IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received October 2, 2020, accepted October 9, 2020, date of publication October 19, 2020, date of current version November 3, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.3031891

Adaptive Online Data-Driven Tracking Control for Highly Flexible Aircrafts With Partial Observability

CHI PENG AND JIANJUN MA, (Member, IEEE)
College of Intelligence Science and Technology, National University of Defense Technology, Changsha 410073, China
Corresponding author: Jianjun Ma (jianjunma@hotmail.com)

This work was supported in part by the National Science Foundation of China under Grant 61203095 and Grant 61403407.

ABSTRACT In this study, an adaptive online data-driven tracking controller for a highly flexible aircraft (HFA) was developed. This control design innovatively combines integral reinforcement learning (IRL) and the optimal control theory to ensure asymptotic tracking performance, even when system dynamics information is difficult to obtain. As online data collection may cause partially observable control problems, this study also incorporates a class of state parameterization method in the proposed controller in order to deal with partial observability. Finally, the proposed controller is demonstrated via a simulation of the longitudinal dynamics of a HFA model.

INDEX TERMS Integral reinforcement learning, optimal control theory, highly flexible aircraft, optimal tracking control.

I. INTRODUCTION

Some modern airplanes feature higher aspect ratios of the wings; these modern designs employ composite materials to reduce the weight of the fuselage, which helps improve aerodynamic efficiency, reduce fuel consumption, and ensure long-term operation. Such aircrafts are called highly flexible aircrafts (HFAs) due to their aforementioned characteristics. However, the special structure of HFAs typically leads to coupling effects between the structural and rigid dynamics. Moreover, large elastic deformations during flight result in unexpected difficulties when modeling aircraft dynamics, which, in turn, complicates the design of the control system [1]. In recent years, however, there have been significant developments in the modeling of HFAs. [2]–[4]. Integrated controllers capable of both rigid-body motion control and aeroelastic mode suppression, such as gust load alleviation and maneuver load alleviation, have been designed for HFAs [5]–[9]. Shearer and Cesnik [2] studied the trajectory tracking control of a very flexible aircraft, based on a dynamic model developed in UM/NAST. They separated the tracking problem into a bi-level control architecture including an inner loop and an outer loop. Considering the previous work by Shearer, Dillsaver [10] addressed the longitudinal trajectory tracking control problem in the presence of gust disturbance. Gibson [11] constructed a simplified aircraft model with three rigid wings in order to approximate the complex nonlinear characteristics of an HFA. Subsequently, a linear LQR/LTR controller was compared with an adaptive LQR/LTR controller for the stabilization of longitudinal flight dynamics [12]; the simulation results indicated that the linear LQR/LTR could not return the aircraft to the initial trim state, whereas the adaptive LQR/LTR controller could [13], [14]. Qu [15] developed an adaptive output-feedback controller for a class of multi-input multi-output linear plant models with a relative degree of three or higher and applied the controller to Gibson's HFA model. Furthermore, the work in [16] considered the combination of a linear quadratic regulator-proportional integral (LQR-PI) controller and an explicit reference governor in order to realize trajectory control with state and input constraints for the HFA. Apart from these previous methods, several adaptive tracking control methods have also been suggested recently for different situations such as switched systems, nonlinear networked systems, etc [17]–[23]. The work in [17] focused on solving some of the challenges faced by a class of uncertain switched nonlinear systems including arbitrary switchings, unmodeled dynamics, input saturation, unknown dead-zone output, etc. Then Ma in [18] continued to extend the adaptive fuzzy tracking control technology to a class of completely non-affine uncertain switched

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

pure-feedback nonlinear systems with unmeasurable states. Wang [24] developed a reliable fuzzy tracking control for Near-space hypersonic vehicle under aperiodic measurement information and stochastic actuator failures, and the tracking control problem was regarded as an optimization problem. Otherwise, the work in [25] concerned with the path following control for a robotic airship subject to sensor faults according to the developed detection and isolation mechanism.

It is also worth noting that the above-mentioned control schemes are essentially based on accurate dynamic models. However, designing accurate models for HFAs is highly challenging owing to the greater structural deformations and higher complex dynamics involved. Therefore, it is necessary to design an HFA controller that does not require the use of dynamic models. In recent years, the combination of reinforcement learning (RL) and optimal theory has gained extensive attention for the design of controllers for both linear and nonlinear systems. In control-related fields, RL is also referred as adaptive dynamic programming (ADP), which can be used to solve optimal control problems involving large or continuous state spaces [26]–[29]. RL is a machine learning technique developed in the field of computer science and engineering. It is closely related to optimal control and adaptive control. In [30], Sutton reported that RL is inherently a direct adaptive optimal control method. The primary concept in RL is to approximately solve the Hamilton Jacobi Bellman (HJB) equation through an iterative method, such as policy iteration (PI) and value iteration (VI) [31]. Furthermore, optimal control solutions can be derived by solving the HJB equation, where the HJB equation is converted to an algebraic Riccati equation (ARE) when considering linear systems. Based on this concept, Kleinman developed an offline PI algorithm for linear continuous-time (CT) systems, with a guaranteed convergence of the optimal solution [32]. Subsequently, Vrabie proposed integral reinforcement learning (IRL) for linear CT systems in order to obviate the requirement of drift dynamics [27]. Furthermore, Jiang *et al.* employed this IRL concept to propose an online model-free RL algorithm for completely unknown CT linear systems [33]. Qin extended the IRL technique to solve the optimal tracking control problem in CT linear systems [34]. Zhu *et al.* [35] introduced the IRL approach to develop an online solution for the suboptimal output feedback control of partially unknown linear CT systems. For solving the output feedback control problem, Rizvi proposed a state parametrization scheme to reconstruct the system state based on the input and output signals [36].

Inspired by these previous studies, this study proposes an online data-driven learning control scheme for tracking desired flight commands based on the aircraft model described by Gibson [11].

The major contributions of this paper include:

1) In order to solve the tracking problem with IRL technique, different from the existing results in [37], the integral of tracking error as a new state to construct an augmented system without requiring a discount factor in cost function. Then a full-state data-driven LQR-PI controller can be developed by employing IRL technique in order to iteratively solve the ARE without requiring a priori knowledge of the system.
2) The research in [35] deal with output tracking problem by introducing the recursive equation of L, but what obtained finally is sub-optimal solution. Herein, a state parameterization method is applied to combine the proposed model-free controller to get a optimal feedback gain of output tracking problem.
3) Aiming at the challenge that the dynamic knowledge of HFA system model with partial observability cannot be obtained, we apply the proposed model-free adaptive tracking controller to track the velocity of the aircraft.

The remainder of this paper is organized as follows. Section II provides an overview of the dynamics of an HFA and details the problem formulation. Section III presents the IRL-based model-free control for an HFA model with full state observations and partial observability. Extensive simulations of the control algorithm are presented in Section IV in order to demonstrate the effectiveness of the proposed method. Finally, conclusions of this study are highlighted in Section V.

II. PROBLEM FORMULATION AND PRELIMINARIES

A. AIRCRAFT DYNAMICS

The simplified HFA model proposed in [11] comprises three identical rigid-body wings, as shown in Fig. 1. Two adjacent wings are connected via hinges, while ailerons are connected behind each wing. Each rigid panel is equipped with a propeller for thrust, an aileron that runs along the aft of the main wing, and an elevator attached at the end of the boom.

A schematic of this aircraft, including the appropriate axes and points, is presented in Fig. 2. The control-oriented model considering the longitudinal dynamics of an HFA is expressed as follows:

$$\begin{cases} \dot{V} = (\mathcal{T}\cos\alpha - \mathcal{D})/m - g\sin\gamma \\ \dot{\alpha} = -(\mathcal{T}\sin\alpha + \mathcal{L})/(mV) + q + g\cos(\gamma/V) \\ \dot{h} = V\sin\gamma \\ \dot{\theta} = q \\ \dot{q} = \dfrac{\mathcal{M} - 2c_2\sin(\eta)\cos(\eta)\dot{\eta}q}{c_1 + c_2\sin^2(\eta)} \\ \ddot{\eta} = \dfrac{\mathcal{H} - \kappa_c\dot{\eta} - \kappa_k\eta + d_1 - d_2}{d_3} \end{cases} \tag{1}$$

where V is the velocity, γ is the flight path angle, h is the altitude, α is the attack angle, θ is the pitch angle with $\gamma = \theta - \alpha$, q is the pitch rate, $\mathcal{T}$ is the total thrust, and $\mathcal{D}$ is the drag force. $\mathcal{M}$ and $\mathcal{H}$ denote the total moment and angular moment, respectively, which are related to the control input. The parameters c_1, c_2, d_1, d_2 and d_3 are as follows:

$$c_1 = 3I_{yy}^*$$

$$c_2 = 2I_{zz}^* - 2I_{yy}^* + m^*\frac{s^2}{6}$$

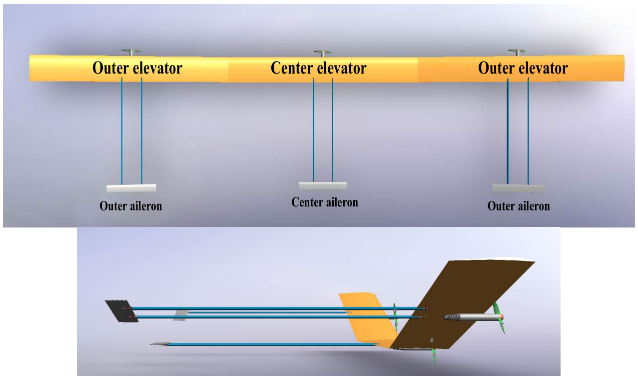


FIGURE 1. Rendering of HFA model.

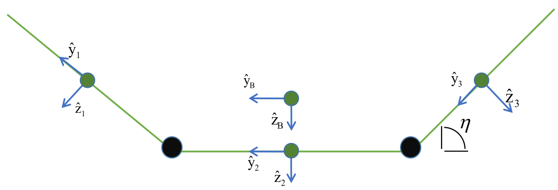


FIGURE 2. Coordinate frames.

$$\begin{aligned} d_1 &= \frac{s}{2}m^* \left((\dot{V}\sin(\alpha) + V\cos(\alpha)\dot{\alpha})\cos(\eta)\right. \\ &\quad \left. - V\sin(\alpha)\sin(\eta)\dot{\eta} - 2\frac{s}{3}\cos(\eta)\sin(\eta)\dot{\eta}^2\right) \\ d_2 &= \left(I^*_{yy} - I^*_{zz} - m^*\frac{s^2}{12}\right)\sin(\eta)\cos(\eta)q^2 \\ &\quad - \frac{s}{2}m^*\cos(\eta)V\cos(\alpha)q \\ d_3 &= I^*_{x_3x_3} + m^*\left(\frac{s^2}{4} + \frac{s^2}{6}\cos^2(\eta)\right) \end{aligned}$$

The nonlinear dynamics in (1) can be concisely expressed as

$$\dot{X} = f(X, U) \tag{2}$$

where $X = \left[V,\ \alpha,\ h,\ \theta,\ q,\ \eta,\ \dot{\eta}\right]^T$ and $U = \left[\delta_t,\ \delta_{a,c},\ \delta_{a,o},\ \delta_{e,c},\ \delta_{e,o}\right]^T$ denote the state vector and the input vector, respectively.

Remark 1: From (1), it is evident that this simplified three-wing aircraft model concentrates the large deformation characteristics at the hinges, which is reflected in the mathematical model as a change in the dihedral angle. The controllability analysis in [11] indicates that, when the control design does not employ an external control surface, the HFA is in the worst controllable state. As the dihedral angle is highly nonlinear with respect to the other states, it is necessary to guarantee the observability of this dihedral angle, in order to ensure quick and stable determination of the dihedral angles. Therefore, the observability of the dihedral angle is maintained in the subsequent controller design.

B. PRELIMINARIES

This section presents the definitions, lemmas, and propositions used in this study. Hereinafter, $\mathbb{R}$ is denoted by sets of real numbers; $\|\cdot\|$ denotes the Euclidean norm for vectors, or the induced matrix norm for matrices; the optimal value of M is denoted by N^*; the ith iteration of M is denoted by M_i; and the ith dimension component of vector M is denoted by M^i. For the matrix $A = [a_1\ a_2 \cdots a_n] \in \mathbb{R}^{m\times n}$, $vec(A)$ is defined as the new $m \times n$ vector formed by the columns of A, that is $vec(A) = [a_1^T\ a_2^T \cdots a_n^T]^T$ [38].

Definition 1: For the real symmetric matrix $P \in \mathbb{R}^{n\times n}$, the new vector $\tilde{P} \in \mathbb{R}^{\frac{1}{2}n(n+1)\times 1}$ is defined as

$$\tilde{P} = \left[p_{11}, 2p_{12}, \ldots, 2p_{1n}, p_{22}, 2p_{23}, \ldots, 2p_{n-1,n}, p_{nn}\right]^T \tag{3}$$

where $P = \begin{bmatrix} p_{11} & p_{12} & \cdots & p_{1n} \\ p_{21} & p_{22} & \cdots & p_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ p_{n1} & p_{n2} & \cdots & p_{nn} \end{bmatrix}$.

Definition 2: For vector $x = [x_1\ x_2 \cdots x_n] \in \mathbb{R}^{n\times 1}$, the new vector $\hat{x} \in \mathbb{R}^{\frac{1}{2}n(n+1)\times 1}$ is defined as

$$\hat{x} = \left[x_1^2, x_1x_2, \ldots, x_1x_n, x_2^2, x_2x_3, \ldots, x_{n-1}x_n, x_n^2\right]^T \tag{4}$$

It should be noted that the mappings defined in Definition 1 and Definition 2 have a one-to-one correspondence with each other.

Definition 3: [39] Kronecker product: For matrices $A \in \mathbb{R}^{m\times n}$ and $B \in \mathbb{R}^{p\times q}$, the Kronecker product M for A and B is defined as follows:

$$M = A \otimes B = \begin{bmatrix} a_{11}B & \cdots & a_{1n}B \\ \vdots & \ddots & \vdots \\ a_{m1}B & \cdots & a_{mn}B \end{bmatrix} \tag{5}$$

According to the above-mentioned definitions, an important identity required for subsequent analyses is proposed as follows:

$$vec(ADB) = (B^T \otimes A)\, vec(D) \tag{6}$$

Remark 2: In order to make the least square calculation in IRL technology more convenient, it is necessary to give the above three definitions. Through these definitions, we transform the calculation between matrices into calculation between vectors, which will contribute to the next work.

Lemma 1: Consider the following linear time invariant CT plant.

$$\begin{aligned} \dot{x} &= Ax + Bu \\ y &= Cx \end{aligned} \tag{7}$$

where $A \in \mathbb{R}^{n\times n}$, $B \in \mathbb{R}^{n\times m}$, $C \in \mathbb{R}^{p\times n}$, and $x \in \mathbb{R}^{n\times 1}$ is the system state; $u \in \mathbb{R}^{m\times 1}$ is the system control input, and $y \in \mathbb{R}^{p\times 1}$ is the system output state. Initializing $K_0 \in \mathbb{R}^{m\times n}$ as any stabilizing feedback gain matrix and assuming P_i as the symmetric positive definite solution of the Lyapunov equation, we obtain

$$(A - BK_i)^T P_i + P_i (A - BK_i) + Q + K_i^T RK_i = 0 \tag{8}$$

where $K_k = R^{-1}B^T P_{k-1}$, with $k = 1, 2, \ldots, n$. Thereafter, the following properties hold: (i) $A-BK_i$ is Hurwitz; (ii) $P^* \leq P_{i+1} \leq P_i$; and (iii) $\lim_{k\to\infty} K_i = K^*$, $\lim_{k\to\infty} P_i = P^*$.

Proof: See [32] for the proof.

Remark 3: As reported in [32], on repeatedly calculating the Lyapunov equation 8 and iterating the feedback gain K, K converges to the optimal feedback gain K^*. This is the earliest idea of iterative integration in optimal control. Subsequent analyses are based on this concept.

Lemma 2: Consider the linear CT system described in (7). The state parametrization is defined as

$$\bar{x}(t) = M_u\zeta_u(t) + M_y\zeta_y(t) \tag{9}$$

where $\zeta_u(t) = \left[\left(\zeta_u^1\right)^T(t) \;\; \left(\zeta_u^2\right)^T(t)\cdots\left(\zeta_u^m\right)^T(t)\right]^T$ and $\zeta_y(t) = \left[\left(\zeta_y^1\right)^T(t) \;\; \left(\zeta_y^2\right)^T(t)\cdots\left(\zeta_y^p\right)^T(t)\right]^T$. Moreover, $\zeta_u(t)$ and $\zeta_y(t)$ are constructed as follows:

$$\begin{aligned}\dot{\zeta}_u^i(t) &= \mathcal{A}\zeta_u^i(t) + bu_i(t), \quad \forall i = 1, 2, \ldots, m\\ \dot{\zeta}_y^i(t) &= \mathcal{A}\zeta_y^i(t) + by_i(t), \quad \forall i = 1, 2, \ldots, p\end{aligned} \tag{10}$$

Subsequently, for any desired Hurwitz matrix $\mathcal{A}$ and $b = [0 \;\; 0\cdots 1]_{1\times n}^T$, $\zeta_u^i(0) = 0$ and $\zeta_y^i(0) = 0$. Thus, we obtain $\lim_{t\to\infty}\|\bar{x}(t) - x(t)\| = 0$.

Proof: See [36] for the proof.

Remark 4: In [36], a filtering-based observer was proposed for parameterizing the state vector in terms of filtered inputs and outputs. Therefore, when the system state cannot be measured precisely, it can instead be decomposed into a linear weighted sum of the filtered inputs and outputs, by means of state parameterization.

III. CONTROL DESIGN

This section discusses the control design to illustrate the primary concept of the proposed controller. First, we perform linearization analyses for different shapes of the HFA nonlinear model. Subsequently, we propose a model-free data-driven optimal tracking controller based on the linear time invariant CT plant. Finally, we integrate a state parameterization method in the proposed controller to address the issue of incomplete state acquisition.

A. HFA SYSTEM ANALYSIS AND CONTROL STRUCTURE

This section discusses the trim analysis for the HFA model. Assuming that small deviations occur at higher altitudes and using the deflection angles $\delta_{a,o}$ and $\delta_{e,c}$ as the control inputs, the other control surfaces, i.e., δ_t, $\delta_{a,c}$, and $\delta_{e,o}$, should be set as a constant based on their steady-state value. Thus, the linear plant can be expressed as

$$\begin{aligned}\dot{x}_p &= A_p x + B_p u\\ y_p &= C_p x\end{aligned} \tag{11}$$

where the state vector and input vector can be rewritten as

$$x_p = [V \;\; \alpha \;\; \theta \;\; q \;\; \eta \;\; \dot{\eta}]^T$$

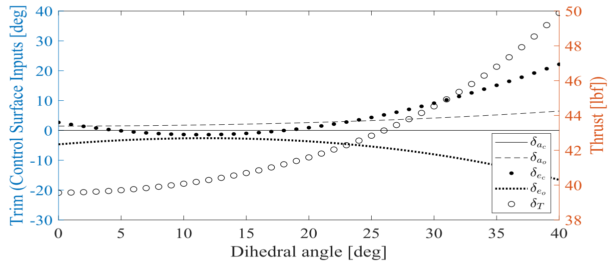
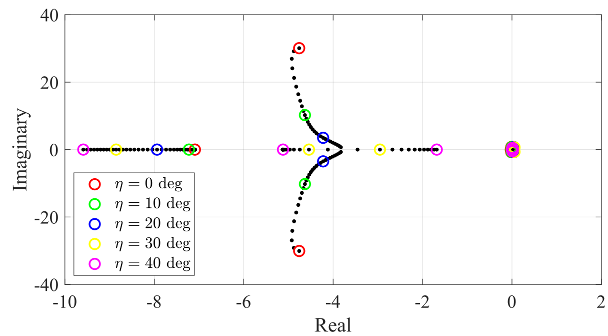
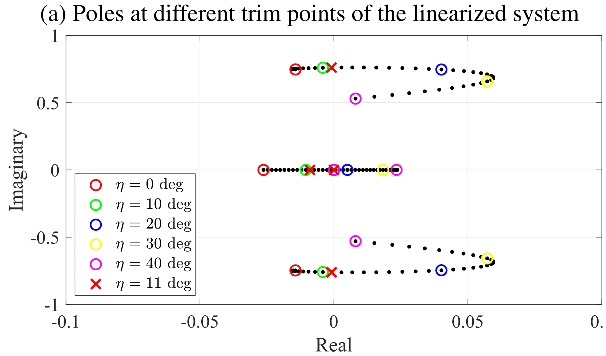


FIGURE 3. Control inputs at different trim points of the linearized plant.



(a) Poles at different trim points of the linearized system



(b) Dominant poles at different trim points of the linearized system

FIGURE 4. (a) Poles at different trim points of the linearized system. (b) Dominant poles at different trim points of the linearized system.

$$\begin{aligned}u_p &= [\delta_{a,o} \;\; \delta_{e,c}]^T\\ y_p &= V\end{aligned} \tag{12}$$

The HFA model in (2) is linearized at $V = 68ft/s$ and $h = 40,000ft$, with the dihedral angle ranging from 0° to 40° in increments of 1°. The variations in the trajectory curve of the control input with respect to the dihedral angle are depicted in Fig. 3, and the evolution of the poles is shown in Fig. 4. Based on these figures, it is evident that the stability of the system gradually decreases as the dihedral angle increases and that an unstable pole begins to appear at $\eta = 11°$.

B. DATA-DRIVEN OPTIMAL TRACKING CONTROL WITH COMPLETELY UNKNOWN DYNAMICS

First, the control objective is defined as "the system output y_p asymptotically tracks the reference signal y_{cmd}." The tracking error is defined as

$$e(t) = y(t) - y_{cmd}(t) \tag{13}$$

The integral of the tracking error can be calculated as

$$e_I(t) = \int e(t) = \int (y(t) - y_{cmd}(t)) \tag{14}$$

Then, the augmented system can be expressed as

$$\underbrace{\begin{bmatrix} \dot{x}_p(t) \\ \dot{e}_I(t) \end{bmatrix}}_{\dot{x}(t)} = \underbrace{\begin{bmatrix} A_p & 0 \\ C_p & 0 \end{bmatrix}}_{A} \underbrace{\begin{bmatrix} x_p(t) \\ e_I(t) \end{bmatrix}}_{x(t)} + \underbrace{\begin{bmatrix} B_p \\ 0 \end{bmatrix}}_{B} u(t) + \underbrace{\begin{bmatrix} 0 \\ -I \end{bmatrix}}_{B_{ref}} y_{cmd}$$

$$y = (C_p \ \ 0) \begin{pmatrix} x_p(t) \\ e_I(t) \end{pmatrix} \tag{15}$$

The linear quadratic method can be used with the proportional-integral feedback connection to obtain the optimal control law. First, we define the following LTI plant:

$$\dot{z} = Az + Bv \tag{16}$$

where $z = \dot{x} = \begin{pmatrix} \dot{x}_p(t) \\ \dot{e}_I(t) \end{pmatrix} \in \mathbb{R}^{(n+p)\times 1}$ and $v = \dot{u}$.

In order to derive the primary results of this study, the following assumptions need to be employed. Considering the linear plant described in (11) and that the reference command is denoted by y_{cmd}, we have

Assumption 1. (A, B) is controllable and A can be stabilized.

Assumption 2. (A, C) is observable.

Assumption 3. The tracking commands y_{cmd} are bounded and constant.

Remark 5: Assumption 1 is universally accepted, and Assumption 2 is ubiquitous for a majority of the practical systems in aerospace and other similar industries, where the outputs (sensors) can be defined by vehicle designers and are placed at specific locations on the vehicle to achieve the desired input–output characteristics [40]. In Assumption 3, a bounded constant signal can generally be selected as the reference command. Therefore, the above-mentioned assumptions for control design are necessary, while also being reasonable.

Thus, the value function or cost function J is given as

$$J = \int_0^\infty (z^T Qz + u^T Ru)dt \tag{17}$$

where $Q = Q^T \geq 0$ and $R = R^T > 0$, with $(A, Q^{1/2})$ being observable. Clearly, the optimal LQR solution in the feedback form is

$$v = \dot{u} = -R^{-1}B^T Pz = -(K_P \ \ K_I) \begin{pmatrix} \dot{x}_p(t) \\ \dot{e}_I(t) \end{pmatrix} \tag{18}$$

where P, which is termed as the cost matrix, is a unique, symmetric positive definite solution of the following ARE:

$$A^T P + PA + Q - PBR^{-1}B^T P = 0 \tag{19}$$

(18) is integrated to obtain the LQR-PI controller:

$$\begin{aligned} u &= -(K_P \ \ K_I) \begin{pmatrix} x_p(t) \\ e_I(t) \end{pmatrix} \\ &= -K_P x_p - K_I e_I = -K_P x_p - \frac{K_I(y - y_{cmd})}{s} \end{aligned} \tag{20}$$

Remark 6: From (20), it is evident that the control input consists of a linear feedback term related with the state as well as an additional feedforward term related with the integral of tracking error. The gain matrices K_P and K_I depend on the solution of the ARE from (19). However, it is necessary to possess accurate knowledge regarding dynamics in order to determine the solution of (19), which can be considerably challenging in practical scenarios. Therefore, subsequent sections discuss measures to circumvent this problem.

Lemma 1 proposes an iterative method to approximately solve P; however, the algorithm is implemented offline and requires knowledge regarding the system drift matrix A_p and the control matrix B_p. Inspired by [27] and [33], we construct a novel data-driven tracking controller, without requiring internal system information (A_p, B_p). For this purpose, the following value function was used:

$$\begin{aligned} V(t) &= \int_t^\infty \left(z_\tau^T Q z_\tau + u^T Ru \right) d\tau \\ &= \int_t^\infty \left(z_\tau^T (Q + K^T RK) z_\tau \right) d\tau \\ &= z_t^T P z_t \end{aligned} \tag{21}$$

Substituting $V(t) = z_t^T P_k z_t$ and $V(t + \delta_T) = z_{t+\delta T}^T P_k z_{t+\delta T}$ in the above-mentioned equation, we obtain

$$\begin{aligned} z_t^T P_k z_t - z_{t+\delta T}^T P_k z_{t+\delta T} &= \int_t^{t+\delta T} \left(z_\tau^T Q z_\tau + u^T Ru \right) d\tau \\ &= \int_t^{t+\delta T} \left(z_\tau^T (Q + K^T RK) z_\tau \right) d\tau \end{aligned} \tag{22}$$

First, we assume that the initial stabilizing feedback matrix K_0 is known. Based on Lemma 1, the iterative feedback gain matrix $K_{i+1} = R^{-1}B^T P_i$. Thus, the plant in (16) can be rewritten as

$$\dot{z} = A_k z + B(K_i x + v) \tag{23}$$

where $A_k = A - BK_k$.

Subsequently, according to Lemma 1, we have

$$\begin{aligned} & z_t^T P_i z_t - z_{t+\delta T}^T P_i z_{t+\delta T} \\ &= \int_t^{t+\delta T} \left[z_\tau^T (A_k^T P_i + P_i A_k) z_\tau + 2(v + K_i x)^T B^T P_i x \right] d\tau \\ &= -\int_t^{t+\delta T} z_\tau^T Q_i z_\tau d\tau + 2 \int_t^{t+\delta T} (v + K_i x)^T RK_{i+1} x d\tau \end{aligned} \tag{24}$$

where $Q_i = Q + K_i^T RK_i$. (24) shows that the matrices P_i and K_i can be iteratively solved, without requiring information regarding (A_p, B_p). Using the Kronecker product representation and the previous definitions, we obtain

$$\begin{aligned} z_t^T P_i z_t &= \hat{z}_t^T \tilde{P}_i \\ z^T Q_i z &= (z^T \otimes z^T)\, vec(Q_i) \\ (v + K_i z)^T R K_{i+1} z &= [(z^T \otimes z^T)(I_n \otimes K_i^T R) \\ &\quad + (z^T \otimes v^T)(I_n \otimes R)]\, vec(K_{i+1}) \end{aligned} \tag{25}$$

For $t \in [t_1, t_l]$, we divide it into several small time intervals $t \in [t_1, t_1+\delta_T] \cup [t_1+\delta_T, t_1+2\delta_T] \cup \cdots \cup [t_l-\delta_T, t_l]$, with $0 \le t_1 < t_2 < \cdots < t_l$, and $t_l = t_1 + N\delta_T$. Therefore, we define data matrices $\delta_{zz} \in \mathbb{R}^{l\times\frac{1}{2}(n+p)(n+p+1)}$, $I_{zz} \in \mathbb{R}^{l\times(n+p)^2}$, and $I_{zv} \in \mathbb{R}^{l\times\frac{1}{2}m(n+p)}$ as follows:

$$\begin{aligned} \delta_{zz} &= [\hat{z}(t_1+\delta_T) - \hat{z}(t_1), \hat{z}(t_1+2\delta_T) - \hat{z}(t_1+\delta_T), \\ &\quad \ldots, \hat{z}(t_1) - \hat{z}(t_1-\delta_T)]^T, \\ I_{zz} &= [\int_{t_1}^{t_1+\delta_T} z\otimes z d\tau, \int_{t_1+\delta_T}^{t_1+2\delta_T} z\otimes z d\tau, \ldots, \int_{t_l-\delta_T}^{t_l} z\otimes z d\tau]^T, \\ I_{zv} &= [\int_{t_1}^{t_1+\delta_T} z\otimes v d\tau, \int_{t_1+\delta_T}^{t_1+2\delta_T} z\otimes v d\tau, \ldots, \int_{t_l-\delta_T}^{t_l} z\otimes v d\tau]^T, \end{aligned}$$

Substituting (25) and δ_{zz}, I_{zz}, and I_{zv} in (24), we obtain

$$\Psi_i \begin{bmatrix} \tilde{P}_i \\ vec(K_{i+1}) \end{bmatrix} = \Phi_i \tag{26}$$

where $\Psi_i = [\delta_{zz}, -2I_{zz}(I_n \otimes K_i^T R) - 2I_{zv}(I_n \otimes R)] \in \mathbb{R}^{l\times[\frac{1}{2}(n+p)(n+p+1)+m(n+p)]}$, and $\Phi_i = -I_{zz}\, vec(Q_i) \in \mathbb{R}^l$. Thus, the iterative matrices P_i and K_i can be calculated from (26). We can summarize the above-mentioned discussion as Algorithm 1.

Algorithm 1 Online Model-Free Tracking Control With Full State Feedback

---

Require:
Initialize the stable control gain K_0, exploration noise e, reference signal y_{cmd}, and expected error ϵ
Consider the plant described in (15)
Let $v = -K_0 z + e$, $t \in [t_1, t_l]$, and compute δ_{zz}, I_{zz}, and I_{zv}
if $\left\|\tilde{P}_{i+1} - \tilde{P}_i\right\| \ge \epsilon$ **then**
  (Policy evaluation) Calculate $\tilde{P}_i$ and $vec(K_{i+1})$ from (26) until $\left\|\tilde{P}_{i+1} - \tilde{P}_i\right\| \le \epsilon$
end if
Calculate K_i from $vec(K_i)$
Calculate optimal feedback control law $v = -Kz = -K_i z$
Calculate optimal feedback control law $u = \int -Kz = -K\begin{pmatrix} x_p(t) \\ e_I(t) \end{pmatrix}$

Remark 7: To ensure that (26) has a suitable solution, the rank condition of $[I_{zz} \;\; I_{zv}]$ should satisfy $rank([I_{zz} \;\; I_{zv}]) = \frac{(n+p)(n+p+1)}{2} + m(n+p)$. The core of Algorithm 1 is aimed at collecting data during the time interval $t \in [t_1, t_l]$; the choice of exploration noise plays a vital role during this process. The exploration noise affects the convergence speed of the calculation process and also determines whether the rank conditions are met. Typically used noises include sinusoidal signals with random amplitudes and frequencies.

The following theorem proves that Algorithm 1 converges to the optimal feedback gain K^*:

Theorem 1: For the linear differential system in (16), initialize the feedback gain matrix K_0, when the rank condition of $[I_{zz} \;\; I_{zv}]$ satisfies $rank([I_{zz} \;\; I_{zv}]) = \frac{(n+p)(n+p+1)}{2} + m(n+p)$. By iteratively calculating (26), we have $\lim_{i\to\infty} \|P_i - P^*\| = 0$, and $\lim_{i\to\infty} \|K_i - K^*\| = 0$.

Proof: First, we prove that, when the rank condition is satisfied, Ψ_i possesses a full column rank for $i = 1, 2, 3, \cdots$. In (26), Ψ_i has a full column rank and is equivalent to the following equation, which has the unique solution x:

$$\Psi_i x = \Phi_i \tag{27}$$

Therefore, we assume that the solution $x = [\tilde{P}_i^T \;\; vec(K_{i+1})^T]$, where $\tilde{P}_i \in \mathbb{R}^{\frac{(n+p)(n+p+1)}{2}}$ and $vec(K_{i+1}) \in \mathbb{R}^{m(n+p)}$. According to Definitions 1 and 2, there exists a symmetric matrix $P_i \in \mathbb{R}^{(n+p)(n+p)}$ and a matrix $K_{i+1} \in \mathbb{R}^{m\times(n+p)}$. Substituting Ψ_i, we obtain

$$\begin{aligned} &[\delta_{zz}, -2I_{zz}(I_n \otimes K_i^T R) - 2I_{zv}(I_n \otimes R)]x \\ &\quad = [\delta_{zz}, -2I_{zz}(I_n \otimes K_i^T R) - 2I_{zv}(I_n \otimes R)] \\ &[\tilde{P}_i^T \;\; vec(K_{i+1})^T] = \Phi_i \end{aligned} \tag{28}$$

According to the Kronecker product representation and (24), $z_t^T P_i z_t - z_{t+\delta T}^T P_i z_{t+\delta T}$ can be rewritten as

$$\begin{aligned} &z_t^T P_i z_t - z_{t+\delta T}^T P_i z_{t+\delta T} \\ &\quad = (z_{t+\delta T}^T \otimes z_{t+\delta T}^T) vec(P_i) - (z_t^T \otimes z_t^T) vec(P_i) \end{aligned} \tag{29}$$

Considering that the time interval δ_T is considerably small, the above equation can be approximately extended as follows:

$$\begin{aligned} &(z_{t+\delta T}^T \otimes z_{t+\delta T}^T) vec(P_i) - (z_t^T \otimes z_t^T) vec(P_i) \\ &\quad = \delta_T \int_t^{t+\delta_T} z \otimes z d\tau \; vec(P_i) \end{aligned} \tag{30}$$

Thus, equation (27) can be rewritten as

$$\begin{aligned} &[\delta_{zz}, -2I_{zz}(I_n \otimes K_i^T R) - 2I_{zv}(I_n \otimes R)][\tilde{P}_i^T \;\; vec(K_{i+1})^T] \\ &= [\delta_T I_{zz}, -2I_{zz}(I_n \otimes K_i^T R) - 2I_{zv}(I_n \otimes R)] \\ &\quad [vec(P_i)^T \;\; vec(K_{i+1})^T] = \Phi_i \end{aligned} \tag{31}$$

As $rank([I_{zz} \;\; I_{zv}]) = rank([I_{zz}, -2I_{zz}(I_n \otimes K_i^T R) - 2I_{zv}(I_n \otimes R)])$, $rank(\Psi_i)$ has a full column rank. Therefore, (28) has a unique solution. Second, we prove that, for the initial stable K_0, the iterative equation (26) is equivalent to (8) in Lemma 1. For a stabilized feedback gain matrix K_i, when $P_i = P_i^T$ is the solution of (8), we obtain $K_{i+1} R^- B^T P_i$, which also satisfies (26). On the contrary, assuming $P_i = P_i^T \in \mathbb{R}^{(n+p)(n+p)}$ and $K_{i+1} \in \mathbb{R}^{m(n+p)}$, the following equation has a unique solution:

$$\Psi_i \begin{bmatrix} \tilde{P}_i \\ vec(K_{i+1}) \end{bmatrix} = \Phi_i$$

Hence, for the initial stable K_0 and $K_{i+1} = R^{-1} B^T P_i$, the iterative equation (26) is equivalent to (8) in Lemma 1.

Based on the conclusion of Lemma 1, we obtain $\lim_{i\to\infty}\|P_i - P^*\| = 0$ and $\lim_{i\to\infty}\|K_i - K^*\| = 0$.

Remark 8: The optimal feedback gain matrix of the augmented system in (15) can be applied to the controlled system, and the LQR-PI controller can be realized after integrating (20). Unlike previous optimal controller designs, the proposed optimal tracking controller does not require any system information.

A schematic of the proposed closed-loop framework in Algorithm 1 is presented in Fig. 5. After initializing with a stable control policy, the online full state information of the HFA model subjected to exploration noise can be obtained within $t \in [t_0, t_l]$. Thereafter, the state gain matrix of the system in (16) is determined via policy evaluation and policy improvement using (26). When the final requirements are satisfied, the iterations are ceased. Subsequently, we apply the gain to the HFA. It is evident that the gain matrix is divided into two components: the state feedback term K_p and the feedforward term for the tracking error K_I.

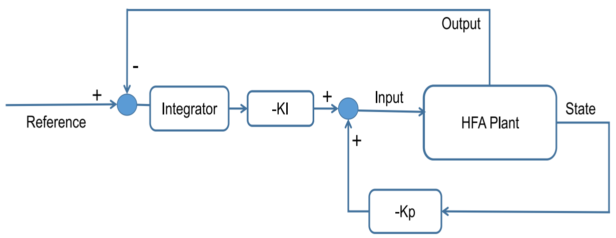


FIGURE 5. Closed-loop control structure of the system in Algorithm1.

C. DATA-DRIVEN OPTIMAL OUTPUT-FEEDBACK TRACKING CONTROL WITH STATE PARAMETRIZATION

Inspired by [36], we consider embedding a method of state parameterization into the controller designed using Algorithm 1, in order to ensure that the system can asymptotically track reference signals under output feedback. For the linear CT plant, we define the new state $\bar{x}_p$ as

$$\bar{x}_p(t) = M_u\zeta_u(t) + M_y\zeta_y(t) \tag{32}$$

Thus, we have $x_p(t) = \bar{x}_p(t) + \sigma(t)$, where $\sigma(t) < \|\sigma\|$ is a small constant. Thereafter, the new integral state $\bar{z} \in \mathbb{R}^{(mn+pn+p)\times 1}$ is defined as

$$\bar{z}(t) = \begin{bmatrix} \dot{\zeta}_u(t) \\ \dot{\zeta}_y(t) \\ \dot{e}_I(t) \end{bmatrix} \tag{33}$$

$$z(t) = \begin{pmatrix} \dot{x}_p(t) \\ \dot{e}_I(t) \end{pmatrix} = \begin{bmatrix} M_u & M_y & 0 \\ 0 & 0 & I \end{bmatrix} \begin{bmatrix} \dot{\zeta}_u(t) \\ \dot{\zeta}_y(t) \\ \dot{e}_I(t) \end{bmatrix} + \begin{bmatrix} \sigma \\ 0 \end{bmatrix} \tag{34}$$

Neglecting σ, the cost function in (21) can be described as

$$\begin{aligned} V &= z^T P z \\ &= \begin{bmatrix} \dot{\zeta}_u(t) \\ \dot{\zeta}_y(t) \\ \dot{e}_I(t) \end{bmatrix}^T \begin{bmatrix} M_u & M_y & 0 \\ 0 & 0 & I \end{bmatrix}^T P \begin{bmatrix} M_u & M_y & 0 \\ 0 & 0 & I \end{bmatrix} \begin{bmatrix} \dot{\zeta}_u(t) \\ \dot{\zeta}_y(t) \\ \dot{e}_I(t) \end{bmatrix} \\ &= \bar{z}^T \bar{P} \bar{z} \end{aligned} \tag{35}$$

where $\bar{P} = \bar{P}^T = \begin{bmatrix} M_u & M_y & 0 \\ 0 & 0 & I \end{bmatrix}^T P \begin{bmatrix} M_u & M_y & 0 \\ 0 & 0 & I \end{bmatrix} \in \mathbb{R}^{(mn+pn+p)\times(mn+pn+p)}$. For simplicity, we denote $\begin{bmatrix} M_u & M_y & 0 \\ 0 & 0 & I \end{bmatrix}^T$ by $\bar{M}$.

Thus, the control law of the augmented system in (15) is expressed as

$$\bar{v} = -\bar{K}\bar{z} \tag{36}$$

where $\bar{K} = K\bar{M} \in \mathbb{R}^{m\times(mn+pn+p)}$.

Substituting $\bar{K}$, $\bar{P}$, and $\bar{z}$ into the ADP equation in (24), we obtain

$$\begin{aligned} &\bar{z}_t^T \bar{P}_i \bar{z}_t - \bar{z}_{t+\delta T}^T \bar{P}_i \bar{z}_{t+\delta T} \\ &= -\int_t^{t+\delta T} \bar{z}_\tau^T \bar{Q}_i \bar{z}_\tau d\tau + 2\int_t^{t+\delta T} (v + \bar{K}_i x)^T R \bar{K}_{i+1} x d\tau \end{aligned} \tag{37}$$

where $\bar{Q}_i = \bar{M}^T Q_i \bar{M} \in \mathbb{R}^{(mn+pn+p)\times(mn+pn+p)}$. (37) shows that the subsequent optimal feedback matrix can be also be designed according to the numerical method in Algorithm 1; this is not repeated here. Thus far, the complete design of the data-driven optimal output-feedback tracking controller has been discussed. This design process is summarized in Algorithm 2.

Algorithm 2 Online Data-Driven Optimal Output-Feedback Tracking Control With IRL

Require:

Initialize the stable control gain K_0, exploration noise e, reference signal y_{cmd}, and expected error ϵ.

Calculate offline state parameterization matrices M_u and M_y. Considering the plant described in (15), let $\bar{v} = -\bar{K}_0\bar{z} + e$, $t \in [t_0, t_1]$ and compute $\delta_{\bar{z}\bar{z}}$, $I_{\bar{z}\bar{z}}$, and $I_{\bar{z}\bar{v}}$.

if

$\left\|\tilde{P}_{i+1} - \tilde{P}_i\right\| \geq \epsilon$ **then**

(Policy evaluation) Calculate $\tilde{P}_i$ and $vec(\bar{K}_{i+1})$ from (26) until $\left\|\tilde{P}_{i+1} - \tilde{P}_i\right\| \leq \epsilon$

end if

Calculate $\bar{K}_{i+1}$ from $vec(\bar{K}_{i+1})$

Calculate optimal feedback control law $v = -\bar{K}\bar{z} = -\bar{K}_{i-+1}\bar{z}$

Theorem 2: On the basis of above assumptions and an appropriate exploration noise e, for the system described in Eq. 16, a stable gain matrix K_0 is initialized, and the new gain matrix K_i is updated using Algorithm 2. Thus, one can obtain $\lim\limits_{i\to\infty}\|P_i - P^*\| = 0$ and $\lim\limits_{i\to\infty}\|K_i - K^*\| = 0$.

Proof: From Lemma 2, the convergence of the proposed state parametrization, which is based on filtered input and output signals, has been proved. The estimated state $\bar{x}_p$ quickly tracks x_p with an exponential convergence rate. Moreover, Theorem 1 completes the proof of the proposed online data-driven tracking controller with full state measurement which is depicted in Algorithm 1. Therefore, the proposed

ADP equation in Algorithm 2 converges to the equation in Algorithm 1, under sufficient excitation and the appropriate exploration noise e. In this manner, the problem to be solved is transformed into the problem in Theorem 1. Thus, the proof is completed.

Remark 9: Unlike Algorithm 1, for Algorithm 2, we use filtered input and output data information ($\delta_{\bar{z}\bar{z}}$, $I_{\bar{z}\bar{z}}$ and $I_{\bar{z}\bar{v}}$) to calculate the required parameters within the time interval $t \in [t_0, t_l]$. Thereafter, we calculate the optimal feedback gain K of the differential system by solving the least squares problem corresponding to (37). Finally, the obtained gain is employed in the augmented system along with the tracking error term; the controller ensures that the system output asymptotically tracks the reference signal.

Remark 10: It is worthy to note that some literatures have recently provided theoretical analysis on the robustness of such adaptive data-driven algorithms based on IRL methods, such as [41]. This paper pointed out that this type of algorithm has a good ability to deal with small interference, but it is difficult to guarantee its performance for large external interference. Therefore, the proposed online controller in our research based on the IRL technique can also show good robustness.

In-line with the previous discussions, we also present the block diagram of the closed-loop control system obtained using Algorithm 2 in Fig. 6.

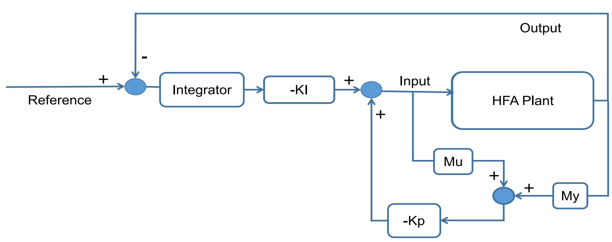


FIGURE 6. Closed-loop control structure of the system in Algorithm2.

Remark 11: Both Algorithm 1 and Algorithm 2 are constructed with PI algorithm, therefore, there is a restrictive assumption that an initial stable control strategy is required. If the system is known to be initially stable, the initial control policy can be selected as 0. For the initial instability of the system, this problem can be overcome through the VI algorithm, although this method requires a large number of iterations.

Remark 12: Note that the proposed adaptive tracking controller does not rely on dynamic knowledge, which is different from the previous controller. It can be seen that both two algorithms contain two main phases. In the first place, an initial stabilizing control strategy under proper exploration noise is injected into the defined augmented system and then the system information is recorded in matrices in δ_{zz}, I_{zz} and I_{zv} ($\delta_{\bar{z}\bar{z}}$, $I_{\bar{z}\bar{z}}$ and $I_{\bar{z}\bar{v}}$). Hereafter, the obtained matrices are applied to calculate the approximate optimal control policy by (26). In the overall control system, the choice of exploration noise has a great influence on the control effect. Therefore, choosing a suitable exploration noise is also one of the difficulties of the control design, especially for high-dimensional systems with more complicated situations. For different systems, we usually need to use different types of exploration noise such as sinusoidal noise, random noise, or exponential noise, etc. In this research, we choose the sum of sinusoidal noises of different frequencies as the exploration noise of the control input.

TABLE 1. Model parameters of the HFA.

Sectional Mass	m^*	100 slugs
Sectional Inertial	I^*_{xx}	200
	I^*_{yy}	20
	I^*_{zz}	160
Sectional Span	s	80 ft
Gravitational Acceleration	g	32.2 ft / s^2
Dihedral Joint Damping	κ_c	140000 lbf / s
Dihedral Joint Stiffness	κ_k	4900 lbf

IV. NUMERICAL EXPERIMENTS AND RESULTS

The overall tracking performance of the proposed controller is applied to the three-wing HFA model. HFAs suffer from highly undesirable, and even unknown, internal dynamics. Although several studies have focused on HFAs, accurate and precise knowledge regarding aircraft dynamics remains elusive. Therefore, this study focuses on a model-free control design based on the online acquisition of data information. Details of the HFA model presented in Section II are listed in Table 1. Although the HFA model is nonlinear under normal operating conditions, a linear model can be employed to develop optimal controllers. The aircraft is linearized at $V = 68ft/s$, $h = 40,000ft$, $\alpha = 2.8°$, $\theta = 2.8°$, $\eta = 5°$, and $\dot{\eta} = 0$. Refer to Appendix A for detailed explanations regarding the matrix. First, the proposed online data-driven adaptive optimal controller employing IRL is applied to the HFA model under full state measurements. The calculation of the gain matrix is divided into two tasks. The first task involves collecting online data by applying exploration noise to the system within a certain time period to calculate the optimal feedback gain; thereafter, the gain matrix is calculated iteratively using (26), and this continues until the error between two iterations is within the expected range. The second task involves conducting a second simulation to validate the state parameterization in order to address the problem that the system state cannot be precisely measured. Finally, a baseline LQR-PI controller is employed for a comparison of tracking performances and to verify the effectiveness of the proposed controller. Throughout the simulation, the choice of exploration noise e is highly critical. Thus, after several trials, we reasonably selected e as

$$e = 1000 \sum_{i=1}^{200} \sin(w_i t) \tag{38}$$

where w_i are randomly selected from $[-200, 200]$.

TABLE 2. Control parameters.

Parameter	Value
Q	diag([1 1 1 1 1 1 1])
R	diag([1 1])
N	200
ϵ	10^{-10}
σ_{max}	50
P_0	$10I$

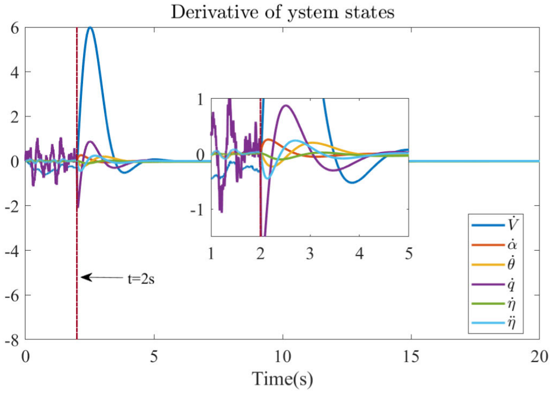


FIGURE 7. Derivative of system state using (16) in Algorithm 1.

A. ONLINE DATA-DRIVEN TRACKING CONTROLLER WITH FULL STATE MEASUREMENTS

The framework's capability of online learning without a priori knowledge of the controller is demonstrated herein. The parameters required for the simulation of Algorithm 1 are listed in Table 2. The weight matrices are denoted by Q and R, respectively. The learning time is denoted by N, which is set to 200. The iterations continue until the maximum number of iterations satisfies $\sigma \geq \sigma_{max}$ or until the difference in P_i for two iterations satisfies $\|P_i - P_{i-1}\| \leq \epsilon$.

The control object is to ensure that the system velocity V tracks the reference signal r under the application of the controller. The initial velocity is chosen as $V = 1$, and the reference signal is chosen as $y_{cmd} = 5$. The state input information of the differential system in (16) is collected during the learning time at $t \in [0, 2]$, where each time interval $\delta_T = 0.01$ s. The control parameters are initialized to zero. Figs. 7 to 9 illustrate the process of the online collection of data for (16) and the calculation of the optimal gain matrices K^* and P^*. Fig. 7 shows that the data matrices δ_{zz}, I_{zz}, and I_{zv} are collected in $t \in [0, 2]$. Thereafter, P_i and K_{i+1} are iteratively updated via (26) until the desired conditions are satisfied. The update processes of the cost matrix P and the gain matrix K are illustrated in Fig. 9; these converge to the steady value. Furthermore, the final matrices P^* and K^* are displayed in Appendix B. Fig. 8 shows that the tracking error converges to 0.

Once the convergence criterion of the cost matrix P_i is met, the exploration noise e can be stopped. We apply the feedback gain matrix K_i, obtained using (26), to the HFA model. The tracking performance and control signals are depicted in Fig. 11.

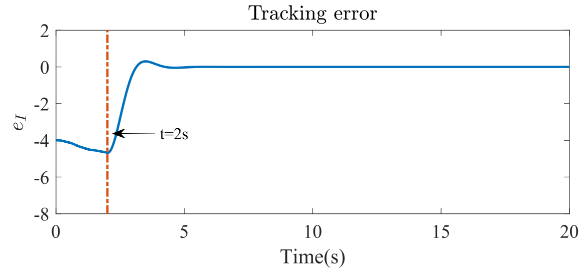


FIGURE 8. Tracking error in Algorithm 1.

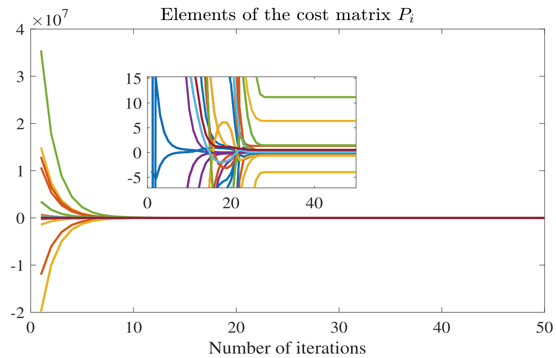


FIGURE 9. Cost matrix P_i at each iteration in Algorithm 1.

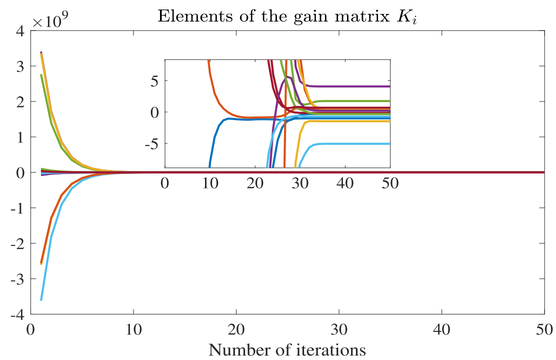


FIGURE 10. Gain matrix K_i at each iteration in Algorithm 1.

B. ONLINE DATA-DRIVEN TRACKING CONTROLLER WITH PARTIAL OBSERVABILITY

To validate the proposed controller in Algorithm 2 with partial observability, a more extensive simulation on the velocity tracking control problem of the HFA model is conducted. In aerospace applications, the output tracking error is commonly the only measurable state. Considering this, we apply state parametrization to filter the input and output information and then perform weighted summation using the method in Lemma 2. The eigenvalues of the observer matrix $\mathcal{A}$ are located at -5.

The objective of this control problem is same as that in case 1, where the initial velocity is $V = 1$, and the reference signal is $y_{cmd} = 5$. To reduce uncertain factors caused by insufficient state information, the learning time parameter N is increased to 400, and δ_T is set as 0.02 s. The remaining control parameters remain unchanged, as shown in Table 2

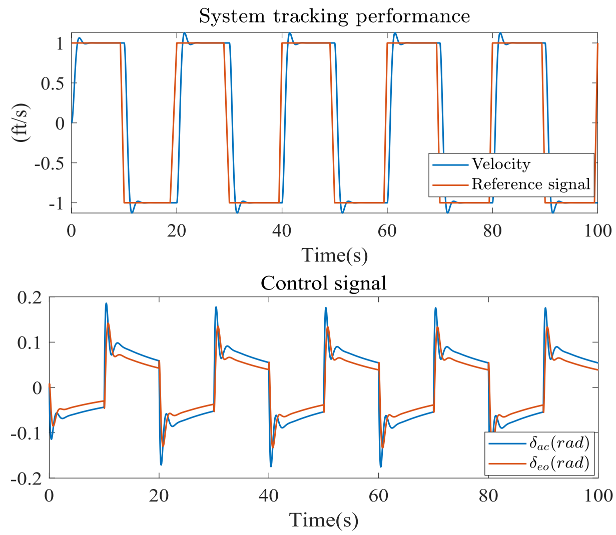


FIGURE 11. Final gain matrix K from Algorithm 1 applied to the HFA model.

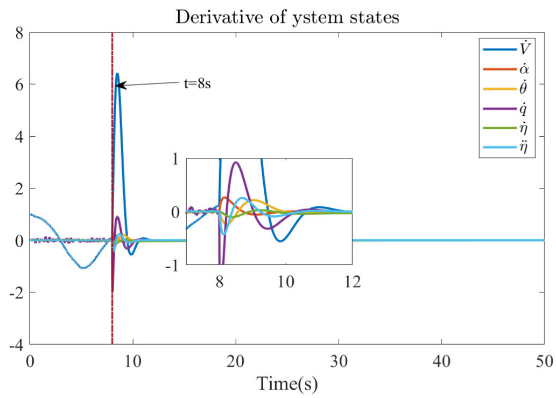


FIGURE 12. Derivative of the system state using (16) in Algorithm 2.

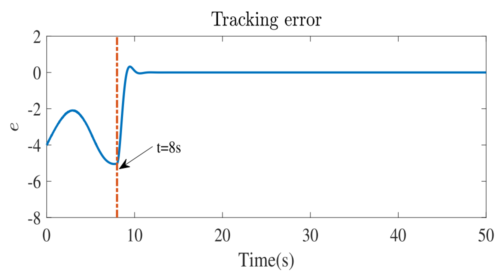


FIGURE 13. Tracking error in Algorithm 2.

Fig. 12 presents the state variables of (16). In $t \in [0, 8]$, the data information matrices δ_{zz}, I_{zz}, and I_{zv} are collected under the effect of the exploration noise e. After 8s, the main feedback matrix K, obtained using (26), is applied to the system in (16); thus, the states quickly converge to 0. Fig. 13 depicts the online tracking performance of the proposed controller with partial observability. Fig. 16 shows the actual velocity and the estimated velocity obtained via state parameterization and by tracking the reference signal, respectively. It is evident that the estimated velocity quickly

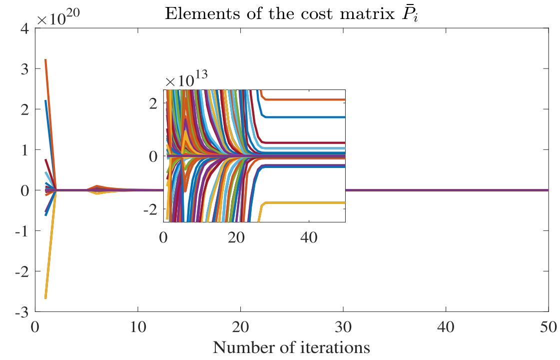


FIGURE 14. Cost matrix $\bar{P}_i$ at each iteration in Algorithm 2.

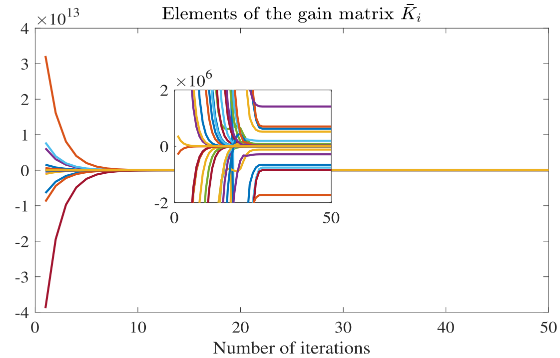


FIGURE 15. Gain matrix $\bar{K}_i$ at each iteration in Algorithm 2.

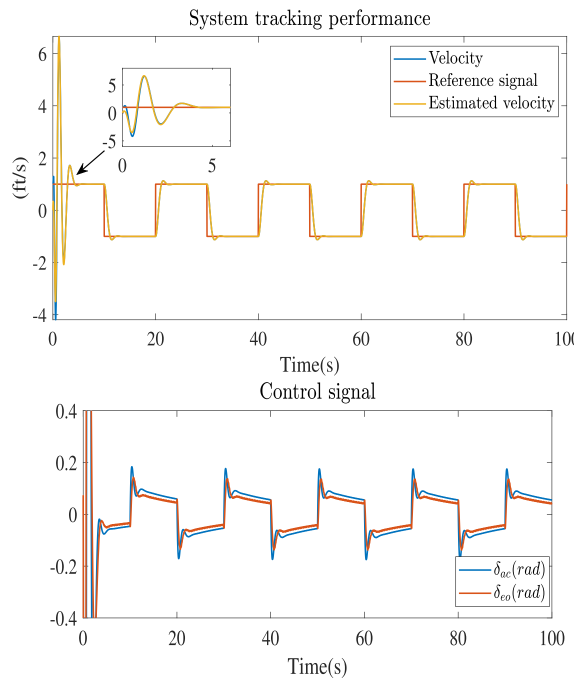


FIGURE 16. Final gain matrix $\bar{K}$ Obtained from Algorithm 2 applied to the HFA model.

becomes consistent with the actual velocity and then tracks the reference signal. Figs. 15 and 14 depict the trend of the matrices $\bar{P}_i = \bar{M}^T P \bar{M} \in \mathbb{R}^{19\times 19}$ and $\bar{K}_i = K\bar{M} \in \mathbb{R}^{2\times 19}$ during the calculation. Due to the relatively high dimensionality of $\bar{P}$ and $\bar{K}$, for the sake of convenience, we only list the final values of P and K for Algorithm 2 in Appendix B.

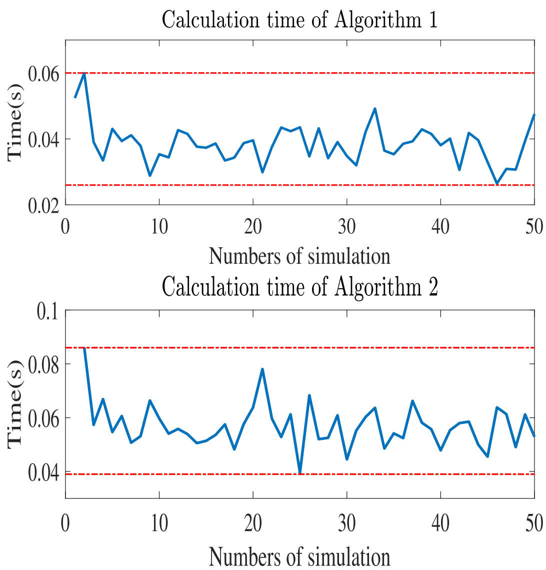


FIGURE 17. **Calculation time per simulation of the two algorithms.**

In the two proposed methods, the core is to use matrix knowledge to calculate the approximate optimal gain through (26). This process may take some time. To better reflect the rationality of the algorithm, we perform 50 simulations on Algorithms 1 and 2 respectively to obtain the average time for calculating the approximate optimal gain process. The details are described in Fig. 17 and the average calculation time of the two algorithms is 0.0385s and 0.0569s respectively. From the picture, we can see that since the value of the detection noise is randomly selected each time, the calculation time is slightly different for each simulation. The simulation calculation time of algorithm one is kept between 0.026s and 0.06, while the value of algorithm two is between 0.039s and 0.08s. In fact, Algorithm 2 needs to collect more data and information than Algorithm 1, so its information matrix has a larger dimension. Therefore, the calculation time of Algorithm 2 is longer. However, the calculation time of this process is acceptable for the control system in general.

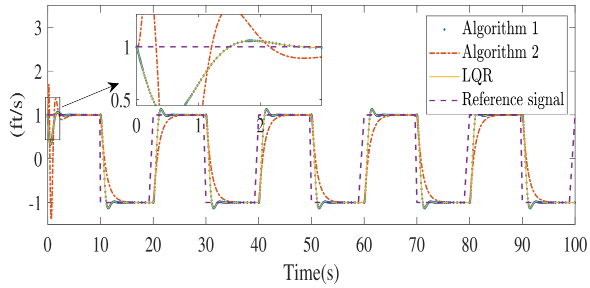


FIGURE 18. **Simulation comparison under different algorithms.**

In the final simulation, we compare the proposed algorithms with the classic LQR algorithm. Through the *lqr* function in MATLAB, we can calculate the optimal feedback gain K_{lqr} after knowing the system matrix knowledge, which is shown in Appendix B. From Appendix B, we can see that the control policy obtained by algorithm 1 and 2 is very close to the optimal gain K_{lqr}, and the simulations described in Fig. 18 confirm the final conclusion. In summary, the proposed method not only does not require the knowledge of system dynamics, but can also obtain an optimal solution almost similar to the LQR method.

V. CONCLUSION

In this study, a novel data-driven adaptive tracking controller employing IRL for a class of HFAs was developed.

$$P_{Algorithm1} = \begin{pmatrix} 0.0662 & 0.0205 & -0.431 & -0.00479 & 0.0668 & 0.0161 & 0.147 \\ 0.0205 & 1.33 & -0.548 & -0.047 & 0.351 & -0.0678 & 0.0128 \\ -0.431 & -0.548 & 6.38 & 0.0577 & -3.96 & -0.575 & -0.635 \\ -0.00479 & -0.047 & 0.0577 & 0.011 & -0.00997 & 0.00303 & -0.00751 \\ 0.0668 & 0.351 & -3.96 & -0.00997 & 11.2 & 1.46 & 0.0894 \\ 0.0161 & -0.0678 & -0.575 & 0.00303 & 1.46 & 0.269 & 0.0253 \\ 0.147 & 0.0128 & -0.635 & -0.00751 & 0.0894 & 0.0253 & 0.533 \end{pmatrix}$$

$$K_{Algorithm1} = \begin{pmatrix} -0.319 & -1.47 & 1.74 & -0.152 & 0.556 & 0.201 & -0.733 \\ 0.432 & 4.07 & -5.07 & -1.0 & 0.588 & -0.314 & 0.68 \end{pmatrix}$$

$$P_{Algorithm2} = \begin{pmatrix} 0.0662 & 0.0204 & -0.4304 & -0.0048 & 0.0646 & 0.0158 & 0.1468 \\ 0.0204 & 1.3346 & -0.5432 & -0.0470 & 0.3386 & -0.0695 & 0.0127 \\ -0.4304 & -0.5432 & 6.3311 & 0.0575 & -3.8244 & -0.5575 & -0.6344 \\ -0.0048 & -0.0470 & 0.0575 & 0.0110 & -0.0096 & 0.0031 & -0.0075 \\ 0.0646 & 0.3386 & -3.8244 & -0.0096 & 10.7982 & 1.4058 & 0.0863 \\ 0.0158 & -0.0695 & -0.5575 & 0.0031 & 1.4058 & 0.2622 & 0.0248 \\ 0.1468 & 0.0127 & -0.6344 & -0.0075 & 0.0863 & 0.0248 & 0.5331 \end{pmatrix}$$

$$K_{Algorithm2} = \begin{pmatrix} -0.3186 & -1.47 & 1.745 & -0.1523 & 0.5372 & 0.1988 & -0.7331 \\ 0.4316 & 4.073 & -5.067 & -1.00 & 0.5673 & -0.3165 & 0.6802 \end{pmatrix}$$

$$K_{lqr} = \begin{pmatrix} -0.3186 & -1.47 & 1.745 & -0.1523 & 0.5362 & 0.1987 & -0.7331 \\ 0.4316 & 4.073 & -5.066 & -1.003 & 0.5663 & -0.3167 & 0.6802 \end{pmatrix}$$

To overcome the lack of system dynamics, IRL was employed to collect data within an expected time interval and utilize these data to calculate the optimal feedback gain. Thereafter, the proposed controller was combined with a state parameterization method in order to overcome the partial observability of the system. This proposed design was validated via numerical simulations involving an HFA model.

In fact, almost all actual engineering systems are subject to various uncertain factors such as actuator saturation [22], [23], sensor faults [25], and actuator failures [24]. When these problems are not fully considered in the overall design of the controller, it will lead to an undesirable impact on the stability and robustness of the closed-loop system. According to the related results on robust adaptive control [42] with consideration of input saturation [43], [44] and state constraints [45], the data-driven robust adaptive fault tolerant control will be focused in future.

APPENDIX A
PLANT MATRICES

$$A = \begin{pmatrix} -0.042 & 4.167 & -32.20 & 0 & 0.08 & 0.18 \\ -0.01 & -9.13 & 0 & 1 & 0.05 & 0.07 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & -446.43 & 0 & -0.18 & 2.24 & 38.46 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0.05 & 0 & 2 & -0.14 & -7.46 \end{pmatrix}$$

$$B = \begin{pmatrix} -6.25 & 0 \\ -1.82 & -0.18 \\ 0 & 0 \\ -24.5 & -91.52 \\ 0 & 0 \\ 0.94 & -0.18 \end{pmatrix}$$

$$C = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

APPENDIX B
CONTROL PARAMETERS OBTAINED FROM ALGORITHM 1 AND 2

$P_{Algorithm1}$, $K_{Algorithm1}$, $P_{Algorithm2}$, $K_{Algorithm2}$, and K_{lqr} are as shown at the bottom of the previous page.

REFERENCES

[1] R. Gadient, E. Lavretsky, and K. Wise, "Very flexible aircraft control challenge problem," in *Proc. AIAA Guid., Navigat., Control Conf.*, Aug. 2012, p. 4973.

[2] C. M. Shearer and C. E. S. Cesnik, "Trajectory control for very flexible aircraft," *J. Guid., Control, Dyn.*, vol. 31, no. 2, pp. 340–357, Mar. 2008.

[3] W. Su and C. E. S. Cesnik, "Dynamic response of highly flexible flying wings," *AIAA J.*, vol. 49, no. 2, pp. 324–339, Feb. 2011.

[4] Y. Wang, A. Wynn, and R. Palacios, "Nonlinear modal aeroservoelastic analysis framework for flexible aircraft," *AIAA J.*, vol. 54, no. 10, pp. 3075–3090, Oct. 2016.

[5] N. Tsushima and W. Su, "A study on adaptive vibration control and energy conversion of highly flexible multifunctional wings," *Aerosp. Sci. Technol.*, vol. 79, pp. 297–309, Aug. 2018.

[6] T. D. S. S. Versiani, F. J. Silvestre, A. B. G. Neto, D. A. Rade, R. G. Annes da Silva, M. V. Donadon, R. M. Bertolin, and G. C. Silva, "Gust load alleviation in a flexible smart idealized wing," *Aerosp. Sci. Technol.*, vol. 86, pp. 762–774, Mar. 2019.

[7] X. Wang, E. Van Kampen, Q. P. Chu, and R. De Breuker, "Flexible aircraft gust load alleviation with incremental nonlinear dynamic inversion," *J. Guid., Control, Dyn.*, vol. 42, no. 7, pp. 1519–1536, Jul. 2019.

[8] J. H. Hansen, M. Duan, I. Kolmanovsky, and C. E. Cesnik, "Control allocation for maneuver and gust load alleviation of flexible aircraft," in *Proc. AIAA Scitech Forum*, Jan. 2020, p. 1186.

[9] W. Fan, H. H. T. Liu, and R. H. S. Kwong, "Gain-scheduling control of flexible aircraft with actuator saturation and stuck faults," *J. Guid., Control, Dyn.*, vol. 40, no. 3, pp. 510–520, Mar. 2017.

[10] M. Dillsaver, C. E. Cesnik, and I. Kolmanovsky, "Trajectory control of very flexible aircraft with gust disturbance," in *Proc. AIAA Atmos. Flight Mech. (AFM) Conf.*, Aug. 2013, p. 4745.

[11] T. Gibson, A. Annaswamy, and E. Lavretsky, "Modeling for control of very flexible aircraft," in *Proc. AIAA Guid., Navigat., Control Conf.*, Aug. 2011, p. 6202.

[12] T. Gibson, A. Annaswamy, and E. Lavretsky, "Improved transient response in adaptive control using projection algorithms and closed loop reference models," in *Proc. AIAA Guid., Navigat., Control Conf.*, Aug. 2012, p. 4775.

[13] T. E. Gibson, A. M. Annaswamy, and E. Lavretsky, "On adaptive control with closed-loop reference models: Transients, oscillations, and peaking," *IEEE Access*, vol. 1, pp. 703–717, 2013.

[14] E. Lavretsky and K. A. Wise, *Robust Adaptive Control*. London, U.K.: Springer, 2013, pp. 317–353.

[15] Z. Qu, B. T. Thomsen, and A. M. Annaswamy, "Adaptive control for a class of multi-input multi-output plants with arbitrary relative degree," *IEEE Trans. Autom. Control*, vol. 65, no. 7, pp. 3023–3038, Jul. 2020.

[16] I. O'Rourke, I. Kolmanovsky, E. Garone, and A. Girard, "Scalar reference governor for constrained maneuver and shape control of nonlinear multibody aircraft," *IFAC-PapersOnLine*, vol. 52, no. 16, pp. 819–824, 2019.

[17] L. Ma, X. Huo, X. Zhao, and G. Zong, "Adaptive fuzzy tracking control for a class of uncertain switched nonlinear systems with multiple constraints: A small-gain approach," *Int. J. Fuzzy Syst.*, vol. 21, no. 8, pp. 2609–2624, Nov. 2019.

[18] L. Ma, N. Xu, X. Huo, and X. Zhao, "Adaptive finite-time output-feedback control design for switched pure-feedback nonlinear systems with average dwell time," *Nonlinear Anal., Hybrid Syst.*, vol. 37, Aug. 2020, Art. no. 100908.

[19] Z.-M. Li and J. H. Park, "Dissipative fuzzy tracking control for nonlinear networked systems with quantization," *IEEE Trans. Syst., Man, Cybern., Syst.*, early access, Sep. 24, 2018, doi: 10.1109/TSMC.2018.2866996.

[20] Z.-M. Li, X.-H. Chang, and J. H. Park, "Quantized static output feedback fuzzy tracking control for discrete-time nonlinear networked systems with asynchronous event-triggered constraints," *IEEE Trans. Syst., Man, Cybern. Syst.*, early access, Aug. 15, 2019, doi: 10.1109/TSMC.2019.2931530.

[21] Y. Chang, Y. Wang, F. E. Alsaadi, and G. Zong, "Adaptive fuzzy output-feedback tracking control for switched stochastic pure-feedback nonlinear systems," *Int. J. Adapt. Control*, vol. 33, no. 10, pp. 1567–1582, 2019.

[22] N. Sun, Y. Fu, T. Yang, J. Zhang, Y. Fang, and X. Xin, "Nonlinear motion control of complicated dual rotary crane systems without velocity feedback: Design, analysis, and hardware experiments," *IEEE Trans. Autom. Sci. Eng.*, vol. 17, no. 2, pp. 1017–1029, Apr. 2020.

[23] T. Yang, N. Sun, H. Chen, and Y. Fang, "Observer-based nonlinear control for tower cranes suffering from uncertain friction and actuator constraints with experimental verification," *IEEE Trans. Ind. Electron.*, early access, May 12, 2020, doi: 10.1109/TIE.2020.2992972.

[24] Y. Wang, X. Yang, and H. Yan, "Reliable fuzzy tracking control of near-space hypersonic vehicle using aperiodic measurement information," *IEEE Trans. Ind. Electron.*, vol. 66, no. 12, pp. 9439–9447, Dec. 2019.

[25] Y. Wang, W. Zhou, J. Luo, H. Yan, H. Pu, and Y. Peng, "Reliable intelligent path following control for a robotic airship against sensor faults," *IEEE/ASME Trans. Mechatronics*, vol. 24, no. 6, pp. 2572–2582, Dec. 2019.

[26] F. L. Lewis and K. G. Vamvoudakis, "Reinforcement learning for partially observable dynamic processes: Adaptive dynamic programming using measured output data," *IEEE Trans. Syst., Man, Cybern. B, Cybern.*, vol. 41, no. 1, pp. 14–25, Feb. 2011.

[27] D. Vrabie, O. Pastravanu, M. Abu-Khalaf, and F. L. Lewis, "Adaptive optimal control for continuous-time linear systems based on policy iteration," *Automatica*, vol. 45, no. 2, pp. 477–484, Feb. 2009.

[28] T. Bian, Y. Jiang, and Z.-P. Jiang, "Adaptive dynamic programming and optimal control of nonlinear nonaffine systems," *Automatica*, vol. 50, no. 10, pp. 2624–2632, Oct. 2014.

[29] Q. Wei, D. Liu, and H. Lin, "Value iteration adaptive dynamic programming for optimal control of discrete-time nonlinear systems," *IEEE Trans. Cybern.*, vol. 46, no. 3, pp. 840–853, Mar. 2016.
[30] R. S. Sutton, A. G. Barto, and R. J. Williams, "Reinforcement learning is direct adaptive optimal control," *IEEE Control Syst.*, vol. 12, no. 2, pp. 19–22, Apr. 1992.
[31] F. L. Lewis, D. Vrabie, and K. G. Vamvoudakis, "Reinforcement learning and feedback control: Using natural decision methods to design optimal adaptive controllers," *IEEE Control Syst.*, vol. 32, no. 6, pp. 76–105, Dec. 2012.
[32] D. Kleinman, "On an iterative technique for Riccati equation computations," *IEEE Trans. Autom. Control*, vol. 13, no. 1, pp. 114–115, Feb. 1968.
[33] Y. Jiang and Z.-P. Jiang, "Computational adaptive optimal control for continuous-time linear systems with completely unknown dynamics," *Automatica*, vol. 48, no. 10, pp. 2699–2704, Oct. 2012.
[34] C. Qin, H. Zhang, and Y. Luo, "Online optimal tracking control of continuous-time linear systems with unknown dynamics by using adaptive dynamic programming," *Int. J. Control*, vol. 87, no. 5, pp. 1000–1009, May 2014.
[35] L. M. Zhu, H. Modares, G. O. Peen, F. L. Lewis, and B. Yue, "Adaptive suboptimal output-feedback control for linear systems using integral reinforcement learning," *IEEE Trans. Control Syst. Technol.*, vol. 23, no. 1, pp. 264–273, Jan. 2015.
[36] S. A. A. Rizvi and Z. Lin, "Reinforcement learning-based linear quadratic regulation of continuous-time systems using dynamic output feedback," *IEEE Trans. Cybern.*, early access, Jan. 3, 2019, doi: 10.1109/TCYB.2018.2886735.
[37] H. Modares and F. L. Lewis, "Linear quadratic tracking control of partially-unknown continuous-time systems using reinforcement learning," *IEEE Trans. Autom. Control*, vol. 59, no. 11, pp. 3051–3056, Nov. 2014.
[38] H. Neudecker, "Some theorems on matrix differentiation with special reference to Kronecker matrix products," *J. Amer. Stat. Assoc.*, vol. 64, no. 327, pp. 953–963, Sep. 1969.
[39] J. Brewer, "Kronecker products and matrix calculus in system theory," *IEEE Trans. Circuits Syst.*, vol. 25, no. 9, pp. 772–781, Sep. 1978.
[40] G. J. Balas and P. M. Young, "Sensor selection via closed-loop control objectives," *IEEE Trans. Control Syst. Technol.*, vol. 7, no. 6, pp. 692–705, 1999.
[41] B. Pang, T. Bian, and Z.-P. Jiang, "Robust policy iteration for continuous-time linear quadratic regulation," 2020, *arXiv:2005.09528*. [Online]. Available: http://arxiv.org/abs/2005.09528
[42] S. Xiao and J. Dong, "Robust adaptive fault-tolerant tracking control for uncertain linear systems with actuator failures based on the closed-loop reference model," *IEEE Trans. Syst., Man, Cybern.*, vol. 50, no. 9, pp. 3448–3455, Sep. 2018.
[43] J. Ma, Z. Zheng, and P. Li, "Adaptive dynamic surface control of a class of nonlinear systems with unknown direction control gains and input saturation," *IEEE Trans. Cybern.*, vol. 45, no. 4, pp. 728–741, Apr. 2015.
[44] J. Ma, S. S. Ge, Z. Zheng, and D. Hu, "Adaptive NN control of a class of nonlinear systems with asymmetric saturation actuators," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 26, no. 7, pp. 1532–1538, Jul. 2015.
[45] K. P. Tee, S. S. Ge, and E. H. Tay, "Barrier Lyapunov functions for the control of output-constrained nonlinear systems," *Automatica*, vol. 45, no. 4, pp. 918–927, Apr. 2009.

CHI PENG was born was born in Anqing, Anhui, China, in 1996. He received the bachelor's degree in electrical engineering and automation from the Hefei University of Technology, Anhui, in 2018. He is currently pursuing the M.S. degree in control science and engineering with the National University of Defense Technology, Changsha, Hunan, China. His current research interests include adaptive control, learning control, and applications of control in highly flexible aircraft systems.

JIANJUN MA (Member, IEEE) received the M.E. and Ph.D. degrees in control science and engineering from the National University of Defense Technology, Changsha, China, in 2004 and 2010, respectively. From 2011 to 2012, he was a Research Fellow with the Interactive Digital Media Institute (IDMI), National University of Singapore, Singapore. He is currently a Professor with the College of Intelligence Science and Technology, National University of Defense Technology. His research interests include guidance and control of aerospace vehicles, adaptive control theory, and fault tolerant flight control.