor matrix and $C = \frac{1}{2}\|\Xi\|^2$ is a bounded constant. This inequality implies $\dot{L}_{\hat{w}} < 0$ whenever $\|\tilde{w}\| > \sqrt{C/\lambda_{\min}}$, proving that the estimation error $\tilde{w}$ converges to a compact set around zero UUB. □