drones

MDPI

Article

Deep Deterministic Policy Gradient (DDPG) Agent-Based Sliding Mode Control for Quadrotor Attitudes

Wenjun Hu , Yueneng Yang * and Zhiyang Liu

College of Aerospace Science and Engineering, National University of Defense Technology, Changsha 410073, China; huwenjun@nudt.edu.cn (W.H.); liuzhiyang@nudt.edu.cn (Z.L.)
* Correspondence: yangyueneng@nudt.edu.cn; Tel.: +86-13548676452

Abstract: A novel reinforcement deep learning deterministic policy gradient agent-based sliding mode control (DDPG-SMC) approach is proposed to suppress the chattering phenomenon in attitude control for quadrotors, in the presence of external disturbances. First, the attitude dynamics model of the quadrotor under study is derived, and the attitude control problem is described using formulas. Second, a sliding mode controller, including its sliding mode surface and reaching law, is chosen for the nonlinear dynamic system. The stability of the designed SMC system is validated through the Lyapunov stability theorem. Third, a reinforcement learning (RL) agent based on deep deterministic policy gradient (DDPG) is trained to adaptively adjust the switching control gain. During the training process, the input signals for the agent are the actual and desired attitude angles, while the output action is the time-varying control gain. Finally, the trained agent mentioned above is utilized in the SMC as a parameter regulator to facilitate the adaptive adjustment of the switching control gain associated with the reaching law. The simulation results validate the robustness and effectiveness of the proposed DDPG-SMC method.

Keywords: quadrotor; attitude control; deep deterministic policy gradient; gain adjusted; sliding mode control

check for updates

Citation: Hu, W.; Yang, Y.; Liu, Z. Deep Deterministic Policy Gradient (DDPG) Agent-Based Sliding Mode Control for Quadrotor Attitudes. *Drones* **2024**, *8*, 95. https://doi.org/10.3390/drones8030095

Academic Editor: Abdessattar Abdelkefi

Received: 15 January 2024
Revised: 3 March 2024
Accepted: 7 March 2024
Published: 12 March 2024

Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

As an unmanned flight platform, a quadrotor UAV has the advantages of a simple structure, lightweight fuselage, and low cost. It is widely used in various tasks such as cargo transportation, aerial photography, agricultural plant protection, rescue and relief operations, remote-sensing mapping, and reconnaissance [1–4]. A wide range of application scenarios also impose strict requirements on its flight control capability, particularly the attitude control during UAV flight [4–6]. However, the lightweight fuselage of a quadrotor leads to its poor ability to resist external disturbances, which reduces the accuracy of its attitude control.

There have been many studies on attitude control methods for quadrotors. Some linear control methods such as proportional integral derivative (PID) control [7–9] and linear quadratic regulation [10] have been widely used in engineering practice, due to the advantages of their simple structure and easy implementation. The PID and LQ methods were applied for the attitude angle control of a micro quadrotor, and the control laws were validated through autonomous flight experiments in the presence of external disturbances [11]. A robust PID control methodology was proposed for quadrotor UAV regulation, which could reduce the power consumption and perform well in the disturbances of parameter uncertainties and aerodynamic interferences [12]. Twelve PID coefficients of a quadrotor controller were optimized using four classical evolutionary algorithms, respectively, and the simulation results indicated that the coefficients optimized from the differential evolution algorithm (DE) could minimize the energy consumption when compared with other algorithms [7]. While linear or coefficient-optimized linear controllers may be suitable for some of the above scenarios, it is often found that the nonlinear effects of the quadrotor dynamics are non-negligible [13], and that the linear control methodologies are incapable due to their reliance on approximately linearized dynamical models. Various control approaches have been used in quadrotors considering the nonlinear dynamics model. One of these approaches is nonlinear dynamic inversion (NDI), which can theoretically eliminate the nonlinearities of the control system [14], but this control method is much dependent on the model accuracies [15]. The incremental nonlinear dynamic inversion (INDI) methodology was used to improve the robustness against the model inaccuracies, which could achieve stable attitude control even though the change in pitch angle was up to 90° [16]. The adaptive control algorithm has also been widely used in quadrotor systems [17,18]. Two adaptive control laws were designed for the attitude stabilization of a quadrotor in order to deal with the problem of parametric uncertainty and external disturbance [18]. A robust adaptive control strategy was developed for tracking the attitude of foldable quadrotors, which were modeled as switched systems [19].

Due to the advantages of fast response times and strong robustness, the sliding mode control (SMC) methodology has been widely applied in the attitude tracking of quadrotors [20,21]. However, the problem of control input chattering is apparent in the traditional reaching law designed in SMC. A fuzzy logic system was developed to adaptively schedule the control gains of the sign function, effectively suppressing the control signal chattering [22]. A novel discrete-time sliding mode control (DSMC) reaching law was proposed based on theoretical analysis, which could significantly reduce chattering [23]. An adaptive fast nonsingular terminal sliding mode (AFNTSM) controller was introduced to achieve attitude regulation and suppress the chattering phenomenon. The effectiveness of this controller was verified through experiments [24]. A fractional-order sliding mode surface was designed to adaptively adjust the parameters of SMC for the fault-tolerant control of a quadrotor model with mismatched disturbances [25].

The above works of research have great significance as references. However, the control signal chattering still needs further improvement and attention when the SMC method is applied in attitude regulation with external disturbances. There are many strategies for reducing the chattering phenomenon of an SMC algorithm [26–31], such as the super-twisting algorithm and sigmoid approximation. The first strategy mainly introduces an integral term into the switching function term [28–30], which is equivalent to low-pass filtering, so as to make the control signal continuous. However, the disadvantage of this strategy is that it needs to select an appropriate integral term coefficient, and the coefficient needs to be adjusted according to the change of the system motion state. The second strategy usually approximates the switch function by constructing a function [31], so as to remove the dependence on the switch function, so as to make the state change of the system smoother. The disadvantage of this strategy is that the lack of the switch function leads to a decline in the robustness of the control system, and eventually the steady-state error becomes larger.

With the development of artificial intelligence technology, more and more reinforcement learning algorithms have been applied to traditional control methodologies [32,33]. Inspired by these studies, a deep deterministic policy gradient (DDPG) [34] agent was introduced to the SMC in this paper. The parameters linked to the sign function can be adaptively regulated by the trained DDPG agent. This adaptive regulation helps to suppress the control input chattering in attitude control, especially in the presence of external disturbances.

The primary contribution of our work can be summarized as follows: a reinforcement learning agent, based on DDPG, is trained to adaptively adjust the switching control gain in the traditional SMC method. This adaptation effectively suppresses the chattering phenomenon in attitude control.

The remainder of this paper is organized as follows: Section 2 introduces the attitude dynamics modeling for a quadrotor UAV. In Section 3, the traditional SMC and the proposed DDPG-SMC are designed for solving attitude control problems. In Section 4, the robustness and effectiveness of the proposed control approach are validated through simulation results, followed by key conclusions in Section 5.

2. Attitude Dynamics Modeling for Quadrotor UAV

The quadrotor is considered a rigid body, and its attitude motion can be described by two coordinate frames: an inertial reference frame (frame I) $O_i x_i y_i z_i$ and a body reference frame (frame B) $O_b x_b y_b z_b$, as shown in Figure 1. The attitude motion of the quadrotor can be achieved by rotating each propeller. The attitude angles can be described as $\boldsymbol{\eta} = [\phi, \theta, \psi]^T$ in frame B, where ϕ, θ, ψ are the roll angle (rotation around the x-axis), pitch angle (rotation around the y-axis), and yaw angle (rotation around the z-axis), respectively. The attitude angular velocities are expressed as $\zeta = [p, q, r]^T$, where p, q, r are the angular velocities in the roll, pitch, and yaw directions, respectively.

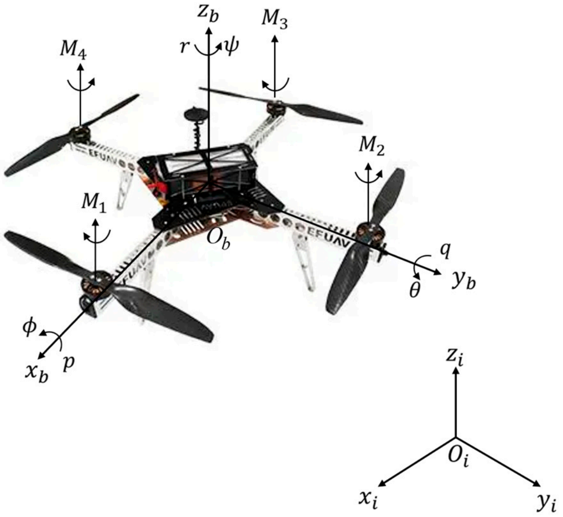


Figure 1. Attitude motion of the quadrotor in coordinate frames.

According to the relationship between the angular velocities and the attitude rate, the attitude kinematics equation of the quadrotor can be expressed as follows [35]:

$$\dot{\boldsymbol{\eta}} = \Phi(\boldsymbol{\eta})\zeta \tag{1}$$

where

$$\Phi(\boldsymbol{\eta}) = \begin{bmatrix} 1 & \tan\theta \sin\phi & \tan\theta\cos\phi \\ 0 & \cos\phi & -\sin\phi \\ 0 & \sec\theta\sin\phi & \sec\theta\cos\phi \end{bmatrix} \tag{2}$$

The attitude dynamics equation of the quadrotor can be written as follows [36]:

$$\boldsymbol{J}\dot{\zeta} + \zeta \times (\boldsymbol{J}\zeta) = \boldsymbol{\tau} \tag{3}$$

where $\boldsymbol{J} = \mathrm{diag}(J_x,\ J_y,\ J_z)$; J_x, J_y, and J_z are the moments of inertia along the $O_b x_b$, $O_b y_b$, and $O_b z_b$ axes, respectively; $\boldsymbol{\tau} = [L, M, N]^T$ denotes the control inputs; L, M, and N are the control torques in the roll, pitch, and yaw directions, respectively. When external disturbances are taken into account, the attitude dynamics Equation (3) can be rewritten as

$$\boldsymbol{J}\dot{\zeta} = -\zeta \times (\boldsymbol{J}\zeta) + \boldsymbol{\tau} + \boldsymbol{\tau}_d, \tag{4}$$

where $\boldsymbol{\tau}_d$ denotes the external disturbances.

3. Control Design for Attitude Control

In consideration of attitude control in the presence of external disturbances, a sliding mode controller, along with its sliding mode surface and reaching law, is chosen for the quadrotor dynamic system. The stability of the designed SMC system is validated using the Lyapunov stability theorem. Then, a reinforcement learning agent based on DDPG is trained and applied to the aforementioned SMC method without compromising the system's stability.

3.1. SMC Design

In this section, a sliding mode controller is designed for attitude regulation of the quadrotor. The control objective can be described as follows: the actual attitude $\boldsymbol{\eta} = [\phi, \theta, \psi]^T$ needs to be regulated to the desired attitude $\boldsymbol{\eta}_d = [\phi_d, \theta_d, \psi_d]^T$ asymptotically, i.e., $\lim_{t\to\infty} \|\boldsymbol{\eta} - \boldsymbol{\eta}_d\| = 0$.

In the controller design process, the sliding mode surface is first selected. Then, the control law is chosen to compute the control signal. Finally, the stability proof of the designed SMC system is validated using the Lyapunov stability theorem. The control scheme of SMC for attitude tracking is depicted in Figure 2.

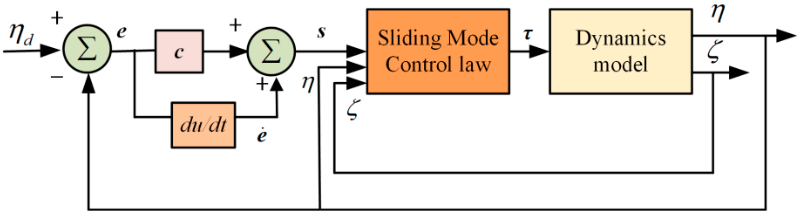


Figure 2. Block diagram of SMC.

The specific design of the sliding mode controller can be expressed as Algorithm 1.

Algorithm 1. Design Methodology of SMC.

```
Input:
    (1) Desired attitude angles $\boldsymbol{\eta}_d$
    (2) Actual attitude angles $\boldsymbol{\eta}$
    (3) Model parameters of the quadrotor
Output: Control signals for the attitude dynamics model
Step 1: Design of the control signal
    (a) Define the sliding mode surface $s$
    (b) Select the reaching law $\dot{s}$
    (c) Compute the control signal $\boldsymbol{\tau}$
Step 2: Proof of the stability of the closed-loop system
    (a) Select a Lyapunov candidate function $V$
    (b) Calculate the first-order derivative of $V$
    (c) Analyze the sign of the above derivative of $V$
    (d) Conclude the convergence of the attitude motion
Step 3: Termination
    If the attitude control errors meet the requirements, conduct the algorithm termination and
output the control signal $\boldsymbol{\tau}$. Otherwise, carry out Step 1 until convergence of the control errors.
```

Step 1 (a):

The control error can be defined as

$$\boldsymbol{e} = \boldsymbol{\eta}_d - \boldsymbol{\eta} \tag{5}$$

Then, the sliding mode surface can be derived as

$$\boldsymbol{s} = \dot{\boldsymbol{e}} + \boldsymbol{c}\boldsymbol{e}, \tag{6}$$

where $\boldsymbol{c} = \mathrm{diag}(c_1, c_2, c_3)$, and $c_1,\ c_2,\ c_3$ are selected positive numbers.

The derivative of $\boldsymbol{s}$ can be expressed as

$$\dot{\boldsymbol{s}} = \ddot{\boldsymbol{e}} + \boldsymbol{c}\dot{\boldsymbol{e}} \tag{7}$$

Substituting Equations (1)–(3) and (5) into (7), we can obtain

$$\dot{\boldsymbol{s}} = \boldsymbol{c}\dot{\boldsymbol{e}} + \ddot{\boldsymbol{e}} = \boldsymbol{c}\dot{\boldsymbol{e}} + \ddot{\boldsymbol{\eta}}_d - \dot{\Phi}(\boldsymbol{\eta})\zeta - \Phi(\boldsymbol{\eta})\boldsymbol{J}^{-1}(-\zeta \times (\boldsymbol{J}\zeta) + \boldsymbol{\tau} + \boldsymbol{\tau}_d), \tag{8}$$

where we can define

$$\boldsymbol{d} = \boldsymbol{\tau}_d \tag{9}$$

Equation (8) can be rewritten as

$$\dot{\boldsymbol{s}} = \boldsymbol{c}\dot{\boldsymbol{e}} + \ddot{\boldsymbol{\eta}}_d - \dot{\Phi}(\boldsymbol{\eta})\zeta - \Phi(\boldsymbol{\eta})\boldsymbol{J}^{-1}(-\zeta \times (\boldsymbol{J}\zeta) + \boldsymbol{\tau} + \boldsymbol{d}) \tag{10}$$

Assumption 1. *The external disturbance $\boldsymbol{d}$ is assumed to be bounded and satisfies*

$$\|\boldsymbol{d}\| \leq D, \tag{11}$$

in which D is a positive finite variable.

Step 1 (b):

The reaching law of the sliding mode surface can be selected as follows [37]:

$$\dot{\boldsymbol{s}} = -\boldsymbol{\lambda}\boldsymbol{s} - \boldsymbol{k}\mathrm{sign}(\boldsymbol{s}), \tag{12}$$

in which $\boldsymbol{\lambda}$ and $\boldsymbol{k}$ are both diagonally positive definite matrices, with $\boldsymbol{\lambda} = \mathrm{diag}(\lambda_1, \lambda_2, \lambda_3)$, and $\lambda_1,\ \lambda_2,\ \lambda_3$ are selected as positive numbers, the same as $\boldsymbol{k} = \mathrm{diag}(k_1, k_2, k_3)$, and $k_i(i = 1, 2, 3)$ is also a selected positive number, and $\mathrm{sign}(\cdot)$ represents the sign function.

Step 1 (c):

Based on calculations of the angular velocity ζ and transformation matrix $\Phi(\boldsymbol{\eta})$, as well as the derivation of Equations (8) and (12), the control signal for attitude dynamics model can be designed as follows:

$$\boldsymbol{\tau} = \zeta \times (\boldsymbol{J}\zeta) + \boldsymbol{J}\Phi^{-1}(\boldsymbol{\eta})\Big(\boldsymbol{\lambda}\boldsymbol{s} + \boldsymbol{k}\mathrm{sign}(\boldsymbol{s}) + \boldsymbol{c}\dot{\boldsymbol{e}} + \ddot{\boldsymbol{\eta}}_d - \dot{\Phi}(\boldsymbol{\eta})\zeta\Big) \tag{13}$$

Step 2:

The stability of the closed-loop system is proven as follows:

Theorem 1. *Considering the attitude dynamics system described in Equation (4), with the sliding mode surface selected as Equation (6), if the exponential reaching law is chosen as Equation (12), and the control signals for the attitude dynamics model are designed according to Equation (13), then the designed SMC system is stable, and the actual attitude can converge to the desired attitude in finite time.*

Proof of Theorem 1. We can select a Lyapunov candidate function as

$$V = \frac{1}{2}\boldsymbol{s}^T\boldsymbol{s} \tag{14}$$

□

Based on Equation (10), taking the derivative of Equation (14) with respect to time, we can obtain

$$\dot{V} = \boldsymbol{s}^T\dot{\boldsymbol{s}} = \boldsymbol{s}^T\left\{\boldsymbol{c}\dot{\boldsymbol{e}} + \ddot{\boldsymbol{\eta}}_d - \dot{\Phi}(\boldsymbol{\eta})\zeta - \Phi(\boldsymbol{\eta})\boldsymbol{J}^{-1}(-\zeta \times (\boldsymbol{J}\zeta) + \boldsymbol{\tau} + \mathbf{d})\right\} \tag{15}$$

Then, substituting (13) into (15), we have

$$\begin{aligned}\dot{V} &= \boldsymbol{s}^T\left\{\begin{array}{l} \boldsymbol{c}\dot{\boldsymbol{e}} + \ddot{\boldsymbol{\eta}}_d - \dot{\Phi}(\boldsymbol{\eta})\zeta - \Phi(\boldsymbol{\eta})\boldsymbol{J}^{-1} \\ \left(\begin{array}{l} -\zeta \times (\boldsymbol{J}\zeta) + \zeta \times (\boldsymbol{J}\zeta) + \boldsymbol{J}\Phi^{-1}(\boldsymbol{\eta}) \\ \left(\boldsymbol{\lambda}\boldsymbol{s} + \boldsymbol{k}\mathrm{sign}(\boldsymbol{s}) + \boldsymbol{c}\dot{\boldsymbol{e}} + \ddot{\boldsymbol{\eta}}_d - \dot{\Phi}(\boldsymbol{\eta})\zeta + \boldsymbol{d}\right) \end{array}\right) \end{array}\right\} \\ &= \boldsymbol{s}^T(-\boldsymbol{\lambda}\boldsymbol{s} - \boldsymbol{k}\mathrm{sign}(\boldsymbol{s}) - \boldsymbol{d}) \\ &= -\boldsymbol{\lambda}\boldsymbol{s}^T\boldsymbol{s} - \boldsymbol{k}\|\boldsymbol{s}\| - \boldsymbol{s}^T\boldsymbol{d} \\ &\leq -\|\boldsymbol{\lambda}\|\|\boldsymbol{s}\|^2 - \|\boldsymbol{k}\|\|\boldsymbol{s}\| - D\|\boldsymbol{s}\| \leq 0 \end{aligned} \tag{16}$$

We can assume that the sliding mode surface $\boldsymbol{s} = 0$ and obtain the following equation:

$$\lim_{t\to\infty} \boldsymbol{s} = \lim_{t\to\infty} (\boldsymbol{e} + \boldsymbol{c}\dot{\boldsymbol{e}}) = \lim_{t\to\infty} \{(\boldsymbol{\eta}_d - \boldsymbol{\eta}) + \mathbf{c}(\dot{\boldsymbol{\eta}}_d - \dot{\boldsymbol{\eta}})\} \tag{17}$$

Based on the selection of the diagonal positive definite matrix c, we can obtain the following expression:

$$\lim_{t\to\infty} \|\boldsymbol{\eta}_d - \boldsymbol{\eta}\| = 0, \ \lim_{t\to\infty} \|\dot{\boldsymbol{\eta}}_d - \dot{\boldsymbol{\eta}}\| = 0, \tag{18}$$

Remark 1. *From Equation (18), the designed control law in Equation (13) can guarantee the stability of the closed-loop system based on the Lyapunov stability theorem. The attitude-tracking error will converge to zero asymptotically if the sliding mode surface is equal to zero. Consequently, proving the stability of the designed SMC system has been completed.*

3.2. DDPG-SMC Design

3.2.1. The Architectural Design of DDPG-SMC

The above derivations have proven that the control error can converge to zero asymptotically in the designed SMC for a nonlinear system (Equation (4)). However, high-frequency chattering of the control signal will appear near the sliding surface due to the selected reaching law (Equation (12)) with a sign function. The intensity of chattering is determined by the parameter associated with the sign function, namely the control gain $\boldsymbol{k}$.

Inspired by the combination of reinforcement learning algorithms and traditional control methodologies, a reinforcement learning agent based on DDPG is trained to adaptively adjust the switching control gain. The trained agent is applied as a parameter regulator for the designed SMC, and the block diagram of this designed DDPG-SMC is shown in Figure 3.

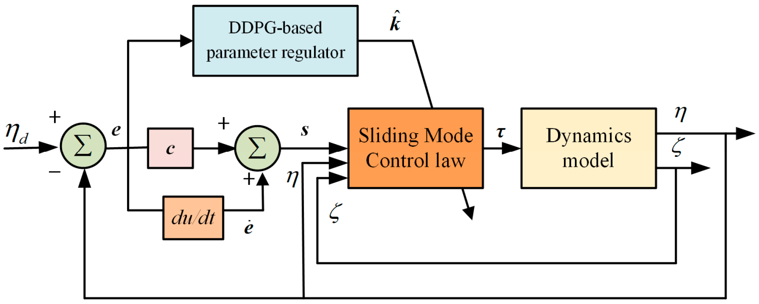


Figure 3. Block diagram of DDPG-SMC.

The architecture of the DDPG-based parameter regulator is shown in Figure 4.

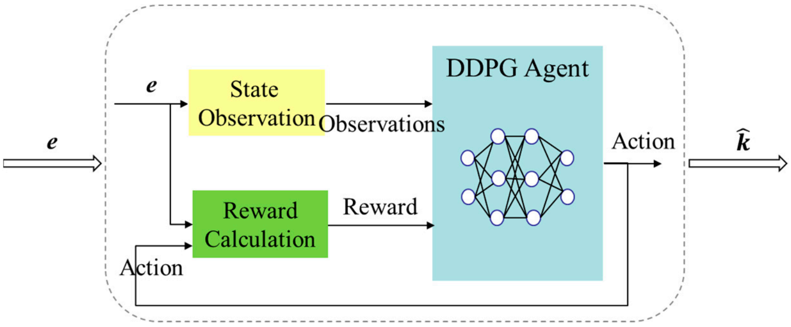


Figure 4. The architecture of the DDPG-based parameter regulator.

3.2.2. The Basic Principle of the DDPG Algorithm

DDPG is an algorithm designed to address continuous action problems within the Actor–Critic (AC) framework [38], In this approach, the policy network parameters are continuously optimized to enhance the output action to achieve higher scores in the value network. In the designed DDPG-SMC approach in this paper, the DDPG agent needs to be trained beforehand. The system described in Figure 3 serves as a training environment, and the training data are derived from multiple flight simulations.

The basic principle of the DDPG algorithm (Algorithm 2) can be introduced as follows.

Algorithm 2. DDPG Algorithm.

```
Input: Experience replay buffer $D$, initial critic networks' Q-function parameters $\theta^{Q}$, actor networks' policy parameters $\theta^{\pi}$, target networks $Q'$ and $\pi'$.
Initialize the target network parameters: $\theta^{Q'} \leftarrow \theta^{Q}$, $\theta^{\pi'} \leftarrow \theta^{\pi}$.
for episode = 1 to $M$ do
  Initialize stochastic process $N$ to add exploration to the action.
  Observe initial state $s_1$.
  for time step = 1 to T do
    Select action $a_t = \pi(s_t|\theta^{\pi}) + N_t$.
    Perform action $a_t$ and transfer to next state $s_{t+1}$, then acquire the reward value $r_t$ and the termination signal $d_t$.
    Store the state transition data $(s_t,\ a_t,\ r_t,\ d_t,\ s_{t+1})$ in experience replay buffer $D$.
    Calculate the target function:
    $y(r_t,\ s_{t+1},\ d_t) = r_t + \gamma(1-d_t)Q'\left(s_{t+1},\ \pi'\left(s_{t+1}\middle|\theta^{\pi'}\right)\middle|\theta^{Q'}\right)$
    Update the critic network using the minimized loss function:
    $L = \frac{1}{B}\sum_{(s_t,\, a_t,\, r_t,\, d_t,\, s_{t+1})\in B}\left(y(r_t,\ s_{t+1},\ d_t) - Q\left(s_t,\ a_t\middle|\theta^{Q}\right)\right)^2$
    Update the actor network using the policy gradient method:
    $\nabla_{\theta^{\pi}} J \approx \nabla_{\theta^{\pi}} \frac{1}{|B|}\sum_{s\in B} Q_{\theta^{\pi}}(s,\ \mu_{\theta^{Q}}(s))$
    Update target networks:
    $\theta^{Q'} \leftarrow \rho\theta^{Q'} + (1-\rho)\theta^{Q}$
    $\theta^{\pi'} \leftarrow \rho\theta^{\pi'} + (1-\rho)\theta^{\pi}$
  end for
end for
```

The design of the DDPG-based parameter regulator consists of two processes: training and validation. During the training process, the quadrotor's flight simulation is conducted to collect all of the state and control data, which amounts to the accumulation of experience. Then, based on the accumulated experience data, the neural network parameters are optimized and updated using gradient calculation, the stochastic gradient descent method, and other techniques. After multiple episodes of iterative training, the policy in the policy function converges to the optimal one. The validation process is used to validate the feasibility and generalization of the trained agent's optimal policy.

3.2.3. Design of the Neural Network and Parameters Related to DDPG

The neural network of DDPG mainly consists of a Critic network and an Actor network. The Critic network consists of a state part and an action part. The state part receives state input and passes through a fully connected layer with 128 and 200 nodes, respectively. The action part receives action input and passes through a fully connected layer with 200 nodes. The activation function of each fully connected layer is Relu. The state part and action part are connected together and form the Critic network through a Relu layer and a full connection layer. The structure of the Critic network is depicted in Figure 5.

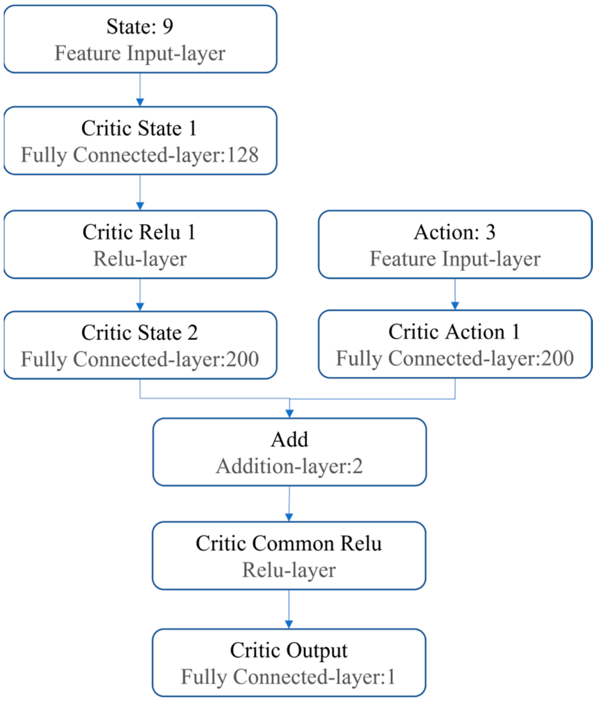


Figure 5. The structure of the Critic network.

The Actor network consists of three fully connected layers with 128, 200, and 3 nodes, respectively, and receives state input and outputs action signals within a set range. The activation functions of fully connected layers are Relu and Tanh, respectively. The structure of the Actor network is shown in Figure 6.

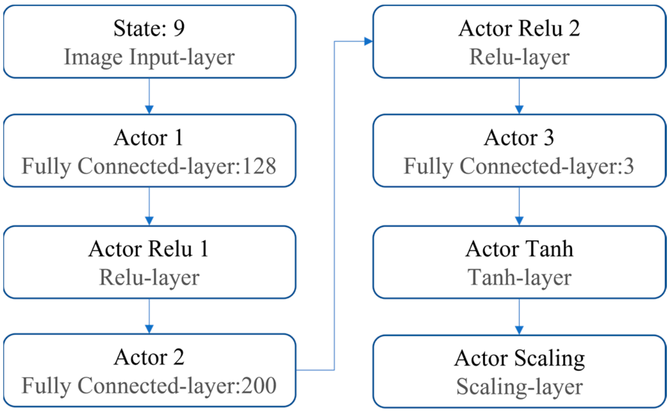


Figure 6. The structure of the Actor network.

To train the DDPG agent to adjust the switching control gain $\boldsymbol{k}$, the training episodes were set as 200, with the simulation time of each episode being 10 s and the time step being 0.02 s. Initial and desired attitude angles during the training were selected as $\boldsymbol{\eta}_0 = [0.1\text{rad}, 0.2\text{rad}, -0.1\text{rad}]^T$ and $\boldsymbol{\eta}_d = [0.2\text{rad}, 0.1\text{rad}, 0.1\text{rad}]^T$, respectively. More parameters related to DDPG are listed in Table 1.

Table 1. Parameters related to the DDPG agent.

Parameter	**Value**
State dimension of the input layer	9
Action dimension of the output layer	3
Reward discount factor	0.995
Minimum batch size	128
Max steps per episode	500
Max episodes	200
Agent sample time	0.02 s
Experience replay buffer	1×10^6
Target smooth factor	1×10^{-3}

The cumulative reward after each episode of training was recorded and output, and the reward at each step could be calculated using the following equation:

$$r(t) = -\varpi_1 t \cdot |\boldsymbol{\eta}(t) - \boldsymbol{\eta}_d(t)|, \tag{19}$$

where ϖ_1 represents the weight matrix, which was selected as $\begin{bmatrix} 1 & 1 & 1 \end{bmatrix}$, and t is the flight time.

3.2.4. The Training Results of DDPG

The training process was stopped when the average reward of cumulative training was less than -1 or the number of training episodes reached 200. The final training result is shown in Figure 7, where the green line represents the expected reward value for each training episode, and the blue and red lines denote the actual and average reward values, respectively. It can be seen that the actual reward value suddenly decreases at the 165th training episode, which is a normal result obtained by the agent in the process of continuous exploration, and the difference between the reward value and the maximum reward value is not obvious. With the increase in the number of training episodes, the actual and average rewards converge to the maximum at the 170th training episode, and the actual reward almost coincides with the expected reward. This indicates that the agent has completed training, and can be introduced as a parameter regulator in the above sliding mode controller.

In order to validate the generalization of the trained agent's optimal policy, it is necessary to test the control performance of the UAV model under various flight conditions. Specifically, it is necessary to evaluate the improvement in control performance by adjusting control parameters adaptively under different flight conditions. The relevant numerical simulation results are presented in Section 4.3.

Remark 2. *By using the designed parameter regulator based on the trained DDPG agent, the switching control gain related to reaching law can be adjusted adaptively according to the attitude control error. Compared with SMC, the only difference of DDPG-SMC is that the parameter k can be adaptively adjusted within the same value range. As a result, the stability proof of DDPG-SMC is the same as that of SMC in Section 3.1, and the closed-loop system of both methods is stable.*

Therefore, the control signal in DDPG-SMC can be represented as

$$\boldsymbol{\tau} = \zeta \times (\boldsymbol{J}\zeta) + \boldsymbol{J}\Phi^{-1}(\boldsymbol{\eta})\Big(\boldsymbol{\lambda s} + \hat{\boldsymbol{k}}\text{sign}(\boldsymbol{s}) + \boldsymbol{c}\dot{\boldsymbol{e}} + \ddot{\boldsymbol{\eta}}_d - \dot{\Phi}(\boldsymbol{\eta})\zeta\Big), \tag{20}$$

in which $\hat{\boldsymbol{k}}$ is the time-varying switching control gain related to reaching law.

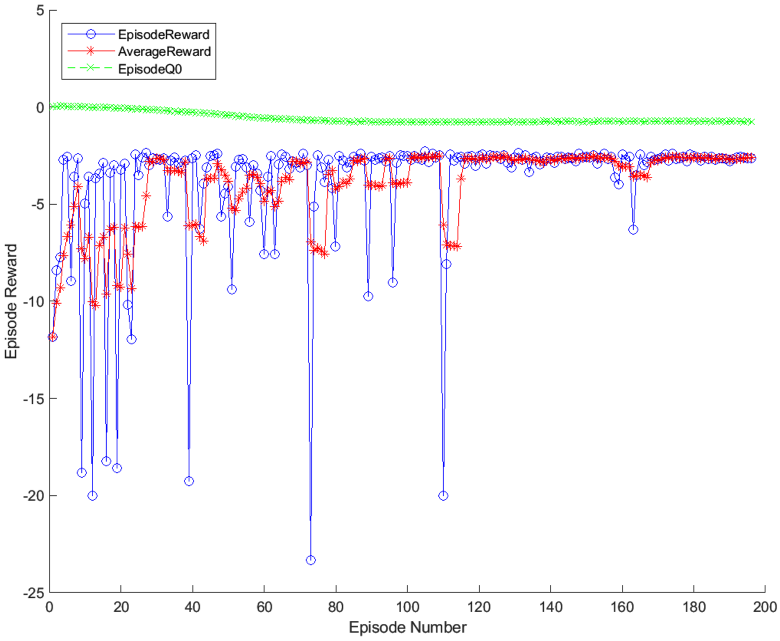


Figure 7. Each episode reward for gain-learning with the DDPG agent.

4. Simulation Results

The robustness and effectiveness of the proposed control approach can be verified via flight simulations. The quadrotor used in our study is modified and designed on the basis of a DJI F450 UAV, and its specific technical parameters are shown in Table 2.

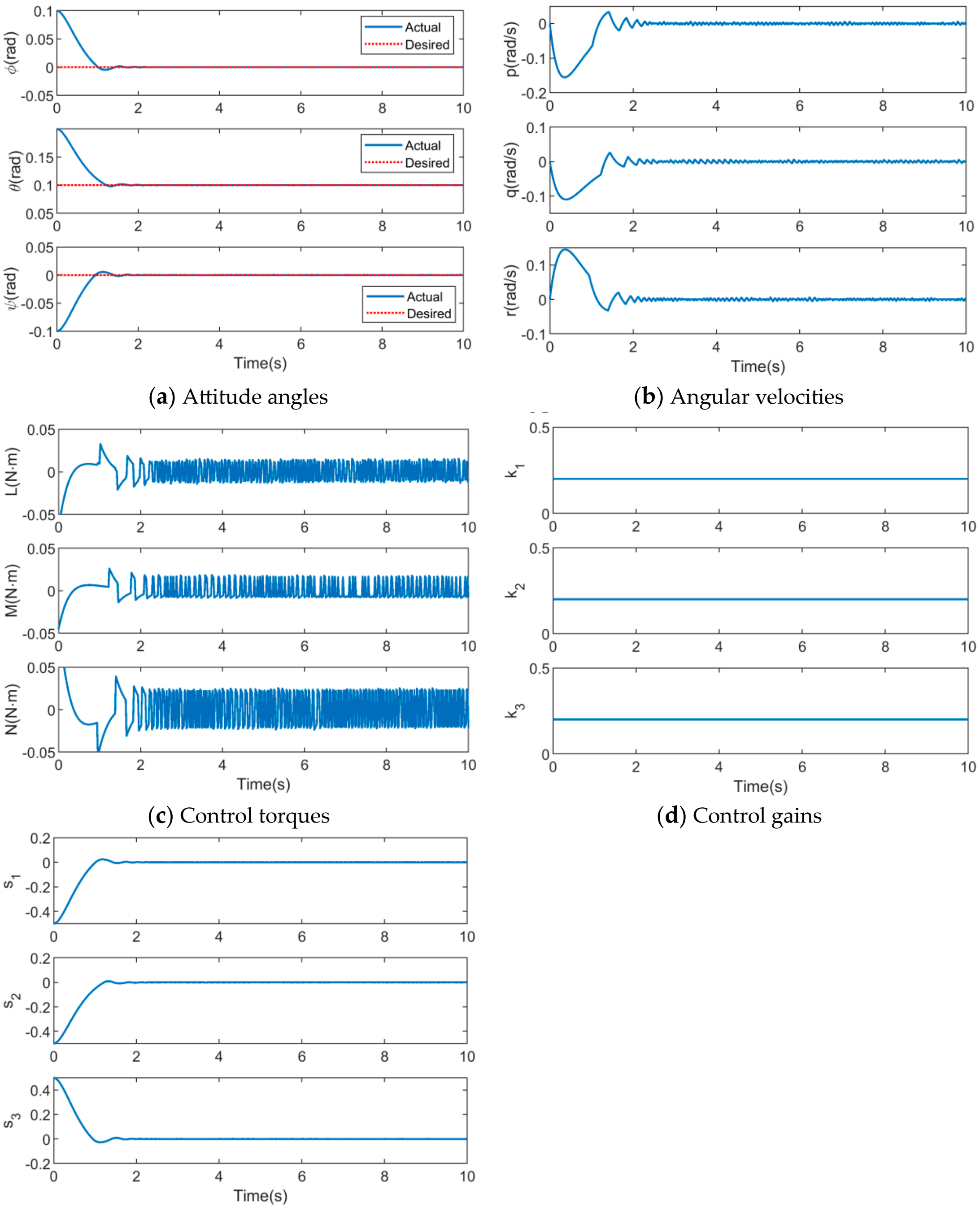
Table 2. Model parameters of the quadrotor.

Parameter	Value
Mass m/kg	3.350
Inertia moment about $o_b x_b$ $J_\text{x}/(\text{kg}\cdot\text{m}^2)$	0.0588
Inertia moment about $o_b y_b$ $J_\text{y}/(\text{kg}\cdot\text{m}^2)$	0.0588
Inertia moment about $o_b z_b$ $J_\text{z}/(\text{kg}\cdot\text{m}^2)$	0.1076
Lift factor b	8.159×10^{-5}
Drag factor d	2.143×10^{-6}
Distance between the center of mass and the rotation axis of any propeller l/m	0.195

The basic simulation conditions are described as follows. Initial attitude angles and angular velocities of the quadrotor are set as $\boldsymbol{\eta}_0 = [0.1\text{rad}, 0.2\text{rad}, -0.1\text{rad}]^T$ and $\zeta_0 = [0\text{rad}/s, 0\text{rad}/s, 0\text{rad}/s]^T$, respectively. The desired attitude angles are selected as $\boldsymbol{\eta}_d = [0\text{rad}, 0.1\text{rad}, 0\text{rad}]^T$. The external disturbances are assumed to act on the system in the form of torques: $\boldsymbol{\tau}_d = 0.005 \times \begin{bmatrix} \sin(\pi/100t) & \cos(\pi/100t) & \sin(\pi/100t) \end{bmatrix}^T \text{N} \cdot \text{m}$. Three control approaches, including SMC, the AFGS-SMC proposed in reference [22], and the DDPG-SMC designed in this paper, are used in the flight simulation, respectively.

4.1. Simulation Results of SMC

The relevant control parameters of the sign function in SMC are designed as follows: $\boldsymbol{k} = \text{diag}(0.2, 0.2, 0.2)$ and $\boldsymbol{\lambda} = \text{diag}(1.5, 1.5, 1.5)$. The numerical simulation results are depicted in Figure 8.



(**a**) Attitude angles

(**b**) Angular velocities

(**c**) Control torques

(**d**) Control gains

(**e**) Sliding surfaces

Figure 8. Simulation results of SMC.

As shown in Figure 8a, the dashed and solid lines represent the desired and actual attitude angles, respectively. The convergence times of attitude angles in three directions (roll, pitch, and yaw) are 1.8 s, 2.1 s, and 1.8 s, respectively, from the initial value to the desired value. This indicates that the quadrotor attitude can be regulated into the desired attitude using the SMC algorithm, in the presence of disturbances.

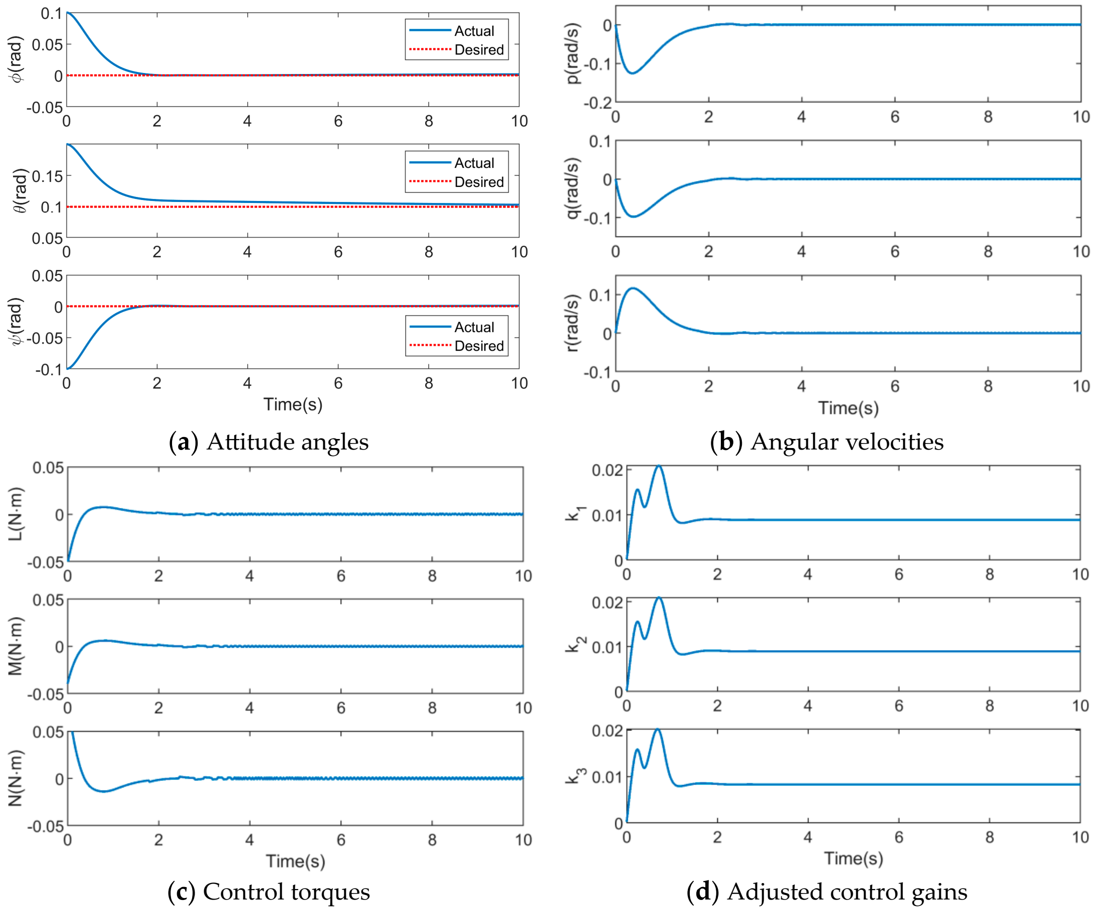
The time histories of angular velocities and control inputs are depicted in Figure 8b,c, respectively. It can be seen that the attitude angular velocities in three directions approach 0 rad/s during time periods of 2.2 s, 2.4 s, and 2.2 s, respectively. The angular velocities oscillate slightly around 0 rad/s to maintain balance in the quadrotor system.

However, the chattering of the control inputs is more severe when the control system is stabilized. The control input signal in the roll direction oscillates in the range of −0.012 N·m to 0.014 N·m, the control input signal in the pitch direction oscillates in the range of −0.008 N·m to 0.016 N·m, and the control input signal in the yaw direction oscillates in the range of −0.018 N·m to 0.024 N·m. Since the control input signals denote the torques generated by the quadrotor propellers, chattering at high frequencies is absolutely unacceptable for the quadrotor's actuators.

Figure 8d,e represent the time evolutions of control gains related to the reaching law and sliding mode surfaces, respectively. It can be seen that sliding mode surfaces converge to zero asymptotically, and the control gains remain constant throughout the entire simulation process. These constant gains lead to the chattering phenomenon of control signals.

4.2. Simulation Results of AFGS-SMC

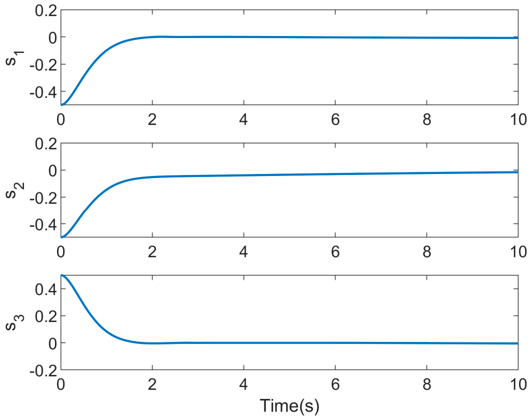
The chattering phenomenon is caused by high-frequency switching around the sliding mode surface, attributed to the term $\boldsymbol{k}\cdot\text{sign}(\boldsymbol{s})$ in SMC. The adaptive fuzzy gain-scheduling sliding mode control (AFGS-SMC) method in the reference [22] was proposed by the authors' team in 2016. This method can effectively suppress the control signal chattering, and the authors would like to compare it with the method proposed in this paper in terms of control performance. The simulation results for this method are depicted in Figure 9.



(**a**) Attitude angles (**b**) Angular velocities

(**c**) Control torques (**d**) Adjusted control gains

Figure 9. *Cont.*



(**e**) Sliding mode surfaces

Figure 9. Simulation results of AFGS-SMC.

As presented in Figure 9a, the dashed and solid lines represent the desired and actual attitude angles, respectively. The convergence times of attitude angles in three directions (roll, pitch, and yaw) are 2.2 s, 2.5 s, and 2.2 s, respectively, from the initial value to the desired value. This demonstrates that the quadrotor attitude can be regulated into the desired attitude using the AFGS-SMC algorithm, in the presence of disturbances.

The time evolutions of the quadrotor's angular velocities and control inputs are depicted in Figure 9b,c, respectively. It can be seen that the attitude angular velocities in the three directions approach 0 rad/s during time periods of 3.1 s, 2.8 s, and 3.1 s, respectively, and the oscillation is significantly reduced. In contrast to the results for SMC, the chattering phenomenon of the control input is significantly reduced.

Figure 9d,e represent the time evolutions of control gains related to the reaching law and sliding mode surfaces, respectively. It can be seen that the sliding mode surfaces converge to zero asymptotically, and the control gains are adjusted adaptively via the associated fuzzy rules in AFGS-SMC. This adaptive adjustment helps reduce the chattering phenomenon of control signals.

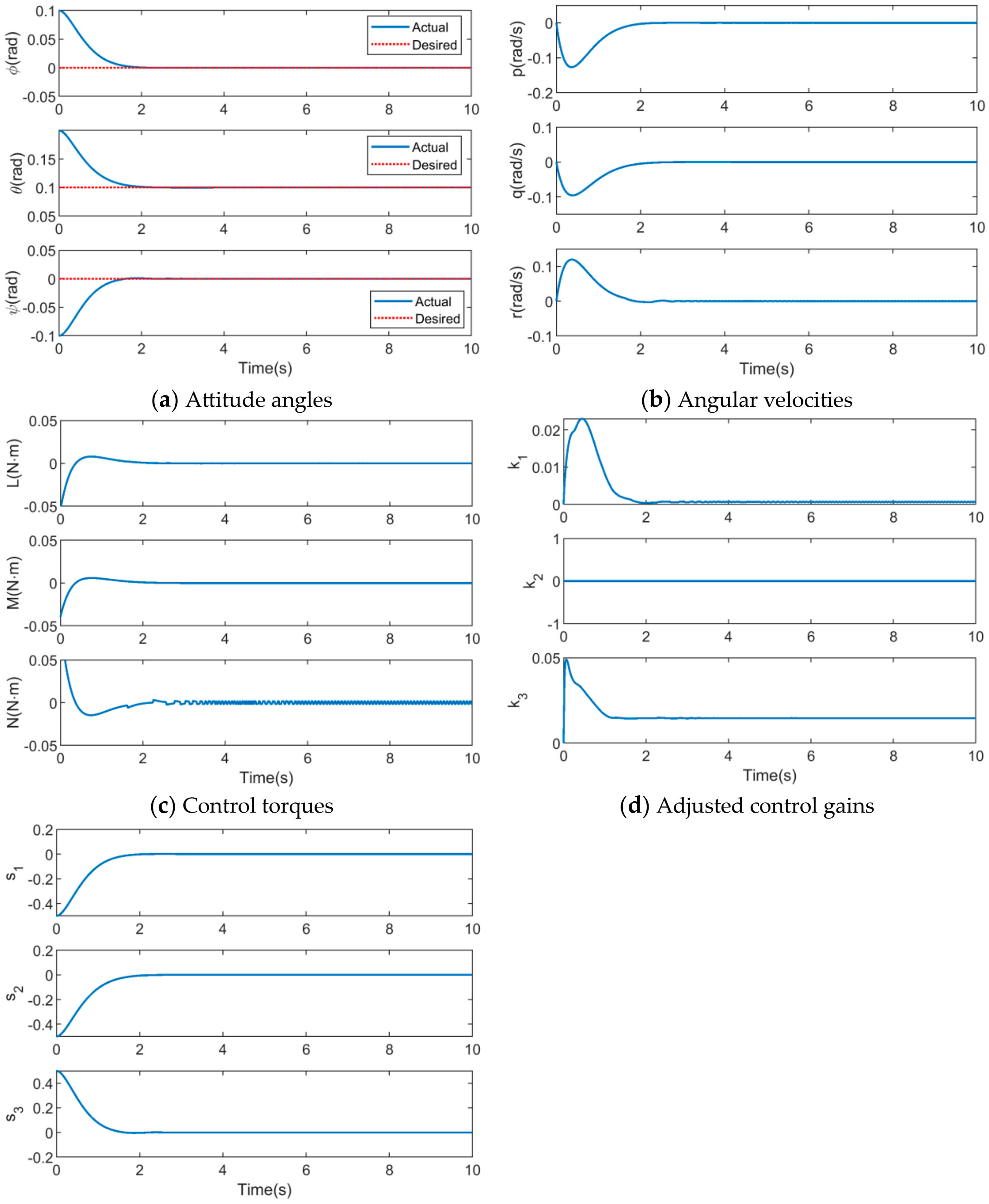
4.3. Simulation Results of DDPG-SMC

Similar to the AFGS-SMC method mentioned above, the control gains of DDPG-SMC are time-varying and can be adaptively scheduled through the DDPG-based parameter regulator. The simulation results for DDPG-SMC are depicted in Figure 10.

As depicted in Figure 10a, the dashed and solid lines represent the desired and actual attitude angles, respectively. The convergence times of attitude angles in three directions (roll, pitch, and yaw) are 2.0 s, 1.9 s, and 2.0 s, respectively, from the initial value to the desired value. This demonstrates that the quadrotor's attitude can be regulated into the desired attitude using the designed DDPG-SMC algorithm, in the presence of disturbances.

The time evolutions of the quadrotor's angular velocities and control inputs are presented in Figure 10b,c, respectively. It can be seen that the attitude angular velocities in the three directions approach 0 rad/s during time periods of 2.1 s, 2.4 s, and 2.6 s, respectively, and the oscillation is much less.

Figure 10d,e represent the time evolutions of control gains related to the reaching law and sliding mode surfaces, respectively. It can be seen that the sliding mode surfaces converge to zero asymptotically, and the control gains related to the reaching law are adjusted adaptively via the trained DDPG agent. This adjustment can help reduce the chattering phenomenon of control signals.



(**a**) Attitude angles

(**b**) Angular velocities

(**c**) Control torques

(**d**) Adjusted control gains

(**e**) Sliding mode surfaces

Figure 10. Simulation results of DDPG-SMC.

4.4. Comparative Analysis of Simulation Results

To compare the control performance of the above three methods, the convergence time and steady-state errors of the attitude angles and the chattering amplitudes of the control signals are selected as performance indicators, which are listed in Table 3.

Table 3. Control performance indicators of the three methods.

Performance Indicators	SMC	AFGS-SMC	DDPG-SMC
Convergence time (s)	2.1	2.5	2.0
Steady-state errors (%)	2	5	1
Chattering amplitudes (N·m)	(−0.018, 0.024)	(−0.006, 0.006)	(−0.005, 0.005)

It can be seen that both AFGS-SMC and DDPG-SMC can greatly reduce the chattering of control signals. However, in the DDPG-SMC method, the convergence time is shorter and the steady-state error is smaller than those of the AFGS-SMC method, indicating that the DDPG-SMC method exhibits better control performance.

Remark 3. *(1) The traditional SMC, referenced AFGS-SMC, and designed DDPG-SMC methods all perform effectively and robustly in attitude control, with the presence of external continuous disturbances. (2) The disadvantage of the traditional SMC is that a high-frequency chattering phenomenon exists in the control input signals. (3) The control gains related to the reaching law in DDPG-SMC can be adjusted adaptively via the trained reinforcement learning agent, where the chattering phenomenon is effectively reduced.*

5. Conclusions

In view of the chattering phenomenon in the traditional SMC for quadrotor attitudes, a novel approach based on reinforcement learning, called DDPG-SMC, is proposed. The attitude dynamics model of the studied quadrotor is derived, and the attitude control problem is described by formulas initially. A traditional sliding mode controller is designed for the nonlinear dynamic system, and the stability of the closed-loop system is ensured via the Lyapunov stability theorem. A reinforcement learning agent, based on DDPG, is trained to adaptively adjust the switching control gain in traditional SMC. This trained agent is then utilized in SMC as a parameter regulator to develop the DDPG-SMC approach. The simulation results indicate that the proposed DDPG-SMC approach demonstrates excellent robustness and effectiveness in attitude control for quadrotors. Compared with the traditional SMC method, the proposed approach can effectively suppress the chattering phenomenon in the presence of external disturbances. The research in this paper can provide a methodological reference for addressing the chattering problem of SMC when the control system is affected by external disturbances. The authors will conduct hardware experiments to verify the feasibility of the proposed method in the future.

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

Funding: This research received no external funding.

Data Availability Statement: Data are contained within the article.

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

References

1. Grima, S.; Lin, M.; Meng, Z.; Luo, C.; Chen, Y. The application of unmanned aerial vehicle oblique photography technology in online tourism design. *PLoS ONE* **2023**, *18*, e0289653.
2. Clarke, R. Understanding the drone epidemic. *Comput. Law Secur. Rev.* **2014**, *30*, 230–246. [CrossRef]
3. Xu, B.; Wang, W.; Falzon, G.; Kwan, P.; Guo, L.; Chen, G.; Tait, A.; Schneider, D. Automated cattle counting using Mask R-CNN in quadcopter vision system. *Comput. Electron. Agric.* **2020**, *171*, 105300. [CrossRef]
4. Idrissi, M.; Salami, M.; Annaz, F. A review of quadrotor unmanned aerial vehicles: Applications, architectural design and control algorithms. *J. Intell. Robot. Syst.* **2022**, *104*, 22. [CrossRef]
5. Adiguzel, F.; Mumcu, T.V. Robust discrete-time nonlinear attitude stabilization of a quadrotor UAV subject to time-varying disturbances. *Elektron. Elektrotechnika* **2021**, *27*, 4–12. [CrossRef]
6. Shen, J.; Wang, B.; Chen, B.M.; Bu, R.; Jin, B. Review on wind resistance for quadrotor UAVs: Modeling and controller design. *Unmanned Syst.* **2022**, *11*, 5–15. [CrossRef]
7. Gün, A. Attitude control of a quadrotor using PID controller based on differential evolution algorithm. *Expert Syst. Appl.* **2023**, *229*, 120518. [CrossRef]
8. Zhou, L.; Pljonkin, A.; Singh, P.K. Modeling and PID control of quadrotor UAV based on machine learning. *J. Intell. Syst.* **2022**, *31*, 1112–1122. [CrossRef]

9. Khatoon, S.; Nasiruddin, I.; Shahid, M. Design and simulation of a hybrid PD-ANFIS controller for attitude tracking control of a quadrotor UAV. *Arab. J. Sci. Eng.* **2017**, *42*, 5211–5229. [CrossRef]
10. Landry, B.; Deits, R.; Florence, P.R.; Tedrake, R. Aggressive quadrotor flight through cluttered environments using mixed integer programming. In Proceedings of the 2016 IEEE International Conference on Robotics and Automation (ICRA), Stockholm, Sweden, 16–21 May 2016.
11. Bouabdallah, S.; Noth, A.; Siegwart, R. PID vs LQ control techniques applied to an indoor micro quadrotor. In Proceedings of the 2004 1EEE/RSJ Internationel Conference On Intelligent Robots and Systems, Sendal, Japan, 28 September–2 October 2004.
12. Miranda-Colorado, R.; Aguilar, L.T. Robust PID control of quadrotors with power reduction analysis. *ISA Trans.* **2020**, *98*, 47–62. [CrossRef] [PubMed]
13. Wang, Z.; Zhao, J.; Cai, Z.; Wang, Y.; Liu, N. Onboard actuator model-based incremental nonlinear dynamic inversion for quadrotor attitude control: Method and application. *Chin. J. Aeronaut.* **2021**, *34*, 216–227. [CrossRef]
14. Smeur, E.J.J.; Chu, Q.; de Croon, G.C.H.E. Adaptive incremental nonlinear dynamic inversion for attitude control of micro air vehicles. *J. Guid. Control. Dyn.* **2016**, *39*, 450–461. [CrossRef]
15. da Costa, R.R.; Chu, Q.P.; Mulder, J.A. Reentry flight controller design using nonlinear dynamic inversion. *J. Spacecr. Rocket.* **2003**, *40*, 64–71. [CrossRef]
16. Yang, J.; Cai, Z.; Zhao, J.; Wang, Z.; Ding, Y.; Wang, Y. INDI-based aggressive quadrotor flight control with position and attitude constraints. *Robot. Auton. Syst.* **2023**, *159*, 104292. [CrossRef]
17. Wang, B.; Zhang, Y.; Zhang, W. A composite adaptive fault-tolerant attitude control for a quadrotor UAV with multiple uncertainties. *J. Syst. Sci. Complex.* **2022**, *35*, 81–104. [CrossRef]
18. Huang, T.; Li, T.; Chen, C.L.P.; Li, Y. Attitude stabilization for a quadrotor using adaptive control algorithm. *IEEE Trans. Aerosp. Electron. Syst.* **2023**, *60*, 334–347. [CrossRef]
19. Patnaik, K.; Zhang, W. Adaptive attitude control for foldable quadrotors. *IEEE Control. Syst. Lett.* **2023**, *7*, 1291–1296. [CrossRef]
20. Chen, J.; Long, Y.; Li, T.; Huang, T. Attitude tracking control for quadrotor based on time-varying gain extended state observer. *Proc. Inst. Mech. Eng. Part I J. Syst. Control. Eng.* **2022**, *237*, 585–595. [CrossRef]
21. Zheng, Z.; Su, X.; Jiang, T.; Huang, J. Robust dynamic geofencing attitude control for quadrotor systems. *IEEE Trans. Ind. Electron.* **2023**, *70*, 1861–1869. [CrossRef]
22. Yang, Y.; Yan, Y. Attitude regulation for unmanned quadrotors using adaptive fuzzy gain-scheduling sliding mode control. *Aerosp. Sci. Technol.* **2016**, *54*, 208–217. [CrossRef]
23. Chen, X.; Li, Y.; Ma, H.; Tang, H.; Xie, Y. A novel variable exponential discrete time sliding mode reaching law. *IEEE Trans. Circuits Syst. II Express Briefs* **2021**, *68*, 2518–2522. [CrossRef]
24. Lian, S.; Meng, W.; Lin, Z.; Shao, K.; Zheng, J.; Li, H.; Lu, R. Adaptive attitude control of a quadrotor using fast nonsingular terminal sliding mode. *IEEE Trans. Ind. Electron.* **2022**, *69*, 1597–1607. [CrossRef]
25. Sun, H.; Li, J.; Wang, R.; Yang, K. Attitude control of the quadrotor UAV with mismatched disturbances based on the fractional-order sliding mode and backstepping control subject to actuator faults. *Fractal Fract.* **2023**, *7*, 227. [CrossRef]
26. Belgacem, K.; Mezouar, A.; Essounbouli, N. Design and analysis of adaptive sliding mode with exponential reaching law control for double-fed induction generator based wind turbine. *Int. J. Power Electron. Drive Syst.* **2018**, *9*, 1534–1544. [CrossRef]
27. Mechali, O.; Xu, L.; Xie, X.; Iqbal, J. Fixed-time nonlinear homogeneous sliding mode approach for robust tracking control of multirotor aircraft: Experimental validation. *J. Frankl. Inst.* **2022**, *359*, 1971–2029. [CrossRef]
28. Kelkoul, B.; Boumediene, A. Stability analysis and study between classical sliding mode control (SMC) and super twisting algorithm (STA) for doubly fed induction generator (DFIG) under wind turbine. *Energy* **2021**, *214*, 118871. [CrossRef]
29. Danesh, M.; Jalalaei, A.; Derakhshan, R.E. Auto-landing algorithm for quadrotor UAV using super-twisting second-order sliding mode control in the presence of external disturbances. *Int. J. Dyn. Control* **2023**, *11*, 2940–2957. [CrossRef]
30. Siddique, N.; Rehman, F.U.; Raoof, U.; Iqbal, S.; Rashad, M. Robust hybrid synchronization control of chaotic 3-cell CNN with uncertain parameters using smooth super twisting algorithm. *Bull. Pol. Acad. Sci. Tech. Sci.* **2023**, *71*, 1–8. [CrossRef]
31. Chen, Y.; Cai, B.; Cui, G. *The Design of Adaptive Sliding Mode Controller Based on RBFNN Approximation for Suspension Control of MVAWT*; 2020 Chinese Automation Congress (CAC): Shanghai, China, 2020.
32. Wang, D.; Shen, Y.; Sha, Q. Adaptive DDPG design-based sliding-mode control for autonomous underwater vehicles at different speeds. In Proceedings of the 2019 IEEE Underwater Technology (UT), Kaohsiung, Taiwan, 16–19 April 2019.
33. Nicola, M.; Nicola, C.-I.; Selișteanu, D. Improvement of the control of a grid connected photovoltaic system based on synergetic and sliding mode controllers using a reinforcement learning deep deterministic policy gradient agent. *Energies* **2022**, *15*, 2392. [CrossRef]
34. Lillicrap, T.P.; Hunt, J.J.; Pritzel, A.; Heess, N.; Erez, T.; Tassa, Y.; Silver, D.; Wierstra, D. Continuous control with deep reinforcement learning. *arXiv* **2015**, arXiv:1509.02971.
35. Mechali, O.; Xu, L.; Huang, Y.; Shi, M.; Xie, X. Observer-based fixed-time continuous nonsingular terminal sliding mode control of quadrotor aircraft under uncertainties and disturbances for robust trajectory tracking: Theory and experiment. *Control. Eng. Pract.* **2021**, *111*, 104806. [CrossRef]
36. Tang, P.; Lin, D.; Zheng, D.; Fan, S.; Ye, J. Observer based finite-time fault tolerant quadrotor attitude control with actuator faults. *Aerosp. Sci. Technol.* **2020**, *104*, 105968. [CrossRef]

37. Nasiri, A.; Kiong Nguang, S.; Swain, A. Adaptive sliding mode control for a class of MIMO nonlinear systems with uncertainties. *J. Frankl. Inst.* **2014**, *351*, 2048–2061. [CrossRef]
38. Silver, D.; Lever, G.; Heess, N. Deterministic policy gradient algorithms. In Proceedings of the 31st International Conference on Machine Learning, Beijing, China, 21–26 June 2014.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.