IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received February 18, 2022, accepted March 4, 2022, date of publication March 14, 2022, date of current version March 25, 2022.

Digital Object Identifier 10.1109/ACCESS.2022.3157878

Incremental Dual Heuristic Dynamic Programming Based Hybrid Approach for Multi-Channel Control of Unstable Tailless Aircraft

HANGXU LI, LIGUO SUN, WENQIAN TAN, XIAOYU LIU, AND WEIGAO DANG
School of Aeronautic Science and Engineering, Beihang University, Beijing 100191, China

Corresponding author: Wenqian Tan (tanwenqian@buaa.edu.cn)

This work was supported in part by the National Natural Science Foundation of China under Grant 11502008, in part by the Aeronautical Science Foundation of China under Grant 20185702003, and in part by the Fundamental Research Funds for the Central Universities of China under Grant YWF-21-BJ-J-809.

ABSTRACT Actor-critic based online reinforcement learning control has been proved to be promising method for control of aerial vehicles. However, it is difficult to guarantee high-level success rate of initial training and to tune the large amount of parameters for actors and critics considering unstable multi-input and multi-output (MIMO) aircraft. In order to facilitate and simplify the training of the actor and critic for unstable aircraft, classic stability augmentation system (SAS) is designed for the open-loop aircraft first. Then the online incremental model based dual heuristic dynamic programming (IDHP) method, which has been proposed recently, is extended in application to design a multi-channel robust adaptive controller, and MIMO form network structures are designed and determined for the actors and critics considering the three-channel coupling issues. Consequently, the classic SAS and the IDHP controller make up a novel hybrid control framework. In this control framework, the SAS takes charge of counteracting the unstable eigenvalues of the open-loop aircraft system, and the IDHP takes charge on guaranteeing robust and adaptive performance for high-performance tailless aircraft equipped with the SAS. Specifically, the introduction of the classic control method decreases the difficulty of the initial training for multi-channel IDHP controller. The tuning process for initial parameters of actor and critic neural networks in multiple channels is greatly facilitated. Without the help of SAS, the initial training for multi-channel IDHP controllers of unstable plants is almost impossible to succeed. Finally, the novel hybrid control architecture and method are validated using the Innovative Control Effectors (ICE) model, which has unstable modes in the longitudinal dynamics. Typical aerodynamic model uncertainties are numerically simulated to demonstrate the effectiveness of the proposed control method.

INDEX TERMS Incremental DHP, actor and critic, unstable aircraft, reinforcement learning, MIMO control.

I. INTRODUCTION

Many aerospace systems, robots, etc., are becoming more and more complex, and are difficult to control. In many cases, modeling the accurate physical models for these plants may be extremely difficult or impossible, but a large amount of measured state data can possibly be easily collected. In view of this, data-driven control (DDC) theory is developed. DDC means a class of methods in which the controller is designed by directly using collected data [1]. Approximate dynamic programming (ADP) is one branch of DDC methods. It combines reinforcement learning with dynamic programming in order to solve optimal control problems forward-in-time [2]. ADP methods use actor-critic structure to perform a learning and adapting process. According to what critic outputs, ADP methods can be generally categorized into three main schemes: heuristic dynamic programming (HDP), dual heuristic dynamic programming (DHP), and global dual heuristic dynamic programming (GDHP) [3]. Several studies have shown that DHP and GDHP outperform

The associate editor coordinating the review of this manuscript and approving it for publication was Shun-Feng Su.

HDP in success rate and precision since they directly output the derivatives of the value function [3], [4]. Furthermore, although GDHP shows a little advantage over DHP in terms of success rate and precision, the complexity of computation and implementation are considerably higher for GDHP when compared to DHP [5], [6].

ADP methods are now developed rapidly and have been applied on many complex nonlinear plants. In Ref. [7], a multi-input multi-output (MIMO) neuro-based controller is designed by Enns to finish the command tracking problem of the Apache helicopter, and it shows a satisfactory results. A goal representation heuristic dynamic programming (GrHDP) based controller is developed and validated by Tang in Ref. [8]. In Ref. [9], Han used HDP methods to control the angle of attack of a nonlinear hypersonic vehicle model. Shayan uses HDP method to design a longitudinal controller for controlling the angle of attack and test it on an over-actuated innovative control effector (ICE) aircraft [10]. The aforementioned methods all need an accurate global model of the controlled plants. Therefore, they are usually sensitive to model uncertainties and environment disturbances when designing controller for high-performance aircraft with highly nonlinear and complex dynamics. Recently, deep deterministic policy gradient (DDPG) method is improved and used to design a controller for a 6-DoF unmanned aircraft model [11]. After being trained with four set of data, this controller is able to well perform a reference tracking task. In the proposed DDPG, four neural networks, i.e., actor, critic, actor target, critic target, are required to perform the learning process. And these networks often have three or more hidden layers. In this way, a large amount of parameters, i.e., weights for neural networks, needs to be tuned when using these methods, and a set of effective initial parameters used to make these methods function properly is usually hard to obtain, expecially for MIMO flight control problems.

Incremental modelling technique is introduced by Zhou *et al.* to reduce the model dependency of ADP methods [6], [12] [13]. The actor-critic based ADP controllers can obtain more accurate and real-time information from the incremental models of the controlled plants, especially when the plants are higly nonlinear or subject to model uncertainties. Therefore, if they are used without a priori training, the incremental model based ADP methods learn faster and have higher success rate than common ADP methods [6], [13]. Among these methods, IDHP method shows better performance from comprehensive consideration of control precision, adaptability, and computational cost [6], [13] [14]. However, current researches about incremental model based ADP methods, especially IDHP method, are still relatively limited. Dias uses iADP method to design a rate controller for a stable fully-actuated F-16 model [15]. Heyer designs longitudinal and lateral controllers separately by using the IDHP method on a stable fully-actuated CS-25 aircraft [16]. However, these studies are limited to control for stable fully-actuated aircraft, and the application and extension of IDHP on unstable and over-actuated aircraft still needs to be investigated. For multi-input multi-output (MIMO) unstable flight control problems, dynamic coupling between different control channels of aircraft needs to be considered in controller design, and the neural networks of actor and critic need to be chosen properly in terms of number of layers, neurons and initial weights tuning. Specifically, for complex controlled plant with unstable eigenvalues, e.g., tailless aircraft, most of ADP methods are usually extremely difficult to achieve successful initial training, because the actor-critic mechanism in this case commonly cannot receive adequate number of expected/effective act-reaction data before corruption.

This paper focuses on flight controller design for over-actuated aircraft with unstable modes, highly nonlinear dynamics, and strong three-axis couplings. The motivation of this paper is to propose an IDHP based hybrid control approach for high-performance and unstable aircraft with the unstable issue and the multi-channel control coupling issues considered. The work of this paper aims at solving the learning inefficiency phenomenon for ADP methods when facing unstable modes. For three-channel control of over-actuated highly maneuverable aircraft, network structure and parameters of multi-channel IDHP controller are designed and determined, and control allocation is introduced. More specifically, to ensure that the IDHP based controller can effectively learn at starting stages, a hybrid control framework synthesizing the IDHP method with classical control measures, i.e., stability augmentation system, and control allocation unit is proposed in this paper. It should be noted that the actor-critic methods such as IDHP will be struggling in their initial training phase if these methods are applied directly due to the unstable characteristics of the aircraft system. In view of nonlinear dynamic characteristics and coupling issues, the IDHP method is used to design a three-axis rate controller for solving the MIMO flight control problem. Furthermore, MIMO form network structures are designed and optimized for the actors and critics considering the three-channel coupling issues. To facilitate the application of the IDHP method on unstable aircraft, the stability augmentation system (SAS) is designed before hand for obtaining an augmented stable plant, and the IDHP controller is then used to design rate controller for the augmented aircraft. By introducing SAS, the actor-critic mechanism is enabled to possibly obtain sufficient number of expected/effective act-reaction data. These data can help the IDHP controller find the right convergent direction and therefore facilitate the use of IDHP controller. In this paper, a 6-DoF nonlinear model of a tailless aircraft (ICE-101) is finally used to validate the proposed hybrid IDHP/SAS method. Considering dynamic couplings between different channels of the tailless aircraft, how to determine the neural network structure for the actor and critic, e.g., design an actor for each channel or different channels share one actor, is investigated. Numerical simulation results show that the proposed hybrid controller can effectively deal with three-axis couplings and model uncertainties.

II. PRELIMINARIES ON INCREMENTAL MODEL BASED DUAL HEURISTIC DYNAMIC PROGRAMMING

DHP is a kind of reinforcement learning based control method, which has high control accuracy and needs low computational cost [17]. Based on it, IDHP is proposed as an improved DHP method which uses an incremental model to provide dynamic information for actor-critic networks. Compared with the DHP method, the learning and adapting ability of the IDHP method is greatly improved. Therefore, the IDHP method has higher training success rate, higher control accuracy, and strong ability to deal with nonlinearity and model uncertainties. However, the IDHP method has not been used to design a MIMO controller and has not been validated on a complex nonlinear aircraft with unstable aircraft dynamic modes.

IDHP controller is consisted of three parts: incremental model, critic, and actor, as shown in Fig. 1. The incremental model is used to approximate dynamics of the controlled plant around current state point; and the critic is used to evaluate the control policy provided by the actor and help actor to update itself; and actor is used to find the optimal control strategy $\boldsymbol{u}^*\left(\boldsymbol{x}_t^*\right)$ according to the evaluation by critic. These three parts make up the IDHP controller and work together to adapt the nonlinear plant.

A. INCREMENTAL MODEL

The incremental model plays an important role in the rapid convergence of the IDHP controller. Compared with global model, incremental model can have a better approximation of the dynamics of nonlinear plant around current state point [18], [19]. This enables the IDHP method to have stronger adaptability and better control performance than common DHP methods.

For a nonlinear system, it can be expressed in a discrete form assuming a high sampling frequency, as follows:

$$\boldsymbol{x}_{t+1} = f\left(\boldsymbol{x}_t, \boldsymbol{u}_t\right) \tag{1}$$

When the sampling time Δt is small enough, the dynamic of the system at current state point can be linearized by Taylor expansion, namely:

$$\boldsymbol{x}_{t+1} \approx \boldsymbol{x}_t + \boldsymbol{F}_{t-1} \cdot \left(\boldsymbol{x}_t - \boldsymbol{x}_{t-1}\right) + \boldsymbol{G}_{t-1} \cdot \left(\boldsymbol{u}_t - \boldsymbol{u}_{t-1}\right) \tag{2}$$

where $\boldsymbol{x} \in \mathbb{R}^{n\times 1}$ is the n-dimensional state vector of the system, and $\boldsymbol{u} \in \mathbb{R}^{m\times 1}$ is the m-dimensional control input vector of the system, $\boldsymbol{F}_{t-1} = \left.\frac{\partial f(\boldsymbol{x},t)}{\partial \boldsymbol{x}}\right|_{\boldsymbol{x}_{t-1},\boldsymbol{u}_{t-1}} \in \mathbb{R}^{n\times n}$ is the state transition matrix, and $\boldsymbol{G}_{t-1} = \left.\frac{\partial f(\boldsymbol{x},t)}{\partial \boldsymbol{u}}\right|_{\boldsymbol{x}_{t-1},\boldsymbol{u}_{t-1}} \in \mathbb{R}^{n\times m}$ is the control effectiveness matrix. The state transition matrix and the control effectiveness matrix are time-varying because of the nonlinearity of the system.

An incremental model to be identified can be derived from Eq. 2, as follows:

$$\Delta\boldsymbol{x}_{t+1}^T = \boldsymbol{X}_t^T \hat{\boldsymbol{\Theta}}_t \tag{3}$$

where $\Delta\boldsymbol{x}_t = \boldsymbol{x}_t - \boldsymbol{x}_{t-1}$, $\Delta\boldsymbol{u}_t = \boldsymbol{u}_t - \boldsymbol{u}_{t-1}$, $\boldsymbol{X}_t = \begin{bmatrix}\Delta\boldsymbol{x}_t \\ \Delta\boldsymbol{u}_t\end{bmatrix} \in \mathbb{R}^{(n+m)\times 1}$ is the measured data, $\boldsymbol{\Theta}_t = \begin{bmatrix}\boldsymbol{F}_{t-1}^{\mathrm{T}} \\ \boldsymbol{G}_{t-1}^{\mathrm{T}}\end{bmatrix} \in \mathbb{R}^{(n+m)\times n}$ is the parameter matrix, and $\hat{\square}$ marks the estimated value. Then, the recursive least square (RLS) method is used to identify parameters in $\boldsymbol{\Theta}$, as follows:

$$\boldsymbol{\varepsilon}_t = \Delta\boldsymbol{x}_{t+1}^T - \Delta\hat{\boldsymbol{x}}_{t+1}^T \tag{4a}$$

$$\hat{\boldsymbol{\Theta}}_{t+1} = \hat{\boldsymbol{\Theta}}_t + \frac{Cov_t\boldsymbol{X}_t}{\kappa + \boldsymbol{X}_t^T Cov_t \boldsymbol{X}_t}\boldsymbol{\varepsilon}_t \tag{4b}$$

$$Cov_{t+1} = \frac{1}{\kappa}\left(Cov_t - \frac{Cov_t\boldsymbol{X}_t\boldsymbol{X}_t^T Cov_t}{\kappa + \boldsymbol{X}_t^T Cov_t \boldsymbol{X}_t}\right) \tag{4c}$$

where $\boldsymbol{\varepsilon}_t \in \mathbb{R}^{1\times n}$ is the prediction error, $Cov_t \in \mathbb{R}^{(n+m)\times(n+m)}$ is the estimated covariance matrix, and κ is the forgetting factor in the RLS method. After parameters in $\hat{\boldsymbol{\Theta}}_t$ is identified by RLS method, the state variables can be predicted, as follows:

$$\hat{\boldsymbol{x}}_{t+1} = \boldsymbol{x}_t + \Delta\hat{\boldsymbol{x}}_{t+1}^T = \boldsymbol{x}_t + \boldsymbol{X}_t^T\hat{\boldsymbol{\Theta}}_t \tag{5}$$

B. CRITIC

The critic in IDHP controller is used to evaluate control policy provided by the actor. Without a good evaluation on the control policy, the actor will not be able to learn correctly and to provide a better and better control signal along with time. In IDHP, a good evaluation can be specified as the partial derivative of the cost function to the state variables, as follows:

$$\boldsymbol{\lambda}^*\left(\boldsymbol{x}_t^*\right) = \frac{\partial V^*\left(\boldsymbol{x}_t^*\right)}{\partial \boldsymbol{x}_t^*} \tag{6}$$

In Eq. 6, the optimal cost function is the sum of the one-step cost from current moment to the infinite time in the future:

$$V^*\left(\boldsymbol{x}_t^*\right) = \sum_{k=t}^{\infty}\gamma^{k-t}c_k \tag{7}$$

where $\gamma \in [0, 1]$ is a scalar called the forgetting factor, and c_t is the one-step cost function at time t. In this paper, the c_t is defined as a quadratic function of the state variables and the control inputs, as follows:

$$\begin{aligned} c_t &= c\left(\boldsymbol{x}_t, \boldsymbol{x}_{ref,t}, \boldsymbol{u}_t\right) \\ &= \left(\boldsymbol{x}_t - \boldsymbol{x}_{ref,t}\right)^{\mathrm{T}}\boldsymbol{Q}\left(\boldsymbol{x}_t - \boldsymbol{x}_{ref,t}\right) + \boldsymbol{u}_t^{\mathrm{T}}\boldsymbol{R}\boldsymbol{u}_t \end{aligned} \tag{8}$$

where $\boldsymbol{Q} \in \mathbb{R}^{n\times n}$ and $\boldsymbol{R} \in \mathbb{R}^{m\times m}$ are given positive definite matrices [20]. In order to implement the reinforcement learning process, the aforementioned optimal cost function $V^*\left(\boldsymbol{x}_t^*\right)$ should be approximated by critic. The approximated one is denoted as $V\left(\boldsymbol{x}_t\right)$.

Generally, a back-propagation (BP) neural network is used as the critic for its strong approximation ability to unknown nonlinear function. In order to approximate $\boldsymbol{\lambda}^*\left(\boldsymbol{x}_t^*\right)$, the critic

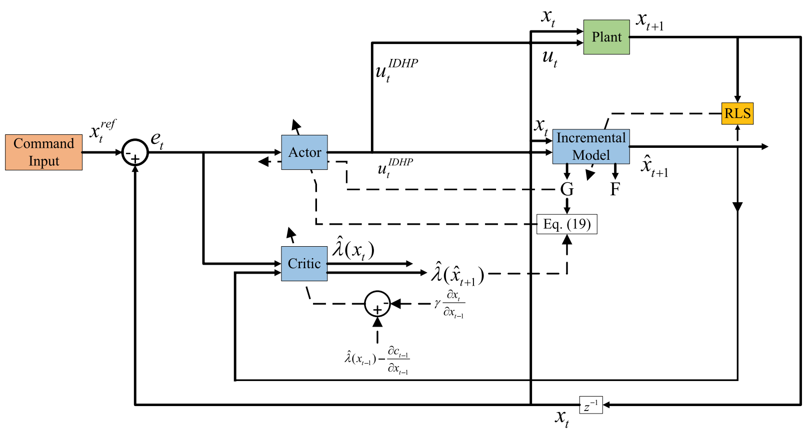

FIGURE 1. Schematic diagram of IDHP (The solid lines represent the feed forward flow of signals, and the dashed lines represent the adaptation pathways.).

network aims at minimizing the temporal difference (TD) error, as follows:

$$E_c(t) = \frac{1}{2}\boldsymbol{e}_c(t)^{\mathrm{T}}\boldsymbol{e}_c(t) \tag{9}$$

where:

$$\begin{aligned}\boldsymbol{e}_c(t) &= \frac{\partial\left[V\left(\boldsymbol{x}_{t-1}\right) - c_{t-1} - \gamma V\left(\boldsymbol{x}_t\right)\right]}{\partial \boldsymbol{x}_{t-1}} \\ &= \boldsymbol{\lambda}\left(\boldsymbol{x}_{t-1}\right) - \left(\frac{\partial c_{t-1}}{\partial \boldsymbol{x}_{t-1}} + \gamma\boldsymbol{\lambda}\left(\boldsymbol{x}_t\right)\frac{\partial \boldsymbol{x}_t}{\partial \boldsymbol{x}_{t-1}}\right)\end{aligned} \tag{10}$$

In Eq. 10, $\boldsymbol{\lambda}(\boldsymbol{x})$ is the approximation result of $\boldsymbol{\lambda}^*\left(\boldsymbol{x}_t^*\right)$ and can be calculated through critic neural network [21], and $\frac{\partial \boldsymbol{x}_t}{\partial \boldsymbol{x}_{t-1}}$ can be obtained using incremental model:

$$\begin{aligned}\frac{\partial \boldsymbol{x}_t}{\partial \boldsymbol{x}_{t-1}} &= \frac{\partial f}{\partial \boldsymbol{x}_{t-1}} + \frac{\partial f}{\partial \boldsymbol{u}_{t-1}} \cdot \left.\frac{\partial \boldsymbol{u}_{t-1}}{\partial \boldsymbol{x}_{t-1}}\right|_{actor} \\ &\approx \hat{\boldsymbol{F}}_{t-1} + \hat{\boldsymbol{G}}_{t-1} \cdot \left.\frac{\partial \boldsymbol{u}_{t-1}}{\partial \boldsymbol{x}_{t-1}}\right|_{actor}\end{aligned} \tag{11}$$

where $\left.\frac{\partial \boldsymbol{u}_{t-1}}{\partial \boldsymbol{x}_{t-1}}\right|_{actor}$ can be calculated through the actor [21]. The TD error is used to calculate the gradient of weights and biases in the critic neural network and then to adjust them with a positive learning rate η_c, as follows:

$$\boldsymbol{w}_c(t+1) = \boldsymbol{w}_c(t) + \Delta\boldsymbol{w}_c(t) \tag{12}$$

where:

$$\begin{aligned}\Delta\boldsymbol{w}_c(t) &= -\eta_c \cdot \frac{\partial E_c(t)}{\partial \boldsymbol{\lambda}(\boldsymbol{x}_{t-1})} \cdot \frac{\partial \boldsymbol{\lambda}(\boldsymbol{x}_{t-1})}{\partial \boldsymbol{w}_c(t)} \\ &= -\eta_c \cdot \frac{\partial E_c(t)}{\partial \boldsymbol{e}_c(t)} \cdot \frac{\partial \boldsymbol{e}_c(t)}{\partial \boldsymbol{\lambda}(\boldsymbol{x}_{t-1})} \cdot \frac{\partial \boldsymbol{\lambda}(\boldsymbol{x}_{t-1})}{\partial \boldsymbol{w}_c(t)} \\ &= -\eta_c \cdot \frac{\partial \frac{1}{2}\boldsymbol{e}_c(t)^{\mathrm{T}}\boldsymbol{e}_c(t)}{\partial \boldsymbol{e}_c(t)} \cdot \frac{\partial \boldsymbol{e}_c(t)}{\partial \boldsymbol{\lambda}(\boldsymbol{x}_{t-1})} \cdot \frac{\partial \boldsymbol{\lambda}(\boldsymbol{x}_{t-1})}{\partial \boldsymbol{w}_c(t)} \\ &= -\eta_c \cdot \boldsymbol{e}_c(t)^{\mathrm{T}} \cdot \frac{\partial \boldsymbol{\lambda}(\boldsymbol{x}_{t-1})}{\partial \boldsymbol{w}_c(t)}\end{aligned} \tag{13}$$

C. ACTOR

The actor in IDHP is used to approximate the optimal control policy. It is able to adapt controlled nonlinear plants with the evaluation from the critic. In general, a BP neural network is chosen as the actor. In order to approximate the optimal control policy, the actor tries to minimize the cost $V(\boldsymbol{x}_t)$ at current moment, as follows:

$$\begin{aligned}\boldsymbol{u}_t^* &= \arg\min_{\boldsymbol{u}_t}\left(V\left(\boldsymbol{x}_t\right)\right) \\ &= \arg\min_{\boldsymbol{u}_t}\left(c_t + \gamma V\left(\boldsymbol{x}_{t+1}\right)\right)\end{aligned} \tag{14}$$

By using the gradient descent method, the weights of the actor network can be modified with a positive learning rate η_a, as follows:

$$\boldsymbol{w}_a(t+1) = \boldsymbol{w}_a(t) + \Delta\boldsymbol{w}_a(t) \tag{15}$$

where:

$$\begin{aligned}\Delta\boldsymbol{w}_a(t) &= -\eta_a \cdot \frac{\partial V\left(\boldsymbol{x}_t\right)}{\partial \boldsymbol{u}_t} \cdot \frac{\partial \boldsymbol{u}_t}{\partial \boldsymbol{w}_a(t)} \\ &= -\eta_a \cdot \left(\frac{\partial c_t}{\partial \boldsymbol{u}_t} + \gamma\boldsymbol{\lambda}\left(\boldsymbol{x}_{t+1}\right)\frac{\partial \boldsymbol{x}_{t+1}}{\partial \boldsymbol{u}_t}\right) \cdot \frac{\partial \boldsymbol{u}_t}{\partial \boldsymbol{w}_a(t)} \\ &\approx -\eta_a \cdot \left(\frac{\partial c_t}{\partial \boldsymbol{u}_t} + \gamma\boldsymbol{\lambda}\left(\hat{\boldsymbol{x}}_{t+1}\right)\hat{\boldsymbol{G}}_{t-1}\right) \cdot \frac{\partial \boldsymbol{u}_t}{\partial \boldsymbol{w}_a(t)}\end{aligned} \tag{16}$$

III. IDHP BASED THREE-AXIS ATTITUDE CONTROL SYSTEM DESIGN FOR TAILLESS AIRCRAFT

The IDHP method is used in this section to design a nonlinear MIMO control system for tailless aircraft. This method has shown strong adaptability, and the ability to resist model uncertainties or disturbances [6]. Therefore, it is chosen to deal with control problems caused by the characteristics of the tailless aircraft, like strong nonlinearity, three-axis couplings, and so on. In this section, the ICE-101 plane is firstly introduced. Then, an IDHP based multi-channel controller is designed in order to deal with control problems existed in the ICE-101plane.

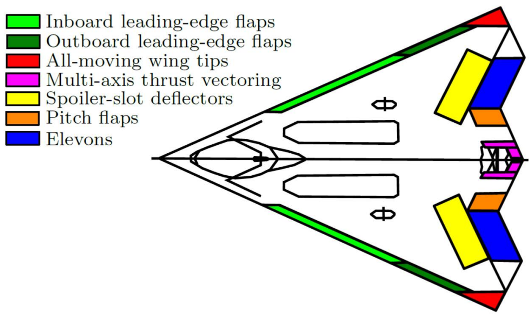

FIGURE 2. Control surface distribution of ICE-101.

A. TAILLESS AIRCRAFT MODEL

The tailless aircraft has both advantages and disadvantages because of its innovative aerodynamic configuration. It has excellent aerodynamic characteristics and stealth performance. Besides, the tailless aircraft is of lighter structure weight, stronger load capacity, higher cruise efficiency, and so on [22]. However, some disadvantages are also existed because of the specific configuration of the tailless aircraft, like the strong nonlinearity and the strong three-axis coupling. The tailless aircraft is also easy to be negatively affected by the model uncertainties or disturbances. A robust adaptive control system is required for controlling the tailless aircraft in view of these characteristics.

The tailless aircraft simulation model used in this paper is a 6-DoF model built based on the aerodynamic data of ICE-101 published by Lockheed Martin Space Systems Company [22], [23] [24]. This model is able to reflect the characters of tailless aircraft, like strong nonlinearity, three-axis couplings, etc. Eleven control surfaces and two thrust vectors can be used to control the ICE-101, as shown in Fig. 2. These control inputs δ_u can be classified into 7 types. They are inboard leading edge flap (IBLEF), outboard leading edge flap (OBLEF), elevon (ELE), pitch flap (PF), all moving wing tip (AMT), spiller-slot deflector (SSD) and thrust vector. The thrust vector includes the pitching thrust vector (PTV) and the yawing thrust vector (YTV). The deflection limits of these control surfaces are obtained from [22] and are shown in Table 1. In this paper, only 7 main control inputs, i.e., left ELE (LELE), right ELE (RELE), left AMT (LAMT), right AMT (RAMT), PF, PTV, and YTV, are used when designing the IDHP based attitude control system. Before going further, it is still worthwhile to mention that each single control input is tightly and nonlinearly coupled with each other control input. Furthermore, each single control input can contribute to nearly all six control channels (6 degree of freedoms).

The moment equations in body-fixed frame can be written as follows:

$$\begin{bmatrix} \dot{p} \\ \dot{q} \\ \dot{r} \end{bmatrix} = -\boldsymbol{J}^{-1}\left(\begin{bmatrix} p \\ q \\ r \end{bmatrix} \times \left(\boldsymbol{J}\begin{bmatrix} p \\ q \\ r \end{bmatrix}\right)\right)$$

$$+\bar{q}S\boldsymbol{J}^{-1}\begin{bmatrix} bC_l \\ \bar{c}C_m \\ bC_n \end{bmatrix}$$

$$= -\boldsymbol{J}^{-1}\left(\begin{bmatrix} p \\ q \\ r \end{bmatrix} \times \left(\boldsymbol{J}\begin{bmatrix} p \\ q \\ r \end{bmatrix}\right)\right) + \bar{q}S\boldsymbol{J}^{-1}\begin{bmatrix} bC_l^{base} \\ \bar{c}C_m^{base} \\ bC_n^{base} \end{bmatrix} + \bar{q}S\boldsymbol{J}^{-1}\begin{bmatrix} bC_l^u \\ \bar{c}C_m^u \\ bC_n^u \end{bmatrix} \tag{17}$$

TABLE 1. Control limits of the control surfaces of the ICE-101.

Control surfaces	Deflection limits (deg)	Rate limits (deg/s)
IBLEF	$0 \sim 40$	$-40 \sim 40$
OBLEF	$-40 \sim 40$	$-40 \sim 40$
ELE	$-30 \sim 30$	$-70 \sim 70$
PF	$-30 \sim 30$	$-70 \sim 70$
AMT	$0 \sim 60$	$-70 \sim 70$
SSD	$0 \sim 60$	$-150 \sim 150$

where p, q, r are the angular velocities along rolling, pitching, and yawing directions respectively, $\boldsymbol{J}$ is the moment of inertia matrix, $\bar{q}$ is the dynamic pressure, S is the reference area, b is the wingspan, $\bar{c}$ is the average aerodynamic chord length, C_l, C_m, C_n are rolling, pitching, and yawing moment coefficients, C_l^{base}, C_m^{base}, C_n^{base} are three-axis moment coefficients generated by the airframe, and C_l^u, C_m^u, C_n^u are three-axis moment coefficients contributed by control inputs. The moment coefficients generated by the frame can be usually expressed as:

$$\begin{cases} C_l^{base} = C_l^{\alpha}(\alpha, Ma) + C_l^{\beta}(\alpha, \beta, Ma) \\ \qquad + \dfrac{pb}{2V}C_l^{p}(\alpha, Ma) + \dfrac{rb}{2V}C_l^{r}(\alpha, Ma) \\ C_m^{base} = C_m^{\alpha}(\alpha, Ma) + C_m^{\beta}(\alpha, \beta, Ma) \\ \qquad + \dfrac{q\bar{c}}{2V}C_m^{q}(\alpha, Ma) \\ C_n^{base} = C_n^{\alpha}(\alpha, Ma) + C_n^{\beta}(\alpha, \beta, Ma) \\ \qquad + \dfrac{pb}{2V}C_n^{p}(\alpha, Ma) + \dfrac{rb}{2V}C_n^{r}(\alpha, Ma) \end{cases} \tag{18}$$

where α is the angle of attack, β is the side slip angle, Ma is the Mach number, V is the airspeed. Eq. 17 and Eq. 18 show that the dynamics of the tailless aircraft have strong nonlinearity and strong three-axis couplings. The moment coefficients contributed by control inputs can be represented as:

$$\begin{bmatrix} C_l^u \\ C_m^u \\ C_n^u \end{bmatrix} = \boldsymbol{B} \cdot \boldsymbol{u} = \begin{bmatrix} C_{l_u} \\ C_{m_u} \\ C_{n_u} \end{bmatrix} \cdot \boldsymbol{u} \tag{19}$$

where $\boldsymbol{B}$ is the control effectiveness matrix, $\boldsymbol{u} \in \mathbb{R}^{13\times 1}$ is the control inputs including 11 control surfaces and 2 thrust vectors, and C_{l_u}, C_{m_u}, C_{n_u} are the derivatives of the three-axis moment coefficients to the control inputs respectively.

The kinematic equations about the rolling angle ϕ, the angle of attack α, and the side slip angle β can be written as follows:

$$\dot{\boldsymbol{x}} = h(\boldsymbol{x}, \xi) + k(\boldsymbol{x}, \xi)\,\boldsymbol{\omega} \tag{20}$$

where $\boldsymbol{x} = [\phi\ \alpha\ \beta]^{\mathrm{T}}$ is the controlled states, $\boldsymbol{\omega} = [p\ q\ r]^{\mathrm{T}}$ is the three-axis angle velocity, ξ is the states of the ICE plane. Besides, in Eq. 20:

$$\begin{cases} h(\boldsymbol{x}, \xi) \\ \quad = \begin{bmatrix} 0 \\ \dfrac{1}{u^2+w^2}[uf_z - wf_x] \\ \dfrac{1}{\sqrt{u^2+w^2}}\left[\dfrac{1}{\sqrt{u^2+w^2}} + \left(1-\dfrac{vv}{V^2}\right)f_y - \dfrac{vw}{V^2}f_z\right] \end{bmatrix} \\ k(\boldsymbol{x}, \xi) = \begin{bmatrix} 1 & \sin\phi\tan\theta & \cos\phi\tan\theta \\ \dfrac{-uv}{u^2+w^2} & 1 & \dfrac{-vw}{u^2+w^2} \\ \dfrac{w}{\sqrt{u^2+w^2}} & 0 & \dfrac{-u}{\sqrt{u^2+w^2}} \end{bmatrix} \end{cases}$$

where θ is the pitching angle of the aircraft, u, v, w are the portion of the aircraft airspeed in body axes, f_x, f_y, f_z are the triaxial specific force of the aircraft caused by the aerodynamic force and thrust, i.e., the triaxial acceleration of the aircraft without the influence of gravity, and g is the gravitational acceleration.

B. CONTROL REQUIREMENTS ANALYSIS AND PARAMETERS DESIGN FOR STABILITY AUGMENTATION SYSTEM

In order to enhance the stability of the ICE plane and facilitate the training of the IDHP controller, the stability augmentation system (SAS) is designed in this paper. In this paper, the attitude control of the tailless aircraft is studied under the condition of 9000m and 0.7Ma. Under this condition, the linearized ICE-101 has two positive eigenvalues, one larger and one smaller, as shown in Table 2. The unstable modes represented by these two eigenvalues have negative influence on the stability and the control of the ICE-101. If excited by the corresponding eigenvector of the dutch roll mode, the states of the ICE plane will diverge violently. Besides, when the reinforcement learning controller is firstly trained or initially used on the ICE plane, it needs a large state space for exploration and learning. If the states of the ICE plane are easy to diverge at certain state point, the immature reinforcement learning controller which has few experience is not able to provide effective control inputs and to pull the aircraft states back to the desired state trajectory. Therefore, in order to enhance the stability of the ICE-101 and to provide the reinforcement learning controller a wide range of exploration space, an SAS is designed for the unstable ICE-101 aircraft.

An SAS is designed for ICE-101 plant model to improve stability. It is developed using a pole placement method, and feedbacks full states to control inputs, as follows:

$$\begin{cases} \boldsymbol{u} = \boldsymbol{u}_0 + \Delta\boldsymbol{u}_x \\ \Delta\boldsymbol{u}_x = \boldsymbol{K}\boldsymbol{x} \end{cases} \tag{21}$$

TABLE 2. Dynamic modes and eigenvalues of the open-loop ICE-101 (9000m, 0.7Ma).

Modes	Short period	Long period	Dutch roll
Eigenvalues	$-0.8541 \pm 1.7572i$	$-0.0039 \pm 0.0327i$	0.8139
Modes	Roll subsidence	Spiral divergence	
Eigenvalues	$-0.8646 \pm 0.5116i$	0.0002	

TABLE 3. Desired eigenvalues of the ICE-101 with SAS (9000m, 0.7Ma).

Modes	Short period	Long period	Dutch roll
Eigenvalues	$-0.8576 \pm 1.7550i$	$-0.0040 \pm 0.0337i$	-0.5718
Modes	Roll subsidence	Spiral divergence	
Eigenvalues	$-0.8506 \pm 0.8528i$	-0.0029	

where $\boldsymbol{K}$ is the feedback gain. And the structure diagram of the SAS is shown in Fig. 3. Since the unstable modes are only existed along rolling and yawing axes, only stability enhancement along lateral directions is considered in the SAS. For using the full state feedback method, a set of desired eigenvalues are determined, as shown in Table 3. Then, the feedback matrix can be obtained after comparing the dynamics of the ICE aircraft with and without SAS, as follows:

$$K = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & -8.17 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 8.17 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & -115 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 115 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \tag{22}$$

If needed, the SAS should be designed carefully on all flying points within the flight envelop. But in this paper, the designed SAS is already enough. Because the motivation of designing an SAS before designing IDHP controller is to only ensure that the eigenvalues of the augmented aircraft are roughly located in the left-half complex plane.

C. IDHP AND SAS BASED HYBRID ATTITUDE CONTROL SYSTEM

In order to enhance the control capability of the tailless aircraft which have strong nonlinearity and strong three-axis couplings, an IDHP based attitude control system is designed in this paper. This system mainly incorporates three parts, the SAS, the inner loop controller, and the outer loop controller, as shown in Fig. 4. Firstly, as shown in Sec. III-B, the SAS is designed to provide an augmented stable plant. Then, a multi-channel IDHP controller and a control allocation module is used to control the angular rate. This IDHP based controller is able to adapt the dynamic characteristics of the controlled

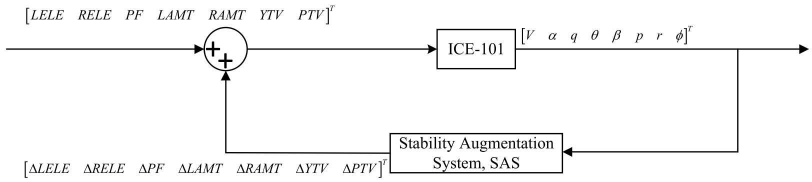

FIGURE 3. SAS design on the ICE-101 aircraft.

plant. Finally, the outer loop control is performed by a nonlinear dynamic inversion (NDI) controller.

1) IDHP BASED THREE-AXIS ANGULAR RATE CONTROLLER

The classic SISO IDHP controller is expanded into a MIMO controller in this paper for controlling the nonlinear tailless aircraft with SAS, as shown in Fig. 5. This multi-channel IDHP controller is able to deal with the strong nonlinearity and the strong three-axis couplings existing in the aircraft dynamics and to reject model uncertainties or disturbances to some extent. The actor-critic updating rules of the proposed MIMO IDHP controller can still be represented as Eq. 12 and Eq. 15. In classic IDHP controller, both the actor and the critic networks have only one input and one output, like the examples in Refs. [6], [14]. In this paper, a MIMO neural network is chosen and optimized to design the actor. This network has three inputs, namely the tracking error of rolling, pitching, and yawing angular rate. The outputs of the actor are the demands of three-axis moment coefficients. Similarly, the critic network has also been designed with three inputs which are the same as the inputs of the actor, and has three outputs, namely the derivatives of the optimal cost function with regard to the rolling, pitching, and yawing angular velocities respectively. In this multi-channel IDHP controller, the incremental model uses the demand of the three-axis moment coefficients and the aircraft states measured by sensors to update its parameters in real time.

In the proposed IDHP based angular rate controller, the critic network has only one hidden layer and one output layer. The hidden layer has 6 neurons, and the activation function of these neurons is the hyperbolic tangent function. The output layer has three neurons, and the activation function is the linear function. Therefore, the outputs of critic can be expressed as follows:

$$\begin{bmatrix} \partial V/\partial p \\ \partial V/\partial q \\ \partial V/\partial r \end{bmatrix} = critic\left(\begin{bmatrix} e_p \\ e_q \\ e_r \end{bmatrix}\right) \tag{23}$$

where $critic(\cdot)$ stands for mapping relation of critic network, and details can be found in [21]. Similarly, the actor network has one hidden layer and one output layer. The hidden layer has 4 neurons and the output layer has 3 neurons. These two layers both use hyperbolic tangent function as their activation function. The mapping function of actor is as follows:

$$\begin{bmatrix} \partial V/\partial p \\ \partial V/\partial q \\ \partial V/\partial r \end{bmatrix} = actor\left(\begin{bmatrix} e_p \\ e_q \\ e_r \end{bmatrix}\right) \tag{24}$$

where $actor(\cdot)$ stands for mapping relation of actor network which can be found in [21]. The initial weights in the actor and the critic are random numbers being evenly distributed between -0.01 and 0.01. These two networks use gradient descent method to update their weights. The learning rate of the actor and the critic network are 0.3 and 0.08 respectively. A smaller critic learning rate is chosen here for a stable learning process and a larger actor learning rate is for quickly updating control law. It is better to update the actor more quickly than the critic otherwise a convergent control law is hard to be obtained under changeable cost function. In addition, the forgetting factor γ in Eq. 7 is 0.4 for focusing on a suitable future time range, and the positive definite matrix $\boldsymbol{Q}$ in Eq. 8 is $diag(5, 5, 8)$ in order to emphasize yaw-channel control which is weak in ICE. When identifying the incremental model, the forgetting factor κ in RLS method is set to 0.99 since a quick response to plant change is required.

Before being applied to the ICE plane, the designed multi-channel IDHP controller needs a priori in-loop training. This is because, for a two-layer neural network, the number of weights in the neural network will be increased several times with the addition of input or output port [7]. Therefore, the neural networks in IDHP based MIMO controller have several times as many neurons as the SISO IDHP controller. If the MIMO IDHP controller is used directly on the controlled plant, the weights in this controller will need a long process to converge to a reasonable range or the MIMO IDHP controller can not work at all. The second reason is that the ICE-101 aircraft is complex. The IDHP controller cannot adapt it within a short time period. To summarize, the designed IDHP based MIMO controller needs an adequate training before using.

A priori training process of the designed IDHP controller is shown in Fig. 6. Some parameters in IDHP controller, like the learning rate and forgetting factor, have significant influence on the control accuracy of the whole controller. Smaller values of these parameters will induce a lower control accuracy but will also relieve the learning burden of the IDHP controller. Therefore, at the beginning of the pre-training process, these parameters are set to lower values in order to make the IDHP controller easy to converge. After the first

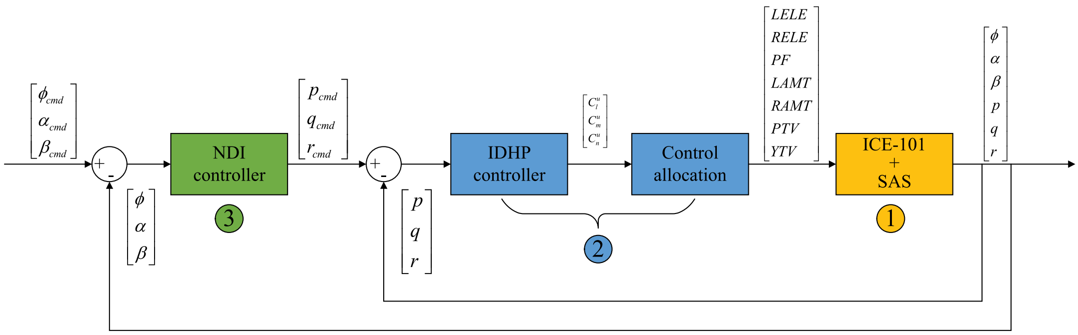

FIGURE 4. IDHP based attitude control system for ICE-101.

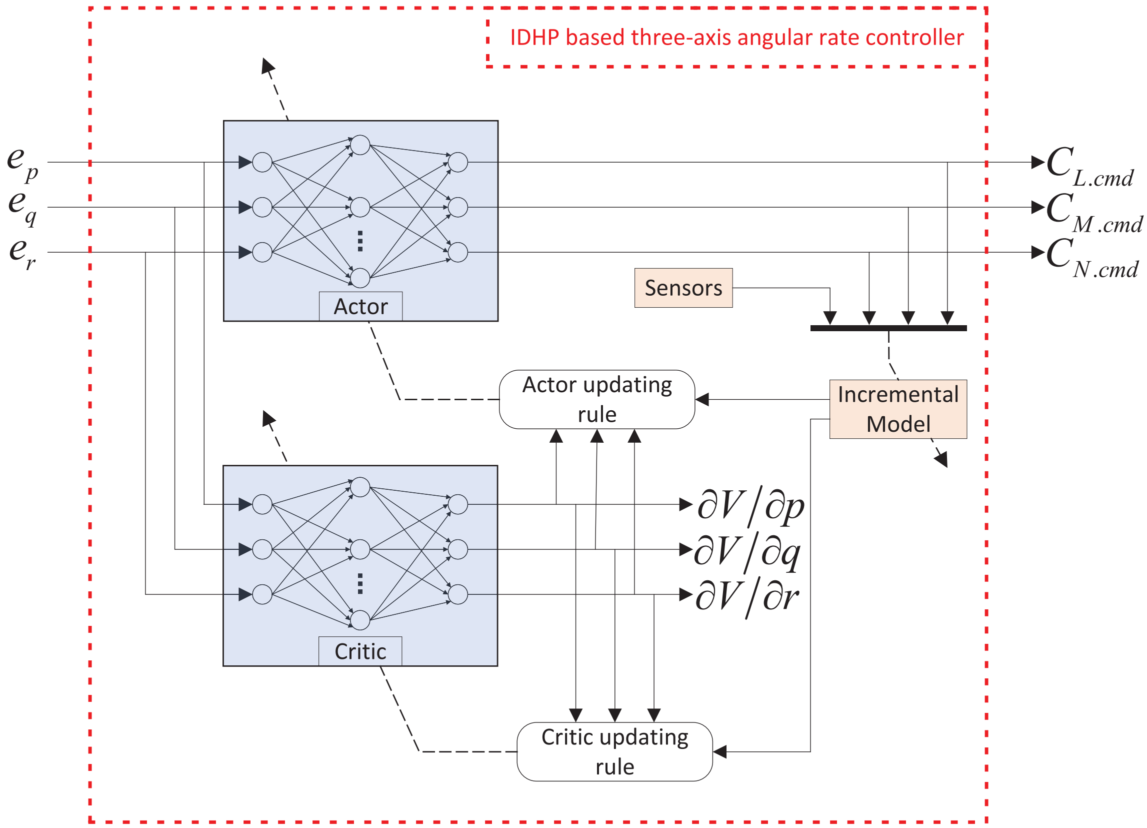

FIGURE 5. IDHP based three-axis angular rate controller.

convergence, a group of actor-critic networks weights can be easily obtained and recorded. These weights will be used as the initial weights of the next training. In the whole process of pre-training, the weights obtained from last training will be used as the initial weights of the next training. Moreover, at the beginning of each new training, the parameters in IDHP controller will be adjusted to a bigger value in order to achieve a higher level of control accuracy. In this way, the control performance of the IDHP controller can be gradually improved as the training goes on. But it should be noted that the parameters should be increased properly in order to balance the control accuracy and the anti-disturbance ability. In this paper, the whole process contains 15 trainings. Every training lasts for 100s and the sample rate is 100Hz. After the training process, the IDHP controller can be used on the tailless aircraft. When the IDHP controller is used, the learning rates of both the actor and the critic can be appropriately set to smaller values than the learning rates used in training process in order to enhance robustness.

2) ACTIVE SET METHOD BASED CONTROL ALLOCATION

After designing the IDHP controller, a control allocation module is necessarily needed since the ICE-101 plane researched in this paper is of redundant control surfaces and thrust vectors [25]. The control allocation module is used to allocate the three-axis moment coefficient commands

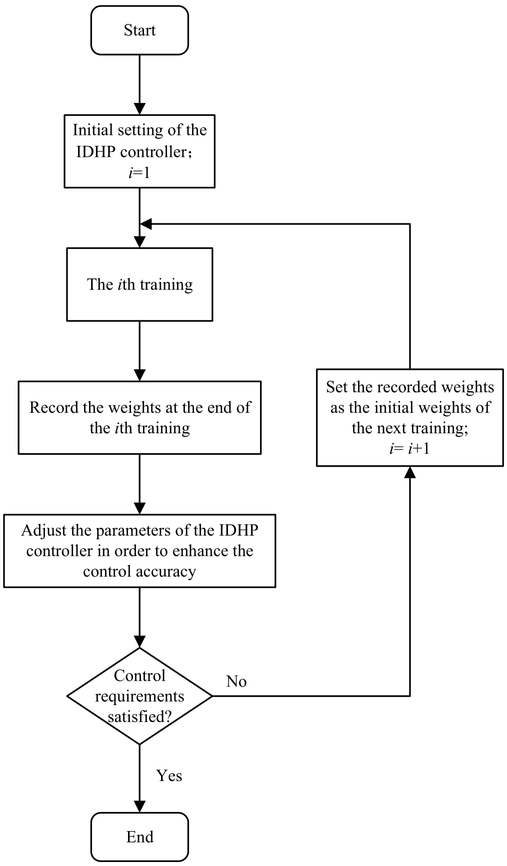

FIGURE 6. The pre-training process of the IDHP based MIMO angular rate controller.

generated by IDHP inner loop controller to every active control input, which, in this paper, are LELE, RELE, PF, LAMT, RAMT, YTV, and PTV.

Generally, the dynamics of the actuators can be ignored when compared to the dynamics of the ICE plane since the actuators respond more quickly. Under such an assumption, the control allocation of the ICE plane is to obtain the control inputs command in real time to satisfy the required three-axis moment coefficients. In this solving process, the limits and speed limits of the actuator of each control input are considered. The control allocation problem can be expressed as follows:

$$\boldsymbol{v} = \boldsymbol{B}\boldsymbol{u} \tag{25}$$

where $\boldsymbol{v} \in \mathbb{R}^n$ is the required three-axis moment coefficients, $\boldsymbol{u} \in \mathbb{R}^m$ is the control inputs, and $\boldsymbol{B} \in \mathbb{R}^{n\times m}$ is the control effectiveness matrix. The position limits and the speed limits of the control inputs can be expressed as follows:

$$\begin{cases} u_i^{min} \le u_i \le u_i^{max} \\ \dot{u}_i^{min} \le \dot{u}_i \le \dot{u}_i^{max} \end{cases} \tag{26}$$

where u_i and $\dot{u}_i$ are the deflection and the deflection speed of the ith control input respectively, as in Table 1. For the 6-DoF ICE plane model in this paper, there are three required moment coefficients to be allocated, i.e., rolling, pitching, and yawing moment coefficients. Therefore, the dimension of the vector v in Eq. 25 is 3. Since 7 control inputs are used in this paper, the dimension of the vector u in Eq. 25 is 7.

The active set method is chosen to solve the control allocation problem expressed by the Eq. 25 because of its small computational cost and convenience to consider control limit and speed limit of every control input. Before using the active set method, the control allocation problem of the ICE plane should be transfered to a standard problem on which the active set method can be directly applied. For this reason, the problem expressed by Eq. 25 and Eq. 26 can be firstly reconstructed as the following weighted least squares (WLS) formulation of the control allocation problem [26]:

$$\begin{aligned} \min_{\underline{\boldsymbol{u}}\le\boldsymbol{u}\le\overline{\boldsymbol{u}}} J &= \|\boldsymbol{W}_u(\boldsymbol{u}-\boldsymbol{u}_d)\|^2 + \gamma\,\|\boldsymbol{W}_v(\boldsymbol{B}\boldsymbol{u}-\boldsymbol{v})\|^2 \\ &= \left\| \begin{pmatrix} \sqrt{\gamma}\boldsymbol{W}_u\boldsymbol{B} \\ \boldsymbol{W}_u \end{pmatrix} \boldsymbol{u} - \begin{pmatrix} \sqrt{\gamma}\boldsymbol{W}_v\boldsymbol{v} \\ \boldsymbol{W}_u\boldsymbol{u}_d \end{pmatrix} \right\|^2 \end{aligned} \tag{27}$$

where $\underline{\boldsymbol{u}}$ and $\overline{\boldsymbol{u}}$ are the lower and upper bound of control inputs which is determined by Eq. 26 and Table 1, $\boldsymbol{W_u}$ is a positive weighting matrix on which the importance of each control input to the control allocation depends, $\boldsymbol{u}_d$ is the desired deflection of control inputs, $\boldsymbol{W}_v$ is a positive weighting matrix for required moment coefficients to be allocated, and γ is a weighting factor which is used to adjust the importance of each item in Eq. 27. In this paper, the desired deflections of control inputs are set to the actual deflections in last moment. For the WLS problem described by Eq. 27, the active set method can be directly used to obtain the optimal control inputs commands. The algorithm and details of this method can be found in Ref. [26].

An effective real-time control allocation module can be designed by using the aforementioned active set method, which is able to deal with the control allocation problem caused by redundant control inputs of the tailless plane. Besides, the physical limits of the actuator of each control input, like the position limit and the speed limit, can be easily added in this active set based control allocation module. In this paper, the position limits of each control inputs are shown in Table 1. The weighting matrix $\boldsymbol{W}_u$ is set to $diag(5, 5, 5, 5, 5, 0.1, 0.1)$, and the weighting matrix $\boldsymbol{W}_v$ is set to $diag(1, 1, 1)$. The weighting factor λ in Eq. 27 is 10^6.

3) NDI BASED THREE-AXIS ANGLE CONTROLLER

In this paper, the NDI method is chosen to design an outer-loop angle controller since it is able to deal with the outer loop control problem and to have a satisfactory control performance. The outer-loop NDI controller can produce effective angular rate commands for inner-loop IDHP controller to track, which is helpful for the learning and adapting of the IDHP controller.

The states in angle loop to be controlled are ϕ, α, and β. The kinematic equations about the aforementioned three angles are expressed in Eq. 20. Since the control inputs $\boldsymbol{\omega}$ are explicitly existed in the Eq. 20, the nonlinear item in this

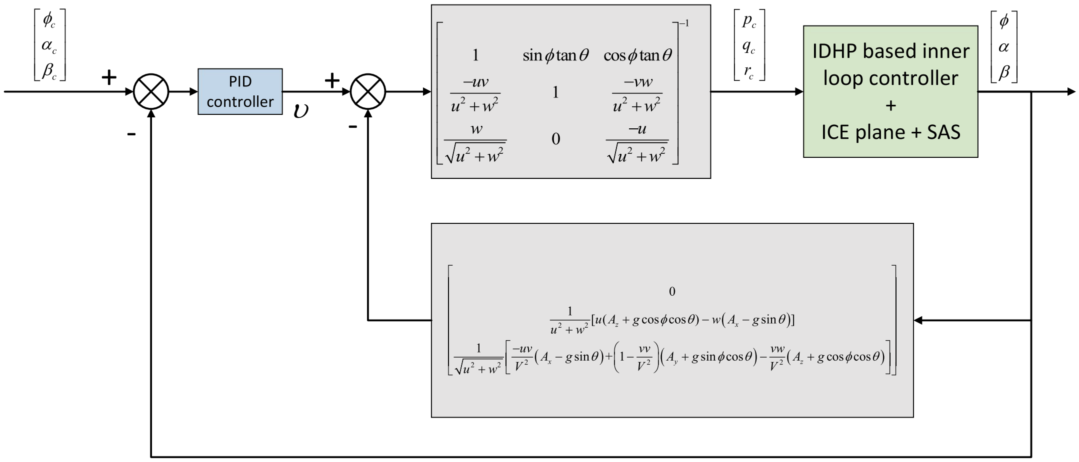

FIGURE 7. The outer loop NDI controller of the attitude control system.

equation can be directly fed back when designing the NDI controller. If using NDI method, the control signals can be expressed as follow:

$$\boldsymbol{\omega} = k\left(\boldsymbol{x}, \xi\right)^{-1}\left(\boldsymbol{v} - h\left(\boldsymbol{x}, \xi\right)\right) \tag{28}$$

where, $\boldsymbol{v}$ is the designed virtual control inputs. Submitting Eq. 28 into Eq. 20, a linear system can be obtained, as follows:

$$\dot{\boldsymbol{x}} = \boldsymbol{v} \tag{29}$$

For the linear system expressed by Eq. 29, some simple linear control method can be used on it. In this paper, PID controller is chosen to control the linear system. Synthesizing Eq. 20, Eq. 28-Eq. 29, the structure of the designed NDI controller can be easily obtained, as shown in Fig. 7.

IV. RESULTS AND ANALYSIS

In order to verify the designed attitude control system based on the IDHP method, a 6-DoF nonlinear ICE-101 aircraft model is used to perform simulation tests. For validating the adaptation ability of the control system designed in this paper, reference tracking task is performed under both nominal case and sudden model uncertainties or disturbances.

A. REFERENCE TRACKING PERFORMANCE UNDER NOMINAL CASE

In this paper, a reference tracking task under nominal case is firstly performed to validate the effectiveness of the designed control system. The step signal is used as the command signal in this task because it is representative. When tracking the step command, the performance of the designed control system can be evaluated from a time domain perspective.

The tracking of the attitude angle is shown in Fig. 8. The nonlinear ICE plane with the IDHP based control system is able to response to the command quickly and has small tracking error. The side slip angle only has a small fluctuation when the roll angle and the angle of attack is stepped. The change of the side slip angle is less than 0.5 (deg). Besides, the roll angle only has a small change when the angle of attack is stepped and vice versa. To summarize, the IDHP based control system has a good performance on the angle tracking.

The angular rate tracking performance of the designed IDHP based control system can be given in Fig. 9. Under the control of the IDHP based inner-loop controller, the angular rate of the ICE plane in three-axis directions can closely track the corresponding command signals generated by the outer-loop NDI controller. When the angle of attack is stepped, the yaw rate and the roll rate change no more than 2 (deg/s). When the roll angle is stepped, the pitch rate changes a little but the yaw rate variates about 3 (deg/s). It is because that the coupling between the roll and yaw axes are more significant than the one between the roll and pitch axes. The good tracking of the yaw rate command shows that the IDHP controller is able to deal with the three-axis couplings existed in the tailless aircraft. To summarize, the designed IDHP based multi-channel angular rate controller accomplishes the angular rate tracking task very well with strong nonlinearity and strong three-axis coupling considered.

The commands of three-axis moment coefficients are given by the IDHP controller and are satisfied through control allocation module, as shown in Fig. 10. In order to fulfill the moment commands, seven control effectors are used as shown in Fig. 11. The usages of these control effectors are within limits shown in Table 1. The nonlinearity and three-axis coupling are more serious in ICE plane than in common plane. When tracking attitude angle command along one axis, the moments in other two axes are also needed simultaneously. Because the attitude angles in these two axes will be affected and need to be adjusted to their own command. In order to satisfy required three-axis moments, all seven control effectors deflect a lot, especially the PTV and YTV which can only provide pitch moment and yaw moment respectively.

The accuracy of the incremental model is important for the inner-loop IDHP controller. This model is used to provide important dynamic information for the update of the actor-critic networks. When the incremental model is used to predict three-axis angular velocities through Eq. 5, residuals

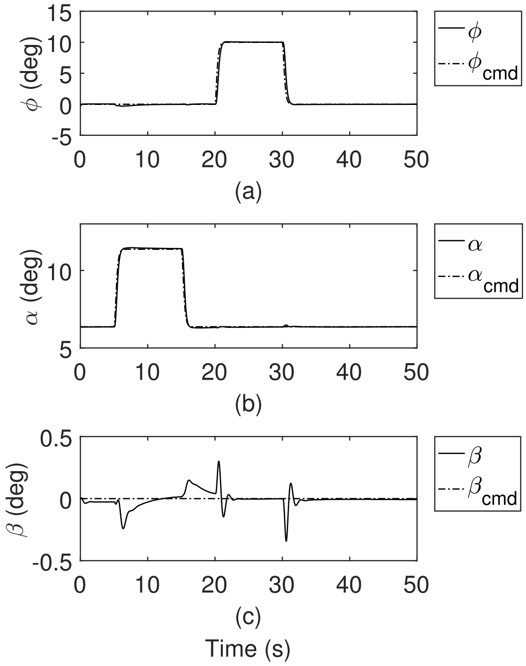

FIGURE 8. Step command tracking performance under nominal case.

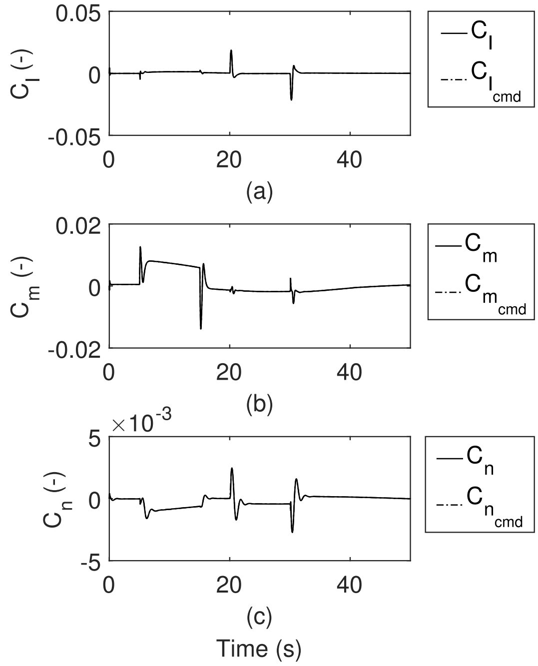

FIGURE 10. Changes of three-axis moment coefficients under nominal case.

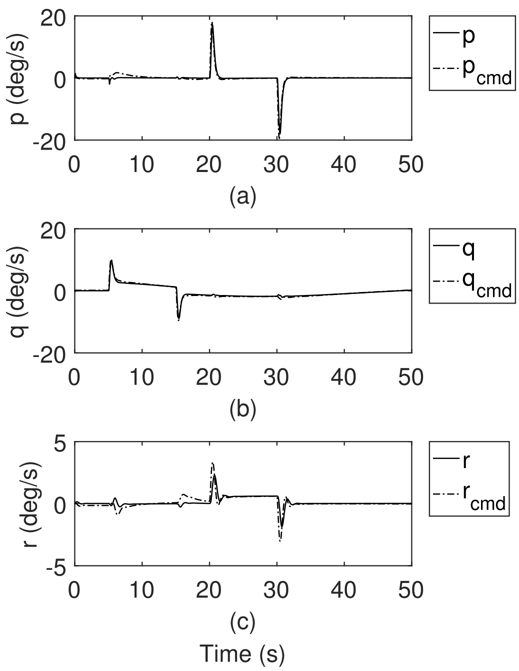

FIGURE 9. Angle velocity tracking performance under nominal case.

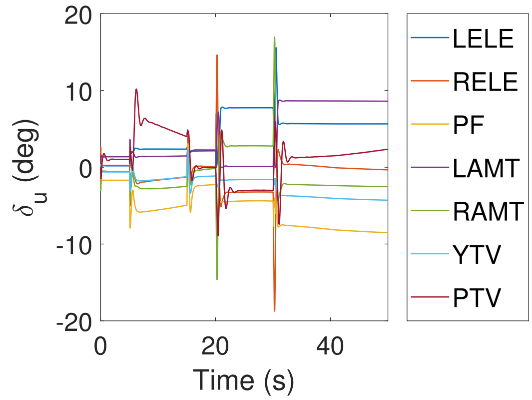

FIGURE 11. Deflections of the control effectors under nominal case.

between the predictions and the corresponding real values are shown in Fig. 12. The residuals are quite small compared to the true values shown in Fig. 9. At the moment when the attitude angle along a certain axis is stepped, the residuals are increased but within ±0.15 (deg/s). And they will soon decrease to zero. Besides, the three-axis angular velocities predictions provided by the incremental model can be shown in Fig. 13. The estimated states by incremental model are almost coincide with true values and the incremental model is therefore proved to be effective. In conclusion, the incremental model is able to provide accurate three-axis angular velocities for actor-critic networks, and then contribute to the good performance of the designed inner-loop IDHP controller.

The changes of the weights in the actor-critic neural networks can be given in Fig. 14 and Fig. 15. At the moment when the attitude angle along a certain axis is stepped, the actor-critic weights change a lot. It is because that the IDHP controller can explore more broad state space at the step moments than common moments when states are steady. Actor-critic networks will try to adjust its weights when more information comes, which indicates that the actor and

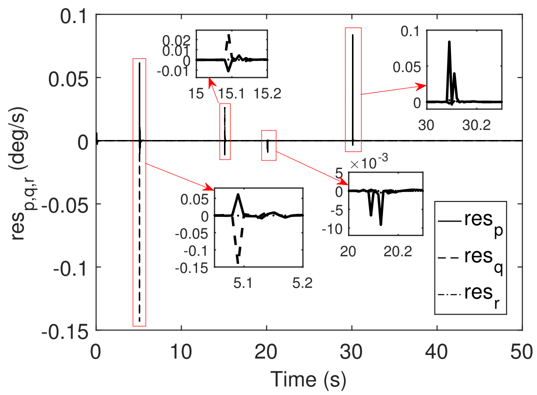

FIGURE 12. Residuals of three-axis angular velocities prediction under nominal case.

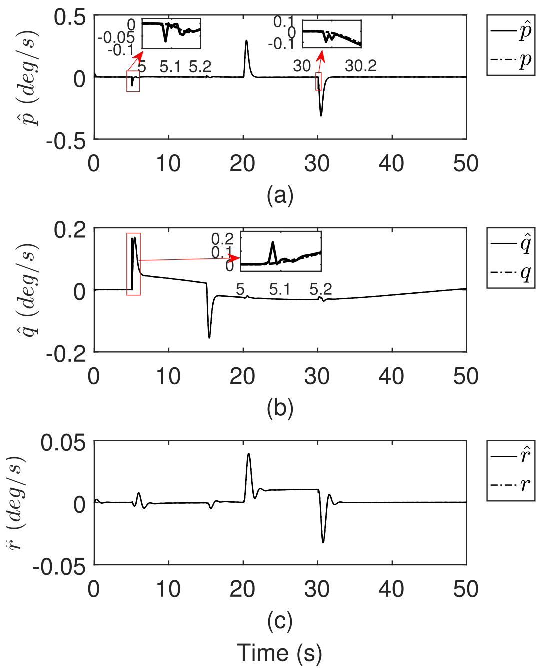

FIGURE 13. Predictions of three-axis angular velocities under nominal case.

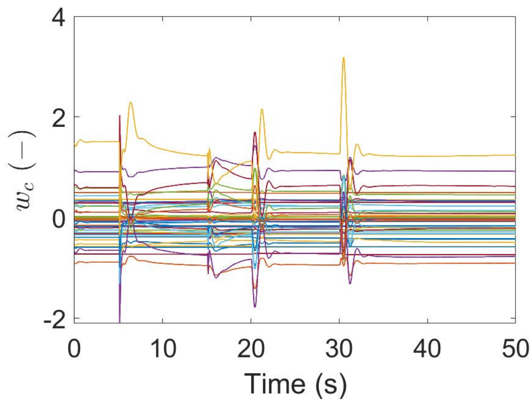

FIGURE 14. Critic weights change under nominal case.

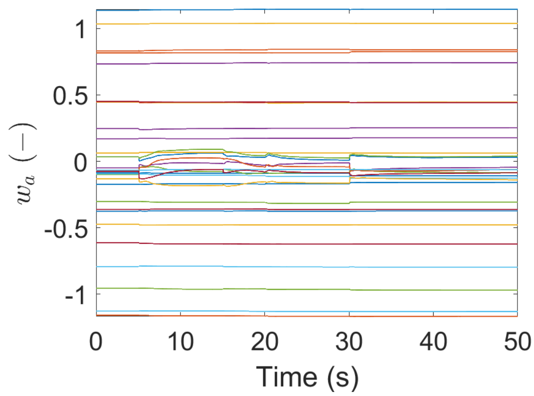

FIGURE 15. Actor weights change under nominal case.

the critic are continuously performing online learning in the control process.

B. REFERENCE TRACKING PERFORMANCE UNDER SUDDEN MODEL UNCERTAINTIES OR DISTURBANCES

In order to verify the adaptability of the proposed IDHP based control system, sudden model uncertainties or disturbances are added when the ICE plane tracks the reference commands. A large elastic deformation of the wings or the damage of a certain control effector will cause the reduction of the lift and the pitching moment of the aircraft. If the reduction is too excessive, it may cause a large attitude change of the aircraft and even cause the aircraft out of control. To simulate the effect of sudden model uncertainties, the lift and the pitching moment generated by the angle of attack are reduced by 40% after 10s. And the control effectiveness of the pitch flap is also reduced by 40% at the same time to simulate the damage of a part of the pitch flap. Under such a bad flight situation, the designed IDHP based control system is validated on the nonlinear ICE model.

The attitude angle tracking performance of the ICE plane under sudden uncertainties or disturbances is as good as the tracking under nominal case, as shown in Fig. 16. At 10s when sudden uncertainties occur, as the red line in Fig. 16 indicates, the lift and the pitching moment of the ICE plane are reduced abruptly. The angle of attack changes about 0.2 (deg) due to the reduction. Then, the designed nonlinear IDHP based control system can adjust it to the angle command in 0.5s, as shown in Fig. 16 (b). At 20s and 30s, the angle of attack changes no more than 0.2 (deg) due to the step of the roll angle and gradually returns to its command vicinity. The changes are a little larger than those under nominal case but are still within reasonable range. Because the IDHP based controller is adapting the change that the plane can not provide the same lift and pitching moment as before because of the efficiency decrease. During the whole task, the variation of the side slip angle is within ± 0.3 (deg). To summarize, when the baseline aerodynamic efficiency of the ICE plane and the control effectiveness of the pitch flap

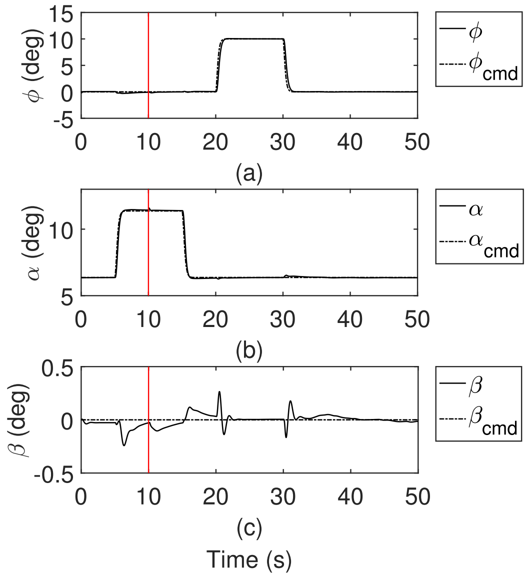

FIGURE 16. Step command tracking performance under sudden uncertainties or disturbances.

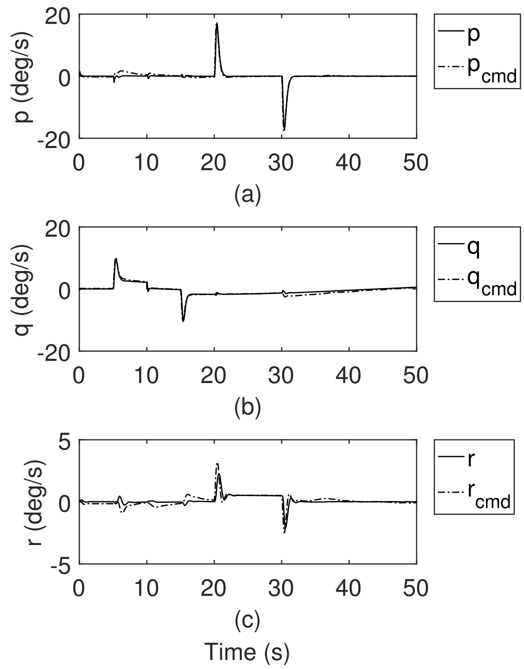

FIGURE 17. Angle velocity tracking performance under sudden uncertainties or disturbances.

decrease by 40%, the designed IDHP based attitude angle control system still has an excellent performance on the step command tracking task.

The angle velocity tracking performance by the inner-loop IDHP controller can be given in Fig. 17. The angle velocities in the rolling and pitching axes are closely coincide with the command provided by the NDI controller. The angle velocity tracking performance in the yawing axis is not as good as in the other two axes, but is still satisfactory enough. The strong nonlinearity and the strong three-axis coupling are mainly

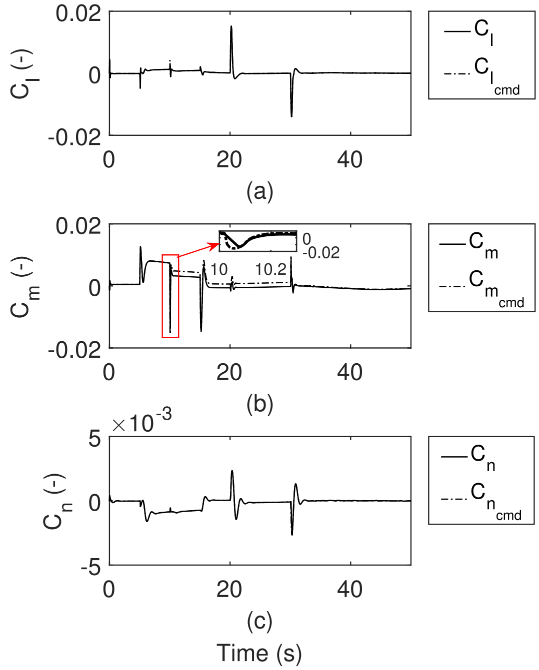

FIGURE 18. Changes of three-axis moment coefficients under sudden uncertainties or disturbances.

existed in the inner-loop dynamic equations. Therefore, the good angular velocities tracking performance demonstrates that the angular rate controller, i.e. the IDHP controller in the inner-loop, has an excellent ability to deal with control problems caused by the nonlinearity and the three-axis coupling of the ICE aircraft.

The three-axis moment coefficients commands can be well satisfied before the sudden uncertainties but not after them, as shown in Fig. 18. After the sudden uncertainties, the control effectiveness of the pitch flap decreases and is not known by the control allocation module. The control allocation module in this paper is not able to update the constant control effectiveness matrix used in this module. Therefore, after 10s, the control allocation module uses a wrong matrix to allocate the desired three-axis moments, and then can not provide good control effectors commands to satisfy the desired moments, as shown in Fig. 18 and Fig. 19.

However, as shown in Fig. 16, the attitude angles can track their commands closely. Because the IDHP controller is able to adapt the sudden uncertainties and to counter the mistake existed in the control allocation module. In about 1s after the sudden uncertainties, the IDHP adapts itself using the information induced by the sudden excitation, as shown in Fig. 20 and Fig. 21. It gives a sudden change of the pitching moment at 10s in order to counter the sudden uncertainties. After the sudden uncertainties, the IDHP controller tries to give a larger pitching moment command to induce a proportional pitching moment, which can guarantee that the ICE plane have enough pitching moment to perfectly finish the angle tracking task. To summary, the IDHP controller in the inner loop is able

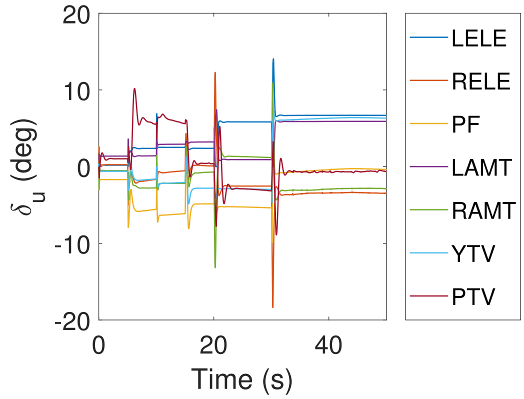

FIGURE 19. Deflections of the control effectors under sudden uncertainties or disturbances.

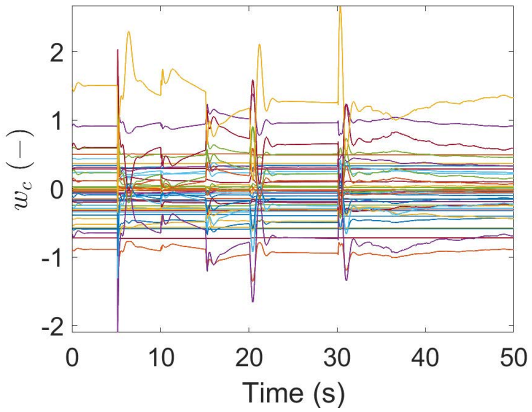

FIGURE 20. Critic weights change under sudden uncertainties or disturbances.

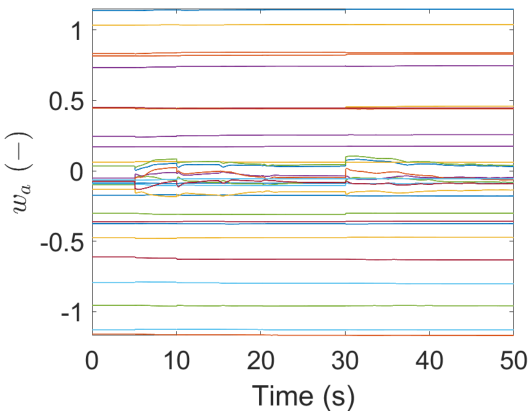

FIGURE 21. Actor weights change under sudden uncertainties or disturbances.

to make the whole three-axis attitude angle control system adaptive and to deal with control problems caused by sudden uncertainties or disturbances.

V. CONCLUSION

High-performance aircraft such as tailless aircraft commonly have strong nonlinearity and strong three-axis coupling characteristics, and even have unstable modes. The contribution of this paper is to propose a novel IDHP based three-axis attitude controller design method for circumventing these problems. The core of this controller is a hybrid IDHP/SAS based angular rate controller. Besides, the attitude controller also includes an NDI based angular controller, and a control allocation module. The IDHP method is used to design a controller for an unstable high-performance aircraft with control input redundancy. And MIMO form network structures are designed and optimized for the actors and critics considering the three-channel coupling issues. To address the unstable mode problem associated with the IDHP type method, the SAS is embedded in the controller for enhancing the stability of the ICE-101 plane and providing a wider exploration space for the IDHP controller. The synthesis of the IDHP with the SAS facilitating the training of the IDHP controller. Besides, a pre-training method is proposed in this paper. The pre-training process can help the proposed multi-channel IDHP controller adapt the controlled plant before using and therefore guarantee the safety of the controlled nonlinear aircraft. The proposed IDHP based control system is validated on a 6-DoF ICE-101 model and is proved to be able to achieve satisfactory performance both in nominal case and under sudden model uncertainties.

REFERENCES

[1] Z. S. Hou, H. J. Gao, and F. L. Lewis, "Data-driven control and learning systems," *IEEE Trans. Ind. Electron.*, vol. 64, no. 5, pp. 4070–4075, May 2017.
[2] Z.-S. Hou and Z. Wang, "From model-based control to data-driven control: Survey, classification and perspective," *Inf. Sci.*, vol. 235, pp. 3–35, Jun. 2013.
[3] D. V. Prokhorov and D. C. Wunsch, "Adaptive critic designs," *IEEE Trans. Neural Netw.*, vol. 8, no. 5, pp. 997–1007, Sep. 1997.
[4] G. K. Venayagamoorthy, R. G. Harley, and D. C. Wunsch, "Comparison of heuristic dynamic programming and dual heuristic programming adaptive critics for neurocontrol of a turbogenerator," *IEEE Trans. Neural Netw.*, vol. 13, no. 3, pp. 764–773, May 2002.
[5] J. Si and Y.-T. Wang, "Online learning control by association and reinforcement," *IEEE Trans. Neural Netw.*, vol. 12, no. 2, pp. 264–276, Mar. 2001.
[6] Y. Zhou, E. J. van Kampen, and Q. P. Chu, "Incremental model based online dual heuristic programming for nonlinear adaptive control," *Control Eng. Pract.*, vol. 73, pp. 13–25, Apr. 2018.
[7] R. Enns and J. Si, "Helicopter trimming and tracking control using direct neural dynamic programming," *IEEE Trans. Neural Netw.*, vol. 14, no. 4, pp. 929–939, Aug. 2003.
[8] Y. Tang, H. He, J. Wen, and J. Liu, "Power system stability control for a wind farm based on adaptive dynamic programming," *IEEE Trans. Smart Grid*, vol. 6, no. 1, pp. 166–177, Jan. 2015.
[9] X. Han, Z. Zheng, L. Liu, B. Wang, Z. Cheng, H. Fan, and Y. Wang, "Online policy iteration ADP-based attitude-tracking control for hypersonic vehicles," *Aerosp. Sci. Technol.*, vol. 106, Nov. 2020, Art. no. 106233.
[10] K. Shayan and E.-J. Van Kampen, "Online actor-critic-based adaptive control for a tailless aircraft with innovative control effectors," in *Proc. AIAA Scitech Forum*, 2021, p. 0884.
[11] A. Tsourdos, I. A. Dharma Permana, D. H. Budiarti, H.-S. Shin, and C.-H. Lee, "Developing flight control policy using deep deterministic policy gradient," in *Proc. IEEE Int. Conf. Aerosp. Electron. Remote Sens. Technol. (ICARES)*, Oct. 2019, pp. 1–7.
[12] Y. Zhou, E.-J. van Kampen, and Q. Chu, "Incremental approximate dynamic programming for nonlinear adaptive tracking control with partial observability," *J. Guid., Control, Dyn.*, vol. 41, no. 12, pp. 2554–2567, Dec. 2018.

[13] Y. Zhou, E.-J. van Kampen, and Q. Chu, "Incremental model based online heuristic dynamic programming for nonlinear adaptive tracking control with partial observability," *Aerosp. Sci. Technol.*, vol. 105, Oct. 2020, Art. no. 106013.
[14] H. Li, L. Sun, W. Tan, B. Jia, and X. Liu, "Switching flight control for incremental model-based dual heuristic dynamic programming," *J. Guid., Control, Dyn.*, vol. 43, no. 7, pp. 1352–1358, Jul. 2020.
[15] P. M. Dias, Y. Zhou, and E. J. Van Kampen, "Intelligent nonlinear adaptive flight control using incremental approximate dynamic programming," in *Proc. AIAA Scitech Forum*, 2019, p. 2339.
[16] S. Heyer, D. Kroezen, and E.-J. Van Kampen, "Online adaptive incremental reinforcement learning flight control for a CS-25 class aircraft," in *Proc. AIAA Scitech Forum*, 2020, p. 1844.
[17] F.-Y. Wang, H. Zhang, and D. Liu, "Adaptive dynamic programming: An introduction," *IEEE Comput. Intell. Mag.*, vol. 4, no. 2, pp. 39–47, May 2009.
[18] L. Sun, L. Shi, W. Tan, and X. Liu, "Flying qualities evaluation based nonlinear flight control law design method for aircraft," *Aerosp. Sci. Technol.*, vol. 106, Nov. 2020, Art. no. 106126.
[19] J. Liu, L. Sun, W. Tan, X. Liu, and G. Li, "Finite time observer based incremental nonlinear fault-tolerant flight control," *Aerosp. Sci. Technol.*, vol. 104, Sep. 2020, Art. no. 105986.
[20] D. E. Kirk, *Optimal Control Theory: An Introduction*. Chelmsford, MA, USA: Courier Corporation, 2012.
[21] E. J. van Kampen, Q. P. Chu, and J. A. Mulder, "Continuous adaptive critic flight control aided with approximated plant dynamics," in *Proc. AIAA Guid., Navigat., Control Conf. Exhibit*, vol. 5, 2006, pp. 2989–3016.
[22] M. A. Niestroy, K. M. Dorsett, and K. Markstein, "A tailless fighter aircraft model for control-related research and development," in *Proc. AIAA Modeling Simulation Technol. Conf.*, Jan. 2017, p. 1757.
[23] K. M. Dorsett and D. R. Mehl, "Innovative control effectors (ICE)," Wright Lab., Air Force Mater. Command, Tech. Rep. TR-WL-TR-96-3043, Jan. 1996.
[24] K. M. Dorsett, S. P. Fears, and H. P. Houlden, "Innovative control effectors (ICE) phase III," Wright Lab., Air Force Mater. Command, Tech. Rep. TR-WL-TR-3059, Aug. 1997.
[25] L. Sun, Q. Zhou, B. Jia, W. Tan, and H. Li, "Effective control allocation using hierarchical multi-objective optimization for multi-phase flight," *Chin. J. Aeronaut.*, vol. 33, no. 7, pp. 2002–2013, Jul. 2020.
[26] O. Härkegåkd, "Efficient active set algorithms for solving constrained least squares problems in aircraft control allocation," in *Proc. 41st IEEE Conf. Decis. Control (CDC)*, vol. 2, Dec. 2002, pp. 1295–1300.

HANGXU LI received the B.S. degree from the Nanjing University of Aeronautics and Astronautics, Nanjing, China, in 2017, and the M.Eng. degree from Beihang University, Beijing, China, in 2020. His research interests include adaptive control, parameter estimation, and intelligent control schemes.

LIGUO SUN received the B.S. and M.S. degrees in control and simulation direction from the College of Energy and Power Engineering, Nanjing University of Aeronautics and Astronautics, Nanjing, China, in 2008 and 2010, respectively, and the Ph.D. degree from the Control and Simulation Division, Faculty of Aerospace Engineering, Delft University of Technology, Delft, The Netherlands, in 2014. He is currently an Associate Professor with the School of Aeronautic Science and Engineering, Beihang University. His research interests include adaptive fault tolerant flight control, intelligent control algorithms, and trajectory optimization methods.

WENQIAN TAN received the Ph.D. degree in aerospace engineering from the Moscow Aviation Institute, Moscow, Russia, in 2008. She is currently an Associate Professor with the School of Aeronautic Science and Engineering, Beihang University, Beijing, China. Her current research interests include flight dynamics and control, man–machine systems, and flying qualities.

XIAOYU LIU received the B.S. degree from Beihang University, Beijing, China, in 2020, where she is currently pursuing the Ph.D. degree with the School of Aeronautic Science and Engineering. Her research interests include adaptive fault-tolerant control, intelligent control schemes, and pilot behavior modeling.

WEIGAO DANG received the B.S. degree from Beihang University, Beijing, China, in 2020, where he is currently pursuing the master's degree in engineering with the School of Aeronautic Science and Engineering. His research interests include adaptive fault-tolerant control and trajectory optimization.