IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received February 5, 2020, accepted February 11, 2020, date of publication February 19, 2020, date of current version March 2, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.2974578

Dual-Loop Robust Attitude Control for an Aerodynamic System With Unknown Dynamic Model: Algorithm and Experimental Validation

ZHI-CHANG QIN[1], YING XIN[2,3], AND JIAN-QIAO SUN[4]

[1]School of Transportation and Vehicle Engineering, Shandong University of Technology, Zibo 255000, China
[2]School of Mechanical Engineering, Tianjin University of Technology, Tianjin 300384, China
[3]Tianjin Key Laboratory for Advanced Mechatronic System Design and Intelligent Control, Tianjin University of Technology, Tianjin 300384, China
[4]School of Engineering, University of California at Merced, Merced, CA 95343, USA

Corresponding authors: Zhi-Chang Qin (qinzhichang123@126.com) and Ying Xin (xinying913@163.com)

This work was supported in part by the National Natural Science Foundation of China under Grant 11572215, and in part by the Natural Science Foundation of Shandong Province under Grant ZR2018LA009.

ABSTRACT This paper proposes two kinds of dual-loop nonlinear robust control strategies that are implemented on an open-loop unstable two-degree-of-freedom (2-DOF) helicopter system with unmodeled dynamics and uncertainties. The inner feedback loop is considered as a nominal controller realized by an existing ''intelligent'' proportional differential controller (iPD) while the outer layer feedback control is regarded as a compensation loop. We study two different forms of outer loop in this paper. One is model-free sliding mode compensator (MFSMC) and another is model-free data-driven compensator (MFDDC). The combination of the shared inner loop and either of the outer loops forms two different kinds of model-free robust control strategies, i.e., iPD-MFSMC and iPD-MFDDC. Both robust control approaches are validated experimentally on the attitude tracking control of a 2-DOF laboratory helicopter, whose control objective is to have the helicopter attitudes, i.e., pitch and yaw motions, track specified trajectories. To demonstrate the utility of the two control approaches, we compare them with linear quadratic regulator (LQR), optimal feedback linearization control (OFLC) and iPD, respectively. The extensive comparison of the simulation and experimental results shows that the dual-loop robust control approaches are quite promising in controlling the systems with unknown dynamical models.

INDEX TERMS Dual-loop nonlinear robust control, sliding mode compensator, data-driven compensator, unmodeled dynamics and uncertainties, small helicopter.

I. INTRODUCTION

Flight control of unmanned aerial vehicles (UAVs) is a popular topic in recent years. Small helicopters can be considered as a special type of UAVs with multi-variable, inherently unstable and strongly coupled nonlinear dynamics. They have been frequently studied as experimental platforms for control designs. In this paper, we conduct two different model-free robust control experiments for attitude control of a laboratory helicopter and compare the proposed controls with model-based control methods.

At present, we can retrieve a large number of papers on helicopter control from various databases. From the perspective of whether the controller design depends on the dynamic model of the system, the control design methods can be roughly divided into two categories: model-based and model-free. Generally, the model-based control method requires the designer to establish a mathematical model that can reflect the laws of helicopter motion, often in the form of differential equations or difference equations. There are many control methods that fall into this category, such as LQR [1]–[4], feedback linearization control [5]–[7], backstepping control [8]–[10], sliding mode control [11]–[13], model predictive control [14]–[16] and so on. When the helicopter works near the hovering point, for the sake of simplicity, a linear model can be built for controller design. The LQR design is an example. To enhance the flight performance of the helicopter in a complex environment or for maneuvering,

The associate editor coordinating the review of this manuscript and approving it for publication was Sun Junwei.

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see http://creativecommons.org/licenses/by/4.0/

a more accurate nonlinear system model for controller design is needed. For a specific control method, the control performance is positively correlated with the accuracy of the model. However, the construction of an accurate helicopter dynamics model is quite expensive, and there is the possibility that the designed controller may not return the optimal set of parameters in model-based design. For these reasons, the model-free control design method has been favored by scholars.

In recent years, researchers are increasingly interested in controller design and optimization based on experimental data. As a result, there are less requirements for the system model, or the model is not even needed at all. Numerous model-free controller designing and tuning techniques have been proposed based on the classical control theory, such as simultaneous perturbations stochastic approximation (SPSA) [17], iterative feedback tuning (IFT) [18], virtual reference feedback tuning (VRFT) [19], model-free adaptive control (MFAC) [20], model-free control (MFC) [21], model-free iterative learning control (MFILC) [22], *etc*. These techniques are general data-driven or data-based control methods with extensive theoretical research and applications in different fields [23]–[26].

Different from other model-free control approaches, the MFC works with an ultra-local model continuously updated according to input-output behavior, which can deal with the unmodeled dynamics, system uncertainties, and external disturbances. It usually works together with a P/PD/PI/PID controller, by which the combination is known as i-P/PD/PI/PID controller [21]. Without knowing the exact dynamic model of the system, MFC only needs general knowledge about system dynamics to determine and tune controller parameters. Besides, it has large advantages in dealing with the unmodeled system dynamics, uncertainties, and external disturbances that could have a crucial effect on system stability. It should be emphasized that this control method has already been most successfully applied in many practical case-studies [27]–[30]. The key to the success of the MFC is to continuously update the unknown ultra-local system dynamics with input-output data, while generating control signals. However, many objective factors, such as derivative estimation error, control signal lag, environmental noise, *etc*., always introduce a certain difference between the estimated value and the real value of the system dynamics. They all affect the control performance of the system and may even lead to the instability of the controller. In this paper, we take a PD controller based on MFC as a nominal controller (iPD) to stabilize the 2-DOF helicopter system firstly. Then, two different compensation strategies are introduced to the iPD-controlled system to improve the attitude control performance of the helicopter.

The first compensator derives from a nonlinear sliding mode control (SMC) design. SMC is an easily understandable nonlinear control technique with the advantage of robustness against system uncertainties, parameter variations, and external disturbances [31]–[34]. To decrease the influence of derivative estimation error, control signal lag, and environmental noise to MFC, the sliding mode controller is used to design compensators that are embedded into the MFC and enhance its control performance [35]. Wang *et al.* [36] design a model-free based sliding mode controller (MFSMC) by combining the MFC and a sliding mode controller together and apply it to the attitude control of a quadrotor. In the work of [37], two model-free sliding mode control system structures are designed and validated by a set of real-time experimental results on a nonlinear laboratory twin-rotor aerodynamic system. In this paper, the model-free sliding mode compensation technique is investigated and used to improve the performance of the nominal controller. The combination between the iPD and MFSMC forms the first dual-loop control structure (i.e. iPD-MFSMC) we studied in this paper.

The second model-free compensator in this paper is a data-driven offline learning technique based on integral reinforcement learning. Reinforcement learning (RL) is known in the control community as adaptive (or approximate) dynamic programming (ADP). RL can be divided into model-based RL and model-free RL according to how much knowledge of the system dynamics is known. In the scope of model-based RL, the online synchronous policy iteration (PI) algorithm [38] is used to update the actor and critic neural networks (NN) [39] based on the full knowledge of system dynamics. The integral RL (IRL) technique [40], [41] is considered as a data-based RL only using partial knowledge of system dynamics. With the combination of IRL and off-policy scheme, the offline iterative learning is used to control the partially unknown system [42]. In the scope of model-free RL, Zhang *et al.* [43] make a significant contribution in the optimal robust tracking control, which provides a solid foundation for using RL in the field of optimal tracking control of unknown general nonlinear systems. The method has two main advantages: 1) only the input-output data are required instead of an exact system model, 2) the tracking error converges to zero asymptotically in an optimal way. Based on the IRL technique and off-policy scheme [44], the data-driven RL has been used in uncertain systems [45], zero-sum games [46], [47], nonzero-sum games [48], [49], H_∞ control [50], etc. With online measurement and off-policy learning, Zhang *et al.* [48] solved the continuous-time unknown nonzero-sum game with partially constrained inputs by a model-free ADP algorithm. However, this method makes sense only if the right-hand side of system $\dot{x} = f(x)+g_1(x)u_1+g_2(x)u_2$ is Lipschitz continuous on a compact set $\Omega \in \Re^n$ containing the origin and the system is stabilizable on Ω. Unfortunately, few real-time applications or actual studies apply this control method to open-loop unstable systems. In this paper, we extend this method to open-loop unstable dynamic systems by introducing a dual-loop feedback control strategy. The model-free iPD is considered as the inner-loop to stabilize the initial system firstly, then a model-free data-driven compensator (MFDDC) is embedded into the iPD controlled system. The mixture of iPD and MFDDC fully exploits their respective advantages, broadens

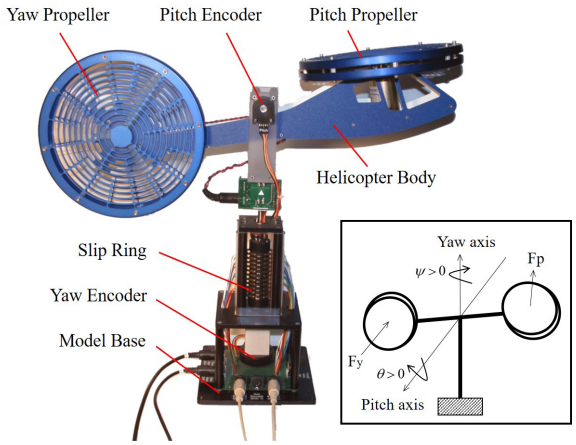

FIGURE 1. The twin-rotor helicopter model made by Quanser company.

their application scope, and improves their control effects. This mixed control approach (i.e., iPD-MFDDC) is the second dual-loop control structure we studied in this paper.

Both dual-loop feedback control strategies are validated on the attitude tracking control of an initially unstable aerodynamic system. Furthermore, this paper offers a thorough discussion of experimental results by a cross-comparison of different control approaches. The contributions of this study are as follows,

(1) Proposed two dual-loop model-free control strategies and apply them to the attitude tracking control of an unstable aerodynamic system;

(2) Carried out a thorough experimental comparison between the dual-loop control approach and other model-based approaches. It is shown that the model-free controller with/without the proposed compensator has better performance than the model-based controller. By adding a compensator to iPD, the tracking control performance can be improved largely. Besides, iPD-MFDDC has superior compensation performance to iPD-MFSMC.

This paper is organized as follows. Section II introduces the helicopter setup and reviews a model-free control strategy, i.e., iPD. The MFSMC and MFDDC are designed in Section III and IV respectively. The simulation and experimental studies are presented in Section V. Finally, Section VI concludes this paper.

II. THE HELICOPTER SYSTEM AND MODEL-FREE CONTROL DESIGN

The Quanser 2-DOF laboratory helicopter shown in Figure 1 consists of a helicopter model mounted on a fixed base with two propellers that are driven by DC motors. The front propeller controls the elevation of the helicopter pitch angle, and the back propeller controls the side to side motion of the helicopter yaw angle. The pitch θ and yaw ψ angles are measured through two high-resolution encoders. The pitch encoder and motor signals are transmitted via a slip ring, which eliminates the possibility of wires tangling and allows the yaw angle to rotate freely 360 degrees.

A. MODELING OF THE HELICOPTER

To establish an accurate model, many factors, such as the thrust force produced by the rotation of the propeller, counter torque produced by the rotation, gyroscopic effect, etc., are usually taken into consideration. However, for convenience, the helicopter body and the propellers are usually assumed to be rigid. The complex rotor aerodynamics and their interaction with the helicopter fuselage are generally simplified. The mathematical model of the 2-DOF helicopter is obtained based on the following conventions [51],

1) The helicopter is horizontal when the pitch angle $\theta = 0$;
2) The pitch angle increases positively $\theta(t) > 0$ when the nose moves upwards and the body moves in counter-clockwise direction;
3) The yaw angle increases positively $\psi(t) > 0$ when the body rotates in counter-clockwise direction;
4) Pitch increases $\theta > 0$ when the pitch thrust force is positive $F_p > 0$;
5) Yaw increases $\psi > 0$ when the yaw thrust force is positive $F_y > 0$.

With the Euler-Lagrange method, the equations of the pitch and yaw motions with the servo motor voltages as inputs can be described as follows [51],

$$\begin{aligned}(J_p + m_h l_{cm}^2)\ddot{\theta} + m_h l_{cm}^2 \dot{\psi}^2 \sin\theta\cos\theta + B_p\dot{\theta} \\ + m_h g l_{cm}\cos\theta = k_{pp}V_\theta + k_{py}V_\psi, \\ (J_y + m_h l_{cm}^2\cos^2\theta)\ddot{\psi} - 2m_h l_{cm}^2\dot{\theta}\dot{\psi}\sin\theta\cos\theta \\ + B_y\dot{\psi} = k_{yp}V_\theta + k_{yy}V_\psi. \end{aligned} \tag{1}$$

where V_θ and V_ψ are the voltage inputs to the motors of the propellers acting on the pitch and yaw respectively.

To consider the extra unmodeled dynamics and system uncertainties, we can write the system model into a state-space form equation as follows,

$$\begin{bmatrix}\dot{x}_1\\ \dot{x}_2\\ \dot{x}_3\\ \dot{x}_4\end{bmatrix} = \begin{bmatrix}x_3\\ x_4\\ f_1(\mathbf{x})\\ f_2(\mathbf{x})\end{bmatrix} + \begin{bmatrix}0 & 0\\ 0 & 0\\ g_{11} & g_{12}\\ g_{21} & g_{22}\end{bmatrix}\begin{bmatrix}u_1\\ u_2\end{bmatrix}, \tag{2}$$

where $\mathbf{x} = [x_1, x_2, x_3, x_4]^{\mathrm{T}}$ is state variable, $x_1 = \theta$, $x_2 = \psi$, $x_3 = \dot{\theta}$ and $x_4 = \dot{\psi}$ represent the pitch angle, yaw angle, pitch angle velocity, and yaw angle velocity respectively. $f_i(\mathbf{x})$, g_{ij} and u_i $(i = 1, 2; j = 1, 2)$ represent the system drift dynamics, input factor and input voltage respectively. They have the following expressions,

$$\begin{aligned} f_1(\mathbf{x}) &= -\frac{1}{\Lambda_1}(m_h l_{cm}^2 x_4^2 \sin x_1\cos x_1 + B_p x_3 \\ &\quad + m_h g l_{cm}\cos x_1) + \Delta f_1, \\ f_2(\mathbf{x}) &= \frac{1}{\Lambda_2}\left(2m_h l_{cm}^2 x_3 x_4 \sin x_1\cos x_1 - B_y x_4\right) + \Delta f_2, \\ g_{11} &= \frac{k_{pp}}{\Lambda_1} + \delta_{11},\ g_{21} = \frac{k_{yp}}{\Lambda_2} + \delta_{21}, \\ g_{12} &= \frac{k_{py}}{\Lambda_1} + \delta_{12},\ g_{22} = \frac{k_{yy}}{\Lambda_2} + \delta_{22}, \\ \Lambda_1 &= J_p + m_h l_{cm}^2,\ \Lambda_2 = J_y + m_h l_{cm}^2\cos^2 x_1, \\ u_1 &= V_\theta,\ u_2 = V_\psi. \end{aligned}$$

where Δf_i $(i = 1, 2)$ represents the unmodeled dynamics and uncertainties in the pitch and yaw channel respectively. The unmodeled dynamics and uncertainties may include the approximation of propeller viscous damping forces and aerodynamic forces, measurement error of system parameters, external disturbances and other factors not mentioned here. δ_{ij} $(i = 1, 2; = 1, 2)$ denote the uncertainties of experimental measurements of the motor voltage-torque constant and propeller torque-thrust constant. However, it is difficult or almost impossible to determine the exact form of expression for Δf_i and δ_{ij} in real applications.

The helicopter system can be simplified into a two-order nonlinear model by considering the system outputs $y_1 = x_1$ and $y_2 = x_2$, and defining two user-defined non-physical constants α_1 and α_2,

$$\begin{bmatrix} \ddot{y}_1 \\ \ddot{y}_2 \end{bmatrix} = \begin{bmatrix} \tilde{f}_1(\cdot) \\ \tilde{f}_2(\cdot) \end{bmatrix} + \begin{bmatrix} \alpha_1 u_1 \\ \alpha_2 u_2 \end{bmatrix}, \tag{3}$$

where α_i $(i = 1, 2)$ are non-zero user-designed parameters, $\tilde{f}_1(\cdot) = f_1(\cdot) + (g_{11} - \alpha_1)u_1 + g_{12}u_2$ and $\tilde{f}_2(\cdot) = f_2(\cdot) + g_{21}u_1 + (g_{22} - \alpha_2)u_2$ are two nonlinear functions that include unknown system dynamics and undesigning control inputs.

Without an accurate dynamic model of the helicopter system, it will be quite difficult to implement the control task according to the traditional model-based optimal control methods. However, building an accurate mathematical description of unmodeled dynamics, system uncertainties, and external disturbances is usually difficult and expensive, even impossible. The conflict between the need to build an accurate mathematical model and its high cost drives our research on model-free control in this paper. We take the initiative to bypass the step of establishing a precise mathematical model of the helicopter and adopt a model-free control method to complete its attitude control task.

B. A MODEL-FREE CONTROL APPROACH

By equivalent conversion, the system (3) can be approximated for a short time window into an ultra-local model,

$$\ddot{y}_i = h_i + \alpha_i u_i \tag{4}$$

where h_i $(i = 1, 2)$ is a continuously updated value that captures all the unknown nonlinearity and uncertainties in the input-output behavior of the system. Since the above equation is valid for a short time window, it must be updated at each sampling time, Δt. At time interval $[k\Delta t, (k+1)\Delta t]$, the value of h_i is updated from the measurement of $\alpha_i u_i$ and $\ddot{y}_i$ in the following manner,

$$\hat{h}_i(k) = [\ddot{y}_i(k)]_{est} - \alpha_i u_i(k-1), \quad i = 1, 2 \tag{5}$$

where $\hat{h}_i(k)$ is the estimated value of h_i at the time point $k\Delta t$. It will be used for the computation of the control input $u_i(k)$ later. The notation $[\ddot{y}_i(k)]_{est}$ is the estimated value of the second-order derivative of the output y_i at the time point $k\Delta t$. In this study, the first and second derivative are estimated by low-pass filter (LPF) to attenuate the noisy signals with the following transfer functions,

$$L_1(s) = \frac{\omega_{cf}^2 s}{s^2 + 2\zeta_f \omega_{cf} s + \omega_{cf}^2}, \tag{6}$$

$$L_2(s) = \frac{\omega_{cf}^2 s^2}{s^2 + 2\zeta_f \omega_{cf} s + \omega_{cf}^2}, \tag{7}$$

where $\omega_{cf} = 20\pi$ and $\zeta_f = 0.85$ are the cutoff frequency and the damping ratio of the low-pass filter respectively. Besides, to get the second derivative of a time sequence of the measured output, we can also take the first derivative twice. The notation $u_i(k-1)$ is the control input of the previous sampling time point. Generally, the model-free control input can be written as

$$u_{mfc,i} = -\frac{\hat{h}_i - \ddot{y}_{di} + u_{ci}}{\alpha_i}, \tag{8}$$

where $\ddot{y}_{di}(i = 1, 2)$ is the second-order derivative of the desired output and $u_{ci}(i = 1, 2)$ is a feedback controller used to stabilize the ultra-local system.

Substituting Equation (8) into (4), and assuming that h_i can be well approximated by functions $\hat{h}_i$ in Equation (5), we have

$$\ddot{e}_i + u_{ci} = 0, \tag{9}$$

where $e_i = y_i - y_{di}$ $(i = 1, 2)$ is the output error. With an iPD control strategy, $u_{ci} = k_{pi}e_i + k_{di}\dot{e}_i$, $i = 1, 2$, the model-free controller (8) reads

$$u_{mfc,i} = -\frac{\hat{h}_i - \ddot{y}_{di} + k_{pi}e_i + k_{di}\dot{e}_i}{\alpha_i}, i = 1, 2. \tag{10}$$

III. A MODEL-FREE SLIDING MODE COMPENSATOR

The key step for the success of the model-free control is lying in updating the unknown system dynamics in real-time through Equation (5). However, many factors, such as derivative estimation error, control signal lag, environmental noise, etc., always introduce a certain deviation between the estimated value and the real value of the system dynamics. In this section, a sliding mode compensator is developed to compensate for the estimation errors of iPD.

A. SLIDING MODE COMPENSATION APPROACH

An augmented sliding mode compensator, u_{smc}, will be designed and added to the model-free controller (10),

$$u_i = -\frac{\hat{h}_i - \ddot{y}_{di} + k_{pi}e_i + k_{di}\dot{e}_i}{\alpha_i} + u_{smc,i}, \quad i = 1, 2 \tag{11}$$

Substituting Equation (5) into (11), the closed-loop control system can be described by the state-space equations through introducing new state variables, $z_1 = e_1$, $z_2 = e_2$, $z_3 = \dot{e}_1$ and $z_4 = \dot{e}_2$,

$$\begin{cases} \dot{z}_1 = z_3, \\ \dot{z}_2 = z_4, \\ \dot{z}_3 = -k_{d1}z_3 - k_{p1}z_1 + \alpha_1 u_{smc,1} + \Delta\varrho_1, \\ \dot{z}_4 = -k_{d2}z_4 - k_{p2}z_2 + \alpha_2 u_{smc,2} + \Delta\varrho_2, \end{cases} \tag{12}$$

where $\Delta\varrho_i = \ddot{y}_i - [\ddot{y}_i]_{est} = h_i - \hat{h}_i, (i = 1, 2)$ represents the unknown estimation errors. Because the state equations in (12) are decoupled, two sliding model compensators can be designed separately.

Define a pair of sliding surfaces for pitch motion and yaw motion respectively,

$$s_1 = z_1 + \lambda_1 z_3, \; s_2 = z_2 + \lambda_2 z_4. \tag{13}$$

For the pitch motion as an example, we consider a Lyapunov function

$$V_1 = \frac{1}{2} s_1^2, \tag{14}$$

According to the Lyapunov stability theory $\dot{V}_1 < 0$, the design of pitch motion sliding mode compensator needs to satisfy the following reaching and existence condition

$$s_1 \dot{s}_1 < 0. \tag{15}$$

On the sliding surface, we impose $s_1 = 0$ and $\dot{s}_1 = 0$,

$$\begin{aligned} \dot{s}_1 &= \dot{z}_1 + \lambda_1 \dot{z}_3 \\ &= z_3 + \lambda_1 \left(-k_{d1} z_3 - k_{p1} z_1 + \alpha_1 u_{smc,1} + \Delta\varrho_1 \right). \end{aligned} \tag{16}$$

An equivalent control law can be solved,

$$u_{smc,eq1} = \frac{1}{\lambda_1 \alpha_1} [\lambda_1 k_{p1} z_1 - (1 - \lambda_1 k_{d1}) z_3 - \lambda_1 \Delta\varrho_1]. \tag{17}$$

Since the estimation error $\Delta\varrho_1$ is unknown and can't be measured accurately, the equivalent controller can't be used to control the system directly. In an actual implementation, $\Delta\varrho_1$ is usually replaced by a hypothetical value function $\Delta\hat{\varrho}_1$, and then an additional switching term $u_{smc,sw1} = -\frac{K_1}{\lambda_1 \alpha_1} \text{sign}(s_1)$ with $K_1 > 0$ a user-defined parameter is added to ensure stable running of the system along the sliding surface. Hence the final sliding mode compensation law reads

$$u_{smc,1} = \hat{u}_{smc,eq1} + u_{smc,sw1}, \tag{18}$$

where $\hat{u}_{smc,eq1} = \frac{1}{\lambda_1 \alpha_1} [\lambda_1 k_{p1} z_1 - (1 - \lambda_1 k_{d1}) z_3 - \lambda_1 \Delta\hat{\varrho}_1]$ and the estimated error and the hypothetical value function satisfies that $\left| \Delta\varrho_1 - \Delta\hat{\varrho}_1 \right| \leq E_1$. Here, $E_1 > 0$ is a known constant. Normally, the discontinuous switching law in Equation (18) may cause chattering phenomenon. To tackle this issue, we follow the traditional practice by introducing a saturation function to replace the sign function in the switching law,

$$\begin{aligned} u_{smc,sw1} &= -\frac{K_1}{\lambda_1 \alpha_1} \text{sat}\left(\frac{s_1}{\phi_1} \right) \\ &= -\frac{K_1}{\lambda_1 \alpha_1} \begin{cases} 1, & s_1 > \phi_1 \\ \dfrac{s_1}{\phi_1}, & -\phi_1 \leq s_1 \leq \phi_1 \\ -1, & s_1 < -\phi_1 \end{cases} \end{aligned} \tag{19}$$

where $0 < \phi_1 < 1$ is the boundary layer thickness of the saturation function. The block diagram of the proposed control structure for the model-free control with the sliding model compensator is shown in Figure 2.

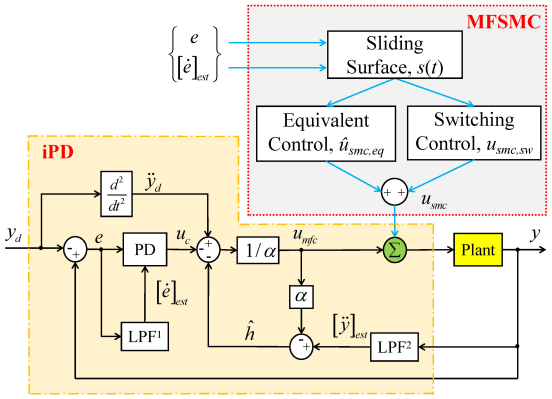

FIGURE 2. The block diagram of the iPD-MFSMC approach. The iPD acts as the inner-loop controller that is used to stabilize the initial system while the MFSMC serves as the outer-loop compensator, which is used to compensate the inner-loop controlled system and finish the attitude tracking task.

B. STABILITY PROOF

The stability of the model-free controller with the proposed sliding mode compensator is discussed by taking different values of the sliding mode parameter, s_1 : $|s_1| \leq \phi_1$ and $|s_1| > \phi_1$.

Case 1: $|s_1| \leq \phi_1$

Substituting Equations (18) and (19) into Equation (16), the existence and reaching condition (15) becomes

$$\begin{aligned} s_1 \dot{s}_1 &= s_1 \left[\lambda_1 \left(\Delta\varrho - \Delta\hat{\varrho}_1 \right) - K_1 \text{sat}\left(\frac{s_1}{\phi_1} \right) \right] \\ &= \lambda_1 \left(\Delta\varrho - \Delta\hat{\varrho}_1 \right) s_1 - \frac{K_1}{\phi_1} s_1^2 \\ &\leq \lambda_1 E_1 |s_1| - \frac{K_1}{\phi_1} s_1^2 \end{aligned} \tag{20}$$

If $K_1 > \frac{\lambda_1 E_1 \phi_1}{|s_1|}$, the reaching and existence condition (15) is guaranteed.

Case 2: $|s_1| > \phi_1$

Substituting Equations (18) and (19) into Equation (16), the existence and reaching condition (15) becomes

$$\begin{aligned} s_1 \dot{s}_1 &= s_1 \left[\lambda_1 \left(\Delta\varrho - \Delta\hat{\varrho}_1 \right) - K_1 \text{sat}\left(\frac{s_1}{\phi_1} \right) \right] \\ &= \lambda_1 \left(\Delta\varrho - \Delta\hat{\varrho}_1 \right) s_1 - K_1 s_1 \text{sign}\left(s_1 \right) \\ &= \lambda_1 \left(\Delta\varrho - \Delta\hat{\varrho}_1 \right) s_1 - K_1 |s_1| \\ &\leq \left(\lambda_1 E_1 - K_1 \right) |s_1| \end{aligned} \tag{21}$$

If $K_1 > \lambda_1 E_1$, the reaching and existence condition (15) is guaranteed yet.

When the value of parameter K_1 is selected to $\max \left\{ \frac{\lambda_1 E_1 \phi_1}{|s_1|}, \lambda_1 E_1 \right\} + \eta_1$ with $\eta_1 > 0$ a positive constant, the Lyapunov stable condition is satisfied, i.e., $\dot{V}_1 = s_1 \dot{s}_1 \leq -\eta_1 |s_1|$. The expression of the sliding mode compensation law added to the model-free control (10) for the pitch motion is

$$\begin{aligned} u_{smc,1} = \frac{1}{\lambda_1 \alpha_1} [&\lambda_1 k_{p1} z_1 - (1 - \lambda_1 k_{d1}) z_3 \\ &-\lambda_1 \Delta\hat{\varrho}_1 - K_1 \text{sat}(s_1/\phi_1), \end{aligned} \tag{22}$$

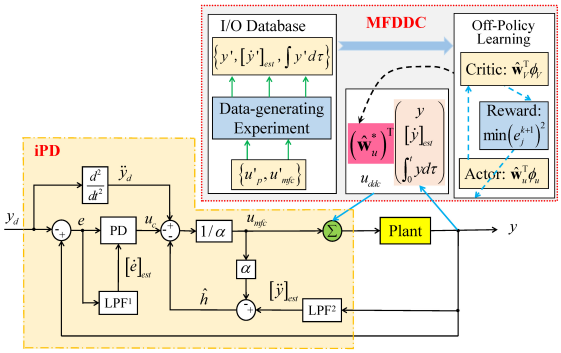

FIGURE 3. The block diagram of the iPD-MFDDC approach. The iPD controller and MFDDC represent the inner-loop and outer-loop respectively.

Similarly, the sliding mode compensation law added to the yaw motion can be written as

$$u_{smc,2} = \frac{1}{\lambda_2\alpha_2}[\lambda_2 k_{p2} z_2 - (1-\lambda_2 k_{d2})\, z_4 - \lambda_2 \Delta\hat{\varrho}_2 - K_2 \mathrm{sat}(s_2/\phi_2)], \quad (23)$$

$$K_2 = \max\left\{\frac{\lambda_2 E_2 \phi_2}{|s_2|}, \lambda_2 E_2\right\} + \eta_2,\ \eta_2 > 0, \quad (24)$$

where $E_2 > 0$ is a known constant. It is noting that the gains K_1 and K_2 are relating to the sliding surface parameters s_1 and s_2 respectively. In practice, we need to calculate the value of each sliding surface at each sampling time t. Refer to section V-B for the constant parameters used in this section.

IV. A DATA-DRIVEN COMPENSATOR

In this section, a model-free data-driven compensator (MFDDC) that serves as the outer-loop control strategy is introduced to iPD control. MFDDC is implemented by an actor-critic neural network (NN), which learns the optimal value function and optimal compensation policy simultaneously. The dual-loop feedback control structure (iPD-MFDDC) is illustrated in Figure 3.

Recalling the helicopter model with unmodeled dynamics and system uncertainty described in Equation (2), substituting the model-free control law (10) into it and considering the data-driven compensator $\mathbf{u}_{ddc}$, we can write the controlled system into the following form,

$$\dot{\mathbf{x}} = \mathbf{f}(\mathbf{x}) + \mathbf{g}(\mathbf{x})\mathbf{u}_{ddc}, \quad (25)$$

where $\mathbf{x} = [x_1, x_2, x_3, x_4]^T$, $\mathbf{f}(\mathbf{x}) = [x_3, x_4, f_1(\mathbf{x}) + g_{11}u_{mfc,1} + g_{12}u_{mfc,2}, f_2(\mathbf{x}) + g_{21}u_{mfc,1} + g_{22}u_{mfc,2}]^{\mathrm{T}}$, $\mathbf{g}(\mathbf{x}) = \begin{bmatrix} 0, & 0, & g_{21}, & g_{22} \\ 0, & 0, & g_{11}, & g_{12} \end{bmatrix}^{\mathrm{T}}$ and $\mathbf{u}_{ddc} = [u_{ddc,1}, u_{ddc,2}]^{\mathrm{T}}$ represents the data-driven compensation law.

Define a continuous value function

$$V(\mathbf{x}) = \int_t^{\infty} \mathcal{L}(\mathbf{x}, \mathbf{u}_{ddc}) d\tau. \quad (26)$$

where $\mathcal{L}(\mathbf{x}, \mathbf{u}) = \frac{1}{2}(\mathbf{x}^{\mathrm{T}}\mathbf{Q}\mathbf{x} + \mathbf{u}_{ddc}^{\mathrm{T}}\mathbf{R}\mathbf{u}_{ddc})$ is the Lagrange function with $\mathbf{Q} \geq 0$ a semi-positive symmetric matrix and $\mathbf{R} > 0$ a positive symmetric matrix. The Hamilton-Jacobi-Bellman (HJB) equation for the system (25) is

$$\mathcal{L}(\mathbf{x}, \mathbf{u}_{ddc}) + \frac{\partial V(\mathbf{x})}{\partial \mathbf{x}}^{\mathrm{T}}[\mathbf{f}(\mathbf{x}) + \mathbf{g}(\mathbf{x})\mathbf{u}_{ddc}] = 0. \quad (27)$$

The optimal control can be solved from the HJB equation,

$$\mathbf{u}^*_{ddc}(t) = -\mathbf{R}^{-1}\mathbf{g}^{\mathrm{T}}(\mathbf{x}^*)\frac{\partial V(\mathbf{x}^*)}{\partial \mathbf{x}}. \quad (28)$$

where $\mathbf{x}^*$ is the optimal state at time t. Let $\mathbf{u}^k_{ddc}$ and $V^k(\mathbf{x})$ denote the control input and value function at the k^{th} iteration and let $\mathbf{u}'_{ddc}$ denote an admissible control at the $(k+1)^{th}$ iteration step, we have

$$\begin{cases} \dfrac{\partial V^{k+1}(\mathbf{x})}{\partial \mathbf{x}}^{\mathrm{T}}[\mathbf{f}(\mathbf{x}) + \mathbf{g}(\mathbf{x})\mathbf{u}^k_{ddc}] + \mathcal{L}(\mathbf{x}, \mathbf{u}^k_{ddc}) = 0, \\ \dot{V}^{k+1}(\mathbf{x}) = \dfrac{\partial V^{k+1}(\mathbf{x})}{\partial \mathbf{x}}^{\mathrm{T}}[\mathbf{f}(\mathbf{x}) + \mathbf{g}(\mathbf{x})\mathbf{u}'_{ddc}], \end{cases} \quad (29)$$

which implies that

$$\dot{V}^{k+1}(\mathbf{x}) = -\mathcal{L}(\mathbf{x}, \mathbf{u}^k_{ddc}) + \frac{\partial V^{k+1}(\mathbf{x})}{\partial \mathbf{x}}^{\mathrm{T}}\mathbf{g}(\mathbf{x})(\mathbf{u}'_{ddc} - \mathbf{u}^k_{ddc}). \quad (30)$$

Besides, Equation (28) implies that $\frac{\partial V^{k+1}(\mathbf{x})}{\partial \mathbf{x}}^{\mathrm{T}}\mathbf{g}(\mathbf{x}) = -(\mathbf{u}^{k+1}_{ddc})^{\mathrm{T}}\mathbf{R}$. So the Equation (30) can be rewritten as

$$\dot{V}^{k+1}(\mathbf{x}) = -\mathcal{L}(\mathbf{x}, \mathbf{u}^k_{ddc}) - (\mathbf{u}^{k+1}_{ddc})^{\mathrm{T}}\mathbf{R}[\mathbf{u}'_{ddc} - \mathbf{u}^k_{ddc}]. \quad (31)$$

According to the integral reinforcement learning [48], [52], integrating both sides of (31) from t to $t+\Delta t$, the following equation is true

$$V^{k+1}(\mathbf{x}(t)) - V^{k+1}(\mathbf{x}(t+\Delta t)) - \int_t^{t+\Delta t}(\mathbf{u}^{k+1}_{ddc}(\tau))^{\mathrm{T}}\mathbf{R}(\mathbf{u}'_{ddc}(\tau) - \mathbf{u}^k_{ddc}(\tau))d\tau = \int_t^{t+\Delta t}\mathcal{L}(\mathbf{x}(\tau), \mathbf{u}^k_{ddc}(\tau))d\tau \quad (32)$$

With this updating rule, the unknown value function V^{k+1} and the compensation law $\mathbf{u}^k_{ddc}$ are no longer relevant to the system model. They both can converge to the optimal ones V^* and $\mathbf{u}^*_{ddc}$ simultaneously [48]. For implementation purposes, the optimal value function V^* and control policy $\mathbf{u}^*_{ddc}$ can be approximated through a critic neural network and an actor neural network respectively. The approximate solutions of (32) based on the actor-critic NN can be written as

$$\hat{V}^{k+1}(\mathbf{x}) = \hat{\mathbf{w}}^{\mathrm{T}}_{V,k+1}\phi_V(\mathbf{x}), \quad (33)$$

$$\hat{\mathbf{u}}^{k+1}_{ddc} = \hat{\mathbf{w}}^{\mathrm{T}}_{u,k+1}\phi_u(\mathbf{x}), \quad (34)$$

where $\phi_V : \mathbf{R}^n \to \mathbf{R}^{K_V}$, $\phi_u : \mathbf{R}^n \to \mathbf{R}^{K_u}$ are linearly dependent basis function vectors, $\hat{\mathbf{w}}_{V,k+1} \in \mathbf{R}^{K_V}$ and $\hat{\mathbf{w}}_{u,k+1} \in \mathbf{R}^{K_u \times m}$ are the estimations of unknown coefficient vector and matrix with K_V and K_u the numbers of hidden neurons. It is known that as $K_V \to \infty$ and $K_u \to \infty$, the approximate solution $\hat{V}(\mathbf{x})$ and $\hat{\mathbf{u}}_{ddc}(t)$ will converge to the true solution $V(\mathbf{x})$ and $\mathbf{u}_{ddc}(t)$ respectively. For the special

case of 2-DOF helicopter system, the parameters satisfies: $m = 4$ and $n = 2$.

Define a time sequence $t_j = j\Delta t$ with $j = 0, 1, \ldots, q$ for a large interval. The residual error of the critic NN is

$$\begin{aligned} e_j^{k+1} &= \hat{V}^{k+1}(\mathbf{x}(t_j)) - \hat{V}^{k+1}(\mathbf{x}(t_{j+1})) \\ &\quad - \int_{t_j}^{t_{j+1}} \mathcal{L}(\mathbf{x}, \hat{\mathbf{u}}_{ddc}^k) d\tau \\ &\quad - \int_{t_j}^{t_{j+1}} (\hat{\mathbf{u}}_{ddc}^{k+1})^{\mathrm{T}} (\mathbf{R}\mathbf{u}'_{ddc} - \hat{\mathbf{u}}_{ddc}^k) d\tau \\ &= \left[\phi_V\left(\mathbf{x}(t_j)\right) - \phi_V\left(\mathbf{x}(t_{j+1})\right)\right]^{\mathrm{T}} \hat{\mathbf{w}}_{V,k+1} \\ &\quad - \int_{t_j}^{t_{j+1}} \phi_u^{\mathrm{T}} \hat{\mathbf{w}}_{u,k+1} \mathbf{R}\left(\mathbf{u}'_{ddc} - \hat{\mathbf{w}}_{u,k}^{\mathrm{T}} \phi_u\right) d\tau \\ &\quad - \frac{1}{2} \int_{t_j}^{t_{j+1}} \left(\mathbf{x}^{\mathrm{T}} \mathbf{Q} \mathbf{x} + \phi_u^{\mathrm{T}} \hat{\mathbf{w}}_{u,k} \mathbf{R} \hat{\mathbf{w}}_{u,k}^{\mathrm{T}} \phi_u\right) d\tau. \end{aligned} \tag{35}$$

The residual error can be written in a compact form by introducing the Kronecker product $\otimes$,

$$e_j^{k+1} = \rho_j^{\mathrm{T}}(\bar{\mathbf{W}}_k)\bar{\mathbf{W}}_{k+1} - \pi_j(\bar{\mathbf{W}}_k) \tag{36}$$

$$\bar{\mathbf{W}}_{k+1} = \left[\hat{\mathbf{w}}_{V,k+1}^{\mathrm{T}}, vec(\hat{\mathbf{w}}_{u,k+1})^{\mathrm{T}}\right]^{\mathrm{T}} \tag{37}$$

where $\bar{\mathbf{W}}_{k+1}^{\mathrm{T}} \in \mathbf{R}^{\bar{K}}$ is the estimated weighting function vector with $\bar{K} = K_V + mK_u$. $vec(\cdot)$ denotes the vectorization of a matrix formed by stacking the columns of the matrix into a single column vector. Besides, the iterative index $k \in \{0, 1, \ldots\}$, the time sequence index $j \in \{0, 1, \ldots, q\}$, and ρ_j, π_j are defined as

$$\rho_j = \begin{bmatrix} \phi_V\left(\mathbf{x}(t_j)\right) - \phi_V\left(\mathbf{x}(t_{j+1})\right), \\ -\int_{t_j}^{t_{j+1}} \mathbf{R}\left(\mathbf{u}'_{ddc} - \hat{\mathbf{w}}_{u,k}^{\mathrm{T}} \phi_u\right) \otimes \phi_u d\tau \end{bmatrix}, \tag{38}$$

$$\pi_j = \frac{1}{2} \int_{t_j}^{t_{j+1}} \mathbf{x}^{\mathrm{T}} \mathbf{Q} \mathbf{x} + \phi_u^{\mathrm{T}} \hat{\mathbf{w}}_{u,k} \mathbf{R} \hat{\mathbf{w}}_{u,k}^{\mathrm{T}} \phi_u d\tau. \tag{39}$$

Based on the least-squares (LS) principle, the estimated weighting function vector $\bar{\mathbf{W}}_{k+1}$ can be determined by minimizing $(e_j^{k+1})^2$. The solution is

$$\bar{\mathbf{W}}_{k+1} = [\mathbf{P}^{\mathrm{T}}(\bar{\mathbf{W}}_k)\mathbf{P}(\bar{\mathbf{W}}_k)]^{-1}\mathbf{P}^{\mathrm{T}}(\bar{\mathbf{W}}_k)\Pi(\bar{\mathbf{W}}_k) \tag{40}$$

with

$$\mathbf{P}(\bar{\mathbf{W}}_k) = [\rho_0, \rho_1, \ldots, \rho_q]^{\mathrm{T}}, \tag{41}$$

$$\Pi(\bar{\mathbf{W}}_k) = [\pi_0, \pi_1, \ldots, \pi_q]^{\mathrm{T}}. \tag{42}$$

The inverse of the matrix $\mathbf{P}^{\mathrm{T}}(\bar{\mathbf{W}}_k)\mathbf{P}(\bar{\mathbf{W}}_k)$ must exist, i.e., the matrix $\mathbf{P}^{\mathrm{T}}(\bar{\mathbf{W}}_k)$ is a full rank matrix. In general, the number of data points should satisfy $q \geq rank(\mathbf{P}(\bar{\mathbf{W}}_k))$. Besides, the terminate condition of the updating rule is set as $\left\|\bar{\mathbf{W}}_{k+1} - \bar{\mathbf{W}}_k\right\| \leq \epsilon$, where ϵ is a very small positive number. So, combining the data-driven compensator with the model-free controller, we obtain the final control input

$$\mathbf{u} = \mathbf{u}_{mfc} + (\hat{\mathbf{w}}_u^*)^{\mathrm{T}} \phi_u(\mathbf{x}), \tag{43}$$

where $\hat{\mathbf{w}}_u^*$ is the optimal gain parameter trained with the input and output data of the system. The model-free control law with a data-driven compensator is completely independent of the system model but only related to the input and output data of the system.

V. RESULTS AND DISCUSSIONS

We first review two kinds of model-based control approaches that are used to compare to the proposed model-free control strategies. Then, a simulation is taken to show the robustness proprieties of the model-free controller with different types of compensation mechanisms. Finally, we designed three experiments to demonstrate the compensation effect of the two designed compensators.

A. MODEL-BASED CONTROL FOR COMPARISON

By neglecting the unmodeled dynamics, system uncertainties, and external disturbances in Equation (2), the 2-DOF helicopter system can be described accurately with the system parameters listed in Table 1. The linear quadratic regulator (LQR) control [51] and an optimal feedback linearization control (OFLC) [7] are taken as two comparison baselines to evaluate the control performances of the two model-free control methods above. It's worth noting that the two model-based control methods are based on the state equation (2) that ignores the unmodeled dynamics, system uncertainties, and external disturbances.

1) LQR DESIGN

According to [51], the pitch angle θ is regulated by a proportional integral differential (PID) with a feed-forward term, meanwhile, the yaw angle ψ is regulated by a PID controller without a feed-forward term. The nonlinear feed-forward term in the pitch angle control compensates the gravitational torque $\tau_g = m_{heli} g l_{cm} \cos\theta$ in Equation (1) and reads

$$u_{ff} = k_{ff} \frac{m_{heli} g l_{cm} \cos\theta_d}{K_{pp}}, \tag{44}$$

where θ_d is the desired pitch angle and $k_{ff} = 1.0$ is the feedforward control gain, which compensates the gravity. The PID feedback control $[u_1, u_2]^{\mathrm{T}} = \mathbf{K}_{pid}(\mathbf{x}_d - \mathbf{x})$ is optimized with LQR by linearizing the nonlinear model into a linear one. The control gain is

$$\mathbf{K}_{pid} = \begin{bmatrix} 18.94, 1.98, 7.49, 1.53, 7.03, 0.77 \\ -2.2, 19.45, -0.45, 11.89, -0.77, 7.03 \end{bmatrix}. \tag{45}$$

So the model-based PID controller reads

$$\mathbf{u} = \begin{bmatrix} u_{ff} \\ 0 \end{bmatrix} + \mathbf{K}_{pid}(\mathbf{x}_d - \mathbf{x}). \tag{46}$$

2) OPTIMAL FEEDBACK LINEARIZATION CONTROL (OFLC)

Let $\mathbf{y} = [y_1, y_2]^{\mathrm{T}} = [\theta(t), \psi(t)]^{\mathrm{T}}$ be the system output, and $\mathbf{y}_d = [y_{d1}, y_{d2}]^{\mathrm{T}} = [\theta_d(t), \psi_d(t)]^{\mathrm{T}}$ be the desired trajectories for the outputs. Neglecting the unmodeled dynamics and system uncertainties in Equation 2 and extracting the last two rows of this equation, we have

$$\ddot{\mathbf{y}} = \mathbf{F}(\mathbf{y}, \dot{\mathbf{y}}) + \mathbf{G}(\mathbf{y})\mathbf{u}, \tag{47}$$

TABLE 1. The parameters of the helicopter system.

Symbol	Description	Value	Unit
m_h	total moving mass of the helicopter	1.3872	kg
l_{cm}	center of mass along helicopter body from pitch axis	0.1860	m
$J_{eq,p}$	total moment of inertia about pitch axis	0.0384	$kg.m^2$
$J_{eq,y}$	total moment of inertia about yaw axis	0.0432	$kg.m^2$
k_{pp}	thrust torque constant of pitch axis from pitch propeller	0.2040	$N.m/V$
k_{py}	thrust torque constant of pitch axis from yaw propeller	0.0068	$N.m/V$
k_{yy}	thrust torque constant of yaw axis from yaw propeller	0.0720	$N.m/V$
k_{yp}	thrust torque constant of yaw axis from pitch propeller	0.0219	$N.m/V$
k_{ff}	control gain of the gravity compensator	1.0	–
B_p	equivalent viscous damping about pitch axis	0.800	N/V
B_y	equivalent viscous damping about yaw axis	0.318	N/V

where $\mathbf{F}(\mathbf{y}, \dot{\mathbf{y}}) = [f_1, f_2]^{\mathrm{T}}$ and $\mathbf{G}(\mathbf{y}) = \begin{bmatrix} g_{11} & g_{12} \\ g_{21} & g_{22} \end{bmatrix}$. Referring to the results in [7], the optimal feedback linearization controller (OFLC) is

$$\mathbf{u}_{oflc} = \mathbf{G}^{-1}[-\mathbf{F}(\mathbf{y}, \dot{\mathbf{y}}_e) + \mathbf{K}_{oflc}(\Gamma_d - \Gamma_e)], \tag{48}$$

where $\dot{\mathbf{y}}_e$ is the numerical estimation of the derivative of the output signal, $\Gamma_d^{\mathrm{T}} = [\mathbf{y}_d^{\mathrm{T}}, \dot{\mathbf{y}}_d^{\mathrm{T}}, \mathbf{0}_2^{\mathrm{T}}]$ and $\Gamma_e^{\mathrm{T}} = [\mathbf{y}^{\mathrm{T}}, \dot{\mathbf{y}}_e^{\mathrm{T}}, \int_0^t (\mathbf{y}^{\mathrm{T}} - \mathbf{y}_d^{\mathrm{T}}) d\tau]$ are two extended vectors, and the control gain reads

$$\mathbf{K}_{oflc} = \begin{bmatrix} 19.02, 0.82, 11.74, 0.47, 7.07, -0.25 \\ 2.14, 19.11, -0.47, 11.75, 0.25, 7.07 \end{bmatrix}. \tag{49}$$

B. PARAMETERS OF THE MODEL-FREE CONTROL WITH/WITHOUT A COMPENSATOR

Before the simulation and experiment, we first give the parameters of each controller and their selection basis. The parameters of the iPD controller (10) are selected as

$$\alpha_1 = 18, k_{p1} = 3.1623, k_{d1} = 3.3652, \tag{50}$$
$$\alpha_2 = 16, k_{p2} = 3.1623, k_{d2} = 3.3652. \tag{51}$$

Based on iPD, the control parameters used in iPD-MFSMC are

$$\lambda_1 = 5, \eta_1 = 2, \phi_1 = 0.1, \tag{52}$$
$$\lambda_2 = 3, \eta_2 = 2, \phi_2 = 0.1. \tag{53}$$

Furthermore, for convenience, we assume that the estimation error is zero, i.e. $\Delta\hat{\varrho}_1 = \Delta\hat{\varrho}_2 = 0$, and after multiple attempts, we set the estimation upper bounds as $E_1 = 2$ and $E_2 = 3$.

The implementation of iPD-MFDDC needs to train an optimal compensator coefficient $\mathbf{w}_u^*$ using the input-output data of the helicopter system. In this study, we artificially choose a set of probing excitation signal $\mathbf{u}_p'(t) = [u_{p,\theta}, u_{p,\psi}]^T$ with the following form,

$$\mathbf{u}_p'(t) = -\frac{1}{2}\begin{bmatrix} \sin(0.4t) + 2\sin(1.6t), \\ \sin(0.5t) + \sin(1.9t) + \sin(9.1t) \end{bmatrix}. \tag{54}$$

The sampling time of each experiment is set as $\Delta t = 0.005$s. When the data generation experiment runs

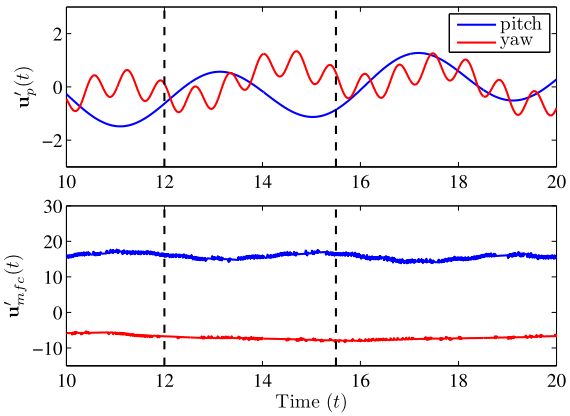

FIGURE 4. The probing and pre-stabilizing control signal used to excite and stabilize the systems, respectively. Only the data located in the time window [12s,15.5s] will be used to train the data-driven compensator.

$T_{run} = 20$s, we totally accumulate 4001 pairs of input-output data. The input data consists of the probing signal $\mathbf{u}_p'(t)$ and the iPD control signal $\mathbf{u}_{mfc}'(t)$ while the output data is composed of the system output $\mathbf{y}'$, the output derivative estimation $[\dot{\mathbf{y}}']_{est}$ and the integral of the system output signal $\int_0^t \mathbf{y}' d\tau$. The input data and the corresponding output data are shown in Figures 4 and 5 respectively. Without loss of generality, we select a narrow time window [12s, 15.5s] that contains 700 pairs of input and output data from the whole database for updating the compensator gain $\mathbf{w}_u$.

To approximate the optimal solutions of the value function and control policy with actor-critic NN, we define the complete basis function vectors as

$$\begin{aligned} \phi_V &= [x_1^2, x_2^2, x_3^2, x_4^2, x_1x_2, x_1x_3, x_1x_4, x_2x_3, x_2x_4, x_3x_4]^T \\ &= [\theta^2, \psi^2, \dot{\theta}^2, \dot{\psi}^2, \theta\psi, \theta\dot{\theta}, \theta\dot{\psi}, \psi\dot{\theta}, \psi\dot{\psi}, \dot{\theta}\dot{\psi}]^T, \end{aligned} \tag{55}$$

$$\begin{aligned} \phi_u &= \left[x_1, x_2, x_3, x_4, \int_0^t x_1 d\tau, \int_0^t x_2 d\tau\right]^T \\ &= \left[\theta, \psi, \dot{\theta}, \dot{\psi}, \int_0^t \theta d\tau, \int_0^t \psi d\tau\right]^T. \end{aligned} \tag{56}$$

The initial weights $\mathbf{w}_{V0}$ and $\mathbf{w}_{u_0}$ of the two NNs are both initialized to zero. In order to get an optimal solution with

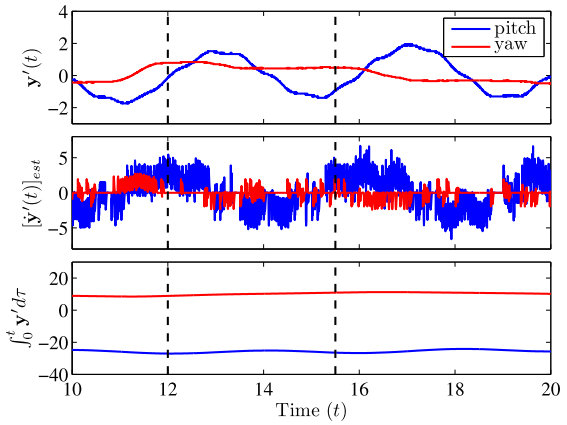

FIGURE 5. The measured output data and its derivative estimation and integral value. Only the data located in the time window [12s,15.5s] will be used to train the data-driven compensator.

sufficiently high precision, we set the iteration termination condition of the LS updating rule as $\left\|\bar{\mathbf{W}}_{k+1}-\bar{\mathbf{W}}_k\right\|<10^{-8}$. After 13 times iteration, the LS updating procedure is terminated. We make the weight matrices $\mathbf{w}_V^{13}$ and $\mathbf{w}_u^{13}$ obtained from the 13^{th} iteration represents the gains of the optimal value function and optimal compensator policy respectively, i.e.,

$$V(\mathbf{x}) \approx \left(\mathbf{w}_V^{13}\right)^T \phi_V, \tag{57}$$

$$\mathbf{u}_{ddc}(t) \approx \left(\mathbf{w}_u^{13}\right)^T \phi_u, \tag{58}$$

with $\mathbf{w}_V^{13} = [0.32, -0.89, -0.0003, 0.0003, -1.07, 0.01, 0.005, -0.024, -0.042, 0.007]^T$ and

$$\mathbf{w}_u^{13} = \begin{bmatrix} -4.24, -13.89, 0.25, -0.12, -1.21, -2.50 \\ -0.87, -5.29, 0.43, -0.14, -1.10, -2.51 \end{bmatrix}^T.$$

C. SIMULATION RESULTS AND DISCUSSION

This section presents the simulation of step response for the helicopter control based on five different control strategies. By comparing the simulation results, we studied the robustness of different control methods against unmodeled dynamics, system uncertainties, and external disturbance. We artificially set up two scenarios for comparison:

(1) Assume that the system model in Equation (2) is completely known and accurate, which means that the system has no unmodeled dynamics, system uncertainties, and external disturbance, i.e., $\Delta f_i = 0, \delta_{ij} = 0$. This scenario serves as a baseline that aims to illustrate all five control approaches are effective.

(2) Assume that the system model has a certain degree of unmodeled dynamics, system uncertainties and external disturbance, i.e., $\Delta f_i \neq 0, \delta_{ij} \neq 0$. In this scenario, we artificially add uncertainty and random noise to the system and take it as the real accurate model of the system. For convenience in this scenario, we assume that the uncertainty δ_{ij} can be written as a percentage of the corresponding item in the original system. For example, we set $\delta_{11} = -0.153 g_{11}$, which means that we take g_{11} in Equation (2) without uncertainty minus 15.3% of the original value as the exact value of this item. Similarly, we add uncertainty to the other terms in turn, i.e., $\delta_{12} = 0.165 g_{12}$, $\delta_{21} = 0.137 g_{21}$ and $\delta_{22} = -0.185 g_{22}$. Different from the handling way on δ_{ij}, we set Δf_i as a serial of random number with a specific mean value and variance to represent the uncertainty and external disturbance. In this study, we set $\Delta f_1 = rand(0.14, 0.27)$ and $\Delta f_2 = rand(0.31, 0.23)$, which means that Δf_1 and Δf_2 are two random sequences with mean value 0.14 and 0.31 and variance 0.27 and 0.23 respectively. Besides, we add two Gaussian white noise signals with mean value zero and variance 0.0141 to the state variables x_1 and x_2 respectively to simulate the sensor noise.

Since LQR and OFLC are model-based methods that require a completely known system model, we design them in scenario (1). On the contrary, the iPD, iPD-MFSMC, and iPD-MFDDC are model-free methods. We design them in scenario (2). After finishing the design of five controllers, we test them in both scenarios (1) and (2). Figures 6 and 7 show the output and input signals of the step responses respectively. Observing the output signals in scenario (1) alone, we can see that each control approach has good tracking control performance. It means that if we have an accurate mathematical model, all five control methods can well realize the control task. However, when we add unmodeled dynamics, system uncertainties and external disturbance to the helicopter system, the tracking control performances of LQR and OFLC become worse. Meanwhile, the control performances of iPD, iPD-MFSMC, and iPD-MFDDC are still satisfied, which means that the three model-free control strategies are quite robust against uncertainties and external disturbances. The simulation results shown in this section imply that the proposed model-free control strategies have stronger robustness compared to the model-based ones.

D. EXPERIMENTAL RESULTS AND DISCUSSION

The controls of the 2-DOF laboratory helicopter are carried out by two servo motors. To protect the system hardware from damage, the input voltages of two servo motors should be limited in the finite intervals. The pitch control voltage of the UPM-2405 DC motor is bounded by the amplifier outputs $V_{p,\max} = 24V$ and $V_{p,\min} = -24V$. The yaw control voltage of the UPM-1503 DC motor is bounded by $V_{y,\max} = 15V$ and $V_{y,\min} = -15V$.

We demonstrate the effectiveness of the proposed control algorithm through three tracking control scenarios:

Tracking a Circle: The reference signals θ_d and ψ_d together form a circular trajectory, whose center coordinate and radius are with $(-10, -20)$ and 20 respectively,

$$(\theta_d + 10)^2 + (\psi_d + 20)^2 = 20^2 \tag{59}$$

Tracking a Square: The reference signals θ_d and ψ_d together form a foursquare trajectory, whose center coordinate and side length are with $(-10, -20)$ and 40 respectively,

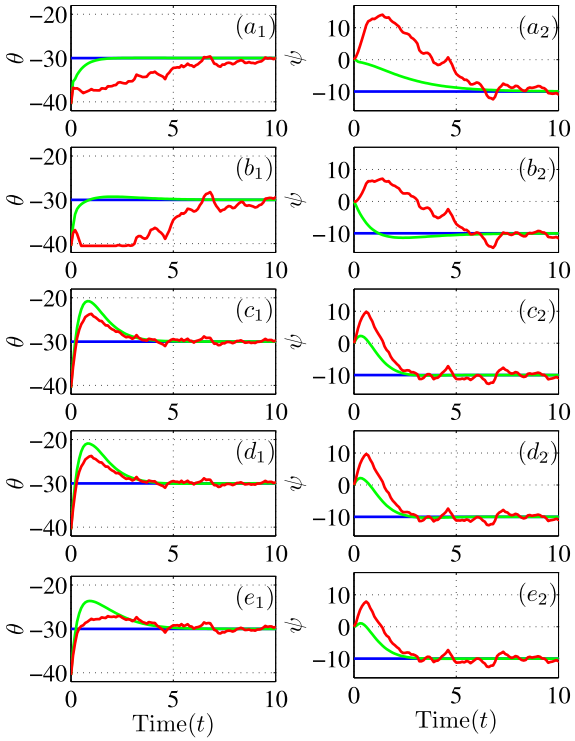

FIGURE 6. The simulation results of the step response of five different kinds of control methods. The subfigures (a_1)-(e_1) represent the pitch motion trajectories under the control input of LQR, OFLC, iPD, iPD-MFSMC, and iPD-MFDDC respectively. The corresponding yaw motion trajectories are shown in subfigures (a_2)-(e_2). In each subfigure, the blue line represents the reference signal, the red (green) curve represents the result when the system model is with (without) uncertainties and disturbances.

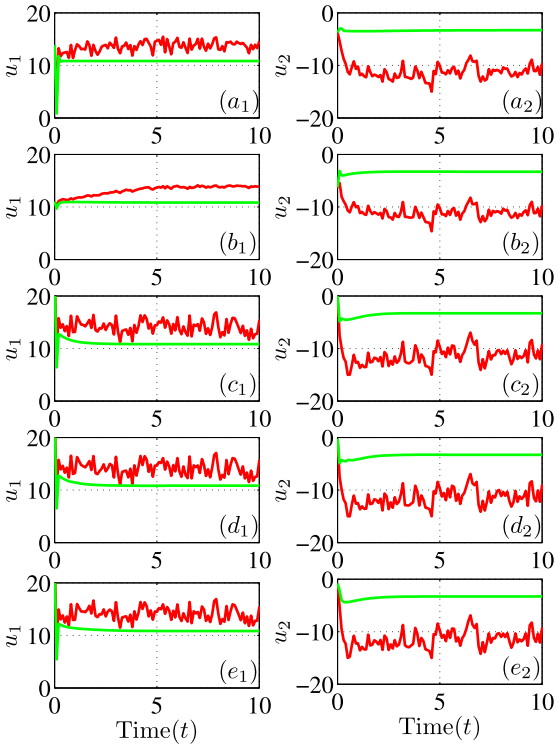

FIGURE 7. The control inputs of five different kinds of control methods. Subfigures (a_1)-(e_1) show the pitch motion control input voltages of the five control methods LQR, OFLC, iPD, iPD-MFSMC, and iPD-MFDDC respectively. The corresponding yaw motion control input voltages are shown in subfigures (a_2)-(e_2). In each subfigure, the red (green) curve represents the control input voltage when the system model is with (without) uncertainties and disturbances.

Tracking Complex Trajectories: The reference signal consist of the summation of different harmonic signals,

$$\theta_d = -8\cos\left(\frac{1}{25}\pi t\right) + 7\sin\left(\frac{2}{25}\pi t + \frac{\pi}{4}\right) + 4\sin\left(\frac{7}{50}\pi t + \frac{\pi}{3}\right) - 10 \tag{60}$$

$$\psi_d = 2\sin\left(\frac{1}{10}\pi t\right) - 6\sin\left(\frac{27}{50}\pi t + \frac{\pi}{9}\right) - 4\cos\left(\frac{16}{25}\pi t + \frac{\pi}{3}\right) - 20 \tag{61}$$

To clearly evaluate the performance difference between model-based and model-free control methods, we define a statistical indicator,

$$J_e(\theta, \psi) = \log_{10}\left(\frac{|e_\theta| + |e_\psi|}{2}\right), \tag{62}$$

where $e_\theta = \theta - \theta_d$ and $e_\psi = \psi - \psi_d$ are the tracking errors of the pitch and yaw motions. Besides, we use the average integral absolute error function to quantitatively discuss the control performance of each control method,

$$J_{IAE} = \frac{1}{T}\int_0^T \left(|e_\theta| + |e_\psi|\right) dt, \tag{63}$$

where $T = 80$s is the total running time of each experiment.

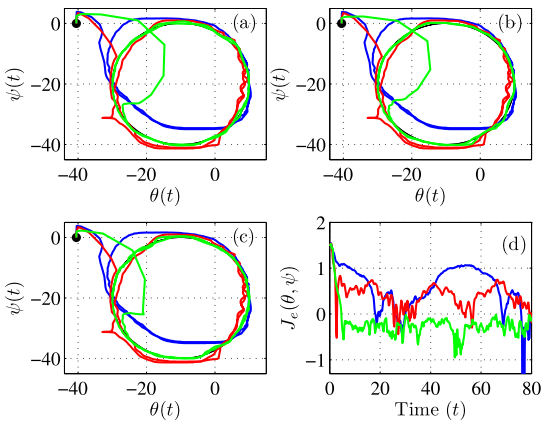

FIGURE 8. The experimental tracking control performances of different control methods when tracking a circle. The blue and red lines represent the result of LQR design and OFLC design respectively. The green lines in subgraphs (a)-(c) represent the results of iPD, iPD-MFSMC, and iPD-MFDDC respectively. The subgraph (d) shows the values of J_e for different control methods, where the green line represents the mean value of three model-free approaches.

Figure 8 shows the experimental output of the first scenario. We can see that the control performance of the OFLC design is better than the LQR design. Compared to the model-based control approaches (i.e., red and blue lines in each subgraph), the model-free control method (i.e., green lines in

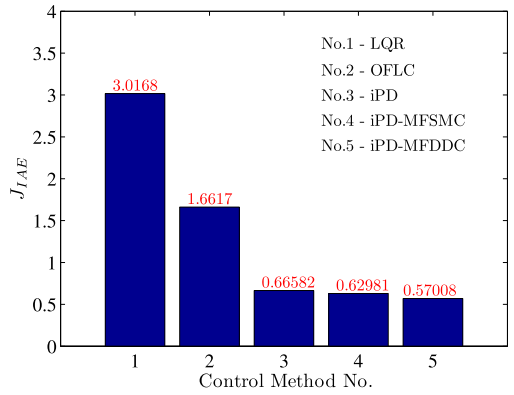

FIGURE 9. The statistical values of the average integral absolute errors for different control methods when tracking a circle.

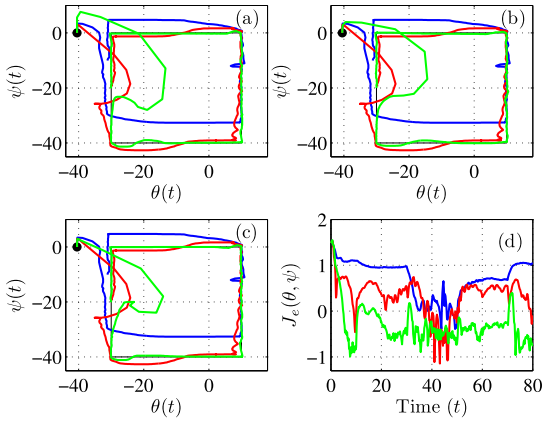

FIGURE 10. The experimental tracking control performances of different control methods when tracking a square. The blue and red lines represent the result of LQR design and OFLC design respectively. The green lines in subgraphs (a)-(c) represent the results of iPD, iPD-MFSMC and iPD-MFDDC respectively. The subgraph (d) shows the values of J_e for different control methods, where the green line represents the mean value of three model-free approaches.

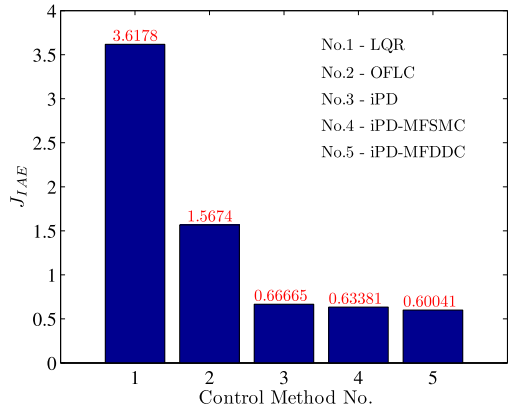

FIGURE 11. The statistical values of the average integral absolute errors fordifferent control methods when tracking a square.

each subgraph) has better control performance because they are closer to the desired trajectory.

The values of (62) for each control method are shown in Figure 8(d). Since there are similar outputs of the three model-free methods, we calculate their average (green line in Figure 8(d)) and compare it with the model-based method.

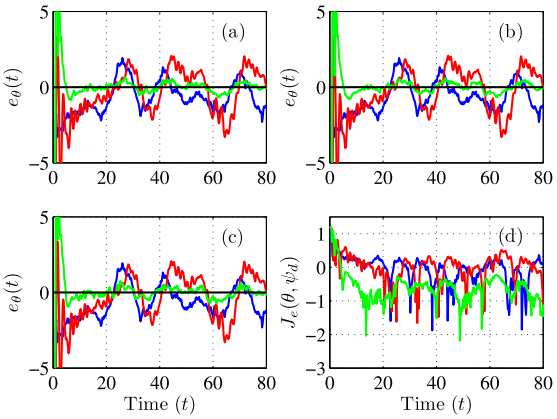

FIGURE 12. The experimental pitch motion tracking error of different control methods when tracking a complex harmonic signal. The blue and red lines represent the result of LQR design and OFLC design respectively. The green lines in subgraphs (a)-(c) represent the results of iPD, iPD-MFSMC, and iPD-MFDDC respectively. The subgraph (d) shows the values of J_e for different control methods, where the green line represents the mean value of three model-free approaches.

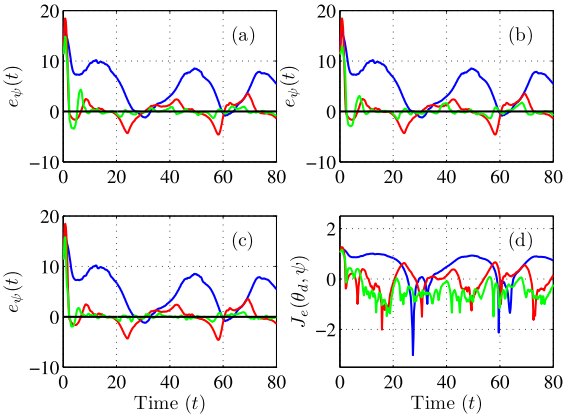

FIGURE 13. The experimental yaw motion tracking error of different control methods when tracking a complex harmonic signal. The blue and red lines represent the result of LQR and OFLC respectively. The green lines in subgraphs (a)-(c) represent the results of iPD, iPD-MFSMC, and iPD-MFDDC respectively. The subgraph (d) shows the values of J_e for different control methods, where the green line represents the mean value of three model-free approaches.

This graph quantitatively shows that the model-free method is superior to the model-based method. To investigate which of the three model-free control methods works best, we calculate the average integral absolute error of each method and present them in Figure 9. It proves again that the control performance is significantly improved with three model-free controls compared to LQR and OFLC. Besides, the control performance slightly improves with iPD-MFSMC and iPD-MFDDC compared to iPD. According to the quantitative calculation, the control effect of iPD-MFSMC is increased by 5.41% compared with iPD. However, the control effect of iPD-MFDDC is 9.48% higher than that of iPD. Figure 10 shows the experimental output results of the second scenario. Similar results of the first scenario can be found in this scenario. Figure 10(a)-(c) qualitatively show that the model-free control has better control performance than the model-based control. Also, Figure 10(d) quantitatively shows

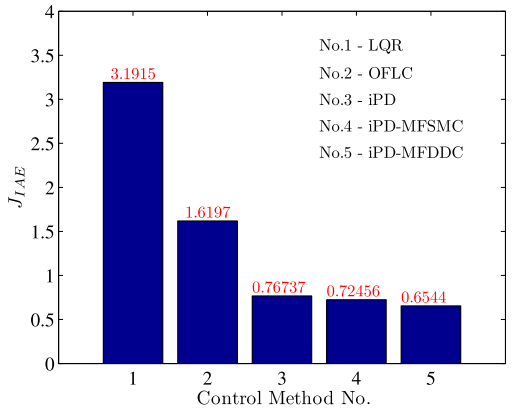

FIGURE 14. The statistical values of the average integral absolute errors for different control methods when tracking a set of harmonic signals.

that model-free control is superior to model-based control. Figure 11 shows the statistical results of the average integral absolute error for each control method. We can see that the model-free control performance improves significantly compared to the model-based. It shows that iPD-MFSMC and iPD-MFDDC are about 4.93% and 5.27% higher than iPD respectively.

Figures 12 and 13 show the experimental outputs of the third scenario. Subfigures (a)-(c) qualitatively indicate that the model-free control has smaller tracking errors than model-based. Subfigure (d) quantitatively shows that model-free control is superior to the model-based control. Figure 14 shows the statistical results of the average integral absolute error for each control method. We can also see that the performance of iPD-MFSMC and iPD-MFDDC are about 5.58% and 9.68% higher than that of iPD respectively.

VI. CONCLUSION

This paper investigates two model-free compensators, the MFSMC and the MFDDC. They are used to compensate for a nominal iPD controller. The compensated nonlinear controller is a dual-loop robust model-free controller with iPD as the inner loop and the compensator as the outer loop. The proposed dual-loop model-free control algorithms are validated by simulations and experiments on a nonlinear laboratory helicopter setup. The cross-comparisons between two robust model-free control methods and two model-based control methods show that: 1) the tracking performance of the dual-loop robust model-free control approaches are superior to that of the model-based method, 2) with a compensator (MFSMC or MFDDC) embedded in the inner loop controller, its tracking performances are improved significantly, and 3) the compensator MFDDC performs better than the compensator MFSMC slightly. The results in this paper indicate that the proposed dual-loop robust model-free control approaches are quite promising in dealing with the control tasks for the systems with unmodeled dynamics and uncertainties, even with unknown dynamical models. This is a very favorable property to potential practical applications.

REFERENCES

[1] Z. Jiang, J. Han, Y. Wang, and Q. Song, "Enhanced LQR control for unmanned helicopter in hover," in *Proc. 1st Int. Symp. Syst. Control Aerosp. Astronaut.*, May 2006, pp. 1–6.

[2] H. Liu, G. Lu, and Y. Zhong, "Robust LQR attitude control of a 3-DOF laboratory helicopter for aggressive maneuvers," *IEEE Trans. Ind. Electron.*, vol. 60, no. 10, pp. 4627–4636, Oct. 2013.

[3] S. Bouabdallah, A. Noth, and R. Siegwart, "PID vs. LQ control techniques applied to an indoor micro quadrotor," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, vol. 3, Sep. 2014, pp. 2451–2456.

[4] E. V. Kumar, G. S. Raaja, and J. Jerome, "Adaptive PSO for optimal LQR tracking control of 2 DoF laboratory helicopter," *Appl. Soft Comput.*, vol. 41, pp. 77–90, Apr. 2016.

[5] M. Chemachema and S. Zeghlache, "Output feedback linearization based controller for a helicopter-like twin rotor MIMO system," *J. Intell. Robot. Syst.*, vol. 80, no. 1, pp. 181–190, Oct. 2015.

[6] M. Mohammadzahei, H. Ziaiefar, M. Ghodsi, and I. B. Bahadur, "Yaw control of an unmanned helicopter with feedback linearization," in *Proc. 1st Int. Conf. Unmanned Vehicle Syst.-Oman (UVS)*, Muscat, Oman, Feb. 2019, pp. 1–5.

[7] Y. Xin, Z.-C. Qin, and J.-Q. Sun, "Input-output tracking control of a 2-DOF laboratory helicopter with improved algebraic differential estimation," *Mech. Syst. Signal Process.*, vol. 116, pp. 843–857, Feb. 2019.

[8] Y. Yu, G. Lu, C. Sun, and H. Liu, "Robust backstepping decentralized tracking control for a 3-DOF helicopter," *Nonlinear Dyn.*, vol. 82, nos. 1–2, pp. 947–960, Oct. 2015.

[9] J. Hu, J. Huang, Z. Gao, and H. Gu, "Position tracking control of a helicopter in ground effect using nonlinear disturbance observer-based incremental backstepping approach," *Aerosp. Sci. Technol.*, vol. 81, pp. 167–178, Oct. 2018.

[10] A. Haruna, Z. Mohamed, M. Ö. Efe, and M. A. M. Basri, "Improved integral backstepping control of variable speed motion systems with application to a laboratory helicopter," *ISA Trans.*, vol. 97, pp. 1–13, Feb. 2020.

[11] P. Lambert and M. Reyhanoglu, "Observer-based sliding mode control of a 2-DOF helicopter system," in *Proc. 44th Annu. Conf. IEEE Ind. Electron. Soc.*, Washington DC, USA, Oct. 2018, pp. 2596–2600.

[12] S. P. Sadala and B. M. Patre, "A new continuous sliding mode control approach with actuator saturation for control of 2-DOF helicopter system," *ISA Trans.*, vol. 74, pp. 165–174, Mar. 2018.

[13] W. Boukadida, A. Benamor, H. Messaoud, and P. Siarry, "Multi-objective design of optimal higher order sliding mode control for robust tracking of 2-DoF helicopter system based on metaheuristics," *Aerosp. Sci. Technol.*, vol. 91, pp. 442–455, Aug. 2019.

[14] A. Ramalakshmi, P. Manoharan, K. Harshath, and M. Varatharajan, "Model predictive control of 2DOF helicopter," *Int. J. Innov. Sci. Res.*, vol. 24, no. 2, pp. 337–346, 2016.

[15] J. Zhang, X. Cheng, and J. Zhu, "Control of a laboratory 3-DOF helicopter: Explicit model predictive approach," *Int. J. Control Automat. Syst.*, vol. 14, no. 2, pp. 389–399, Apr. 2016.

[16] S. Jalili, B. Rezaie, and Z. Rahmani, "A novel hybrid model predictive control design with application to a quadrotor helicopter," *Optim. Control Appl. Meth.*, vol. 39, no. 4, pp. 1301–1322, Jul. 2018.

[17] J. C. Spall and J. A. Cristion, "Model-free control of nonlinear stochastic systems with discrete-time measurements," *IEEE Trans. Autom. Control*, vol. 43, no. 9, pp. 1198–1210, 1998.

[18] H. Hjalmarsson, M. Gevers, S. Gunnarsson, and O. Lequin, "Iterative feedback tuning: Theory and applications," *IEEE Control Syst.*, vol. 18, no. 4, pp. 26–41, Aug. 1998.

[19] M. C. Campi, A. Lecchini, and S. M. Savaresi, "Virtual reference feedback tuning: A direct method for the design of feedback controllers," *Automatica*, vol. 38, no. 8, pp. 1337–1346, Aug. 2002.

[20] Z. Hou and S. Jin, "Data-driven model-free adaptive control for a class of MIMO nonlinear discrete-time systems," *IEEE Trans. Neural Netw.*, vol. 22, no. 12, pp. 2173–2188, Dec. 2011.

[21] M. Fliess and C. Join, "Model-free control," *Int. J. Control*, vol. 86, no. 12, pp. 2228–2252, 2013.

[22] M.-B. Radac, R.-E. Precup, and E. M. Petriu, "Model-free primitive-based iterative learning control approach to trajectory tracking of MIMO systems with experimental validation," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 26, no. 11, pp. 2925–2938, Nov. 2015.

[23] A. S. Bazanella, L. Campestrini, and D. Eckhard, *Data-Driven Controller Design: The H2 Approach*. Springer, 2011.

[24] D. Eckhard and A. S. Bazanella, "Robust convergence of the steepest descent method for data-based control," *Int. J. Syst. Sci.*, vol. 43, no. 10, pp. 1969–1975, Oct. 2012.
[25] Y. Liu, H.-W. Chen, and J.-Q. Lu, "Data-based controllability analysis of discrete-time linear time-delay systems," *Int. J. Syst. Sci.*, vol. 45, no. 11, pp. 2411–2417, Nov. 2014.
[26] S. Yin, X. Li, H. Gao, and O. Kaynak, "Data-based techniques focused on modern industry: An overview," *IEEE Trans. Ind. Electron.*, vol. 62, no. 1, pp. 657–667, Jan. 2015.
[27] P.-A. Gédouin, C. Join, E. Delaleau, J.-M. Bourgeot, S. Arbab-Chirani, and S. Calloch, "A new control strategy for shape memory alloys actuators," in *Proc. 8th Eur. Symp. Martensitic Transf.*, Prague, Czech Republic, 2009, pp. 1–9.
[28] L. Michel, C. Join, M. Fliess, P. Sicard, and A. Cheriti, "Model-free control of DC/DC converters," in *Proc. IEEE 12th Workshop Control Modeling Power Electron. (COMPEL)*, Jun. 2010, pp. 1–8.
[29] R. Madonski and P. Herman, "Model-free control of a two-dimensional system based on uncertainty reconstruction and attenuation," in *Proc. Conf. Control Fault-Tolerant Syst. (SysTol)*, Oct. 2013, pp. 542–547.
[30] J. De Miras, C. Join, M. Fliess, S. Riachy, and S. Bonnet, "Active magnetic bearing: A new step for model-free control," in *Proc. 52nd IEEE Conf. Decis. Control*, Dec. 2013, pp. 7449–7454.
[31] J.-J. E. Slotine, *Applied Nonlinear Control*. Englewood Cliffs, NJ, USA: Prentice-Hall, 1991.
[32] J. Sun, Y. Wang, Y. Wang, and Y. Shen, "Finite-time synchronization between two complex-variable chaotic systems with unknown parameters via nonsingular terminal sliding mode control," *Nonlinear Dyn.*, vol. 85, no. 2, pp. 1105–1117, Jul. 2016.
[33] J. Sun, Y. Wu, G. Cui, and Y. Wang, "Finite-time real combination synchronization of three complex-variable chaotic systems with unknown parameters via sliding mode control," *Nonlinear Dyn.*, vol. 88, no. 3, pp. 1677–1690, May 2017.
[34] J. Sun, X. Zhao, J. Fang, and Y. Wang, "Autonomous memristor chaotic systems of infinite chaotic attractors and circuitry realization," *Nonlinear Dyn.*, vol. 94, no. 4, pp. 2879–2887, Dec. 2018.
[35] R.-E. Precup, M.-B. Radac, C.-A. Dragos, S. Preitl, and E. M. Petriu, "Model-free tuning solution for sliding mode control of servo systems," in *Proc. IEEE Int. Syst. Conf.*, Mar. 2014, pp. 30–35.
[36] H. Wang, X. Ye, Y. Tian, and N. Christov, "Attitude control of a quadrotor using model free based sliding model controller," in *Proc. 20th Int. Conf. Control Syst. Comput. Sci.*, May 2015, pp. 149–154.
[37] R.-E. Precup, M.-B. Radac, R.-C. Roman, and E. M. Petriu, "Model-free sliding mode control of nonlinear systems: Algorithms and experiments," *Inf. Sci.*, vol. 381, pp. 176–192, Mar. 2017.
[38] S. P. Meyn, "The policy iteration algorithm for average reward Markov decision processes with general state space," *IEEE Trans. Autom. Control*, vol. 42, no. 12, pp. 1663–1680, Dec. 1997.
[39] J. Sun, G. Han, Z. Zeng, and Y. Wang, "Memristor-based neural network circuit of full-function pavlov associative memory with time delay and variable learning rate," *IEEE Trans. Cybern.*, to be published.
[40] D. Vrabie and F. Lewis, "Adaptive dynamic programming for online solution of a zero-sum differential game," *J. Control Theory Appl.*, vol. 9, no. 3, pp. 353–360, Aug. 2011.
[41] H.-N. Wu and B. Luo, "Neural network based online simultaneous policy update algorithm for solving the HJI equation in nonlinear H_∞ control," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 23, no. 12, pp. 1884–1895, Dec. 2012.
[42] B. Luo, H.-N. Wu, and T. Huang, "Off-policy reinforcement learning for H_∞ control design," *IEEE Trans. Cybern.*, vol. 45, no. 1, pp. 65–76, Jan. 2015.
[43] H. Zhang, L. Cui, X. Zhang, and Y. Luo, "Data-driven robust approximate optimal tracking control for unknown general nonlinear systems using adaptive dynamic programming method," *IEEE Trans. Neural Netw.*, vol. 22, no. 12, pp. 2226–2236, Dec. 2011.
[44] R. Song, F. L. Lewis, Q. Wei, and H. Zhang, "Off-policy actor-critic structure for optimal control of unknown systems with disturbances," *IEEE Trans. Cybern.*, vol. 46, no. 5, pp. 1041–1050, May 2016.
[45] Y. Jiang and Z.-P. Jiang, "Robust adaptive dynamic programming and feedback stabilization of nonlinear systems," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 25, no. 5, pp. 882–893, May 2014.
[46] H. Li, D. Liu, and D. Wang, "Integral reinforcement learning for linear continuous-time zero-sum games with completely unknown dynamics," *IEEE Trans. Autom. Sci. Eng.*, vol. 11, no. 3, pp. 706–714, Jul. 2014.
[47] Y. Zhu, D. Zhao, and X. Li, "Iterative adaptive dynamic programming for solving unknown nonlinear zero-sum game based on online data," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 28, no. 3, pp. 714–725, Mar. 2017.
[48] Q. Zhang, D. Zhao, and Y. Zhu, "Data-driven adaptive dynamic programming for continuous-time fully cooperative games with partially constrained inputs," *Neurocomputing*, vol. 238, pp. 377–386, May 2017.
[49] Q. Zhang and D. Zhao, "Data-based reinforcement learning for nonzero-sum games with unknown drift dynamics," *IEEE Trans. Cybern.*, vol. 49, no. 8, pp. 2874–2885, Aug. 2019.
[50] B. Luo, T. Huang, H. Wu, and X. Yang, "Data-driven H_∞ control for nonlinear distributed parameter systems," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 26, no. 11, pp. 2949–2961, Nov. 2015.
[51] *2-DOF Helicopter: Laboratory Manual*. Quanser, Markham, ON, Canada, 2011.
[52] H. Modares, F. L. Lewis, and M.-B. Naghibi-Sistani, "Integral reinforcement learning and experience replay for adaptive optimal control of partially-unknown constrained-input continuous-time systems," *Automatica*, vol. 50, no. 1, pp. 193–202, Jan. 2014.

ZHI-CHANG QIN was born in Xintai, Shandong, China, in 1989. He received the B.S. degree in process equipment and control engineering from the Shandong University of Science and Technology, in 2011, and the Ph.D. degree in dynamics and control from Tianjin University, in 2017. He has been an Assistant Professor with the School of Transportation and Vehicle Engineering, Shandong University of Technology, since 2017. His research interests are time-delay control, motion control, and numerical optimization.

YING XIN was born in Anqiu, Shandong, China, in 1989. She received the B.S. degree in mechanical design manufacture and automation from Yantai University and the Ph.D. degree in dynamics and control from Tianjin University, in 2018. She has been an Assistant Professor with the School of Mechanical Engineering, Tianjin University of Technology, since 2019. Her research interests are data-driven control and multiobjective optimization.

JIAN-QIAO SUN was born in Wuhan, Hubei, China, in 1956. He received the B.S. degree in solid mechanics from the Huazhong University of Science and Technology, in 1982, and the Ph.D. degree in mechanical engineering from UC Berkeley, in 1988. In 1994, he joined the University of Delaware as a Faculty Member, until 2007, when he moved to the University of California at Merced. He is currently a Professor and the Chair of mechanical engineering. His research interests include vibrations, controls, energy harvesting, and data-driven modeling and analysis of complex systems. He serves as the Editor-in-Chief for the *International Journal of Dynamics and Control*.