Actor-Critic-Based Optimal Adaptive Control Design for Morphing Aircraft ★

Hanna Lee * **Seong-hun Kim** ** **Youdan Kim** ***

* *Department of Aerospace Engineering, Seoul National University, Seoul, Korea, (e-mail: hn.lee@snu.ac.kr)*
** *Department of Aerospace Engineering, Seoul National University, Seoul, Korea, (e-mail: bgbgof@snu.ac.kr)*
*** *Department of Aerospace Engineering, Seoul National University, Seoul, Korea, (e-mail: ydkim@snu.ac.kr)*

Abstract: An online actor-critic-based control design strategy is proposed for a variable span and sweep morphing wing aircraft considering the morphing parameters as control effectors, which makes the system non-affine in control. By adopting the dynamic property of the morphing system, the augmented morphing aircraft system is formulated to be affine in control input. Through the online actor-critic-based control design for the augmented system, the proposed method has an advantage in terms of control design for the non-affine complex system with uncertainty, because the time-varying internal dynamic model caused by the morphing system is not required. From the augmented dynamic model, the control input frequency constraints of the morphing system, which are generally considered slow can also be addressed. Numerical simulation is performed to demonstrate the effectiveness of the proposed scheme.

Keywords: Optimal control, Adaptive control, Learning control, Intelligent control, Aircraft control, Function approximation

1. INTRODUCTION

Recently, lots of studies on morphing aircraft have been conducted because of its ability that to change its shape during flight, which enables the aircraft to perform various tasks efficiently (Ajaj et al. (2015)). Owing to the large-scale shape change, morphing aircraft has some advantages in extending the flight envelope and improving the performance of the aircraft, such as maneuverability and controllability (Prabhakar et al. (2015, 2016)). Several control strategies have been developed for the morphing aircraft using linear (Zhang and Wu (2014)) and nonlinear (Young et al. (2006)) control design approaches. However, most of those approaches have considered the morphing parameter as an open-loop command, which makes it difficult to take advantage of the benefits of the morphing aircraft. Once considering the morphing parameters as additional control inputs, that is, enabled to control the morphing parameters actively, then the system could have redundancy in control inputs. Then morphing aircraft requires flight control laws capable of high performance while maintaining stability in the presence of large variations in aerodynamic forces, moments of inertia, and mass center. The issues are the identification of the transient aerodynamic forces and moments of significant consequence to design the suitable controller. However, it is difficult to obtain an accurate aerodynamic model of the morphing system, especially for low-cost morphing aircraft under consideration in this study. In addition, the over-actuated systems with redundancy in control inputs require optimal control laws to exploit the better performance of the control system. For this reason, an actor-critic-based online control design method that enables model-free control design can be considered, which can be implemented by using reinforcement learning (RL). In terms of the optimal control design, solving the Hamilton-Jacobi-Bellman (HJB) equation is required in general (Bellman (1957)). Usually, in the case that the system is modeled by linear dynamics and the cost function to be minimized is quadratic in the state and control, the optimal control is obtained by solving a standard Riccati equation. On the other hand, when the system is modeled by nonlinear dynamics, the optimal control is determined by the solutions to the HJB equation given by a nonlinear partial differential equation (Abu-Khalaf and Lewis (2005)). It is often computationally untreatable or impossible to solve. The problem can be dealt with an idea known as adaptive or approximate dynamic programming (ADP) (Lewis and Vrabie (2009); Wang et al. (2009)). ADP is based on value function approximation (VFA) that approximates the cost function by using function approximation, such as neural networks or linear regression, to obtain the solution of the HJB equation. It is an extension of adaptive control that draws optimal online control design techniques and is directly related to the feedback control systems (Lewis et al. (2012)). This approach has received a lot of attention from many researchers in recent years, and the analysis of stability has been actively carried out. The actor-critic structure consists of following two steps: i) policy evaluation by the critic followed by ii) policy improvement. The actor applies an control policy to the environment, and the critic assesses the value of that control policy. Based on this assessment, various schemes can be used to improve the control policy in the sense that the new policy yields better value than previous value. In contrast to the generalized policy iteration (PI) based on a greedy policy improvement using the value

★ This work was supported by the advanced Research Center Program (NRF-2019R1A2C2083946) through the National Research Foundation of Korea (NRF) grant funded by the Korean Government (MSIP) contracted through the the Institute of Advanced Aerospace Technology at Seoul National University.

function, which can be highly problematic in aircraft system, the actor-critic structure generates a smooth control policy by using the function approximation. Note that applying RL to the continuous-time systems is considerably more difficult than applying it to the discrete-time systems, and fewer results are available. A method known as integral reinforcement learning (IRL) allows the application of the RL formulate online optimal adaptive control methods for the continuous-time systems. This method finds real-time solutions to the optimal HJB equations online without knowing the complete system dynamics.

Another challenging issue with the use of the morphing parameters as control input is that the dynamic model of the system is represented as a non-affine in control system, because the aerodynamic effects of the morphing aircraft are non-affine with respect to the morphing parameters. In view of this, due to the dynamic nonlinearities which are dependent not only on the states of the system but also on the control inputs, the control design for non-affine in control system is a difficult problem (Boskovic et al. (2004)). Solving this problem with model approximation may lead to fatal degradation of control performance. On the other hand, in the control design based on the exact dynamic model, the nonlinear dynamics caused by the effect depending on control input should be included. However, it is practically problematic because of the complexity of the system, and obtaining the exact dynamic model of the morphing system is difficult in general. Moreover, most of the control design methods developed in the past decades are applicable to nonlinear and affine in control systems, characterized by control input appearing linearly in the state equation (Isidori (2013)). The same is true of the control design methods in RL.

The objective of this study is to propose an online adaptive optimal control design based on an actor-critic structure, which can be applied to a morphing aircraft system considering morphing parameters as additional control inputs. By using an adaptive control design scheme, the stability of the closed-loop system with respect to the morphing parameter variations is guaranteed. The main contributions of this paper can be summarized as follows. (1) The optimal control design of the morphing aircraft is proposed using the morphing system control actively. The morphing parameters are considered as additional control inputs not the open-loop command, which makes the system have non-affine dependency on control input. Therefore, a virtual dynamic model is augmented with respect to the original system, which can alleviate the difficulties in the control problem. In addition, the augmented system can take into account the limit of the slow morphing actuator dynamics, which is the general assumption on the morphing aircraft. (2) Even though the state-of-the-art online ADP algorithms have been actively studied, the experiments with the complex nonlinear aircraft plant are rare. Motivated by the result of Vamvoudakis and Lewis (2010), the problem is extended to more complex and general system with multiple control inputs including time-varying morphing parameters and then proved its closed-loop stability and the convergence towards the optimal solution with regard to the augmented dynamic system.

This paper is organized as follows. The problem considered in this study is stated in Sec. 2 and the dynamic model of morphing aircraft is included. Section 3 contains the actor-critic-based control design for the continuous-time system and algorithm. Numerical simulation result is shown in Sec. 4, and concluding remark is given in Sec. 5.

2. PROBLEM STATEMENTS

The morphing aircraft model considered in this study is shown in Fig. 1, which has variable-span and variable-sweep wing. Variable-span and variable sweep morphing are parameterized by two morphing parameters, η_1 and η_2. Span and sweep angle variations are linearly mapped onto $[-0.5, 0.5]$, as summarized in Tables 1 and 2.

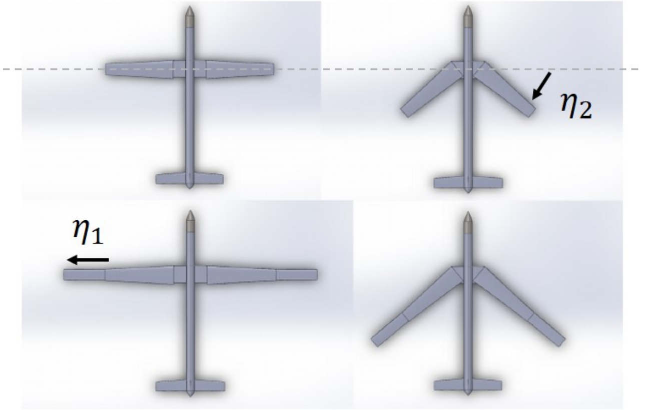

Fig. 1. Morphing parameter definition

Table 1. Span morphing parameter definition

	Min. Span	Max. Span
Value	1.7 m	2.8 m
Variable	$\eta_1 = -0.5$	$\eta_1 = 0.5$

Table 2. Sweep morphing parameter definition

	Min. Sweep	Max. Sweep
Value	0 deg	40 deg
Variable	$\eta_2 = -0.5$	$\eta_2 = 0.5$

2.1 *Dynamic Model of Morphing Aircraft*

In this study, the longitudinal motion of morphing aircraft is considered. The nominal dynamic model is obtained at the flight condition of airspeed 20 m/s with the altitude 300 m, where both morphing parameters, denoted as η, are zero. The longitudinal motion of conventional aircraft is governed by the following dynamic equations.

$$m\dot{V} = F_T \cos(\alpha + \alpha_T) - D - mg \sin\gamma \tag{1}$$

$$m\dot{\gamma}V = F_T \sin(\alpha + \alpha_T) + L - mg \cos\gamma \tag{2}$$

$$\dot{\alpha} = q - \dot{\gamma} \tag{3}$$

$$\dot{q} = M/J_y \tag{4}$$

where the state variables are airspeed V, angle of attack α, pitch rate q, and flight path angle γ, and the control variables are throttle command δ_t, elevator deflection δ_e. The F_T is thrust force and L, D and M is the aerodynamic forces and moment, which is given as functions of states and morphing parameters (Lee and Kim (2019)). Note that the morphing parameters are considered as additional control effectors in this study. Since the aerodynamic effect of the morphing aircraft is represented as non-affine with respect to the morphing parameters, the dynamic model of the system cannot be represented as an affine in control input form anymore. Therefore, it is inevitable that bringing in a technical approach to deal with the difficulty.

By leveraging the advantages of the model-free structure of ADP, a simple first-order controller dynamics can be considered as an augmented system, which enables the affine form of the total system. As the morphing system dynamics (6) is introduced, the whole dynamic model of the morphing aircraft can be represented as

$$\dot{X} = F(X,\eta) + G(X,\eta)U \tag{5}$$

$$\dot{\eta} = A(\eta) + B(\eta)\eta_c \tag{6}$$

where the state vectors are $X = [V,\ \alpha,\ q,\ \gamma]^T$ and $\eta = [\eta_1,\ \eta_2]^T$, and the control vectors are $U = [\delta_t,\ \delta_e]^T$ and $\eta_c = [\eta_{1c},\ \eta_{2c}]^T$.

As a result, the derived augmented dynamic model of the morphing aircraft can be represented as

$$\dot{x} = f(x) + g(x)u \tag{7}$$

where the new state vector is $x = [X^T,\ \eta^T]^T$ and the new control input vector is $u = [U^T,\ \eta_c^T]^T$. Note that the original system (5) is non-affine in control due to η considered as the control input, whereas the augmented system (7) is affine in control by considering η as the state, and η_c as the control input. The optimal control problem now finds an optimal control input for U and η_c with a different performance index including η_c term, which may lead to a different response from the original problem. Let us assume that the unknown morphing actuator dynamics can be represented as follows.

$$\eta_c = l(v) \tag{8}$$

where v is actuator command, and $l(v)$ is unknown non-affine function. In this study, the actor-critic structure does not require the exact model to solve the optimal control problem. However, the augmented dynamic model is needed for the smooth formulation in this structure.

2.2 Optimal Control Problem

In this study, considering the above dynamic system, it is assumed that $f(x)$ and $g(x)$ in (7) is Lipschitz continuous, and the solution $x(t)$ is unique. The objective of the optimal control problem is to minimize the following value function.

$$V(x(t)) = \int_t^\infty r(x(\tau), u(\tau))d\tau \tag{9}$$

where the cost function is $r(x,u) = x^T Qx + u^T Ru$ with Q and R being symmetric positive definite matrices with appropriate dimensions. Note that the optimal control $u(t)$ must not only stabilize the system but also guarantee that (9) is finite, which called the control is admissible. For the admissible control, an infinitesimal equivalent to (9) is the following Bellman equation.

$$0 = V_x^T(f(x) + g(x)u) + x^T Qx + u^T Ru,\ \ V(0) = 0 \tag{10}$$

where V_x is the partial derivative of the value function $V(x)$ with respect to x. In fact, the value function of the original system is given as follows.

$$V_o(t) = \int_t^\infty X^T QX + U^T RU + \eta_c^T R_{\eta_c}\eta_c d\tau. \tag{11}$$

The value function of the augmented system can be rewritten as

$$V(t) = \int_t^\infty X^T QX + U^T RU + \eta^T Q_\eta \eta + v^T R_v v d\tau. \tag{12}$$

According to the value function of the augmented system, the optimal performance of the original system would be decreased. It is assumed that v and η have the similar value with η_c in this study. Note that the further study is needed to select the weighting parameters appropriately for the value function to guarantee the optimal solution. Let us define the following continuous-time Hamiltonian (Lewis and Vrabie (2009)).

$$H(x,u,V) = V_x^T(f(x) + g(x)u) + x^T Qx + u^T Ru. \tag{13}$$

Then, the optimal value function is given as

$$V^*(x) = \min_u \left(\int_t^\infty r(x(\tau), u(\tau))d\tau \right) \tag{14}$$

and it satisfies the following HJB equation.

$$\begin{aligned} 0 &= \min_u H(x,u^*,V^*) \\ &= V_x^{*T}(f(x) + g(x)u^*) + x^T Qx + u^{*T} Ru^* \end{aligned} \tag{15}$$

where $V^*(0) = 0$. Then, by solving the optimal control problem given as following equation.

$$\frac{\partial H(x,u^*,V^*)}{\partial u^*} = 0. \tag{16}$$

Then, the optimal control u^* can be derived as follows.

$$u^* = -\frac{1}{2}R^{-1}g^T(x)V_x^*(x). \tag{17}$$

In practical, for continuous-time nonlinear system, obtaining the optimal control is difficult due to that the HJB equation cannot be solved analytically for the general nonlinear system, and the complete system dynamics should be known, which is inevitable to obtain the optimal value function.

In solving the optimal control problem, the policy iteration (PI) is a typical method of reinforcement learning (RL), which solves the algorithm consisting of policy evaluation and policy improvement iteratively.

2.3 Value Function Approximation

The PI algorithm given in Algorithm 1 proceeds by alternately updating the value (critic) and the control policy (actor) by solving (15) and (17), respectively. To implement PI online for a dynamic system with the continuous-time system, the value function approximation (VFA) can be used that approximates the value with unknown parameters. And the unknown parameters are tuned online exactly as in system identification. According to the Weierstrass higher-order approximation theorem (Hornik et al. (1990)), there exists a dense basis set $\{\phi_i(x)\}$ such that

$$\begin{aligned} V(x) &= \sum_{i=1}^{\infty} w_i\phi_i(x) \\ &= \sum_{i=1}^{L} w_i\phi_i(x) + \sum_{i=L+1}^{\infty} w_i\phi_i(x) \\ &= W_c^T\Phi(x) + \varepsilon(x) \end{aligned} \tag{18}$$

where $\Phi(x) = [\phi_1(x),\ \phi_2(x),\ \cdots,\ \phi_L(x)]^T$ is the basis vector, W_c is the weights vector, and $\varepsilon(x)$ converges uniformly to zero as the number of terms retained $L \to \infty$.

3. CONTROL DESIGN BASED ON ACTOR-CRITIC STRUCTURE

In this section, an online actor-critic-based control system, described in Fig. 2, is designed so that the closed-loop system

Algorithm 1 Policy Iteration (PI)

```
► Policy Evaluation.
  Given policies, solve for the value using (15).
► Policy Improvement.
  Update the control policy using (17).
```

is asymptotically stable. RL is considerably more difficult to be applied for continuous-time systems than for discrete-time systems. As ealier stated, using IRL allows the application of the RL to formulate an online optimal adaptive control method for continuous-time systems (Vamvoudakis et al. (2014)). The control policy as the solution to optimal HJB equation can be found online in real-time without knowing the internal dynamics of the system, $f(x)$.

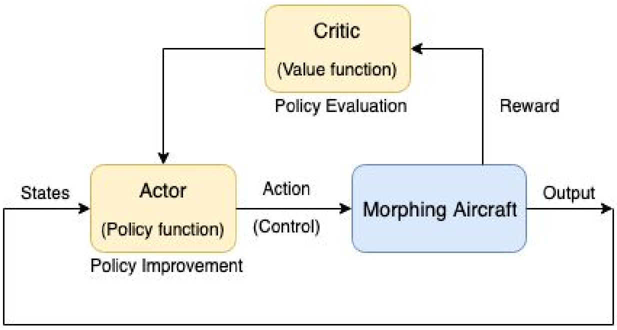

Fig. 2. Actor-critic-based control system

3.1 Critic Network using Value Function Approximation

The weights of the critic network, W_c, are unknown and therefore the weight estimates can be used. Then, the output of the critic network as follows.

$$\hat{V}(x) = \hat{W}_c^T \Phi(x) \tag{19}$$

where $\hat{W}_c$ is the currently known value of the critic network weight. Recall that $\Phi(x)$ is the predefined basis vector with N the number of elements. Then, the approximation error can be represented as follows.

$$\varepsilon = \rho + \hat{W}_c^T \Phi(x(t)) - \hat{W}_c^T \Phi(x(t-T)) \tag{20}$$

where $\rho = \int_{t-T}^{t} r(x(\tau), u(\tau))d\tau$. Equation (20) can be rewritten as

$$\hat{W}_c^T \Delta\Phi(x(t)) = \varepsilon - \rho \tag{21}$$

where $\Delta\Phi(x(t)) = \Phi(x(t)) - \Phi(x(t-T))$. It is desired to select $\hat{W}_c$ to minimize the following squared residual error.

$$E_c = \frac{1}{2}\varepsilon^T \varepsilon. \tag{22}$$

Then, $\hat{W}_c(t) \rightarrow W_c$. The tuning law for the critic network weights as the normalized gradient descent algorithm is selected as follows.

$$\begin{aligned} \dot{\hat{W}}_c &= -a_c \frac{\partial E_c}{\partial \hat{W}_c} \\ &= -a_c \frac{\Delta\Phi^T(x(t))}{\left(1 + \Delta\Phi^T(x(t))\Delta\Phi(x(t))\right)^T}\left(\rho + \Delta\Phi^T(x(t))\hat{W}_c\right). \end{aligned} \tag{23}$$

Note that the data $(\Delta\Phi(t), \rho(t))$ are required at each time in the tuning algorithm.

Let us define the critic network weight estimation error is as $\tilde{W}_c = W_c - \hat{W}_c$. By substituting (20) in (23), the dynamics of the critic network weight estimation error can be obtained as

$$\dot{\tilde{W}}_c = -a_c \bar{\Delta}\Phi(t)\bar{\Delta}\Phi^T(t)\tilde{W}_c + a_c \bar{\Delta}\Phi(t)\frac{\varepsilon}{1 + \Delta\Phi^T(t)\Delta\Phi(t)} \tag{24}$$

where $\bar{\Delta}\Phi(t) = \Delta\Phi(t)/(1 + \Delta\Phi^T(t)\Delta\Phi(t))$. To guarantee the convergence of $\hat{W}_c$ to W_c, following assumptions are required.

Assumption 1. Persistence of excitation (PE). Let the signal $\bar{\Delta}\Phi(t)$ be persistently excited over the interval $[t-T, t]$, i.e., there exist constants $\beta_1 > 0, \beta_2 > 0$, and $T > 0$ such that, for all t,

$$\beta_1 I \le S_0 \equiv \int_{t-T}^{t} \bar{\Delta}\Phi(\tau)\bar{\Delta}\Phi^T(\tau)d\tau \le \beta_2 I \tag{25}$$

Assumption 2. For a given compact set, the approximation error and its gradient are bounded.

Assumptions 1 and 2 are standard assumptions in approximation based control. Note that, from (21), the regression vector $\Delta\Phi(t)$, or the normalized vector $\bar{\Delta}\Phi(t)$ must be persistently excited to solve for $\hat{W}_c$ in a least squares sense. Note that the basis set should be defined appropriately including the morphing parameter η, by using random noise or sinusoidal signal, in order that the whole system may activate (Assumption 1), in this study.

3.2 Actor Network using Adaptive Control Approach

The actor network for policy improvement step in PI by using VFA is given as

$$u(x) = -\frac{1}{2}R^{-1}g^T(x)\nabla\Phi^T W_c \tag{26}$$

with the critic network weights W_c unknown. Therefore, the control policy using the estimate of the actor network is defined as

$$u_2(x) = -\frac{1}{2}R^{-1}g^T(x)\nabla\Phi^T \hat{W}_a \tag{27}$$

where $\hat{W}_a$ is the current estimated value of the actor network weight. The approximate HJB equation can be represented as

$$\begin{aligned} \int_{t-T}^{t} \left(-Q(x) - \frac{1}{4}W_c^T D(x) W_c + \varepsilon(x)\right)d\tau \\ = W_c^T \Delta\Phi(x(t)) \end{aligned} \tag{28}$$

where $D(x) = \nabla\Phi(x)g(x)R^{-1}g^T(x)\nabla\Phi^T(x)$, and W_c is the ideal unknown weights of the critic and actor networks that solve the HJB. The tuning laws for the critic and actor networks can be obtained that guarantee the convergence to the optimal control along with closed-loop stability, given by the following theorems.

Theorem 1. Let tuning law for the critic network be provided by

$$\begin{aligned} \dot{\hat{W}}_c = -a_c &\frac{\Delta\Phi(x(t))^T}{\left(1 + \Delta\Phi(x(t))^T\Delta\Phi(x(t))\right)^2} \\ &\left(\int_{t-T}^{t}\left(-Q(x) - \frac{1}{4}\hat{W}_a^T D(x)\hat{W}_a\right)d\tau + \Delta\Phi(x(t))^T\hat{W}_c\right) \end{aligned} \tag{29}$$

where $\Delta\Phi(x(t)) = \int_{t-T}^{t} \nabla\Phi(x)(f + gu_2)d\tau$, and by assumptions, u_2 is persistently excited. Let the tuning law of the actor network is selected as

$$\begin{aligned} \dot{\hat{W}}_a &= -a_a(k_a\hat{W}_a - k_c\hat{W}_c) \\ &\quad + \frac{1}{4}a_a D(x)\hat{W}_a \frac{\Delta\Phi(x(t))^T}{(1 + \Delta\Phi(x(t))^T\Delta\Phi(x(t)))^2}\hat{W}_c \end{aligned} \tag{30}$$

where a_c and a_a are the learning rates, and k_c and k_a are positive tuning parameters, which can be chosen appropriately to ensure stability. Then, the closed-loop system state is uniformly ultimately bounded (UUB), and the critic network weights error $\tilde{W}_c = W_c - \hat{W}_c$ and the actor network weights error $\tilde{W}_a = W_c - \hat{W}_a$ are UUB.

Proof. The convergence proof is based on Lyapunov analysis. Let us consider the Lyapunov function.

$$L = V(x) + \frac{1}{2}(\tilde{W}_c{}^T a_c^{-1} \tilde{W}_c) + \frac{1}{2}(\tilde{W}_a{}^T a_a^{-1} \tilde{W}_a) \tag{31}$$

Then, the derivative of the L is given by

$$\dot{L} = \dot{V}(x) + \tilde{W}_c{}^T a_c^{-1} \dot{\tilde{W}}_c + \tilde{W}_a{}^T a_a^{-1} \dot{\tilde{W}}_a = \dot{L}_v(x) + \dot{L}_c(x) + \dot{L}_a(x) \tag{32}$$

For the first term,

$$\begin{aligned} \dot{L}_v(x) &= -x^T Qx - \frac{1}{4} W_c^T D W_c + \frac{1}{2} W_c^T D \tilde{W}_a + \varepsilon_{HJB}(x) + \varepsilon_1(x) \\ &\equiv \dot{\bar{L}}_v(x) + \frac{1}{2} W_c^T D \tilde{W}_a + \varepsilon_1(x) \end{aligned} \tag{33}$$

For the second term,

$$\begin{aligned} \dot{L}_c(x) &= \tilde{w}_c^T \frac{\sigma}{(1+\sigma^T\sigma)^2} \left(-\sigma^T \tilde{W}_c + \frac{1}{4} \tilde{W}_a^T D \tilde{W}_a + \varepsilon_{HJB}(x) \right) \\ &\equiv \dot{\bar{L}}_c(x) + \frac{1}{4} \tilde{W}_c^T \frac{\sigma}{(1+\sigma^T\sigma)^2} \tilde{W}_a^T D \tilde{W}_a \end{aligned} \tag{34}$$

Finally, by adding the first and second terms

$$\begin{aligned} \dot{L}(x) = {} & \dot{\bar{L}}_v + \dot{\bar{L}}_c + \varepsilon_1(x) - \tilde{W}_a^T \left[a_a^{-1} \dot{\hat{W}}_a - \frac{1}{4} D \hat{W}_a \frac{\sigma}{(1+\sigma^T\sigma)^2} \hat{W}_c \right] \\ & + \frac{1}{2} \tilde{W}_a^T D W_c + \frac{1}{4} \tilde{W}_a^T D W_c \frac{\sigma}{(1+\sigma^T\sigma)^2} \tilde{W}_c \\ & - \frac{1}{4} \tilde{W}_a^T D W_c \frac{\sigma}{(1+\sigma^T\sigma)^2} W_c + \frac{1}{4} \tilde{W}_a^T D \tilde{W}_a \frac{\sigma}{(1+\sigma^T\sigma)^2} W_c \end{aligned} \tag{35}$$

and we can define the actor tuning law as

$$\dot{\hat{W}}_a = -a_a \left((k_a \hat{W}_a - k_c \hat{W}_c) - \frac{1}{4} D \hat{W}_a \frac{\sigma}{(1+\sigma^T\sigma)^2} \hat{W}_c \right) \tag{36}$$

By choosing the parameter such that L exceeds a certain bound, then, it is shown that $\dot{L}$ is negative. Therefore the closed-loop system and weight parameters are UUB according to the standard Lyapunov extension theorem. The rest of the proof is similar with Vamvoudakis and Lewis (2010).

Theorem 2. Optimal solution. Suppose the hypotheses of *Theorem 1* hold. Then,
$H(x, \hat{u}_2, \hat{W}_c) \equiv \int_{t-T}^{t} (\rho - \varepsilon) d\tau + \hat{W}_c^T \Phi(x(t)) - \hat{W}_c^T \Phi(x(t-T))$ is UUB, where $\hat{u} = -\frac{1}{2} R^{-1} g^T(x) \nabla\Phi^T \hat{W}_c$.
That is, $\hat{W}_c$ converges to the approximate HJB solution.
Also, $\hat{u}_2(x)$ converges to the optimal solution, where $\hat{u}_2 = -\frac{1}{2} R^{-1} g^T(x) \nabla\Phi^T \hat{W}_a$.

Proof. See Vamvoudakis et al. (2011)

4. NUMERICAL SIMULATION

Numerical simulation is performed to demonstrate the performance of the proposed control design. In this study, the variable-span and variable-sweep morphing wing aircraft is modeled as a nonlinear system consisting of the typical aircraft dynamics and morphing system dynamics. The actuator dynamics are modeled considering the saturation of the actuator, and the simulation time and the convergence criterion is set to 100 seconds and 10^{-5}, respectively. The nominal flight condition is a trim condition with $V = 20\,\text{m/s}$ at $h = 300\,\text{m}$, and the initial condition is given by that the flight path angle is perturbed at the trim condition. Simulations are conducted for the optimal control problem with the perturbed initial condition. The basis function Φ is chosen as the quadratic vector in the original states and morphing parameters including ϵ which denotes some exciting terms.

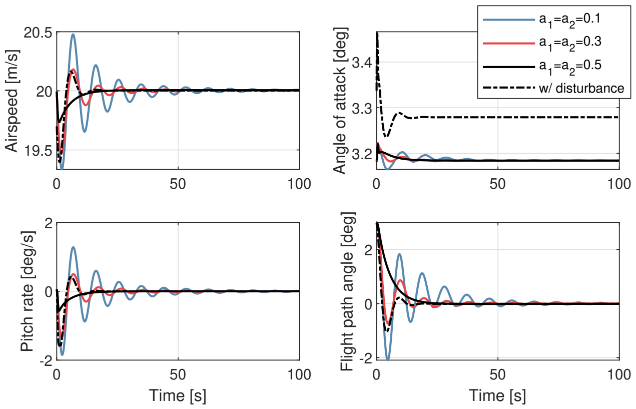

Fig. 3. State trajectories with several learning rates

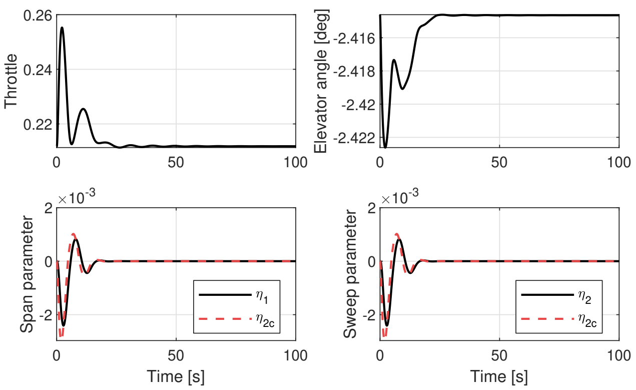

Fig. 4. Control input trajectories (scenario 1)

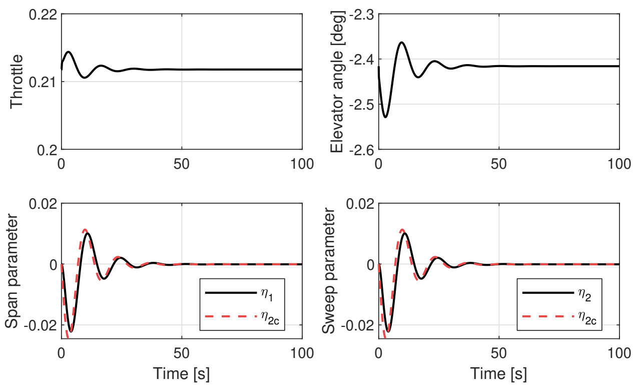

Fig. 5. Control input trajectories (scenario 2). By using morphing parameters more actively, the use of the throttle can be reduced.

Figures 3 - 5 show that the state trajectories with several learning rates and control input trajectories with a fixed learning rate, respectively. Through the state trajectories, it can be seen that the closed-loop system achieves the UUB condition in this regulation problem. It is shown that the solution converges to the trim value according to the smooth control trajectories and the control performance can be achieved even when the constant wind disturbance applied to the system, which is a typical advantage compared to the model-based controller. Figure 3 shows that the performance of learning is improved when the learning rate increases. In figs. 4 and 5, two scenarios that

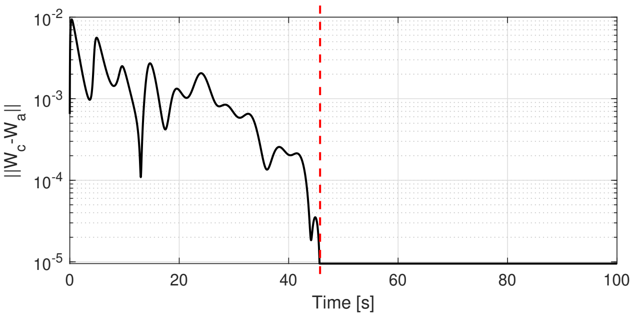

Fig. 6. Actor and critic parameter error trajectory

have different weighting matrices on the control are compared. By considering the morphing parameters as control input, it is shown that the use of the throttle can be reduced according to the pre-design weighting parameters of the cost function. Comparing to the generalized PI that takes the greedy action to maximize the value, which becomes problematic, in this paper, by using the function approximation, the continuous policy trajectory can be generated via the actor-critic structure. The rate limit of the morphing system can also be covered by choosing the parameter of the morphing system dynamics appropriately. The norm of the difference between actor and critic networks weight parameters are shown in Fig. 6. Both of the critic and actor parameters converge after about 40 seconds to the optimal values by arriving at the convergence criterion denoted as the broken line. The critic and actor networks converge to the same values, which shows that the actor network also converges to the optimal values. By adjusting the tuning parameters, while these transient properties can be changed, the actor network always converges to the value of the critic network.

5. CONCLUSION

Online optimal control design strategy for a variable-span and variable-sweep morphing wing aircraft was proposed based on the actor-critic structure. Considering the morphing parameters as control input, the dynamic model of the morphing aircraft has a dependency on the varying control input. Therefore, the augmented dynamic model was derived, which is an affine in control input form, using first-order dynamics for the morphing system. By adjusting the system parameter, the slow dynamic property of the morphing system can be treated. The system internal dynamics should be computed carefully according to the variation of control input due to the effect of morphing parameter variation as a disturbance on the internal system. Comparing with the model-based control design where the dynamic model of the system should be derived by depending on the morphing parameter variation, the online actor-critic method, which does not require the internal dynamics, has a great advantage of the computational efficiency. The stability of the closed-loop system with the adaptation law derived by the proposed control design was proved by the Lyapunov theory. Numerical simulation was conducted to demonstrate the effectiveness of the proposed approach by applying it to the morphing aircraft without knowing the internal system dynamics. Simulation results showed that the proposed control design provides good performance with the actor and critic update laws, even when the internal system dynamics is unknown.

REFERENCES

Abu-Khalaf, M. and Lewis, F.L. (2005). Nearly optimal control laws for nonlinear systems with saturating actuators using a neural network HJB approach. *Automatica*, 41(5), 779–791.

Ajaj, R.M., Beaverstock, C.S., and Friswell, M.I. (2015). Morphing aircraft: The need for a new design philosophy. *Aerospace Science and Technology*, 49, 154–166.

Bellman, R.E. (1957). *Dynamic Programming*. Princeton, NJ: Princeton Univ. Press.

Boskovic, J.D., Chen, L., and Mehra, R.K. (2004). Adaptive control design for nonaffine models arising in flight control. *Journal of Guidance, Control, and Dynamics*, 27(2), 209–217.

Hornik, K., Stinchcombe, M., and White, H. (1990). Universal approximation of an unknown mapping and its derivatives using multilayer feedforward networks. *Neural Networks*, 3(5), 551–560.

Isidori, A. (2013). *Nonlinear Control Systems*. Communications and Control Engineering. Springer, London, UK.

Lee, J. and Kim, Y. (2019). Neural network-based nonlinear dynamic inversion control of variable-span morphing aircraft. *Proceedings of the Institution of Mechanical Engineers, Part G: Journal of Aerospace Engineering*, accpeted for publication.

Lewis, F.L. and Vrabie, D. (2009). Reinforcement learning and adaptive dynamic programming for feedback control. *IEEE Circuits and Systems Magazine*, 9(3), 32–50.

Lewis, F.L., Vrabie, D., and Vamvoudakis, K.G. (2012). Reinforcement learning and feedback control: Using natural decision methods to design optimal adaptive controllers. *IEEE Control Systems*, 32(6), 76–105.

Prabhakar, N., Prazenica, R.J., and Gudmundsson, S. (2015). Dynamic analysis of a variable-span, variable-sweep morphing UAV. In *IEEE Aerospace Conference*. Big Sky, MT.

Prabhakar, N., Prazenica, R.J., Gudmundsson, S., and Balas, M.J. (2016). Transient dynamic analysis and control of a morphing UAV. In *AIAA Guidance, Navigation, and Control Conference*. San Diego, CA.

Vamvoudakis, K.G. and Lewis, F.L. (2010). Online actor-critic algorithm to solve the continuous-time infinite horizon optimal control problem. *Automatica*, 46(5), 878–888.

Vamvoudakis, K.G., Vrabie, D., and Lewis, F.L. (2011). Online adaptive algorithm for optimal control with integral reinforcement learning. In *IEEE Symposium on Adaptive Dynamic Programming and Reinforcement Learning (ADPRL)*. Paris, France.

Vamvoudakis, K.G., Vrabie, D., and Lewis, F.L. (2014). Online adaptive algorithm for optimal control with integral reinforcement learning. *International Journal of Robust and Nonlinear Control*, 24(17), 2686–2710.

Wang, F., Zhang, H., and Liu, D. (2009). Adaptive dynamic programming: An introduction. *IEEE Computational Intelligence Magazine*, 4(2), 39–47.

Young, A., Cao, C., Hovakirnyan, N., and Lavretsky, E. (2006). An adaptive approach to nonaffine control design for aircraft applications. In *AIAA Guidance, Navigation, and Control Conference*. Keystone, CO.

Zhang, J. and Wu, S.T. (2014). Dynamic modeling and control for a morphing aircraft. In *The 26th Chinese Control and Decision Conference*. Changsha, China.