International Journal of Control, Automation and Systems 20(1) (2022) 349-363
http://dx.doi.org/10.1007/s12555-020-0812-z
ISSN:1598-6446 eISSN:2005-4092
http://www.springer.com/12555

Tracking Control for a Quadrotor via Dynamic Surface Control and Adaptive Dynamic Programming

Qiang Gao, Xin-Tong Wei, Da-Hua Li*, Yue-Hui Ji, and Chao Jia

Abstract: In this paper, a data-driven control algorithm based on the Dynamic Surface Control and the Action-Dependent Heuristic Dynamic Programming is proposed to realize the stable tracking control of the quadrotor. Firstly, the dynamic surface control is addressed for the nonlinear model of the quadrotor, which can overcome the "explosion of complexity" problem encountered in traditional back-stepping method inevitably. The controller designed by Dynamic Surface Control is served as the main controller in the total control structure. Secondly, the Action-Dependent Heuristic Dynamic Programming is investigated to construct a complementary attitude controller by involving the learning mechanism. The adoption of Action-Dependent Heuristic Dynamic Programming can provide the capability of adaptation and disturbance rejection to improve the tracking control performance effectively. The overall closed-loop system is proved to be asymptotically stable by the Lyapunov theorem. Finally, the numerical simulation and flight experiments are presented to demonstrate that the proposed tracking control scheme exhibits an excellent tracking performance in the case of external disturbances.

Keywords: Adaptive dynamic programming, dynamic surface control, quadrotor, tracking control.

1. INTRODUCTION

In recent years, the quadrotor has attracted more and more attention in the robotics community due to their compact size, low noise and agile maneuverability [1]. Quadrotor can accomplish dangerous missions, such as the surveillance, the rescue, the photography, the traffic monitoring, the homeland security and the damage assessment in intricate environments [2,3]. Therefore, the tracking control performance of the quadrotor can directly affect the actual application, which makes the tracking control a hot and challenging issue.

To solve the tracking problems of the quadrotor, many control strategies have been utilized, e.g., linear matrix inequalities (LMI) [4], linear parameter varying (LPV) [5], linear-quadratic-regulator (LQR) [6], active disturbance rejection control (ADRC) [7], SMC [8], neural network [9] and back-stepping control [10]. In [4], a nonlinear adaptive robust control algorithm based on LMI is introduced to design the attitude and position controller of the quadrotor. Gao and Fu [5] use LPV modeling and control methods to achieve attenuation of interference in aircraft flight. In [6], a robust fuzzy controller for quadrotor based on LQR is designed. Zhang *et al*. [7] put forward an ADRC scheme to solve the trajectory tracking control problem of a quadrotor. In [8], an approach based on SMC for UAV tracking trajectory is proposed. The work of [9] uses an adaptive neural network control with a neural state observer for quadrotors. Aiming at the parameter uncertainty and external interference of the quadrotor, a robust backstepping output feedback trajectory tracking controller is designed in [10].

The back-stepping control has been widely adopted because of its potential application value in the nonlinear fields, i.e., hypersonic vehicles [11], flexible manipulators [12], quadrotors [13], intelligent vehicles [14] and servo systems [15]. However, the aforementioned control suffers from the problem of "explosion of complexity" caused by high-order analytical derivations of the virtual control. To overcome the "explosion of complexity" issue, the Dynamic Surface Control (DSC) has been introduced, by adopting a first-order filter to estimate virtual control and the derivative [16–18]. The DSC algorithm has been applied in various areas, e.g., floating production storage and offloading vessels [19], underwater vehicles [20], spacecrafts [21], quadrotors [22] and PWM rectifiers [23]. However, the DSC technology is a predefined controller based on an accurate system model, which can-

Manuscript received November 2, 2020; revised March 16, 2021; accepted April 1, 2021. Recommended by Associate Editor Tae-Koo Kang under the direction of Editor Myo Taeg Lim. This work was supported by the Research Project of Tianjin Municipal Education Commission 2017KJ249.

Qiang Gao is with Tianjin Key Laboratory for Control Theory & Applications in Complicated Systems, Tianjin University of Technology, Tianjin, China (e-mail: gaoqiang@tjut.edu.cn). Xin-Tong Wei, Da-Hua Li, Yue-Hui Ji, and Chao Jia are with the School of Electrical and Electronic Engineering, and Tianjin Key Laboratory for Control Theory & Applications in Complicated Systems, Tianjin University of Technology, Tianjin, China (e-mails: 2549928293@qq.com, {lidah2005, jiyuehuitju}@163.com, Jacky20042004@126.com).
* Corresponding author.

not provide excellent control performance in the presence of uncertainty and interference, such as load change and wind shear [24].

As an adaptive evaluation algorithm, adaptive dynamic programming (ADP) is proposed based on dynamic programming (DP), neural networks and the reinforcement learning method. As a data-driven online-learning control scheme, ADP does not rely on the precise mathematical model. Moreover, the parameters are updated iteratively when the system is subjected to external disturbances [25–34]. The ADP approach can provide approximate optimal control according to Bellman's optimal principle [35–44]. The compensation terms are estimated by the approximators (such as neural networks) to compensate dynamic uncertainty or nonlinear effects. Hence, the ADP algorithm can improve the robustness of the quadrotor effectively. In general, ADP mainly includes five basic types, i.e., heuristic dynamic programming (HDP), dual HDP (DHP), globalized dual HDP (GDHP), action-dependent HDP (ADHDP), and action-dependent dual HDP (ADDHP). As one of the most prevalent and powerful algorithms, ADHDP updates the parameters in the critic Neural Networks (NNs)and action NNs iteratively, aiming to minimize the cost functions [45,46], which has been utilized in a still camera [47], static var compensators [48], and chaos systems [49].

Considering that the DSC technology is a predefined controller based on the exact system model, and fails to provide an excellent control performance in the presence of uncertainties and disturbances. Therefore, a complementary controller designed by ADHDP can improve the adaptive tracking control capability for the quadrotor under uncertainties and noise conditions. Lin *et al*. [50] propose a decoupling tracking controller for quadrotors using dynamic surface control (DSC) and second-order sliding mode disturbance observer (SMDO), and the SMDO is utilized to restrain the influence of system uncertainties and external disturbances. Wang *et al*. [51] divide the control loop of the quadcopter into position loop and attitude loop. In the position loop, the adaptive controller is used to estimate the upper limit of external interference online, and the dynamic surface control (DSC) technology is dedicated to solving the "complexity explosion" problem in the traditional reverse design process. The design of the attitude loop controller adopts event-triggered control. In [52], an improved dynamic surface control (DSC) method based on fast terminal sliding mode is developed. In order to eliminate the inherent "complexity explosion" problem of the controller based on the backstepping method, a finite time command filter and error compensation signal are used in the design of the dynamic surface controller.

Inspired by the above status, in this paper, a synthetical controller based on DSC and ADHDP is designed to improve the tracking performance of the quadrotor. DSC is utilized as the main attitude controller, and ADHDP is adopted as a complementary attitude controller. The DSC controller can provide the control signal to force the system operating in the normal condition, and the ADHDP is applied to provide complementary terms around the normal operation condition to improve the tracking performance. When the quadrotor subjects to external disturbances, the ADHDP control policy can adjust controller parameters adaptively, reduce tracking errors, and provide a satisfying control performance.

Remark 1: Although the proposed method suffers from certain complexity, it is indeed feasible for general quadrotor platforms with powerful computing capabilities, such as the Pixhawk autopilot.

In contrast with the traditional back-stepping method, the proposed control method possesses better control performance and can achieve satisfactory tracking performance even under external interference, proved by the normal numerical simulation experiment and the disturbance simulation experiment. In the flight experiments, the proposed synthetical controller and the cascade PID controller are applied to the same quadrotor to conduct the flight experiment and the wind disturbance experiment, the experimental results show that the synthetical controller involved in this article has better control accuracy and interference suppression capabilities than the cascade PID controller.

Compared with the existing results, our design offers some new features:

1) A synthetical attitude controller is designed by combining DSC and ADHDP to further improve the anti-disturbance and control accuracy of the quadrotor tracking control.
2) The application of DSC avoids the analytical derivation of virtual control, thereby overcoming the problem of "explosion of complexity". The ADHDP structure added to the auxiliary attitude controller enables the quadrotor control system to learn online and improve tracking control performance.

The remaining contents are outlined as follows: Section 2 introduces the synthetical control scheme. Section 3 describes the mathematical model of the quadrotor. Section 4 states the attitude control and stability analysis of the DSC. Section 5 describes the design of ADHDP-based controller, the specific iterative processes and the stability analysis. Sections 6 and 7 present the simulation and experiment results for the quadrotor tracking control. Concluding remarks are stated in Section 8.

2. SYNTHETICAL CONTROL SCHEME

To realize tracking control and facilitate an improved control performance for the quadrotor, a control strategy combining DSC with ADHDP is proposed in the paper. The DSC controller can provide the control signal that

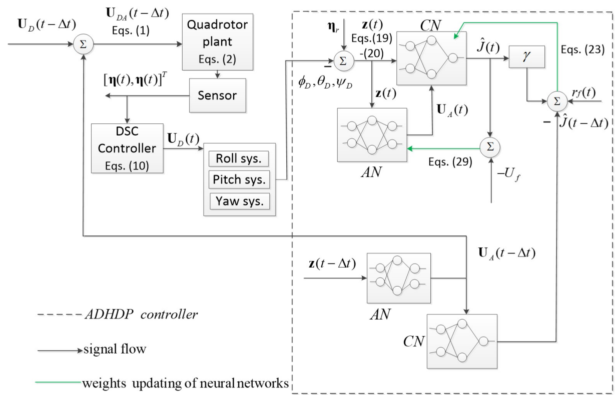


Fig. 1. DSC-ADHDP-based control diagram.

enables the system in the normal operation condition, and the ADHDP controller provides supplementary adjustment signals around the normal operation condition to improve the tracking performance. The control flow chart is shown in Fig. 1. The ADHDP controller is mainly composed of action neural network (AN) and critic neural network (CN). Note that the two action networks (critic network) in Fig. 1 are actually one neural network, the difference is that the timestep is t or $t-\Delta t$. The sum of outputs of the DSC and the AN works as the control input signal. The attitude tracking errors will be reduced remarkably after self-learning iterative process of the action-critic (AC) network.

Denote U_{DA} as the total control law as

$$U_{DA} = U_D + U_A, \tag{1}$$

where U_D is the control law provided by the DSC algorithm, U_A is the control law obtained by the ADP algorithm, $U_{DA} = [u_1, u_2, u_3]^T$.

3. THE MATHEMATICAL MODEL OF A QUADROTOR

By the Newton-Euler equation, the attitude dynamics of the quadrotor are derived as

$$\begin{cases} \ddot{\phi} = \dfrac{u_1}{I_{xx}} + \dot{\theta}\dot{\psi}\left(\dfrac{I_{yy}-I_{zz}}{I_{xx}}\right), \\ \ddot{\theta} = \dfrac{u_2}{I_{yy}} + \dot{\phi}\dot{\psi}\left(\dfrac{I_{zz}-I_{xx}}{I_{yy}}\right), \\ \ddot{\psi} = \dfrac{u_3}{I_{zz}} + \dot{\phi}\dot{\theta}\left(\dfrac{I_{xx}-I_{yy}}{I_{zz}}\right), \end{cases} \tag{2}$$

where ϕ is the roll angle, θ is the pitch angle, and ψ is the yaw angle, u_1, u_2, u_3 are the torques, I_{xx}, I_{yy}, I_{zz} are the moments of inertia.

The attitude vector of the quadrotor is denoted as $\eta = [\phi,\ \theta,\ \psi]^T \in R^3$. Subsequently, the dynamic model of quadrotor can be expressed in vector form

$$J\ddot{\eta} = U_D + \Gamma, \tag{3}$$

where

$$J = diag\{I_{xx}, I_{yy}, I_{zz}\} \in R^{3\times 3},$$
$$\Gamma = \left[\dot{\theta}\dot{\psi}(I_{yy}-I_{zz}), \dot{\phi}\dot{\psi}(I_{zz}-I_{xx}), \dot{\phi}\dot{\theta}(I_{xx}-I_{yy})\right] \in R^3.$$

Assumption 1: The desired reference input signals $\eta_r = [\phi_r,\ \theta_r,\ \psi_r]^T \in R^3$ and the first derivative signals $\dot{\eta}_r = [\dot{\phi}_r,\ \dot{\theta}_r,\ \dot{\psi}_r]^T \in R^3$ are bounded and available. Denote that $X_r = [\eta_r,\ \dot{\eta}_r] \in \Omega_r$, where Ω_r is a bounded compact set $\Omega_r = \{X_r : \eta_r^2 + \dot{\eta}_r^2 \le H_r\}$ and H_r is a known positive constant.

The control objective is for the mathematical model of the quadrotor (2) in presence of unknown external disturbances, the attitude tracking controller is designed to ensure that all the signals in the closed-loop system are uniform and ultimately bounded.

4. ATTITUDE CONTROL BY DYNAMIC SURFACE CONTROL

In this section, the design process of the main controller based on DSC will be described in detail. The position of the DSC controller in the proposed synthetical controller framework is shown in Fig. 1. A first-order filter is introduced to estimate the virtual control and the derivatives [16–18], and the problem of "complexity explosion" is overcome. Because the complicated differentiation process is avoided, the design process of the DSC controller is simple and effective.

4.1. Attitude control laws

According to the attitude dynamics of the quadrotor (3), the dynamic surface control laws U_D are designed. Similar to the traditional back-stepping method, the coordinate transformation is introduced

$$\begin{aligned} s_1 &= \eta - \eta_r, \\ s_2 &= \dot{\eta} - u_{D2}^f, \\ \tilde{u}_{D2} &= u_{D2}^f - u_{D2}, \end{aligned} \tag{4}$$

with s_1, s_2 being the tracking errors, η_r is the desired reference signal, u_{D2}, u_{D2}^f are intermediate virtual feedback control law and filtered virtual control law, respectively, $\tilde{u}_{D2}$ is filtered error. At Step 1, the desired virtual feedback control u_{D2} is defined, and the filtered virtual control function u_{D2}^f is derived by a first-order filter (8). At Step 2, the control law U_D is designed.

Step 1: The dynamics of attitude tracking error s_1 is

$$\dot{s}_1 = \dot{\eta} - \dot{\eta}_r. \tag{5}$$

To stabilize (5), a smooth virtual control input u_{D2} is designed as

$$u_{D2} = -P_1 s_1 + \dot{\eta}_r, \tag{6}$$

where $P_1 \in R^3$ is a positive definite to be designed later.
Yielding that

$$\dot{s}_1 = s_2 + \tilde{u}_{D2} - P_1 s_1. \tag{7}$$

The filtered virtual control vector u_{D2}^f is obtained by a first-order filter

$$\begin{aligned} &\dot{u}_{D2}^f = (u_{D2} - u_{D2}^f)/\tau = -\tilde{u}_{D2}/\tau, \\ &u_{D2}^f(0) = u_{D2}(0), \end{aligned} \tag{8}$$

where τ is the filter gain.

Step 2: The derivative of s_2 along the trajectory (4) is

$$\dot{s}_2 = \ddot{\eta} - \dot{u}_{D2}^f = J^{-1}U_D + J^{-1}\Gamma - \dot{u}_{D2}^f. \tag{9}$$

The control input U_D is designed as

$$U_D = -P_2 J s_2 - \Gamma - J s_1 + J\dot{u}_{D2}^f, \tag{10}$$

where $P_2 \in R^3$ is a positive definite to be specified.
Yielding that

$$\dot{s}_2 = -P_2 s_2 - s_1. \tag{11}$$

In summary, the state error subsystem $s := [s_1^T, s_2^T]^T$ is denoted as

$$\begin{aligned} &\dot{s}_1 = -P_1 s_1 + s_2 + \tilde{u}_{D2}, \\ &\dot{s}_2 = -P_2 s_2 - s_1. \end{aligned} \tag{12}$$

4.2. Stability analysis

The following Theorem 1 is introduced to facilitate system stability analysis.

Theorem 1: Provided with Assumptions 1 satisfied, considering the system (4) under the dynamic surface controller (10) and filter (8), for the proper control gain matrixes P_1 and P_2, the closed loop system is uniformly ultimately bounded (UUB).

Proof: Considering (8), we can obtain that

$$\dot{\tilde{u}}_{D2} = \dot{u}_{D2}^f - u_{D2} = -\frac{1}{\tau}\tilde{u}_{D2} - \dot{u}_{D2}. \tag{13}$$

To stabilize (4), define the Lyapunov function candidate

$$V_D = \frac{1}{2}(s_1^T s_1 + s_2^T s_2 + \tilde{u}_{D2}^T \tilde{u}_{D2}). \tag{14}$$

Then the time derivative of $\dot{V}_D$ is given by

$$\begin{aligned} \dot{V}_D =& s_1^T \dot{s}_1 + s_2^T \dot{s}_2 + \tilde{u}_{D2}^T \dot{\tilde{u}}_{D2} \\ =& s_1^T(s_2 + \tilde{u}_{D2} - P_1 s_1) + s_2^T(-P_2 s_2 - s_1) \\ &+ \tilde{u}_{D2}^T\left(-\frac{1}{\tau}\tilde{u}_{D2} - \dot{u}_{D2}\right) \\ =& -s_1^T P_1 s_1 - s_2^T P_2 s_2 + s_1 \tilde{u}_{D2} + \tilde{u}_{D2}^T\left(-\frac{1}{\tau}\tilde{u}_{D2} - \dot{\tilde{u}}_{D2}\right) \\ =& -s_1^T P_1 s_1 - s_2^T P_2 s_2 - \frac{1}{\tau}\tilde{u}_{D2}^T \tilde{u}_{D2} - \tilde{u}_{D2}^T \dot{u}_{D2}. \end{aligned} \tag{15}$$

Define $\mathrm{H}(s_1, \tilde{u}_{D2}) = \dot{u}_{D2}$, where $\mathrm{H}(\cdot)$ is continuous function, Ω_1 is a bounded compact set, $\Omega_1 = \{[s_1, \tilde{u}_{D2}]^T : V_D \leq \bar{\omega}_1\}$, $\bar{\omega}_1$ is a known positive constant, and $\Omega_1 \times \Omega_r$ is a compact set. Assume that the maximum value of H on the compact set $\Omega_1 \times \Omega_r$ is $\mathrm{H}_{\max}$. By Young's inequality, we can deduce that

$$-\tilde{u}_{D2}^T \dot{u}_{D2} \leq \left\|\tilde{u}_{D2}^T\right\| \mathrm{H} \leq \left\|\tilde{u}_{D2}^T\right\|^2 + \frac{1}{4}\mathrm{H}_{\max}^2. \tag{16}$$

Substituting (16) into (15), yields

$$\begin{aligned} \dot{V}_D =& -s_1^T P_1 s_1 - s_2^T P_2 s_2 - \frac{1}{\tau}\tilde{u}_{D2}^T \tilde{u}_{D2} - \tilde{u}_{D2}^T \dot{u}_{D2} \\ \leq& -s_1^T P_1 s_1 - s_2^T P_2 s_2 - \frac{1}{\tau}\tilde{u}_{D2}^T \tilde{u}_{D2} + \left\|\tilde{u}_{D2}^T\right\|^2 + \frac{1}{4}\mathrm{H}_{\max}^2 \\ \leq& -\bar{k}V_D + C, \end{aligned} \tag{17}$$

$$\bar{k} = \min\{\mathrm{diag}(P_1), \mathrm{diag}(P_2), \frac{1}{\tau} - 1\},\ P_1 > 0,\ P_2 > 0,$$

$$0 < \tau < 1,\ C = \frac{1}{4}\mathrm{H}_{\max}^2.$$

Yields that

$$\dot{V}_D(t) \leq 0, \tag{18}$$

as long as the following inequality holds

$$V_D(t) > C/\bar{k}. \tag{19}$$

According to the Lyapunov stability theorem, the tracking errors and filtered error s_1, s_2, $\tilde{u}_{D2}$ are UUB.

5. ATTITUDE CONTROL BY ADHDP

ADHDP is utilized to generate a complementary data-driven control signal for the total attitude controller, to reduce tracking errors and improve control accuracies. As is shown in Fig. 1, the ADHDP controller is mainly composed of action neural network (AN) and critic neural network (CN). The sum of outputs of the DSC and the AN works as the control input signal. In Section 5.1, the ADHDP control algorithm is proposed. Section 5.2 presents the feed-forward and feed-back learning of the action network and the critic network, and explains the adaptive gradient-based policy to update the weight coefficients of the neural network. Section 5.3 gives the Lyapunov stability analysis of the proposed ADHDP algorithm.

5.1. ADHDP control algorithm

The roll angle, pitch angle and yaw angle responses under the DSC attitude control law $U_D(t)$ are denoted as $\phi_D(t)$, $\theta_D(t)$ and $\psi_D(t)$ at time t. The remainder errors at time t, are defined as

$$\begin{aligned} \tilde{e}_\phi(t) &= \phi_r(t) - \phi_D(t), \\ \tilde{e}_\theta(t) &= \theta_r(t) - \theta_D(t), \\ \tilde{e}_\psi(t) &= \psi_r(t) - \psi_D(t), \\ \tilde{e}_\psi(t) &= \psi_r(t) - \psi_D(t), \end{aligned} \tag{20}$$

which are further reduced by the ADHDP control.

The utility function in ADHDP at time t is defined as

$$r_f(t) = f_r(t) C_r f_r^T(t), \tag{21}$$

where $f_r(t) = [z^T(t), U_A^T(t)]$, C_r is a positive-definite diagonal matrix with corresponding dimensions, and $z_1(t) = [\tilde{e}_\phi(t), \dot{\tilde{e}}_\phi(t), \tilde{e}_\theta(t), \dot{\tilde{e}}_\theta(t), \tilde{e}_\psi(t), \dot{\tilde{e}}_\psi(t)]$.

As the input vector of the action network in ADHDP, $z(t)$ is composed of $z_1(t)$ and the corresponding one-time-step-delay

$$z(t) = [z_1(t-\Delta t), z_1(t)]^T. \tag{22}$$

Define the minimization of the cost function as

$$\begin{aligned} J(z(t)) &= \min_{U_A(t)} \{rf(z(t), U_A(t)) + rf(z(t+\Delta t), \\ &U_A(t+\Delta t)) + rf(z(t+2\Delta t), U_A(t+2\Delta t)) + \cdots\cdots\} \\ &= \min_{U_A(t)} \{rf(z(t), U_A(t)) + \gamma J(z(t+\Delta t))\}, \end{aligned} \tag{23}$$

where $J(z(t))$ is the total cost value, γ, $0 < \gamma < 1$, is a discount factor. If $z(t) \neq 0$ and $U_a \neq 0$, $r_f(z(t), U_A(t))$ is positive-definite, as C_r is the positive-definite matrix, and only when $z(t) = 0$ and $U_a = 0$, $r_f(z(t), U_a(t))$ satisfy $r_f(z(t), U_A(t)) = 0$.

The optimal cost function $J^*(z(t))$ is the exclusive solution of (18), satisfying the following Bellman's equation

$$J^*(z(t)) = \min_{U_A(t)} \{r_f(z(t), U_A(t)) + \gamma J(z(t+\Delta t))\}. \tag{24}$$

The main idea of the ADHDP method is to solve Bellman's equation approximately. $J^*(z(t))$ can convergent to $\hat{J}(z(t))$, which is the output of critic network. That is, when the ADHDP control policy $U_A(t) = 0$, $J^*(z(t)) = \hat{J}(z(t)) = 0$, the roll, pitch and yaw can track to the desired reference signal favorablely when $z_1(t) = [\tilde{e}_\phi(t), \dot{\tilde{e}}_\phi(t), \tilde{e}_\theta(t), \dot{\tilde{e}}_\theta(t), \tilde{e}_\psi(t), \dot{\tilde{e}}_\psi(t)]$ all converge to zero.

5.2. Feed-forward and feed-back learning of action network and critic network

The estimated outputs of the critic network are the cost function $\hat{J}(t)$. There exists an approximation error between $\hat{J}(t)$ and the real cost function $J(t)$. The weights of critic network are updated iteratively to minimize the errors between the predictive cost function $\hat{J}(t)$ and real counterpart $J(t)$. Note that the critic network is the action network dependent. In detail, the inputs and outputs of the action network are chosen as the inputs of critic-network.

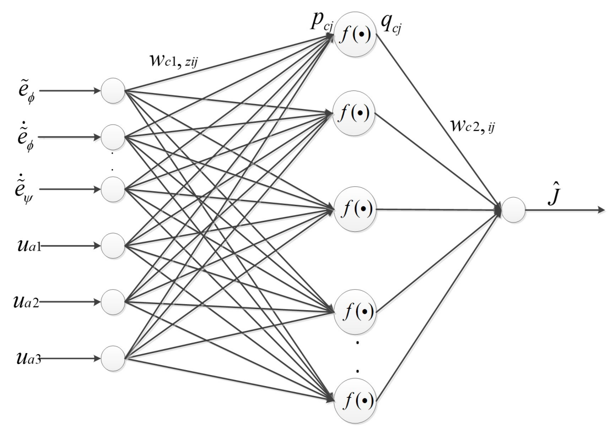


Fig. 2. Critic neural network.

The critic network is the function of $z(t)$, $U_A(t)$, $w_c(t)$, where $w_c(t)$ is the weight vector. In the critic network, the input vector $c_i(t)$ and $c_o(t)$ the output vector are defined as

$$\begin{aligned} c_i(t) &= [z^T(t),\ U_A^T]^T, \\ c_o(t) &= \hat{J}(z(t)). \end{aligned} \tag{25}$$

Defined the error function as

$$e_c(t) = \gamma \hat{J}(t) - [\hat{J}(t-\Delta t) - r_f(t)]. \tag{26}$$

Hence, the cost function to update the weights in the critic network is defined as

$$\min_{\omega_c(t)} E_c(t) = \min_{\omega_c(t)} \frac{1}{2} e_c^T(t) e_c(t). \tag{27}$$

Fig. 2 states the neural networks scheme of the critic network in ADHDP. Assume that the action network has n inputs and m outputs, the critic network has N_{ci} input nodes ($N_{ci} = n + m$), N_{ch} hidden nodes, and one output node. Choosing the hyperbolic tangent threshold function $f_c(t) = f_a(t) = (1-e^t)/(1+e^t)$ as the activation function in the critic network and the action network. The intermediate variables $p_{cj}(t)$ and $q_{cj}(t)$ for the j-th hidden node can be described as

$$\begin{aligned} p_{cj}(t) &= \sum_{i=1}^{n} z_i(t) w_{c1z,ij}(t) + \sum_{i=1}^{m} u_{ai}(t) w_{c1u,ij}(t), \\ j &= 1,\ \cdots,\ N_{ch}, \\ q_{cj}(t) &= f(p_{cj}(t)) = \frac{1-e^{-p_{cj}(t)}}{1+e^{-p_{cj}(t)}},\quad j = 1,\ \cdots,\ N_{ch}, \\ \hat{J}(t) &= \sum_{j=1}^{N_{ch}} w_{c2,j}(t) q_{cj}(t), \end{aligned} \tag{28}$$

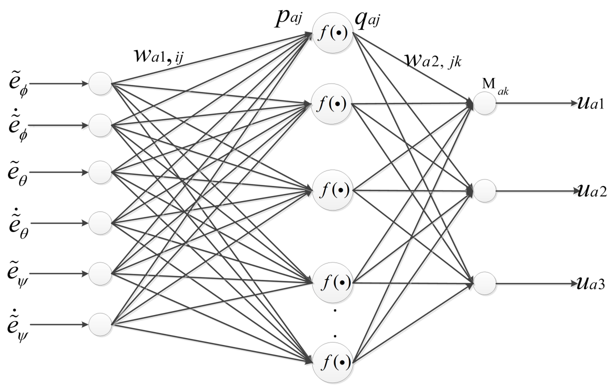


Fig. 3. Action neural network.

where $w_{c1,ij}(t)$ and $w_{c2,j}(t)$ denote the weights from the i-th input node to the j-th hidden node and from the j-th hidden node to output node, $p_{cj}(t)$ and $q_{cj}(t)$ are the input and the output of the j-th hidden node, respectively.

The adaptive gradient-based policy is introduced to update the weight coefficients. Subsequently, the weights updating algorithm in the critic network is expressed as

$$\begin{aligned}
&\Delta w_{c2,j}(t)=-\eta_c(t)\frac{\partial Ec(t)}{\partial \hat{J}(t)}\frac{\partial \hat{J}(t)}{\partial w_{c2,j}(t)},\\
&\Delta w_{c1,zij}(t)=-\eta_c(t)\frac{\partial Ec(t)}{\partial \hat{J}(t)}\frac{\partial \hat{J}(t)}{\partial q_{cj}(t)}\frac{\partial q_{cj}(t)}{\partial p_{cj}(t)}\frac{\partial p_{cj}(t)}{\partial w_{c1,zij(t)}},\\
&\Delta w_{c1,uij}(t)=-\eta_c(t)\frac{\partial Ec(t)}{\partial \hat{J}(t)}\frac{\partial \hat{J}(t)}{\partial q_{cj}(t)}\frac{\partial q_{cj}(t)}{\partial p_{cj}(t)}\frac{\partial p_{cj}(t)}{\partial w_{c1,uij(t)}},\\
&w_{c2,j}(t+\Delta t)=w_{c2,j}(t)+\Delta w_{c2,j}(t),\\
&w_{c1,zij}(t+\Delta t)=w_{c1,zij}(t)+\Delta w_{c1,zij}(t),\\
&w_{c1,uij}(t+\Delta t)=w_{c1,uij}(t)+\Delta w_{c1,uij}(t),
\end{aligned}\tag{29}$$

where $\eta_c(t)$ is the learning rate in the critic network at time t. In the attitude tracking problem, $\phi(t)$, $\theta(t)$ and $\psi(t)$ are system responds under the DSC-ADHDP-based control action.

The action network is a function of $z(t)$, $w_a(t)$, and $w_a(t)$ is the weight vectors of the action network from hidden-to-output layer. The input and the output of the action network are defined as

$$a_i(t)=z(t),\ \ a_0(t)=U_A(t). \tag{30}$$

Fig. 3 shows the structure diagram of the action neural network in ADHDP. The notations N_{ai},N_{ah} and N_{ao} are the quantities of input nodes, hidden nodes and output nodes respectively. And u_{ak}, $k=1,\cdots,N_{ao}$, i.e., $m=N_{ao}$ is the outputs of action network which are also chosen as the input vectors of critic network.

The intermediate variables $p_{aj}(t)$, $q_{aj}(t)$, $M_{ak}(t)$ and the output variable $u_{ak}(t)$ can be expressed as

$$\begin{aligned}
&p_{aj}(t)=\textstyle\sum_{i=1}^{Nai} z_i(t)w_{a1,zij}(t),\ \ j=1,\cdots,N_{ah},\\
&q_{aj}(t)=f(p_{aj}(t))=\frac{1-e^{-p_{aj}(t)}}{1+e^{-p_{aj}(t)}},\ \ j=1,\cdots,N_{ah},\\
&M_{ak}(t)=w_{a2,jk}(t)q_{aj}(t),\ \ k=1,\cdots,N_{ao},\\
&u_{ak}(t)=f(M_{ak}(t))=\frac{1-e^{-M_{ak}(t)}}{1+e^{-M_{ak}(t)}},\ \ k=1,\cdots,N_{ao},
\end{aligned}\tag{31}$$

where $w_{a1,zij}(t)$ and $w_{a2,jk}(t)$ are the weights from the i-th input node to the j-th hidden node and from the j-th hidden node to the k-th output node, $p_{aj}(t)$ and $q_{aj}(t)$ are the input and the output for the j-th hidden node of the action neural network. The action network is trained by back-propagating the error between the ultimate objective U_f and the approximate value $\hat{J}(t)$ from the critic network

$$e_a(t)=\hat{J}(t)-U_f, \tag{32}$$

where $U_f$is the desired ultimate cost objective.

Generally speaking, $U_f=0$ means a success learning implement for all t. The cost function in the action network is designed as

$$\min_{wa(t)} E_a(t)=\min_{wa(t)}\frac{1}{2}e_a^T(t)ea(t). \tag{33}$$

Similarly, an adaptive gradient-based strategy is utilized to update the weight coefficients.

Considering

$$\Delta w_a(t)=-\eta_a(t)\frac{\partial E_a(t)}{\partial w_a(t)}=-\eta_a(t)\frac{\partial E_a(t)}{\partial \hat{J}(t)}\frac{\partial \hat{J}(t)}{\partial U_A(t)}. \tag{34}$$

By the chain derivation rule, the weights are expressed by

$$\begin{aligned}
&\Delta w_{a2,jk}(t)=-\eta_a(t)\frac{\partial Ea(t)}{\partial \hat{J}(t)}\frac{\partial \hat{J}(t)}{\partial u_{ak}(t)}\frac{\partial u_{ak}(t)}{\partial M_{ak}(t)}\frac{\partial M_{ak}(t)}{\partial w_{a2,jk}(t)},\\
&\Delta w_{a1,ij}(t)=-\eta_a(t)\frac{\partial Ea(t)}{\partial \hat{J}(t)}\frac{\partial \hat{J}(t)}{\partial u_{ak}(t)}\frac{\partial u_{ak}(t)}{\partial M_{ak}(t)}\frac{\partial M_{ak}(t)}{\partial q_{aj}(t)}\\
&\qquad\qquad\times\frac{\partial q_{aj}(t)}{\partial p_{cj}(t)}\frac{\partial p_{cj}(t)}{\partial w_{a1,ij(t)}},\\
&w_{a1,ij}(t+\Delta t)=w_{a1,ij}(t)+\Delta w_{a1,ij}(t),\\
&w_{a2,jk}(t+\Delta t)=w_{a2,jk}(t)+\Delta w_{a2,jk}(t),
\end{aligned}\tag{35}$$

where $\eta_a(t)$ is the learning rate in the action network at time t.

5.3. Stability analysis

The uniformly ultimately bounded (UUB) stability for the ADHDP algorithm has been provided theoretically in [46,53–57]. For the sake of clarity, the following Lyapunov stability analysis for the ADP algorithm is provided.

Recalling the universal approximation theorem for neural networks, if the quality of hidden-layer neurons is abundant enough, the approximation error can be arbitrarily small. Hence the weights of the input-to-hidden layer are initialized randomly in the action network and the critic network. $\hat{J}(t)$ can be expressed as $\hat{J}(t) = w_c^T f_c(t)$. Similarly, $U_A(t)$ can also be expressed as $U_A(t) = w_a^T(t) f_a(t)$. $w_c(t)$ is the critic network's weight vectors from the hidden-to-output layer, and $w_a(t)$ is the action network's weight vectors from the hidden-to-output layer.

Assumption 2: Defining the optimal weight vectors of the hidden-to-output layer for critic and action networks as $w_c^*(t)$ and $w_a^*(t)$. All the weight vectors are bounded, $\|w_c(t)\| \le w_{cm}(t)$, $\|w_c^*(t)\| \le w_{cm}(t)$, $\|w_a(t)\| \le w_{am}(t)$ and $\| w_a^*(t)\| \le w_{am}(t)$. Therefore, the activation functions are bounded, $\|f_c(t)\| \le f_{cm}(t)$, and $\|f_a(t)\| \le f_{am}(t)$.

Theorem 2: Assume that the weights of the critic network and the action network are updated according to the gradient descent algorithm, and the reinforcement signal is bounded within $0 \le r_f \le 1$. The critic network weights are given by (37). Then the weight estimation errors between optimal weights w_c^*, w_a^*, and the counterpart estimations $w_c(t)$, $w_a(t)$ are UUB, if the following conditions are fulfilled

$$\begin{aligned} &\eta_c \le \frac{1}{\gamma\|f_c\|^2}, \ \ \eta_a \le \frac{1}{\|f_a\|^2}, \\ &\|\zeta_c(t)\| > (\Upsilon_{1m}^2 + \gamma^{-1}\Upsilon_{2m}^2\Lambda_{2m}^2)^{\frac{1}{2}}. \end{aligned} \tag{36}$$

Proof: The critic network weights are

$$\begin{aligned} w_c(t+\Delta t) =& w_c(t) - \eta_c(t)\frac{\partial E_c(t)}{\partial w_c(t)} \\ =& w_c(t) - \eta_c f_c(t)[\gamma w_c^T(t) f_c(t) \\ &+ r_f(t) - w_c^T(t-\Delta t) f_c(t-\Delta t)]^T. \end{aligned} \tag{37}$$

The action network weights are chosen as

$$\begin{aligned} &w_a(t+\Delta t) \\ &= w_a(t) - \eta_a(t)\frac{\partial E_a(t)}{\partial w_a(t)} \\ &= w_a(t) - \eta_a f_a(t)[w_c^T(t)D(t)]^T[w_c^T(t)f_c(t)]^T, \end{aligned} \tag{38}$$

where $D(t)$ is the matrix with dimension $N_h^c \times m$, and the component is expressed as

$$\begin{aligned} &D_{jk}(t) = 0.5(1-f_{cj}^2(t))w_{cj,n+k}, \\ &j = 1,\cdots,N_h^c, \ \ k = 1,\cdots,m. \end{aligned} \tag{39}$$

Denote the weight error and the approximation error as $\tilde{w}_c(t) = w_c(t) - w_c^*(t)$ and $\zeta_c(t) = \tilde{w}_c(t) f_c(t)$ in the critic network. Denote the weight error as $\tilde{w}_a(t) = w_a - w_a^*(t)$ in the action network. The Lyapunov functions are chosen as $Y(t) = Y_1(t) + Y_2(t)$, $Y_1(t) = \frac{1}{h_c} tr(\tilde{w}_c^T(t)\tilde{w}_c(t))$ and $Y_2(t) = \frac{1}{\eta_a} tr(\tilde{w}_a^T(t)\tilde{w}_a(t))$. The first difference of $Y(t)$ is $\Delta Y(t) = Y(t+\Delta t) - Y(t) = \Delta Y_1(t) + \Delta Y_2(t)$.

In particular,

$$\Delta Y_1(t) = \frac{1}{\eta_c} tr(\tilde{w}_c^T(t+\Delta t)\tilde{w}_c^(t+\Delta t) - \tilde{w}_c^T(t)\tilde{w}_c^(t)), \tag{40}$$

$$\begin{aligned} &\tilde{w}_c(t+\Delta t) \\ &= w_c(t+\Delta t) - w_c^* \\ &= \tilde{w}_c(t) - \eta_c f_c(t)[\gamma(\tilde{w}_c(t) + w_c^*)^T f_c + r_f(t) \\ &\quad - \tilde{w}_c^T(t-\Delta t) f_c(t-\Delta t)]^T \\ &= (1-\eta_c\gamma f_c(t) f_c^T(t))\tilde{w}_c(t) - \eta_c f_c(t) \\ &\quad \times [\gamma(w_c^{*T} f_c(t) + r_f(t) - \tilde{w}_c^T(t-\Delta t) f_c(t-\Delta t))]^T. \end{aligned} \tag{41}$$

Substituting (41) into (40), we obtain

$$\begin{aligned} \Delta Y_1(t) =& \frac{1}{\eta_c}\{(1-\eta_c\gamma f_c(t) f_c^T(t))^2\tilde{w}_c^2(t2\eta_c\tilde{w}_c^T(t) f_c^T(t) \\ &\times (1\eta_c\gamma f_c(t) f_c^T(t))(\gamma w_c^{*T} f_c(t) + r_f(t) \\ &- \tilde{w}_c^T(t-\Delta t) f_c(t-\Delta t))^T \\ &+ \eta_c^2\gamma^2 f_c(t) f_c^T(t)(w_c^{*T} f_c(t+\gamma^{-1} r_f(t) \\ &-\gamma^{-1}\tilde{w}_c^T(t-\Delta t) f_c(t-\Delta t))^2 - \tilde{w}_c^T(t)\tilde{w}_c^(t)))\}. \end{aligned} \tag{42}$$

Denote that

$$\begin{aligned} X_1 &= (1-\eta_c\gamma f_c(t) f_c^T(t))^2\tilde{w}_c^2(t) \\ &= \tilde{w}_c^T(t)\tilde{w}_c(t) - 2\eta_c\gamma\|\zeta_c(t)\|^2 \\ &\quad + \eta_c^2\gamma^2\tilde{w}_c^T(t) f_c(t) f_c^T(t)\tilde{w}_c(t) \\ &= \tilde{w}_c^T(t)\tilde{w}_c(t) - \eta_c\gamma\|\zeta_c(t)\|^2 \\ &\quad - \eta_c\gamma\|\zeta_c(t)\|^2(1-\eta_c\gamma f_c(t) f_c^T(t)), \end{aligned} \tag{43}$$

$$\begin{aligned} X_2 =& -2\eta_c\tilde{w}_c^T(t) f_c(t)(1-\eta_c\gamma f_c(t) f_c^T(t)) \\ &\times (\gamma w_c^{*T} f_c(t) + r_f(t) - \tilde{w}_c^T(t-\Delta t) f_c(t-\Delta t))^T, \end{aligned} \tag{44}$$

$$\begin{aligned} X_3 &= \eta_c^2\gamma^2 f_c(t) f_c^T(t)(w_c^{*T} f_c(t) + \gamma^{-1} r_f(t) \\ &\quad - \gamma^{-1}\tilde{w}_c^T(t-\Delta t) f_c(t-\Delta t)^2 \\ &= \eta_c^2\gamma^2 f_c(t) f_c^T ||w_c^{*T} f_c(t) + \gamma^{-1} r_f(t) \\ &\quad - \gamma^{-1}\tilde{w}_c^T(t-\Delta t) f_c(t-\Delta t)||^2, \end{aligned} \tag{45}$$

using $\Upsilon_1(t) = w_c^{*T} f_c(t) + \gamma^{-1} r_f(t) - \gamma^{-1}\tilde{w}_c^T(t-\Delta t) f_c(t-\Delta t)$ and $\Lambda_1(t) = \gamma(1-\eta_c\gamma f_c(t) f_c^T(t))$.

Considering the expression of X_1, X_2 and X_3, $\Delta Y_1(t)$ can be expressed as

$$\begin{aligned} \Delta Y_1(t) =& \frac{1}{\eta_c}(X_1 + X_2 + X_3 - \tilde{w}_c^T(t)\tilde{w}_c(t)) \\ =& -\gamma\|\zeta_c(t)\|^2 - \Lambda_1(t)\|\zeta_c(t)\|^2 \\ &- 2\Lambda_1(t)\zeta_c(t)\Upsilon_1(t) \\ &- \Lambda_1(t)\|\Upsilon_1(t)\|^2 + \gamma\|\Upsilon_1(t)\|^2 \\ =& -\gamma\|\zeta_c(t)\|^2 - \Lambda_1(t)\|\zeta_c(t) + \Upsilon_1(t)\|^2 \end{aligned}$$

$$+\gamma\|\Upsilon_1(t)\|^2. \tag{46}$$

Similarly,

$$\Delta Y_2(t) = \frac{1}{\eta_a} tr(\tilde{w}_a^T(t+\Delta t)\tilde{w}_a(t+\Delta t) - \tilde{w}_a^T(t)\tilde{w}_a(t)), \tag{47}$$

$$\begin{aligned}\tilde{w}_a(t+\Delta t) &= w_a(t+\Delta t) - w_a^* \\ &= \tilde{w}_a(t) - \eta_a f_a(t) w_x^T(t) D(t)[w_c^T(t) f_c(t)]^T.\end{aligned} \tag{48}$$

Substituting (48) into (47), we can deduce that

$$\begin{aligned}\Delta Y_2(t) =& -2 f_a(t) w_c^T(t) D(t)[w_c^T(\eta_a f_a(t) f_a^T(t) \\ &\times \left\|w_c^T(t) D(t)\right\|^2 \left\|w_c^T(t) f_c(t)\right\|^2)].\end{aligned} \tag{49}$$

Denoting $\Upsilon_2(t) = w_c^T(t)D(t)$ and $\Lambda_2(t) = w_c^T(t) f_c(t)$, (43) can be expressed as

$$\begin{aligned}\Delta Y_2 =& -2 f_a(t)\Upsilon_2(t)\Lambda_2^T(t) \\ &+ \eta_a f_a(t) f_a^T(t)\|\Upsilon_2(t)\|^2 \left\|\Lambda_2^(t)\right\|^2 \\ =& -\|\Upsilon_2(t)\|^2\|\Lambda_2(t)\|^2 \\ &+ \eta_a f_a(t) f_a^T(t)\|\Upsilon_2(t)\|^2\|\Lambda_2(t)\|^2 \\ &+ (\|\Upsilon_2(t)\|^2 \left\|\Lambda_2^(t)\right\|^2 - 2 f_a(t)\Upsilon_2(t)\Lambda_2^T(t)) \\ =& -(1-\eta_a f_a(t) f_a^T(t)\|\Upsilon_2(t)\|^2\|\Lambda_2(t)\|^2 \\ &+ \left\|\Upsilon_2(t)\Lambda_2^T(t) - f_a(t)\right\|^2 - \|f_a(t)\|^2).\end{aligned} \tag{50}$$

The first differential of Lyapunov function $Y(t)$ is

$$\begin{aligned}\Delta Y(t) =& \Delta Y_1(t) + \Delta Y_2(t) \\ =& -\gamma\|\zeta_c(t)\|^2 - \Lambda_1(t)\|\zeta_c(t)+\Upsilon_1(t)\|^2 \\ &+ \gamma\|\Upsilon_1(t)\|^2 \\ &- (1-\eta_a f_a(t) f_a^T(t)\|\Upsilon_2(t)\|^2\|\Lambda_2(t)\|^2 \\ &+ \left\|\Upsilon_2(t)\Lambda_2^T - f_a(t)\right\|^2 - \|f_a(t)\|^2).\end{aligned} \tag{51}$$

According to Assumption 2, we have $\|\Upsilon_1\| \le \Upsilon_{1m}$, $\|\Upsilon_2\| \le \Upsilon_{2m}$ and $\|\Lambda_2\| \le \Lambda_{2m}$, and $\Delta Y(t)$ can be deduced as

$$\begin{aligned}\Delta Y(t) \le& -\gamma\|\zeta_c(t)\|^2 \\ &- \gamma(1-\eta_c\gamma f_c(t) f_c^T(t))\|\zeta_c(t)+\Upsilon_1(t)\|^2 \\ &- (1-\eta_a f_a(t) f_a^T(t))\|\Upsilon_2(t)\|^2\|\Lambda_2(t)\|^2 \\ &+ \gamma\|\Upsilon_1(t)\|^2 + \|\Upsilon_2(t)\|^2\|\Lambda_2(t)\|^2 \\ \le& -\gamma\|\zeta_c(t)\|^2 \\ &- \gamma(1-\eta_c\gamma f_c(t) f_c^T(t))\|\zeta_c(t)+\Upsilon_1(t)\|^2 \\ &- (1-\eta_a f_a(t) f_a^T(t)) \times \|\Upsilon_2(t)\|^2\|\Lambda_2(t)\|^2 \\ &+ \gamma\Upsilon_{1m}^2 + \Upsilon_{2m}^2\Lambda_{2m}^2.\end{aligned} \tag{52}$$

We can deduce that $\Delta Y(t) \le 0$, as long as

$$\eta_c \le \frac{1}{\gamma\|f_c\|^2}, \quad \eta_a \le \frac{1}{\|f_a\|^2},$$

$$\|\zeta_c(t)\| > (\Upsilon_{1m}^2 + \gamma^{-1}\Upsilon_{2m}^2\Lambda_{2m}^2)^{\frac{1}{2}} \tag{53}$$

According to the Lyapunov stability theorem, w_c^* and w_a^*, $w_c(t)$ and $w_a^(t)$ are UUB. Therefore, all signals in the closed loop system are UUB.

6. NUMERICAL SIMULATION

Numerical simulations are carried out to demonstrate the effectiveness of the proposed control strategy. The algorithm designed in this paper is compared with the traditional back-stepping control algorithm, and the quadrotor attitude curve that reflects the control accuracy and anti-disturbance ability is obtained through Simulink, a visual simulation tool in MATLAB.

Considering accuracy and timeliness, and after a lot of experimental data analysis, we choose the controller gains to be as follows: in ADHDP controller, the power matrix in the utility function is $C_r = 0.15 I_{15\times15}$ where I expresses the identity matrix. The initial and final learning rates are setting to $\eta_c(0) = 0.1$, $\eta_a(0) = 0.1$, $\eta_c(\infty) = 0.006$, $\eta_a(\infty) = 0.006$ in the critic and action networks. The desired reference command signal in the simulations is chosen as $\phi_r(t) = \theta_r(t) = \psi_r(t) = \sin(t)$.

Before the actual roll, pitch and yaw track to the desired reference commands, the utility function $r_f(t)$ is greater than 0. At each time step, DSC-ADHDP-based controller is trained based on two stop criterions. One is the accuracy of the tolerance errors which are set to 10^{-5} and denoted as T_c and T_a in the critic and action networks. The others are the maximal backpropagation cycles in the critic and action networks, denoted as n_c and n_a. If any one of the two stop criterions is fulfilled, the critic and action networks are considered to reach the proper weights.

The model and controller parameters listed in Tables 1-5 are also referred to [58–60].

Table 1. Model parameters.

I_{xx}	0.0033 kg·m^2
I_y	0.0033 kg·m^2
I_z	0.0058 kg·m^2
m	1.5 kg
d	1 m

Table 2. Traditional backstepping controller parameters.

p_1	$20 \times diag[0.5, 0.45, 0.45]$
p_2	$2.83 \times diag[3.5, 3.5, 2.5]$
τ	0.5

Table 3. DSC controller parameters (attitude controller).

p_1	$20 \times diag[0.5, 0.5, 0.5]$
p_2	$2.81 \times diag[3.5, 3.5, 2.5]$
τ	0.6

Table 4. ADHDP controller parameters (attitude controller).

Variables	Significance	Values
N_c^h	hidden nodes of the critic network	11
N_a^h	hidden nodes of the action network	9
N_{ao}	Outputs of the action network	3
γ	discount factor	0.9
$\eta_c(0)$	initial learning rate of critic network	0.1
$\eta_a(0)$	initial learning rate of action network	0.1
$\eta_c(\infty)$	final learning rate of critic network	0.006
$\eta_a(\infty)$	final learning rate of action network	0.006
T_c	the accuracy of critic network's tolerance error	10^{-5}
T_a	the accuracy of action network's tolerance error	10^{-5}
n_c	the maximal backpropagation cycles of the critic network	100
n_a	the maximal backpropagation cycles of the action network	80

Table 5. PID controller gains (position controller).

Channels	K_p	K_i	K_d
Roll	0.10	0.00	0.00
Yaw	0.09	0.00	0.00
Pitch	0.10	0.00	0.00

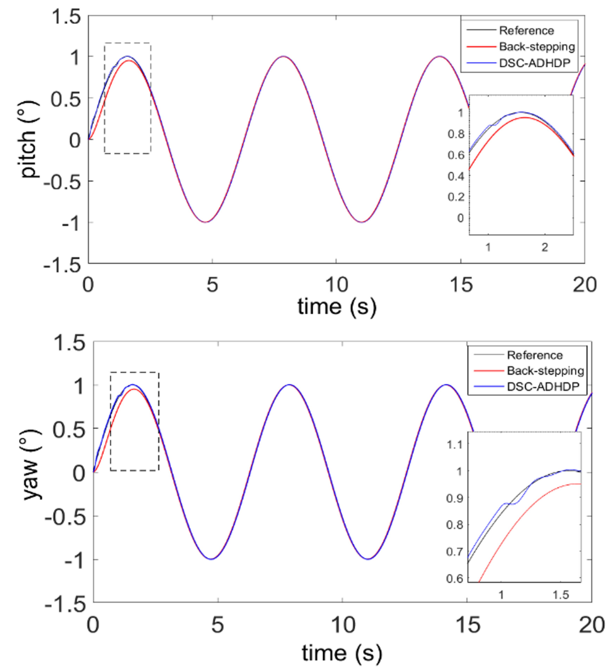
6.1. Simulation of nominal tracking control

Fig. 4(a) shows the attitude tracking under the DSC-ADHDP-based control and traditional back-stepping control. It demonstrates that the attitude tracking signals under synthetical controller can converge to desired reference signals in 1.8 s, and the tracking errors can reach within, whereas the convergence time in traditional back-stepping is 3.1 s, tracking errors are within 5.8°. Fig. 4(b) shows that the simulation results of control input U_1, U_2, U_3, which are continuous and smooth under DSC-ADHDP-based control. Note that the amplitude of U_3 ranges from 6.5×10^{-3} N·m down to 4.6×10^{-3} N·m. Therefore, the attitude system under DSC-ADHDP-based control possesses better control performance than back-stepping technology.

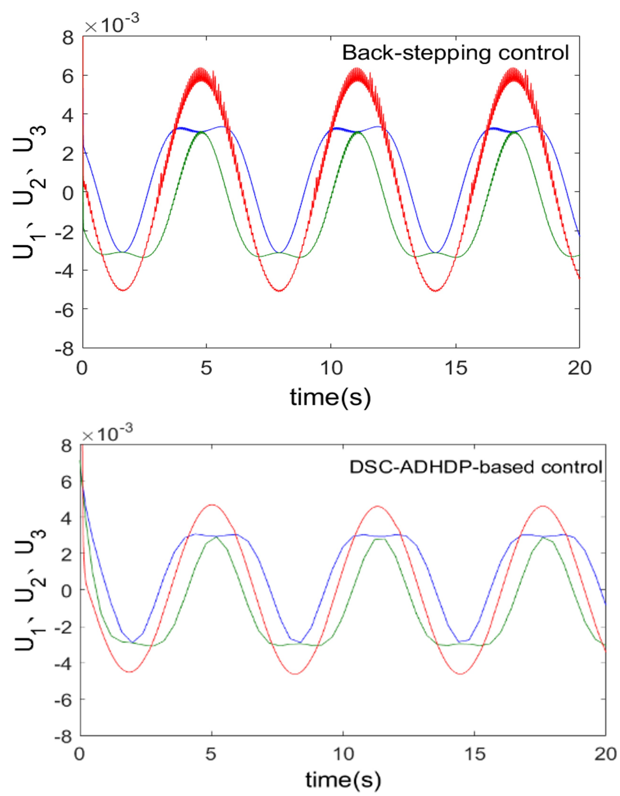
6.2. Tracking control with external disturbances

An additional disturbance signal $0.3\sin(t)$ is set in 8-12 s in roll, pitch and yaw channel. The compared simulation results between the DSC-ADHDP-based control and traditional back-stepping control are present in Fig. 5. Referring to the convergence speed and the effect of anti-disturbance, the DSC-ADHDP-based control is obviously superior to the traditional back-stepping control.

Under DSC-ADHDP-based control, there is a slight deviation between actual trajectory and desired reference at 8 s, then the actual tracking trajectory can track the desired



(a) Attitude tracking under reference signal.



(b) Control input under two policy.

Fig. 4. Simulation results of quadrotor attitude tracking and control input u_1, u_2, u_3 (N·m).

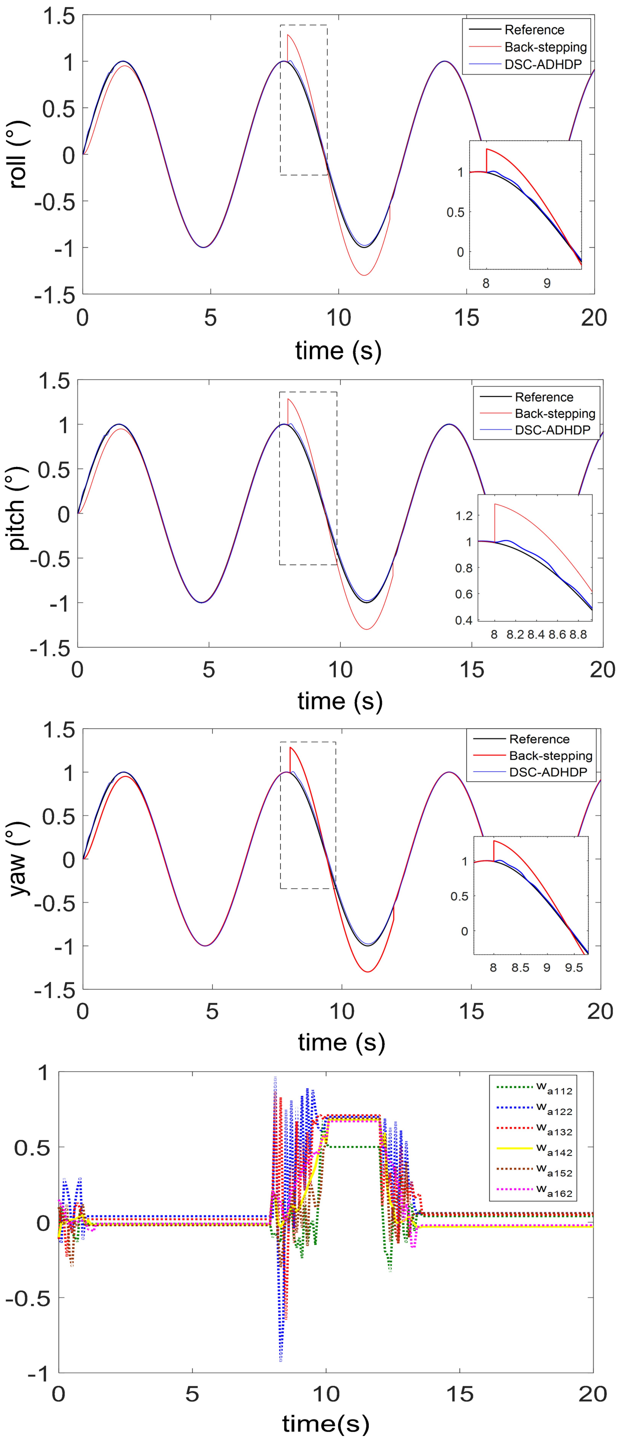


Fig. 5. Simulation results of quadrotor attitude tracking under external disturbance and weights updating of the action network.

reference signal perfectly at 9.7 s and tracking errors can reach within 4.5°. Under traditional back-stepping control, tracking performances are inferior to DSC-ADHDP-based control, as the bigger deviations are generated at the same time. The fourth figure in Fig. 5 shows the weights updating trajectory from all the inputs to the second hidden node in the action network, which illustrates that the weight can be adjusted adaptively.

The simulation results illustrate that the convergence speed and disturbance rejection of the attitude controller are improved due to the introduction of ADHDP complementary control, and the synthetical controller can achieve satisfactory tracking performances even under external disturbances.

7. FLIGHT EXPERIMENTS

To verify the effectiveness of the proposed algorithm, a flight experiment with the Pixhawk autopilot is accomplished, under the proposed algorithm in the attitude system and the PID algorithm in the position system. The controller parameters are shown in Table 7. The cascade PID algorithm was adopted to complete the comparison experiment.

The Pixhawk autopilot is provided with superior computing power, equipped with two processor (specific parameters are shown in Table 6). The proposed control algorithm can be performed in real-time by Pixhawk.

As is shown in Fig. 6, a quadrotor simulation model was established based on Simulink, a visual simulation tool in MATLAB, and communicated with mission planner ground station shown in Fig. 7. The proposed algorithm was transplanted to the attitude controller of the Pixhawk autopilot by using Mavlink communication protocol. The PID algorithm was still used in the position controller. After algorithm transplantation, FlightGear software can be employed for experiment.

When good simulation results are obtained, we begin to carry out hardware flight experiments. In order to validate the disturbance rejection of the proposed algorithm, the

Table 6. Autopilot selection.

Main FMU professor	STM32F765, 32-bit Arm Cortex®-M7, 216MHz, 2MB memory, 512KB RAM
IO professor	STM32F100, 32-bit Arm Cortex®-M3, 24MHz, 8KB SRAM
Control frequency	1000Hz

Table 7. Attitude controller gains based on PID / Position controller gains based on PID.

	Channels	K_p	K_i	K_d
Attitude controller gains based on PID	Roll	0.08	0.03	0.02
	Yaw	0.07	0.01	0.00
	Pitch	0.12	0.02	0.01
Position controller gains based on PID	Roll	0.10	0.00	0.00
	Yaw	0.09	0.00	0.00
	Pitch	0.10	0.00	0.00

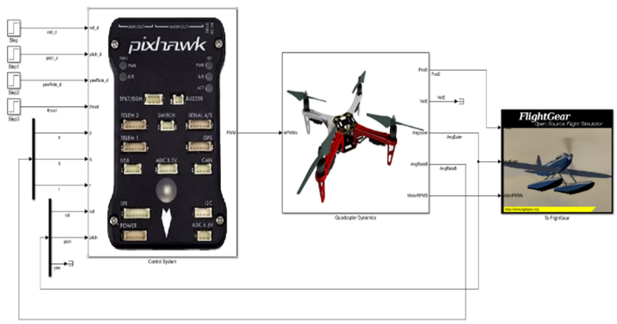


Fig. 6. Pixhawk autopilot simulation experiment.

Fig. 7. Ground station and flight test.

experiments are performed outdoors under the wind speed of about 5.5-7.9 m/s which is measured by an anemoscope.

7.1. Tracking trajectory comparison experiment

In the same ground station, the flight point and route are set. The following flight experiments are conducted based on two algorithms. The attitude tracking results are as follows:

As is shown in Figs. 8-12, Table 8, Table 9, under wind disturbance, the attitude error range and maximum overshoot in DSC+ADHDP algorithm are better than PID. The maximum attitude error in DSC+ADHDP is15.3°, while the maximum attitude error in PID is 17.2°. The maximum flight position error in DSC+ADHDP algorithm is 0.9 m, which is less than 1.4 m in PID. In summary, the system based on DSC+ADHDP controller has better control accuracy and disturbance rejection than Cascade PID.

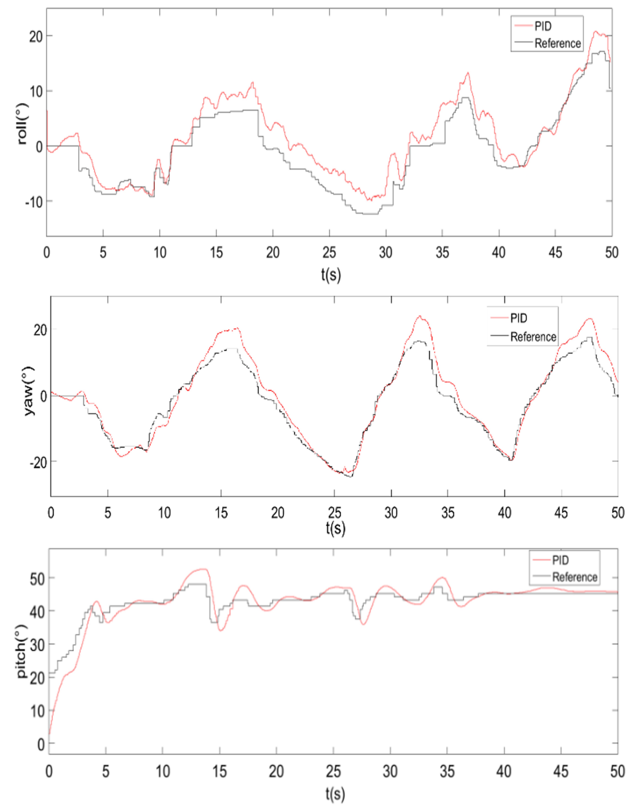


Fig. 8. Attitude tracking trajectory based on PID.

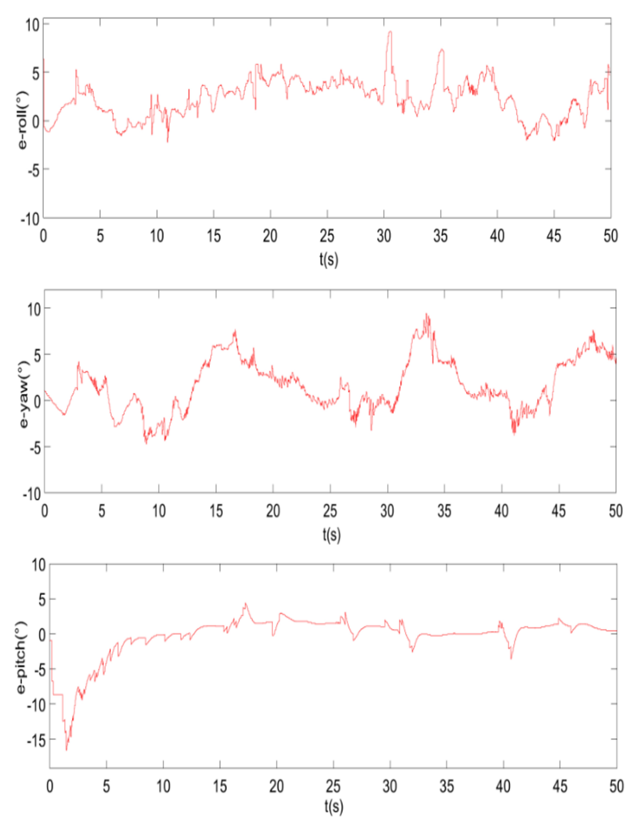


Fig. 9. Attitude tracking errors based on PID.

8. CONCLUSION

In this paper, a synthetical controller is designed to improve the tracking performance of the quadrotor. The designed synthetical controller provides a new research reference method for stable tracking of the quadrotor. In the future, the effect of parameter variations on the dynamic response for the proposed quadrotor control system and the trajectory tracking of multi-quadrotor are worthy of further studies.

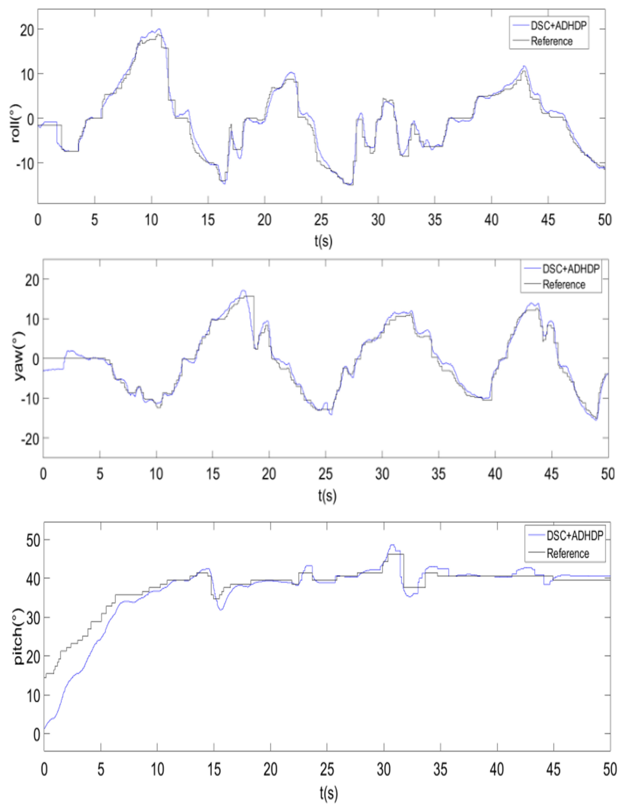


Fig. 10. Attitude tracking trajectory based on DSC+ ADHDP.

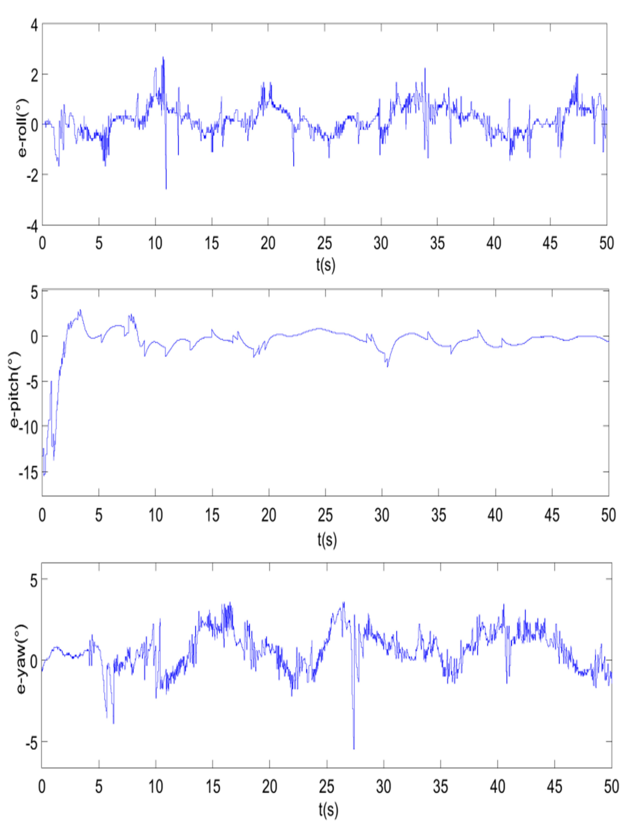


Fig. 11. Attitude tracking errors based on DSC+ADHDP.

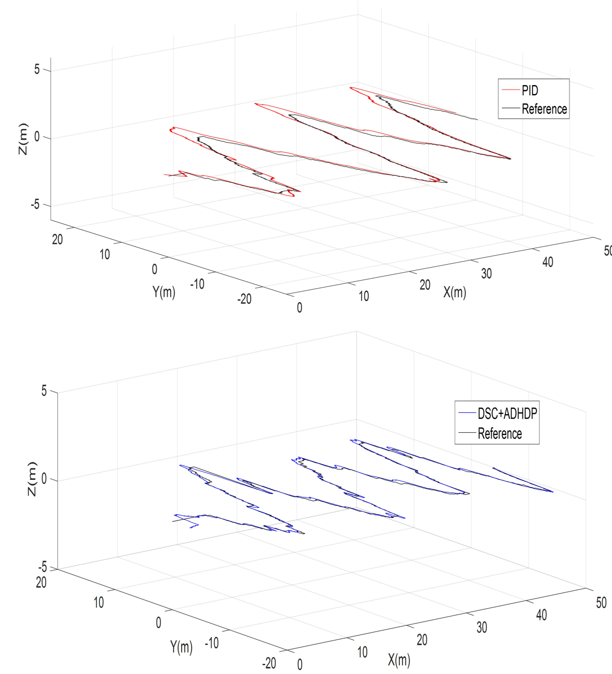


Fig. 12. Flight trajectory.

Table 8. Performance indexes of PID and DSC+ADHDP control.

Control channel	Control algorithm	Error range (°)	Maximum overshoot
Roll	Cascade PID	-4.3∼9.6	42.1%
	DSC+ADHDP	-3.6∼3.5	29.3%
Yaw	Cascade PID	-6.5∼8.9	39.8%
	DSC+ADHDP	-5.1∼4.6	30.2%
Pitch	Cascade PID	-17.2∼5.1	15.4%
	DSC+ADHDP	-15.3∼3.7	11.7%

Table 9. Performance indexes of PID and DSC+ADHDP control.

Coordinate axis	Control algorithm	Error range (°)	Maximum overshoot
X	Cascade PID	-1.1∼0.8	15.1%
	DSC+ADHDP	-0.7∼0.6	10.4%
Y	Cascade PID	-1.4∼0.9	13.8%
	DSC+ADHDP	-0.9∼0.6	9.23%
Z	Cascade PID	-1.2∼1.1	15.8%
	DSC+ADHDP	-0.8∼0.6	11.7%

REFERENCES

[1] M. Basri, "Design and application of an adaptive backstepping sliding mode controller for a six-DOF quadrotor aerial robot," *Robotica*, vol. 36, no. 11, pp. 1701-1727, 2018.

[2] R. H. Ramirez, V. V. Parra, O. A. Sanchez, and S. O. Gar-

cia, "Robust backstepping control based on integral sliding modes for tracking of quadrotors," *Journal of Intelligent & Robotic Systems*, vol. 73, pp. 51-66, 2014.

[3] O. Fritsch, D. Tromba, and B. Lohmann, "Cascaded energy based trajectory tracking control of a quadrotor," *AT-Automatisierungstechnik*, vol. 62, pp. 408-422, 2014.

[4] D. W. Kun and I. Hwang, "Linear matrix inequality-based nonlinear adaptive robust control of quadrotor," *Journal of Guidance Control and Dynamics*, vol. 39, no. 5, pp. 996-1008, 2016.

[5] Z. Gao and J. Fu, "Robust LPV modeling and control of aircraft flying through wind disturbance," *Chinese Journal of Aeronautics*, vol. 32, no. 7, pp. 1588-1602, 2019.

[6] M. J. Mahmoodabadi and N. R. Babak, "Robust fuzzy linear quadratic regulator control optimized by multi-objective high exploration particle swarm optimization for a 4 degree-of-freedom quadrotor," *Aerospace Science and Technology*, vol. 97, 105598, 2020.

[7] Y. Zhang, Z. Q. Chen, and M. W. Sun, "Trajectory tracking control of a quadrotor UAV based on sliding mode active disturbance rejection control," *Nonlinear Analysis-Modelling and Control*, vol. 24, no. 4, pp. 545-560, 2019.

[8] M. Reinoso, L. I. Minchala, and J. P. Ortiz, "Trajectory tracking of a quadrotor using sliding mode control," *IEEE Latin America Transactions*, vol. 14, no. 5, pp. 2157-2166, 2016.

[9] H. Boudjedir, O. Bouhali, and N. Rizoug, "Adaptive neural network control based on neural observer for quadrotor unmanned aerial vehicle," *Advanced Robotics*, vol. 28, no. 17, pp. 1151-1164, 2014.

[10] X. L. Shao, J. Liu, and H. L. Wang, "Robust back-stepping output feedback trajectory tracking for quadrotors via extended state observer and sigmoid tracking differentiator," *Mechanical Systems and Signal Processing*, vol. 104, pp. 631-647, 2018.

[11] Q. Zong, Y. H. Ji, F. Zeng, and H. Liu, "Output feedback back-stepping control for a generic Hypersonic Vehicle via small-gain theorem," *Aerospace Science and Technology*, vol. 23, no. 1, pp. 409-417, 2012.

[12] H. Yang, Y. Yu, Y. Yuan, and X. Fan, "Back-stepping control of two-link flexible manipulator based on an extended state observer," *Advances in Space Research*, vol. 56, no. 10, pp. 2312-2322, 2015.

[13] R. Zhang, Y. L. Ma, Z. X. Li, R. Malekian, and M. A. Sotelo, "Energy dissipation based longitudinal and lateral coupling control for intelligent vehicles," *IEEE Intelligent Transportation Systems Magazine*, vol. 10, no. 2, pp. 121-133, 2018.

[14] C. S. Tang and Z. Y. Duan, "Direct thrust-controlled PMSLM servo system based on back-stepping control," *IEEE Transactions on Electrical and Electronic Engineering*, vol. 13, pp. 785-790, 2018.

[15] F. Chen, W. Lei, K. Zhang, G. Tao, and B. Jiang, "A novel nonlinear resilient control for a quadrotor UAV via back-stepping control and nonlinear disturbance observer," *Nonlinear Dynamics*, vol. 85, no. 2, pp. 1281-1295, 2016.

[16] D. Swaroop, J. K. Hedrick, P. P. Yip, and J. C. Gerdes, "Dynamic surface control for a class of nonlinear systems," *IEEE Transactions on Automatic Control*, vol. 45, pp. 1893-1899, 2002.

[17] J. A. Farrell, M. Polycarpou, and M. Sharma, "Command filtered backstepping," *IEEE Transactions on Automatic Control*, vol. 54, no. 6, pp. 1391-1395, 2009.

[18] Q. Shen and P. Shi, "Distributed command filtered backstepping consensus tracking control of nonlinear multiple-agent systems in strict-feedback form," *Automatica*, vol. 53, pp. 120-124, 2015.

[19] Y. L. Tuo, Y. H. Wang, S. X. Yang, M. Biglarbegian, and M. Y. Fu, "Robust adaptive dynamic surface control based on structural reliability for a turret-moored floating production storage and offloading vessel," *International Journal of Control, Automation, and Systems*, vol. 1, pp. 1-12, 2018.

[20] A. Baldini, L. Ciabattoni, R. Felicetti, F. Ferracuti, A. Freddi, and A. Monteriu, "Dynamic surface fault tolerant control for underwater remotely operated vehicles," *ISA Transactions*, vol. 78, pp. 10-20, 2018.

[21] G. Q. Wu, S. M. Song, and J. G. Sun, "Adaptive dynamic surface control for spacecraft terminal safe approach with input saturation based on tracking differentiator," *International Journal of Control, Automation, and Systems*, vol. 16, pp. 1129-1141, 2018.

[22] X. L. Shao, J. Liu, H. L. Cao, C. Shen, and H. L. Wang, "Robust dynamic surface trajectory tracking control for a quadrotor UAV via extended state observer," *International Journal of Robust and Nonlinear Control*, vol. 28, pp. 2700-2719, 2018.

[23] L. Diao, D. Wang, Z. H. Peng, and L. Guo, "Predictor-based iterative neural dynamic surface control for three-phase voltage source PWM rectifier," *IEEE Transactions on Electrical and Electronic Engineering*, vol. 12, pp. 942-951, 2017.

[24] C. X. Mu, Z. Ni, and C. Y. Sun, "Air-breathing hypersonic vehicle tracking control based on adaptive dynamic programming," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 28, no. 3, pp. 584-598, 2017.

[25] D. V. Prokhorov, R. A. Santiago, and D. C. W. Ii, "Adaptive critic designs: A case study for neuro control," *Neural Networks*, vol. 8, no. 9, pp. 1367-1372, 1995.

[26] J. Si and Y. T. Wang, "Online learning control by association and reinforcement," *IEEE Transactions on Neural Networks*, vol. 12, no. 2, pp. 264-276, 2001.

[27] Z. Ni, H. He, X. Zhong, and D. V. Prokhorov, "Model-free dual heuristic dynamic programming," *IEEE Transactions on Neural Networks*, vol. 26, no. 8, pp. 1834-1839, 2017.

[28] D. Wang, D. R. Liu, Q. Wei, D. Zhao, and N. Jin, "Optimal control of unknown nonaffine nonlinear discrete-time systems based on adaptive dynamic programming," *Automatica*, vol. 48, no. 8, pp. 1825-1832, 2012.

[29] M. Fairbank, D. Prokhorov, and E. Alonso, "Clipping in neurocontrol by adaptive dynamic programming," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 25, no. 10, pp. 1909-1920, 2014.

[30] D. Liu, X. Yang, D. Wang, and Q, Wei, "Reinforcement-learning-based robust controller design for continuous-time uncertain nonlinear systems subject to input constraints," *IEEE Transactions on Cybernetics*, vol. 45, no. 7, pp. 1372-1385, 2015.

[31] D. Liu, D. Wang, F. Wang, H. Li, and X. Yang, "Neural-network-based online HJB solution for optimal robust guaranteed cost control of continuous-time uncertain nonlinear systems," *IEEE Transactions on Cybernetics*, vol. 44, no. 12, pp. 2834-2847, 2014.

[32] Z. Yang, S. Burger, and B. Lohmann, "Automation of the driving using the dynamic programming," *AT-Automatisierungstechnik*, vol. 57, pp. 377-385, 2009.

[33] T. Bruns and A. Trachtler, "Intersection management: Trajectory planning by means of dynamic programming," *AT-Automatisierungstechnik*, vol. 57, pp. 253-261, 2009.

[34] R.-Z. Song, W.-D. Xiao, and Q.-L. Wei, "A new approach of optimal control for a class of continuous-time chaotic systems by an online ADP algorithm," *Chinese Physics B*, vol. 23, no. 5, pp. 138-144, 2014.

[35] R. E. Bellman and S. E. Dreyfus, "Applied dynamic programming," *Journal of the American Statistical Association*. vol. 15, no. 59, pp. 366, 1962.

[36] A. Altamimi, F. L. Lewis, and M. Abukhalaf, "Discrete-time nonlinear HJB solution using approximate dynamic programming: Convergence proof," *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, vol. 38, no. 4, pp. 943-949, 2008.

[37] Q. L. Wei, F. Y. Wang, D. Liu, and X. Yang, "Finite-approximation-error-based discrete-time iterative adaptive dynamic programming," *IEEE Transactions on Cybernetics*, vol. 44, no. 12, pp. 2820-2833, 2014.

[38] X. Yang, D. Liu, D. Wang, and Q. Wei, "Discrete-time online learning control for a class of unknown nonaffine nonlinear systems using reinforcement learning," *Neural Networks*, vol. 55, no. 3, pp. 30-41, 2014.

[39] X. Yang, D. Liu, H. Ma, and Y. C. Xu, "Online approximate solution of HJI equation for unknown constrained-input nonlinear continuous-time systems," *Information Sciences*, vol. 328, pp. 435-454, 2016.

[40] R. Song, F. Lewis, Q. Wei, H. G. Zhang, Z. P. Jiang, and D. Lenvine, "Multiple actor-critic structures for continuous-time optimal control using input-output data," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 26, no. 4, pp. 851-865, 2015.

[41] W. Gao and Z. P. Jiang, "Nonlinear and adaptive suboptimal control of connected vehicles: A global adaptive dynamic programming approach," *Journal of Intelligent & Robotic Systems*, vol. 85, pp. 597-611, 2017.

[42] Y. Wu, Y. Wang, and X. Liu, "Multi-population based univariate marginal distribution algorithm for dynamic optimization problems," *Journal of Intelligent & Robotic Systems*, vol. 59, pp. 127-144, 2010.

[43] G. Xu, A. Song, and H. Li, "Adaptive impedance control for upper-limb rehabilitation robot using evolutionary dynamic recurrent fuzzy neural network," *Journal of Intelligent & Robotic Systems*, vol. 62, pp. 501-525, 2011.

[44] L. Yang, J. Si, S. Konstantinos, K. S. Tsakalis, and A. A. Rodriguez, "Performance evaluation of direct heuristic dynamic programming using control-theoretic measures," *Journal of Intelligent & Robotic Systems*, vol. 55, no. 3, pp. 177-201, 2009.

[45] H. B. He, Z. Ni, and J. Fu, "A three-network architecture for on-line learning and optimization based on adaptive dynamic programming," *Neurocomputing*, vol. 78, no. 1, pp. 3-13, 2012.

[46] Y. Sokolov, R. Kozma, L. D. Werbos, and P. J. Werbos, "Complete stability analysis of a heuristic approximate dynamic programming control design," *Automatica*, vol. 59, pp. 9-18, 2015.

[47] S. Mohagheghi, G. K. Venayagamoorthy, and R. G. Harley, "Adaptive critic design based neuro-fuzzy controller for a static compensator in a multimachine power system," *IEEE Transactions on Power Systems*, vol. 21, no. 4, pp. 1744-1754, 2006.

[48] J. Sun, F. Liu, J. Si, W. Guo, and S. Mei, "An improved approximate dynamic programming and its application in SVC control," *Journal of Electric Machines and Control*, vol. 15, no. 5, pp. 95-102, 2011.

[49] Z.-Q. Wu, W.-J. Jia, L.-R. Zhao, and C.-H. Wukey, "Maximum wind power tracking for PMSG chaos systems-ADHDP method," *Applied Soft Computing*, vol. 36, no. 2015, pp. 204-209, 2015.

[50] X. Lin, Y. Yu, and C. Y .Sun, "A decoupling control for quadrotor UAV using dynamic surface control and sliding mode disturbance observer," *Nonlinear Dynamics*, vol. 97, no. 1, pp. 781-795, July 2019.

[51] Z. W. Wang, X. Wang, H. Sun, and P. Xin, "Static output feedback predictive control for cyber-physical system under denial of service attacks," *Mathematical Problems in Engineering*, vol. 2020, Article ID 5412170, August 2020.

[52] B. Li, H. Y. Ban, W. Q. Gong, and B. Xiao, "Extended state observer-based finite-time dynamic surface control for trajectory tracking of a quadrotor unmanned aerial vehicle," *Transactions of the Institute of Measurement and Control*, vol. 42, no. 15, pp. 2956-2968, November 2020.

[53] Z. Ni, H. He, and J. Wen, "Adaptive learning in tracking control based on the dual critic network design," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 24, no. 6, pp. 913-928, 2013.

[54] P. He and S. Jagannathan, "Reinforcement learning neural-network-based controller for nonlinear discrete-time systems with input constraints," *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, vol. 37, no. 2, pp. 425-436, 2007.

[55] L. Yang, J. Si, K. S. Tsakalis, and A. A. Rodriguez, "Direct heuristic dynamic programming for nonlinear tracking control with filtered tracking error," *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*. vol. 39, no. 6, pp. 1617-1622, 2009.

[56] F. Liu, J. Sun, J. Si, W. Guo, and S. Mei, "A boundedness result for the direct heuristic dynamic programming," *Neural Networks*, vol. 32, pp. 229-235, 2012.

[57] P. G. He and S. Jagannathan, "Reinforcement learning-based output feed-back control of nonlinear systems with input constraints," *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, vol. 35, no. 1, pp. 150-154, 2015.

[58] S. Zeghlache, A. Djerioui, L. Benyettou, and T. Benslimane, "Fault tolerant control for modified quadrotor via adaptive type-2 fuzzy backstepping subject to actuator faults," *ISA Transactions*, vol. 95, pp. 330-345, 2019.

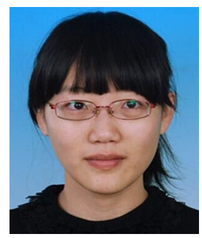
[59] Z. Jia, J. Yu, and X. Ai, "Integral backstepping control for quadrotor helicopters," *Proc. of International Conference on Computer & Automation Engineering*, vol. 68, pp. 299-307, 2017.

[60] R. Wang and J. Liu, "Trajectory tracking control of a 6-DoF quadrotor UAV with input saturation via backstepping," *Journal of the Franklin Institute*, vol. 355, no. 7, pp. 3288-3309, 2018.

Qiang Gao received his B.S. degree from the School of Electrical Engineering at Tianjin University of Technology, China, in 1996. He is currently working as a Professor in the School of Electrical and Electronic Engineering, Tianjin University of Technology, China. He has undertaken a number of vertical scientific research projects such as the National Natural Science Foundation of China and the Tianjin Young and Middle-aged Key Projects, and has carried out a number of major horizontal projects with CNOOC, Sinopec, and China National Automotive Engineering Corporation. His research interests include ADRC control, Nonlinear Control for Quadrotor.

Xin-Tong Wei received his B.S. degree in Tianjin University of Technology majoring in automation, Tianjin, in 2018, which he is pursuing a master's degree. His research interests include adaptive dynamic programming, UAV control and nonlinear control.

Da-Hua Li received his M.S. degree in mechanical manufacturing and automation from Guangxi University, Guangxi, China, in 2004. He had been an Associate Professor at the School of Electrical and Electronic Engineering, Tianjin University of Technology, Tianjin, China, since 2004. He is also the Director in Electrical and Electronic Center, Tianjin University of Technology, China. His current research interests include machine vision, 3D inspection system, and binocular ranging.



Yue-Hui Ji received her B.S. and Ph.D. degrees from the School of Electrical Engineering and Automation at Tianjin University, China, in 2009 and 2012, respectively. She is currently working as a lecturer in the School of Electrical and Electronic Engineering, Tianjin University of Technology, China. Her research interests include nonlinear adaptive control, decentralized control for interconnected systems.

Chao Jia received his B.E. and M.E. degrees from Tianjin University of Technology, Tianjin, China, in 2002 and 2008, respectively, and a Ph.D. degree from Tianjin University, Tianjin, China, in 2013. He joined the School of Electrical and Electronic Engineering, Tianjin University of Technology, China, in 2002, becoming an Associate Professor in 2013. He was a Visiting Scholar with Columbia University, NY, USA, from 2016 to 2017. His current research interests include nonlinear control, adaptive control, repetitive control, as well as their applications.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.