Reinforcement learning based - sliding mode control for trajectory tracking of quadrotor unmanned aerial vehicles under disturbances

Tran Thai Duong[1], Do Duc Manh[1], Nguyen Chi Nhan[1],
Le Duc Thinh[2], Nguyen Tung Lam[1], Nguyen Danh Huy[1*]

[1]School of Electrical and Electronic Engineering, Hanoi University of Science and Technology, 1 Dai Co Viet, Hai Ba Trung, Hanoi, Vietnam;
[2]Faculty of Electrical and Electronics Engineering, Thuyloi University, 175 Tay Son, Dong Da, Hanoi, Vietnam.
[*]Corresponding author: huy.nguyendanh@hust.edu.vn
Received 4 Sep. 2024; Revised 15 Nov.2024; Accepted 5 Feb. 2025; Published 25 Feb. 2025.
DOI: https://doi.org/10.54939/1859-1043.j.mst.101.2025.39-46

ABSTRACT

In this article, a reinforcement learning (RL)-based sliding mode control (SMC) is proposed for trajectory tracking of a quadrotor unmanned aerial vehicle (QUAV) under external disturbances. First, an actor-critic RL framework sliding mode control is provided to tackle the optimal control problem without external disturbances. Secondly, the simulation in an environment with disturbances is carried out to show the robustness of the proposed controller. Theoretical analysis shows that the position and attitude tracking errors converge to a preset region, and the weight estimation errors of the actor-critic networks are uniformly ultimately bounded (UUB). Finally, a comparison of the numerical simulations between the proposed controller and traditional sliding mode controller and the Backstepping (BSP) technique is provided to indicate the advantages and improved performance of the RL-based SMC.

Keywords: Reinforcement learning; Sliding mode control; Optimal control; Actor/critic structure; Quadrotor unmanned aerial vehicle (QUAV).

1. INTRODUCTION

Unmanned aerial vehicles (UAVs) have recently garnered significant interest and attention due to their extensive applications in logistics delivery, environmental monitoring, and various civil and military uses [1]. Quadrotor UAVs, in particular, are popular among users because of their simple structure, flexible flight capabilities, strong environmental adaptability, high maneuverability, and numerous other advantages [2]. Compared to other UAVs, quadrotor UAVs excel in vertical takeoff and landing.

However, designing a control scheme for a quadrotor UAV under external disturbances is challenging. The system's nonlinearity and strong coupling characteristics further complicate achieving high-precision control. Various control schemes, such as adaptive control, sliding mode control (SMC), and robust control, have been developed to address control issues involving disturbances and uncertainties [3]. Among these, SMC offers a faster response [4]. It is important to note that the energy resources UAVs consume are valuable and limited. Thus, designing a high-precision control scheme for a UAV that balances control performance and control cost is crucial.

From a mathematical perspective, establishing the Hamilton–Jacobi–Bellman (HJB) function and its solution is essential for solving optimal control problems [5]. However, deriving an analytical solution from the HJB function is challenging due to the non-linear continuous-time systems [6]. To address this, Werbos [7] initially developed a reinforcement learning (RL) scheme with an actor-critic (AC) structure. In this scheme, a critic network approximates the value function, while an actor network determines the optimal control policy. Building on Werbos' work, Vamvoudakis et al. created an online AC algorithm to solve the continuous-time infinite horizon optimal control problem [8]. Ma et al. developed a learning-based adaptive sliding mode control scheme for a tethered space robot with limited inputs [9]. Given the unknown disturbances, accurately

determining the system dynamics for UAVs is practically difficult, limiting the methods' applicability. Thus, enhancing the robustness of these methods for practical systems with unknown disturbances remains a further challenge.

Motivated by these considerations, a novel RL-based tracking controller for quadrotors that accounts for external disturbances is being investigated. The main contributions and features of the proposed method can be summarized as follows:

- An online RL-based, nearly optimal controller is developed by combining the actor-critic (AC) structure with the hyperbolic tangent performance index. Additionally, the robustness of the learning-based controller is enhanced by integrating a super-twisting-like sliding mode control.

- Comparative numerical simulations demonstrate that the proposed method can enhance performance in terms of transient response and steady-state accuracy while requiring less control effort.

2. PROBLEM

2.1. Dynamic models and problem statement

2.1.1. Dynamic model

The Quad-rotor UAV is considered under two reference frames. The body-fixed reference frame is denoted by $\mathbf{B}=\{O_b,x_b,y_b,z_b\}$, and $\mathbf{E}=\{O_e,x_e,y_e,z_e\}$ is the earth-fixed reference frame. In $\mathbf{B}$ frame, $\mathbf{P}=[x,y,z]^T$ is the position vector, and $\mathbf{\Theta}=[\phi,\theta,\psi]^T$ is Euler angles attitude vector (where ϕ is roll angle, θ is pitch angle, and ψ is yaw angles). The linear velocity vector of the QUAV in the earth-fixed frame and the angular velocity vector of QUAV in the body-fixed frame are denoted as $\mathbf{V}=[v_x,v_y,v_z]^T$ and $\boldsymbol{\omega}=[\omega_\phi,\omega_\theta,\omega_\psi]^T$. The position and attitude dynamics of UAVs are considered as follows:

$$\begin{cases}\dot{\mathbf{P}}=\mathbf{V}\\ m\dot{\mathbf{V}}=\mathbf{RF}-\mathbf{KV}-mg\mathbf{E}_3+\mathbf{d}\end{cases} \tag{1}$$

$$\begin{cases}\dot{\mathbf{\Theta}}=\mathbf{T}\boldsymbol{\omega}\\ \mathbf{I}\dot{\boldsymbol{\omega}}=-\boldsymbol{\omega}\times(\mathbf{I}\boldsymbol{\omega})+\mathbf{F}_{\mathbf{\Theta}}-\mathbf{G}-\mathbf{K}_{\mathbf{\Theta}}\dot{\mathbf{\Theta}}+\mathbf{d}_{\mathbf{\Theta}}\end{cases} \tag{2}$$

where $\mathbf{R}$ is the kinematic rotation matrix, $\mathbf{F}=[0,0,u_m]^T$ is the force vector in $\mathbf{B}$ frame. $\mathbf{K}=\mathrm{diag}(k_x,k_y,k_z)$ is the aerodynamic matrix, m is the mass of the QUAV, $\mathbf{E}_3$ is the unit vector defined as $\mathbf{E}_3=[0,0,1]^T$, and $\mathbf{d}=[d_x,d_y,d_z]^T$ is the disturbance vector. $\mathbf{T}$ is the attitude kinematic transformation matrix, $\mathbf{I}=\mathrm{diag}(I_x,I_y,I_z)$ is the inertia matrix, $\mathbf{F}_{\mathbf{\Theta}}=[u_\phi,u_\theta,u_\psi]^T$ is the attitude control torque vector. $\mathbf{G}=\left(\sum_{i=1}^{4}(-1)^i w_i\right)I_i(\boldsymbol{\omega}\times\mathbf{E}_3)$ is the gyroscopic torque vector, $\mathbf{K}_{\mathbf{\Theta}}=\mathrm{diag}(k_\phi,k_\theta,k_\psi)$ is the attitude aerodynamic matrix. $\mathbf{d}_{\mathbf{\Theta}}=[d_\phi,d_\theta,d_\psi]^T$ is the attitude disturbance vector. The kinematic rotation matrix $\mathbf{R}$ and the attitude kinematic transformation matrix are defined as:

$$\mathbf{R}=\begin{bmatrix}c_\psi c_\theta & c_\psi s_\theta s_\phi-s_\psi c_\phi & c_\psi s_\theta c_\phi+s_\psi s_\phi\\ s_\psi c_\theta & s_\psi s_\theta s_\phi+c_\psi c_\phi & s_\psi s_\theta c_\phi-c_\psi s_\phi\\ -s_\theta & c_\theta s_\phi & c_\theta c_\phi\end{bmatrix} \tag{3}$$

$$\mathbf{T}=\begin{bmatrix}1 & s_\phi \tan\theta & c_\phi \tan\theta\\ 0 & c_\phi & -s_\phi\\ 0 & s_\phi \sec\theta & c_\phi \sec\theta\end{bmatrix} \tag{4}$$

Let $\mathbf{f}=\left(-\mathbf{K}\mathbf{V}-g\mathbf{E}_3\right)/m$, (1) can be written as [10]:

$$\begin{cases}\dot{\mathbf{P}}=\mathbf{V}\\ \dot{\mathbf{V}}=\mathbf{R}\mathbf{F}/m-\mathbf{f}+\mathbf{d}/m\end{cases} \tag{5}$$

Let $\mathbf{f}_\Theta=\mathbf{T}^{-1}\mathbf{I}^{-1}\left(-\boldsymbol{\omega}\times\left(\mathbf{I}\boldsymbol{\omega}\right)-\mathbf{G}-\mathbf{K}_{\boldsymbol{\Theta}}\dot{\boldsymbol{\Theta}}\right)$, equation (2) can be written as [10]:

$$\begin{cases}\dot{\boldsymbol{\Theta}}=\boldsymbol{\Omega}\\ \boldsymbol{\Omega}=\mathbf{f}_\Theta+\mathbf{T}^{-1}\mathbf{I}^{-1}\mathbf{F}_{\boldsymbol{\Theta}}+\mathbf{T}^{-1}\mathbf{I}^{-1}\mathbf{d}_{\boldsymbol{\Theta}}\end{cases} \tag{6}$$

where $\boldsymbol{\Omega}=\left[\dot{\phi},\dot{\theta},\dot{\psi}\right]^T$ is the angular velocity vector of QUAV in the body-fixed frame.

Assumption 1: UAV can be considered a rigid body in flight missions, and its rotation is limited as $-\pi/2<\phi<\pi/2$, $-\pi/2<\theta<\pi/2$, $-\pi\le\psi\le\pi$.

Assumption 2: The desired reference $\mathbf{P}_d=\left[x_d,y_d,z_d\right]^T$, ψ_d, its first and second derivations $\dot{\mathbf{P}}_d$, $\dot{\psi}_d$, $\ddot{\mathbf{P}}_d$, $\ddot{\psi}_d$ are known and bounded. They are available signals in the UAV control.

Remark 1: In fact, the rotational velocities $\omega_i, i=\overline{1,4}$ of four rotors are the direct control input of QUAV. The relationship between both total thrust u_m and control toque u_ϕ,u_θ,u_ψ:

$$\begin{aligned}&u_m=k_t\left(w_1^2+w_2^2+w_3^2+w_4^2\right);u_\phi=k_t l\left(w_2^2-w_4^2\right)\\ &u_\theta=k_t l\left(-w_1^2+w_3^2\right);u_\psi=k_d\left(w_1^2-w_2^2+w_3^2-w_4^2\right)\end{aligned} \tag{7}$$

where k_t, k_d, and w_i are the thrust factor, drag factor, and rotational velocity of the ith rotor, respectively, l is the distance from the rotors to the mass center.

2.1.2. Preliminaries

Lemma 1 [11]. Considering a nonlinear function $f\left(x\right)=\ln\left[1-\tanh^2\left(x\right)\right]$, the equation: $f\left(x\right)=\ln\left(4\right)-2x\,\mathrm{sign}\left(x\right)+k_f$ always holds, k_f is bounded by a real positive constant.

Lemma 2 [12]. For any $\alpha>0$, the following inequality holds $0\le|x|-x\tanh\left(\frac{x}{\alpha}\right)\le k\alpha$. Where k satisfies $k=e^{-k+1}$.

2.2. Optimized UAV position control

In this subsection, the position control problem will be tackled with the RL algorithm, the attitude loop will be carried out by the sliding mode controller. For simplicity, the (5) can be rewritten as:

$$\begin{cases}\dot{\mathbf{P}}=\mathbf{V}\\ \dot{\mathbf{V}}=-\begin{bmatrix}0 & 0 & g\end{bmatrix}^T+\mathbf{U}\end{cases} \tag{8}$$

with $\mathbf{U}=\begin{bmatrix}U_1 & U_2 & U_3\end{bmatrix}^T=\mathbf{R}\mathbf{F}/m$. The relationship between U and F is:

$$\begin{aligned} U_1 &= (\mathrm{c}_\phi \mathrm{s}_\theta \mathrm{c}_\psi + s_\phi s_\psi) u_m / m \\ U_2 &= (\mathrm{c}_\phi \mathrm{s}_\theta \, s_\psi - s_\phi c_\psi) u_m / m \\ U_3 &= (\mathrm{c}_\phi \, c_\theta) u_m / m \end{aligned} \tag{9}$$

Therefore, the control signal u_m can be solved as:

$$u_m = m\sqrt{U_1^2 + U_2^2 + U_3^2} \tag{10}$$

with the desired position trajectory $\mathbf{P}_{re}(t) = \left[x_{re}(t),\, y_{re}(t),\, z_{re}(t) \right]^T$, the tracking error can be obtained as $\mathbf{e}_{p1} = \mathbf{P}(t) - \mathbf{P}_{re}(t)$, and $\mathbf{e}_{p2} = \mathbf{V}(t) - \dot{\mathbf{P}}_{re}(t)$, which yields:

$$\dot{\mathbf{e}}_{p1}(t) = \mathbf{e}_{p2}(t);\;\; \dot{\mathbf{e}}_{p2}(t) = \mathbf{U} - \begin{bmatrix} 0 & 0 & g \end{bmatrix}^T - \ddot{\mathbf{P}}_{re}(t) + \mathbf{d} / m \tag{11}$$

The sliding variable $\mathbf{s}$ is defined as:

$$\mathbf{s} = \mathbf{e}_{p2} + \mathbf{c}.\mathbf{e}_{p1}$$

where $\mathbf{c}$ is a positive diagonal matrix.

$$\dot{\mathbf{s}} = \mathbf{c}.\mathbf{e}_{p2}(t) - \begin{bmatrix} 0 & 0 & g \end{bmatrix}^T - \ddot{\mathbf{P}}_{re}(t) + \mathbf{U} + \mathbf{d} / m \tag{12}$$

The conventional super-twisting controller is designed as follows:

$$\mathbf{U} = -\left(\left(\mathbf{c}.\mathbf{e}_{p2} - \begin{bmatrix} 0 & 0 & g \end{bmatrix}^T - \ddot{\mathbf{P}}_{re}(t) \right) + \lambda_1 \operatorname{sig}(\mathbf{s}) - \mathbf{u}_d \right) \tag{13}$$

where $\operatorname{sig}(\mathbf{s}) = C\left[|s_i|^{1/2} \operatorname{sign}(s_i) \right]$; $\dot{\mathbf{u}}_d = -C\left[\lambda_2 \operatorname{sign}(s_i) \right]$. With the super-twisting sliding mode controller, the chattering phenomenon happening in traditional SMC can be attenuated thanks to the 2 continuous and smooth terms $\operatorname{sig}(\mathbf{s})$ and $\mathbf{u}_d = \int -C\left[\lambda_2 \operatorname{sign}(s_i) \right] dt$ instead of function $sign(s_i)$ that can discontinuously change [17]. However, the cost index cannot be optimized, to tackle this problem, a RL-based controller is provided to improve the performance, assuming there is a family of admissible control inputs:

$$\mathbf{U} = \mathbf{u}_0 - \lambda_1 \operatorname{sig}(\mathbf{s}) + \mathbf{u}_d; \tag{14}$$

Consider the following cost function:

$$V(s) = \int_t^\infty \left\{ V_s(\tau) + J\left[\mathbf{u}_0(\tau) \right] \right\} d\tau \tag{15}$$

where $V_s = \mathbf{s}^T \mathbf{Q} \mathbf{s}$ with $\mathbf{Q}$ is a positive diagonal matrix, $J(\mathbf{u}_0)$ is specified as:

$$J(\mathbf{u}_0) = 2\int_0^{\mathbf{u}_0} \left[\lambda_1 \tanh^{-1}\left(\frac{\mathbf{\upsilon}}{\lambda_{\partial 1}} \right)^T \boldsymbol{\omega} \right] d\mathbf{\upsilon} \tag{16}$$

where $\boldsymbol{\omega}$ is a positive diagonal matrix, the Lyapunov function of (15) is evaluated as:

$$V_s(t) + J\left[\mathbf{u}_0(t) \right] + \nabla V_s^T \dot{\mathbf{s}} = 0 \tag{17}$$

Therefore, the corresponding HJB equation can be formulated as:

$$\min_{\mathbf{u}_0} \left\{ V_s + J(\mathbf{u}_0) + \nabla V_s^{*T} \dot{\mathbf{s}} \right\} = 0 \tag{18}$$

the $\mathbf{u}_0$ that minimizes the cost function is the solution of the following equation:

$$\frac{\partial}{\partial \mathbf{u}_0^*}\left[V_s + J\left(\mathbf{u}_0^*\right) + \nabla V_s^{*T}\dot{\mathbf{s}}\right] = 0 \tag{19}$$

Solving (21) yields:

$$\mathbf{u}_0^* = -\lambda_1 \tanh\left(\bar{\mathbf{u}}_0^*\right), \bar{\mathbf{u}}_0^* = \frac{1}{2\lambda_1}\boldsymbol{\omega}^{-1}\nabla V_s^* \tag{20}$$

Substituting $\mathbf{u}_0^*$ into (16), the following equation can be obtained:

$$J\left(\mathbf{u}_0^*\right) = -\nabla V_s^{*T}\mathbf{I}_3\mathbf{u}_0^* + \lambda_1^2\bar{\boldsymbol{\omega}}\ln\left[\mathbf{1}_c - \tanh^2\left(\bar{\mathbf{u}}_0^*\right)\right] \tag{21}$$

From (19) and (21), it can be derived that:

$$V_s + \nabla V_s^{*T}\dot{\mathbf{s}} + \lambda_1^2\bar{\boldsymbol{\omega}}\ln\left[\mathbf{1}_c - \tanh^2\left(\bar{\mathbf{u}}_0^*\right)\right] = 0 \tag{22}$$

Nevertheless, it is impossible to obtain the analytical solutions of the HJB equations in (22), which are nonlinear partial differential equations. Based on the fine approximation ability of NN, two RL-based control policies will be proposed to deal with this intractable problem. The following approximation of V_s^* can be established

$$V^*\left(\mathbf{s}\right) = \mathbf{W}^{*T}\boldsymbol{\sigma}\left[\mathbf{s}(t)\right] + \varepsilon\left[\mathbf{s}(t)\right] \tag{23}$$

where $\boldsymbol{\sigma}\left[\mathbf{s}(t)\right]$ are the basis function vectors and $\mathbf{W}^*$ denotes the optimal weight vector, $\varepsilon\left[\mathbf{s}(t)\right]$ represent the approximation errors. Subsequently, the gradient of $V^*\left(\mathbf{s}\right)$ is

$$\nabla V_s^* = \left(\nabla\boldsymbol{\sigma}_s\left[\mathbf{s}(t)\right]\right)^T\mathbf{W}^* + \nabla\varepsilon_s\left[\mathbf{s}(t)\right] \tag{24}$$

where $\nabla\boldsymbol{\sigma}_s = \partial\boldsymbol{\sigma}/\partial\mathbf{s}, \nabla\varepsilon_s = \partial\varepsilon/\partial\mathbf{s}$. In light of the universal approximation property of NN for smooth functions on prescribed compact sets, the approximation errors $\varepsilon\left[\mathbf{s}(t)\right]$ and $\nabla\varepsilon\left[\mathbf{s}(t)\right]$ are bounded with a finite dimension of $\boldsymbol{\sigma}\left[\mathbf{s}(t)\right]$. Moreover, it is assumed that $\|\mathbf{W}^*\|$, $\boldsymbol{\sigma}\left[\mathbf{s}(t)\right]$ and $\nabla\boldsymbol{\sigma}\left[\mathbf{s}(t)\right]$ are bounded. Recalling (20), the NN-based nearly optimal control law can be formulated as:

$$\begin{cases} \mathbf{U} = \hat{\mathbf{U}}_0^* - \mathbf{I}_3^{-1}\left\{\lambda_1 sig\left(\mathbf{s}\right) - \mathbf{u}_d\right\} \\ \hat{\mathbf{U}}_0^* = -\lambda_1\tanh\left[\frac{1}{2\lambda_1}\boldsymbol{\omega}^{-1}\mathbf{I}_3^T\left(\nabla\boldsymbol{\sigma}_s^T\mathbf{W}^* + \nabla\varepsilon_s\right)\right] \end{cases} \tag{25}$$

Defining the Bellman error as:

$$B_e = \lambda_1^2\bar{\boldsymbol{\omega}}\left\{\ln\left[\mathbf{1}_c - \tanh^2\left(\bar{\mathbf{U}}_0^*\right)\right] - \ln\left[\mathbf{1}_c - \tanh^2\left(\hat{\bar{\mathbf{U}}}_0^*\right)\right]\right\} \tag{26}$$

with $\hat{\bar{\mathbf{U}}}_0^* = \frac{1}{2\lambda_1}\boldsymbol{\omega}^{-1}\mathbf{I}_3^T\nabla\boldsymbol{\sigma}_s^T\mathbf{W}^*$, (19) can be rewritten as:

$$V_s + \left(\nabla\boldsymbol{\sigma}_s\right)^T\mathbf{W}^*\boldsymbol{\Gamma} + \lambda_1^2\bar{\boldsymbol{\omega}}\ln\left[\mathbf{1}_c - \tanh^2\left(\hat{\bar{\mathbf{U}}}_0^*\right)\right] + \varepsilon_H = 0 \tag{27}$$

where $\varepsilon_H = \nabla\varepsilon_s^T\boldsymbol{\Gamma} + B_\varepsilon + \mathbf{W}^*\nabla\boldsymbol{\sigma}_s D_2$ is the bounded HJB error. 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:

$$\begin{cases} \mathbf{U} = \hat{\mathbf{U}}_a^* - \mathbf{I}_3^{-1}\left\{\lambda_1 sig\left(\mathbf{s}\right) - \mathbf{u}_d\right\} \\ \hat{\mathbf{U}}_a^* = -\lambda_1\tanh\left(\hat{\bar{\mathbf{U}}}_a^*\right), \hat{\bar{\mathbf{U}}}_a^* = \frac{1}{2\lambda_1}\boldsymbol{\omega}^{-1}\mathbf{I}_3^T\nabla\boldsymbol{\sigma}_s^T\hat{\mathbf{W}}_a \end{cases} \tag{28}$$

where $\hat{\mathbf{W}}_a$ is the weight of the actor network. Moreover, the performance index (15) can be

estimated as: $\hat{V}_s = \hat{\mathbf{W}}_c^T \boldsymbol{\sigma}_s \left[\mathbf{s}(t) \right]$. Where $\hat{\mathbf{W}}_c$ is the weight of the critic network. The adaptation laws for $\hat{\mathbf{W}}_c$ and $\hat{\mathbf{W}}_a$ are designed as:

$$\begin{aligned} \dot{\hat{\mathbf{W}}}_c &= -A_1 \left\{ \bar{\boldsymbol{\theta}} \left[\boldsymbol{\theta}^T \hat{\mathbf{W}}_c + V_s + J\left(\hat{\mathbf{U}}_a^* \right) \right] + p_c \left(\hat{\mathbf{W}}_c - \hat{\mathbf{W}}_a \right) \right\} \\ \dot{\hat{\mathbf{W}}}_a &= -A_2 \left[p_a \left(\hat{\mathbf{W}}_a - \hat{\mathbf{W}}_c \right) - \frac{\lambda_1 p_a}{p_c} \nabla \boldsymbol{\sigma}_s \mathbf{I}_3 \boldsymbol{\Psi} \bar{\boldsymbol{\theta}}^T \hat{\mathbf{W}}_c \right] \end{aligned} \tag{29}$$

where $\boldsymbol{\theta} = \nabla \boldsymbol{\sigma}_s \left(\boldsymbol{\Gamma} + \mathbf{I}_3 \hat{\mathbf{U}}_a^* \right)$, $\bar{\boldsymbol{\theta}} = \boldsymbol{\theta} / \left(\boldsymbol{\theta}^T \boldsymbol{\theta} + 1 \right)^2$, $\Psi = \tanh\left(\hat{\bar{\mathbf{U}}}_a^* \right) - \tanh\left(\frac{\hat{\bar{\mathbf{U}}}_a^*}{\kappa} \right)$, p_c, p_a, $k > 0$, A_1, A_2 are positive diagonal matrices. Assuming $\boldsymbol{\theta}_1 = \boldsymbol{\theta} / \left(\boldsymbol{\theta}^T \boldsymbol{\theta} + 1 \right)$ is persistently excited.

Theorem 1. Considering the systems in (8), if the control policies are designed as (28), with the weight update rules in (29), then the weight errors and the tracking errors will be guaranteed to be ultimately uniformly bounded under the proposed RL-based control policies. Based on SMC, attitude control for the angle can be calculated as:

$$\mathbf{F}_\Theta = \mathbf{B}^{-1} \left(-\varepsilon \text{sat}(\mathbf{s}) - k\mathbf{s} - c\dot{\mathbf{e}} - \mathbf{f}_\Theta + \ddot{\mathbf{n}}_{re} \right) \tag{30}$$

If the control policies are designed as (30), the angular errors e_ϕ, e_θ, e_ψ will converge to the origin. Therefore, the angles converge to the reference values.

3. NUMERICAL SIMULATION RESULTS

3.1. Simulation setting

To demonstrate the performance of the control system, Matlab/Simulink enviroment is employed, parameters of the drone is given as follows: $m = 1.776$ kg, $g = 9.81$ m/s^2, $\mathbf{I} = \text{diag}\left(I_x, I_y, I_z \right) = \text{diag}\left(35, 35, 55 \right) \times 10^{-4}$ kg.m^2, $l = 0.225$ m, $I = 2.8 \times 10^{-6}$ kg.m^2, Drag coefficient $k_c = 5.567 \times 10^{-4} s^{-1}$ [14]. The following disturbances affecting the system is provided: $D = -5\left(0.8\sin\left(0.7t \right) + 0.3\sin\left(t \right) + 0.15\sin\left(2t \right) + 0.055\sin\left(\pi t / 2 \right) \right)$

3.2. Simulation results

In this subsection, a comparison between the proposed controller and the other 2 controllers, which are the sliding mode controller (SMC) and the Backstepping-SMC (BSP) is provided. As can be seen in figures 1-3, the trajectory error controlled by RL-SMC is the smallest among the 3 controllers, the error along the x-axis and y-axis being approximately 10^{-4} and along the z-axis being 10^{-5}.

In comparison in figure 5, the other 2 controllers are affected by the disturbances and, therefore, can not converge to the origin. Regarding figure 4 yaw angles, the proposed controller also takes a shorter time to converge than the other 2, approximately 0.45 seconds. This study employs three performance indices to assess tracking performance in quantitation: ISE, IAE, and ITAE, as described in [15]. The ISEp and ISEa indices represent the integral squared errors for position and attitude, respectively. IAEp and IAEa are the integral absolute errors for position and attitude, respectively. The ITAEp and ITAEa indices, which measure the integral time-weighted absolute errors for position and attitude. Therefore, the following index table is obtained as in table 1. According to these performance indices shown in table 1, the position of the designed controller in this paper provides the smallest statistics, which show that the proposed controller can maintain its robustness and effectiveness.

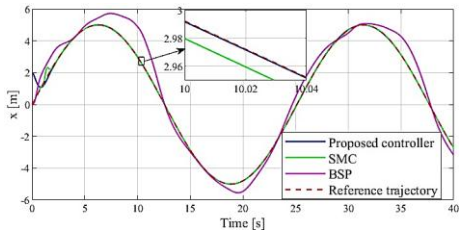

Figure 1. *Response of x position.*

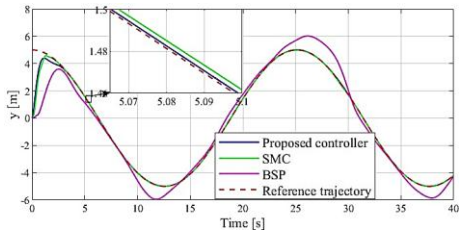

Figure 2. *Response of y position.*

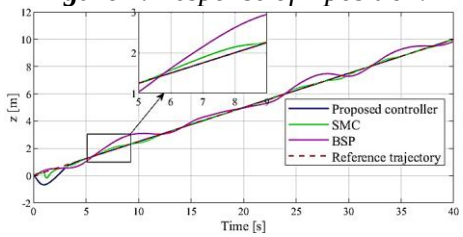

Figure 3. *Response of z position.*

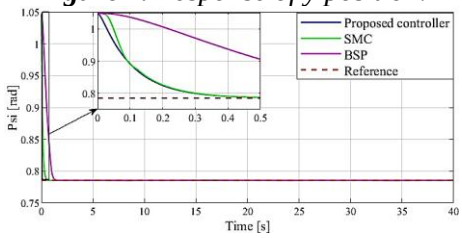

Figure 4. *Response of yaw angle.*

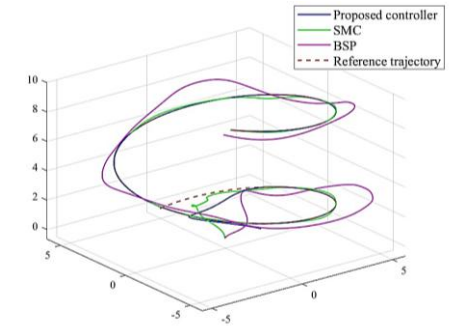

Figure 5. *Trajectory in 3D space.*

Table 1. *The performance index table of the 3 controllers.*

INDEX	ISE_p	ISE_a	IAE_p	IAE_a	$ITAE_p$	$ITAE_a$
SMC	0.081	14.923	0.559	8.795	7.535	71.995
BSP	2.179	59.801	7.558	55.624	137.48	1027.12
RL-SMC	1.923	**10.768**	0.62	**6.029**	0.206	**8.306**

4. CONCLUSIONS

In this article, the model of the UAV with disturbances is demonstrated, based on the combination of the RL and SMC for the position loop, the simulated results on Matlab-Simulink proved that the proposed controller can effectively control the UAV tracking the reference trajectory in the appearance of disturbances.

Given the constraints of time and financial resources, a real-life quadrotor system has not been built. As a result, future work could focus on developing a practical model to validate the proposed controllers and improve the control algorithm for the inner loop as well.

REFERENCES

[1]. Mohsan, Syed Agha Hassnain, et al. *"Unmanned aerial vehicles (UAVs): Practical aspects, applications, open challenges, security issues, and future trends."* Intelligent Service Robotics, 16.1, 109-137, (2023).

[2]. Liu, Hui, et al. *"Reinforcement learning-based tracking control for a quadrotor unmanned aerial vehicle under external disturbances."* International Journal of Robust and Nonlinear Control, 33.17, 10360-10377, (2023).

[3]. Li, Bo, et al. *"Fixed-time integral sliding mode control of a high-order nonlinear system."* Nonlinear Dynamics, 107, 909-920, (2022).
[4]. Qi, Wenhai, Guangdeng Zong, and Wei Xing Zheng. *"Adaptive event-triggered SMC for stochastic switching systems with semi-Markov process and application to boost converter circuit model."* IEEE Transactions on Circuits and Systems I: Regular Papers, 68.2, 786-796, (2020).
[5]. Yong, Jiongmin, et al. *"Dynamic programming and HJB equations."* Stochastic controls: Hamiltonian systems and HJB equations, 157-215, (1999).
[6]. Evans, L. C., and M. R. James. *"The Hamiltonian–Jacobi–Bellman equation for time-optimal control."* SIAM journal on control and optimization, 27.6, 1477-1489, (1989).
[7]. Werbos, Paul. *"Approximate dynamic programming for real-time control and neural modeling."* Handbook of intelligent control, (1992).
[8]. Vamvoudakis, Kyriakos G., and Frank L. Lewis. *"Online actor-critic algorithm to solve the continuous-time infinite horizon optimal control problem."* Automatica, 46.5, 878-888, (2010).
[9]. Ma, Zhiqiang, Panfeng Huang, and Yuxin Lin. *"Learning-based sliding-mode control for underactuated deployment of tethered space robot with limited input."* IEEE Transactions on Aerospace and Electronic Systems, 58.3, 2026-2038, (2021).
[10]. Modares, Hamidreza, Mohammad-Bagher Naghibi Sistani, and Frank L. Lewis. *"A policy iteration approach to online optimal control of continuous-time constrained-input systems."* ISA transactions, 52.5, 611-621, (2013).
[11]. Tan, Jian, and Shijun Guo. *"Backstepping control with fixed-time prescribed performance for fixed wing UAV under model uncertainties and external disturbances."* International Journal of Control, 95.4, 934-951, (2022).
[12]. Wen, Guoxing, et al. *"Optimized backstepping tracking control using reinforcement learning for quadrotor unmanned aerial vehicle system."* IEEE Transactions on Systems, Man, and Cybernetics: Systems, 52.8, 5004-5015, (2021).
[13]. Xu, Shihao, et al. *"Reinforcement-learning-based tracking control with fixed-time prescribed performance for reusable launch vehicle under input constraints."* Applied Sciences, 12.15, 7436, (2022).
[14]. Nguyen, H. S., et al. *"Advanced Motion Control of a Quadrotor Unmanned Aerial Vehicle based on Extended State Observer."* International Conference on System Science and Engineering. IEEE, (2023).
[15]. Liu, K., Wang, R., Zheng, S., Dong, S., & Sun, G. *"Fixed-time disturbance observer-based robust fault-tolerant tracking control for uncertain quadrotor UAV subject to input delay"*. Nonlinear Dynamics, 107(3), 2363-2390, (2022).
[16]. Chen, Fuyang, et al. *"Robust backstepping sliding-mode control and observer-based fault estimation for a quadrotor UAV."* IEEE Transactions on Industrial Electronics, 63.8, 5044-5056, (2016).
[17]. Tan, Lingwei, et al. *"Super-twisting sliding mode control with defined boundary layer for chattering reduction of permanent magnet linear synchronous motor."* Journal of Mechanical Science and Technology, 35: 1829-1840, (2021).

TÓM TẮT

Điều khiển trượt dựa trên học tăng cường trong điều khiển bám quỹ đạo của máy bay không người lái dưới các nhiễu bất định

Bài báo đề xuất phương pháp điều khiển trượt (SMC) dựa trên học tăng cường (RL) để điều khiển bám quỹ đạo của quadrotor UAV (QUAV) dưới tác động của nhiễu bên ngoài. Đầu tiên, một bộ điều khiển trượt dựa trên học tăng cường actor-critic (actor-critic RL) được giới thiệu để giải quyết bài toán điều khiển tối ưu trong điều kiện không có nhiễu. Tiếp theo, mô phỏng trong môi trường có nhiễu được thực hiện nhằm chứng minh tính bền vững của bộ điều khiển được đề xuất. Phân tích lý thuyết cho thấy sai số vị trí và góc của UAV hội tụ về một miền đặt trước, trong khi sai số ước lượng của mạng actor-critic được giới hạn cuối cùng một cách thống nhất (uniformly ultimately bounded - UUB). Cuối cùng, một phân tích so sánh mô phỏng số giữa bộ điều khiển đề xuất, bộ điều khiển trượt truyền thống và bộ điều khiển trượt kết hợp kỹ thuật Backstepping (BSP) được thực hiện để làm rõ các ưu điểm và hiệu suất cải thiện của SMC dựa trên RL.

Từ khóa: Học tăng cường (RL); Cấu trúc Actor/Critic; Điều khiển trượt (SMC); Điều khiển tối ưu; Máy bay không người lái 4 cánh (QUAV).