Model-Based Approximate Optimal Feedback Control of a Hypersonic Vehicle

Max L. Greene*,
University of Florida, Gainesville, Florida 32611
Patryk Deptula†,
The Charles Stark Draper Laboratory, Inc., Cambridge, Massachusetts 02139
Brendan J. Bialy‡,
U.S. Air Force Research Laboratory, Eglin Air Force Base, Florida, 32542
and
Warren E. Dixon§
University of Florida, Gainesville, Florida 32611

A model-based approximate dynamic programming (ADP) controller is applied to a hypersonic vehicle (HSV) with time-varying aerothermoelasatic effects for approximate optimal state regulation. To account for aerothermoelastic parameter variations, the nominal HSV dynamic model is discretely switched over time to better reflect changes caused by the parameters. A Lypunov-based analysis is leveraged to design the actor-critic update laws for the reinforcement learning ADP approach and to prove uniformly ultimately bounded convergence of the switched dynamic system. Simulation results are included, which indicate that the switched model-based ADP controller yields HSV state regulation in the presence of disturbances.

I. Introduction

The design of control algorithms for hypersonic vehicles (HSVs) is complicated by the complex coupling of the HSV states (e.g., pitch, angle of attack, etc.) with aeroelastic structural dynamics and parametric changes caused by effects such as temperature variation.[1,2] Such aerothermoelastic effects can destabilize the aircraft;[2] some previous control designs are based on linear-parameter-varying (LPV) HSV models.[3–5] In the LPV HSV framework, a set of linear nominal matrices represent the actual HSV drift dynamics and control effectiveness; the aerothermoelastic, time-varying parameters are accounted for with additional disturbance terms.[3,6]

Optimal control has been commonly used in aerospace control applications to provide system stability in the presence of model uncertainty and disturbances.[7] One approach to solve optimal control problems is to solve the Hamilton-Jacobi-Bellman (HJB) equation. For the linear quadratic regulator (LQR) optimal control problem, the HJB reduces to the algebraic Riccati equation (ARE).[7] Using computer-aided design tools, LQR problems can be solved at many points in an aircraft's flight envelope. For real-time operation across multiple operating domains, different optimal control policies are determined that yield desirable flight characteristics via gain scheduling.[8,9] However, HSV controllers may become unstable as the dynamics change with aerothermoelastic effects;[9] therefore gain scheduling should be based on an accurate reflection of the actual plant,[9] i.e., the nominal plant must be updated to reflect the slowly time-varying parameters in the plant dynamics.

*Graduate Research Assistant, Department of Mechanical and Aerospace Engineering; maxgreene12@ufl.edu.
†Senior Robotics Researcher, Perception and Autonomy Group; pdeptula@draper.com.
‡Research Aerospace Engineer, Munitions Directorate, Eglin Air Force Base; brendan.bialy@us.af.mil.
§Professor, Department of Mechanical and Aerospace Engineering; wdixon@ufl.edu.

Robust adaptive HSV controllers that yield exponential tracking have been investigated.[6] Despite using a nonlinear control design strategy, such results assume that the nominal matrices are time-invariant. If the time-varying parameters sufficiently deviate from the nominal matrix values, then the system may become unstable.[9] While it may be possible to develop a set of scheduled gains for many sets of nominal dynamics, this requires significant offline computational resources to compensate for the large number of flight conditions and parameter variations.

Reinforcement learning (RL) can be used to approximate solutions of optimal control problems.[10] Approximate dynamic programming (ADP) is a type of RL that has been applied to continuous- and discrete-time deterministic systems to solve optimal control problems via value function approximation.[11–13] In many ADP-based approaches, the optimal value function is approximated in real-time with a single-layer neural network (NN). Continuous-time update laws are used to update the NN weights online, which improves the value function approximation. The Bellman error (BE) is a metric that indirectly evaluates the quality of the value function approximation; the weights of the value function approximation are updated to minimize the BE. Traditional ADP results[14,15] evaluate the BE at on-trajectory points. However, to facilitate improved value function approximation across regions of the state state, other model-based results evaluate the BE at on- and off-trajectory points: this technique is called BE extrapolation.[16,17] BE extrapolation uses the system model to perform simulation of experience[18,19] at user-defined points in the state-space.

An existing model-based ADP result[20] analyzes the stability of a switched system with an arbitrary switching sequence. That is, multiple model-based ADP controllers are applied to a family of switched subsystems. The user selects the active subsystem and when to switch between the subsystems such that some sufficient conditions are satisfied (e.g., gain and dwell-time conditions). Switching instances introduce discontinuities into the closed-loop signals, which must be accounted for in the stability analysis. Event-triggered ADP results[21,22] also introduce discontinuities into the closed-loop signals; however, the aforementioned event-triggered ADP methods do not consider switching between sets of nominal dynamics.

The switched system ADP result[20] can be extended to the HSV control case by implementing a model-based ADP controller to the nominal model of the LPV HSV dynamics. As the nominal LPV HSV matrices are switched, then the ADP update policy, similarly, switches. However, the switched model-based ADP result[20] does not address disturbances or robustness to model uncertainty. This paper applies the switched ADP controller[20] to switch between multiple linear, nominal models of the HSV to perform state regulation (i.e., steady, level flight) for a HSV. Simulation results are included to show the performance of the switched model-based ADP controller in the presence of disturbances and model uncertainty.

Notation

For notational brevity, time-dependence is omitted while denoting trajectories of the dynamical systems. For example, the trajectory $x(t)$, where $x : \mathbb{R}_{\geq 0} \to \mathbb{R}^n$, is denoted as $x \in \mathbb{R}^n$ and referred to as x instead of $x(t)$. Equations of the form $f + h(y,t) = g(x)$ should be interpreted as $f(t) + h(y(t),t) = g(x(t)) \; \forall t \in \mathbb{R}_{\geq 0}$. The gradient $\left[\frac{\partial f(x,y)}{\partial x_1}^T, \ldots, \frac{\partial f(x,y)}{\partial x_n}^T\right]^T$ is denoted by $\nabla_x f(x,y)$. The Euclidean norm for vectors and Frobenius norm for matrices is denoted with $\|\cdot\|$. The $n \times n$ identity matrix is denoted by $I_{n\times n}$. The left-hand limit $\lim_{x\to a^-} f(x)$ is denoted as $f(a^-)$, and the right-hand limit $\lim_{x\to a^+} f(x)$ is denoted as $f(a^+)$.

II. HSV Dynamics

The subsequent development is based on a HSV model with aerothermoelastic effects and structural dynamics given by[23]

$$\dot{V} = \frac{T\cos(\alpha) - D}{m} - g\sin(\theta - \alpha) \tag{1}$$

$$\dot{\alpha} = \frac{L + T\sin(\alpha)}{mV} + q + \frac{g}{V}\cos(\theta - \alpha) \tag{2}$$

$$\dot{q} = \frac{M}{I_{yy}} \tag{3}$$

$$\dot{h} = V \sin(\theta - \alpha) \tag{4}$$

$$\dot{\theta} = q \tag{5}$$

$$\ddot{\eta}_{s,i} = -2\zeta_{s,i}\omega_{s,i}\dot{\eta}_{s,i} - \omega_{s,i}^2\eta_{s,i} + N_i, \tag{6}$$

where $V \in \mathbb{R}$ is the forward velocity of the HSV, $T \in \mathbb{R}_{\geq 0}$ is the thrust, $\alpha \in \mathbb{R}$ is the angle of attack, $D \in \mathbb{R}_{\geq 0}$ is the drag, $h \in \mathbb{R}_{\geq 0}$ is the altitude, $m \in \mathbb{R}_{>0}$ is the HSV mass, $g \in \mathbb{R}_{>0}$ is the gravitational constant, $\theta \in \mathbb{R}$ is the pitch angle, $L \in \mathbb{R}_{\geq 0}$ is the lift, $q \in \mathbb{R}$ is the pitch rate, $M \in \mathbb{R}$ is the pitching moment about the HSV body y-axis, $I_{yy} \in \mathbb{R}_{>0}$ is the moment of inertial about the body y-axis, $\eta_{s,i} \in \mathbb{R}$ is the i^{th} flexible structural mode displacement for $i \in \{1, 2, 3\}$, $\zeta_{s,i}$, $\omega_{s,i}$, $N_{s,i} \in \mathbb{R}_{\geq 0}$ are the damping factor, natural frequency, and generalized elastic forces of the i^{th} structural mode, respectively. The states can be concatenated into a vector $x \in \mathbb{R}^{11}$ that captures the flight dynamic and structural dynamic states

$$x \triangleq \begin{bmatrix} \Delta V & \Delta\alpha & q & \Delta h & \Delta\theta & \eta_{s,1} & \dot{\eta}_{s,1} & \eta_{s,2} & \dot{\eta}_{s,2} & \eta_{s,3} & \dot{\eta}_{s,3} \end{bmatrix}^T,$$

where Δ denotes the difference between the state and its respective trim condition.

The HSV aerodynamic and structural modes are coupled.[1] Specifically, T, L, and D, depend on the structural modes $\eta_{s,i}$. As the HSV temperature increases, the modulus of elasticity linearly decreases.[6] The change in the modulus of elasticity alters the structural damping ratio $\zeta_{s,i}$ and natural frequency $\omega_{s,i}$, which, in turn, significantly alters the structural dynamic responses (i.e., $\ddot{\eta}_{s,i}$).[6]

A. LPV Model

The HSV dynamics can be modeled as a controllable linear-parameter-varying (LPV) system with uncertainty from unmodeled effects as[3,6]

$$\dot{x} = A(\rho(t))\, x + B(\rho(t))\, u + d(t) \tag{7}$$

$$y = Cx, \tag{8}$$

where $A(\rho(t)) \in \mathbb{R}^{11\times 11}$ denotes a LPV state matrix, $\rho(t) \in \mathbb{R}_{\geq 0}$ denotes the unknown time-dependent temperature profile of the HSV, $B(\rho(t)) \in \mathbb{R}^{11\times 2}$ denotes a LPV control effectiveness matrix, $u \in \mathbb{R}^2$ denotes the subsequently defined control input, $d(t) \in \mathbb{R}^{11}$ denotes a time-varying uncertainty and disturbance term, $y \in \mathbb{R}^5$ denotes the measurable states, and $C \in \mathbb{R}^{11\times 5}$ denotes an output matrix.

In this example, the control inputs are the deflection angle of the elevators and canards from their trim condition such that $u \triangleq \begin{bmatrix} \delta_e & \delta_c \end{bmatrix}^T$. The HSV fuel equivalence ratio and diffuser area ratio are fixed at their operational trim condition. The outputs are selected as $y \triangleq \begin{bmatrix} \Delta V & \Delta\alpha & q & \Delta h & \Delta\theta \end{bmatrix}^T$.

Let $j \in \mathbb{S} \subset \mathbb{N}$ represent the total number of nominal dynamic models. The state matrix and control effectiveness matrix can be represented as[3]

$$A(\rho(t)) = A_j + w_j(\rho(t)) \tag{9}$$

$$B(\rho(t)) = B_j + v_j(\rho(t)), \tag{10}$$

where $A_j \in \mathbb{R}^{11\times 11}$ is the j^{th} nominal state matrix, $B_j \in \mathbb{R}^{11\times 2}$ is the j^{th} nominal control effectiveness matrix, and $w_j(\rho(t)) \in \mathbb{R}^{11\times 11}$ is an unknown parameter-varying disturbance term associated with the j^{th} nominal state matrix, and $v_j(\rho(t)) \in \mathbb{R}^{11\times 2}$ is an unknown parameter-varying disturbance term associated with the j^{th} nominal control effectiveness matrix.

B. Switched Nominal Matrices

Let $\sigma : \mathbb{R}_{\geq 0} \to \mathbb{S}$ denote a time-varying switching signal that dictates the active nominal model used for control. As the parameters vary, the time-varying signal switches to a different nominal model. The condition that determines which model and controller is active can be triggered by a sensor-based strategy (e.g., altitude, speed, temperature) and informed by predefined flight plans and computational models; however, the specific design of such triggers is beyond the scope of this paper.

The temperature profile $\rho(t)$ in (9) and (10) is time-varying; hence, the matrices $A(\rho(t))$ and $B(\rho(t))$ are similarly time-varying. As A and B vary with respect to time, their respective nominal matrices A_j and B_j become less accurate. For example, A_1 may approximate $A(\rho(0))$; however, as the temperature changes with time t, the elastic HSV modes $\eta_{s,i}$ have a different response, degrading the approximation.

To account for the time-varying nature of $A(\rho(t))$, the nominal matrix A_j is updated from A_j to A_{j+1}; i.e., the controller switches from the j^{th} nominal model to the $j+1^{\text{th}}$ nominal model. The motivation for switching between nominal models is that the updated models can approximate $A(\rho(t))$ without explicitly determining the time-varying behavior of each parameter. Instead, each nominal matrix is discretely modified.

C. Measurable States

The matrices A_j and B_j represent the nominal HSV dynamics. Not all states in x may be measurable, particularly the structural dynamic modes $\eta_{s,i}$ and $\dot{\eta}_{s,i}$ for all i. Recall, $y \triangleq \begin{bmatrix} \Delta V & \Delta\alpha & q & \Delta h & \Delta\theta \end{bmatrix}^T$.

To facilitate the development of the subsequent control development, the disturbances are assumed to be negligible and are the focus of future research. Specifically, let $\|d(t)\| = 0$, $\|w_j(\rho(t))\| = 0$, $\|v_j(\rho(t))\| = 0$ for all $j \in \mathbb{S}$. The reduced-order model used for the control development is

$$\dot{y} = CA_j y + CB_j u. \tag{11}$$

Some ADP-based results propose different update laws to account for disturbances in the system dynamics.[24] Such robust update laws have not been investigated in the context of switched ADP.[20] Future work will investigate the application of robust ADP update laws and resulting stability of the overall switched system.

III. Control Development

A. Approximate Dynamic Programming

The control objective is to determine a control policy u that minimizes the LQR cost performance index[a]

$$J_j(y,u) = \int_{t_0}^{\infty} \left(y(\tau)^T Q y(\tau) + u(\tau)^T R u(\tau) \right) \mathrm{d}\tau \tag{12}$$

subject to (11) (i.e., evaluated along the trajectories of the j^{th} subsystem) while regulating the output y to the origin, i.e., steady, level flight about the HSV trim condition, where $Q \in \mathbb{R}^{5\times 5}$ and $R \in \mathbb{R}^{2\times 2}$ are constant positive definite (PD) cost matrices.[b] Note that Q satisfies $\lambda_{\min}(Q) \leq Q \leq \lambda_{\max}(Q)$. The infinite horizon value function, or the cost-to-go, of the j^{th} nominal model, denoted as $V_j^* : \mathbb{R}^n \to \mathbb{R}_{\geq 0}$, is

$$V_j^*(y) = \min_{u(\tau)\in U} \int_t^{\infty} \left(y(\tau)^T Q y(\tau) + u(\tau)^T R u(\tau) \right) \mathrm{d}\tau, \tag{13}$$

where $U \subseteq \mathbb{R}^2$ denotes the controller action space. Note that the control policy u, which is defined in the subsequent section, is distinct for each subsystem j.

[a] (12) includes a subscript j to indicate that the objective is to minimize the infinite horizon cost for each subsystem. While (12) takes the same form for each j, generally, it will have different values, which is due to the different state trajectories induced by each subsystem's nominal model.

[b] Q and R can be set to different values for each nominal model. Without loss of generality, let Q and R be constant for each nominal model, i.e., $Q = Q_j$ and $R = R_j$ for all $j \in \mathbb{S}$.

Assumption 1. *Each value function V_j^* is continuously differentiable for all $j \in \mathbb{S}$.*

If the optimal value function is continuously differentiable, then the optimal value function V_j^* is a solution to the corresponding HJB equation

$$0 = \nabla_y V_j^*(y)\left(CA_j y + CB_j u_j^*\right) + y^T Q y + u_j^{*T} R u_j^*, \tag{14}$$

with the boundary condition $V_j^*(0) = 0 \; \forall j \in \mathbb{S}$, where $u_j^* : \mathbb{R}^n \to \mathbb{R}^m$ is the minimizing closed-loop policy for the j^{th} system defined as

$$u_j^* \triangleq \arg\min_u \left(\nabla_y V_j^*(y)\left(CA_j y + CB_j u\right) + y^T Q y + u^T R u\right)$$
$$u_j^* = -\frac{1}{2} R^{-1} (CB_j)^T \left(\nabla_y V_j^*(y)\right)^T, \tag{15}$$

provided the HJB in (14) admits a continuously differentiable PD solution.

B. Value Function Approximation

The solution to the HJB in (14) is the optimal value function. The solution to an LQR optimal control problem can be obtained via the ARE. To circumvent the need to solve the ARE online and at a large number of flight conditions, NNs can be used to approximate the value function in real-time. Although approximating the value function results in a suboptimal solution of the ARE, an online adaptive method is preferable because it can compensate for time-varying HSV model uncertainty. Parametric methods can be used to approximate the value function over a compact domain $y \in \Omega \subset \mathbb{R}^5$.[c] Since the function V_j^* is continuous and an approximation is sought on the compact set Ω, the Stone-Weierstrass Theorem is used to express the value function in (13) in Ω as

$$V_j^*(y) = W_j^T \phi(y) + \epsilon_j(y), \tag{16}$$

where $W_j \in \mathbb{R}^{15}$ is an unknown bounded vector of weights, $\phi : \mathbb{R}^5 \to \mathbb{R}^{15}$ is a user-defined vector of basis functions,[d] and $\epsilon_j : \mathbb{R}^5 \to \mathbb{R}$ is the bounded function approximation error. Substituting (16) into (15), the optimal control policy of the j^{th} subsystem, u_j^*, can be expressed in terms of the gradient of the value function V_j^* as

$$u_j^*(y) = -\frac{1}{2} R^{-1} (CB_j)^T \left(\nabla_y \phi(y)^T W_j + \nabla_y \epsilon_j(y)^T\right). \tag{17}$$

Assumption 2. *There exists a set of constants that bound the unknown weight vector W_j, the user-defined basis vector ϕ, and approximation error ϵ_j, from above such that $\sup_{j\in\mathbb{S}} \|W_j\| \leq \overline{W}$, $\sup_{y\in\Omega} \|\phi(y)\| \leq \overline{\phi}$, $\sup_{y\in\Omega} \|\nabla_y \phi(y)\| \leq \overline{\nabla_y \phi}$, $\sup_{y\in\Omega,\, j\in\mathbb{S}} \|\epsilon_j(y)\| \leq \overline{\epsilon}$, $\sup_{y\in\Omega,\, j\in\mathbb{S}} \|\nabla_y \epsilon_j(y)\| \leq \overline{\nabla_y \epsilon}$, where $\overline{W}$, $\overline{\phi}$, $\overline{\nabla_y \phi}$, $\overline{\epsilon}$, $\overline{\nabla_y \epsilon} \in \mathbb{R}_{>0}$.*[11, 14]

Since the ideal weights are unknown, a parametric estimate, called a critic weight vector $\hat{W}_{c,j} \in \mathbb{R}^{15}$, is substituted to estimate the optimal value function in (16) as $\hat{V}_j : \mathbb{R}^5 \times \mathbb{R}^{15} \to \mathbb{R}$, where

$$\hat{V}_j\left(y, \hat{W}_{c,j}\right) = \hat{W}_{c,j}^T \phi(y). \tag{18}$$

An actor weight vector $\hat{W}_{a,j} \in \mathbb{R}^{15}$, is used to provide an approximate version of (17). The approximate optimal control policy $\hat{u}_j : \mathbb{R}^5 \times \mathbb{R}^{15} \to \mathbb{R}$ is given by

$$\hat{u}_j\left(y, \hat{W}_{a,j}\right) = -\frac{1}{2} R^{-1} (CB_j)^T \left(\nabla_y \phi(y)^T \hat{W}_{a,j}\right). \tag{19}$$

The controller u applied to the system in (11) is $u \triangleq \hat{u}_j\left(y, \hat{W}_{a,j}\right)$.

[c] The subsequent stability analysis, which follows an existing result,[20] proves that if y is initialized within a compact set, then it will remain in a compact set.

[d] To facilitate the Lyapunov-based switching analysis,[20] there is no subscript j for the basis function because each subsystem uses the same basis function.

C. Bellman Error

The HJB equation in (14) is equal to zero under optimal conditions; however, substituting (18) and (19) into (14) results in a residual term $\delta_j : \mathbb{R}^5 \times \mathbb{R}^{15} \times \mathbb{R}^{15} \to \mathbb{R}$, which is referred to as the BE, defined as

$$\delta_j\left(y, \hat{W}_{c,j}, \hat{W}_{a,j}\right) \triangleq \nabla_y \hat{V}_j\left(y, \hat{W}_{c,j}\right)\left(CA_j y + CB_j \hat{u}_j\left(y, \hat{W}_{a,j}\right)\right) + \hat{u}_j\left(y, \hat{W}_{a,j}\right)^T R \hat{u}_j\left(y, \hat{W}_{a,j}\right) + y^T Q y, \tag{20}$$

where $\nabla_y \hat{V}_j\left(y, \hat{W}_{c,j}\right) = \hat{W}_{c,j}^T \nabla_y \phi(y)$. The BE is indicative of how close the actor and critic weight estimates are to the ideal weights (i.e., a measure of suboptimality). The mismatch between the estimates and the ideal values are defined as $\tilde{W}_{c,j} \triangleq W_j - \hat{W}_{c,j}$ and $\tilde{W}_{a,j} \triangleq W_j - \hat{W}_{a,j}$.

At each time instant, the BE in (20) is calculated using the current system state and weight estimates to obtain the instantaneous BE, denoted as $\delta_j \triangleq \delta_j\left(y, \hat{W}_{c,j}, \hat{W}_{a,j}\right)$. To alleviate the classical exploration versus exploitation problem in RL, the BE can also be evaluated along the system trajectory at user-defined points in the state space.[16] This process is called BE extrapolation, and it provides simulation of experience to avoid injecting an exploration signal. The BE is extrapolated at a user-specified number of points and locations $\{y_e : y_e \in \Omega\}_{e=1}^N$, where $N \in \mathbb{N}_{<\infty}$ denotes a user-specified number of BE extrapolation points in Ω.[e] The extrapolation trajectory data is represented as $\delta_{je} \triangleq \delta_j\left(y_e, \hat{W}_{c,j}, \hat{W}_{a,j}\right)$, $\omega_{je} \triangleq \omega_j\left(y_e, \hat{W}_{a,j}\right)$, where $\omega_j\left(y, \hat{W}_{a,j}\right) \triangleq \nabla_y \phi(y)\left(CA_j y + CB_j \hat{u}_j\left(y, \hat{W}_{a,j}\right)\right)$. To facilitate improved learning in comparison to a standard gradient update policy, let $\Gamma_j \in \mathbb{R}^{15\times 15}$ be a time-varying least-squares gain matrix of the j^{th} subsystem.

D. Update Policies

Based on the subsequent Lyapunov-based stability analysis, critic and actor weights are updated with the update laws[20]

$$\dot{\hat{W}}_{c,j} \triangleq \operatorname{proj}\left\{-\eta_{c1}\Gamma_j \frac{\omega_j}{1+\nu\omega_j^T\Gamma_j\omega_j}\delta_j - \eta_{c2}\frac{1}{N}\sum_{e=1}^{N}\frac{\omega_{je}}{1+\nu\omega_{je}^T\Gamma_j\omega_{je}}\delta_{je}\right\}, \tag{21}$$

$$\begin{aligned}\dot{\hat{W}}_{a,j} \triangleq \operatorname{proj}\Bigg\{&-\eta_{a1}\left(\hat{W}_{a,j} - \hat{W}_{c,j}\right) - \eta_{a2}\hat{W}_{a,j} \\ &+\eta_{c1}\frac{G_{\phi j}^T \hat{W}_{a,j}\omega_j^T}{4\left(1+\nu\omega_j^T\Gamma_j\omega_j\right)}\hat{W}_{c,j} + \eta_{c2}\frac{1}{N}\sum_{e=1}^{N}\frac{G_{\phi je}^T\hat{W}_{a,j}\omega_{je}^T}{4\left(1+\nu\omega_{je}^T\Gamma_j\omega_{je}\right)}\hat{W}_{c,j}\Bigg\},\end{aligned} \tag{22}$$

$$\dot{\Gamma}_j \triangleq \left(\lambda\Gamma_j - \eta_{c1}\frac{\Gamma_j\omega_j\omega_j^T\Gamma_j}{\left(1+\nu\omega_j^T\Gamma_j\omega_j\right)^2} - \eta_{c2}\Gamma_j\left(\frac{1}{N}\sum_{e=1}^{N}\frac{\omega_{je}\omega_{je}^T}{1+\nu\omega_{je}^T\Gamma_j\omega_{je}}\right)\Gamma_j\right)\cdot \mathbf{1}_{\{\underline{\Gamma}\le\|\Gamma_j\|\le\overline{\Gamma}\}}, \tag{23}$$

where $\operatorname{proj}\{\cdot\}$ is a smooth projection operator,[f] η_{c1}, η_{c2}, η_{a1}, η_{a2}, ν, $\lambda \in \mathbb{R}$ are positive constant learning gains, $G_{\phi j} \triangleq G_{\phi j}(y) = \nabla_y\phi(y) C_j B_j R^{-1} B_j^T C_j^T \nabla_y \phi(y)^T$, $G_{\phi je} \triangleq G_{\phi j}(y_e)$, $\sup_{y\in\Omega,\, j\in\mathbb{S}} \|G_{\phi j}(y)\| \le \overline{G_\phi}$, and $\underline{\Gamma}, \overline{\Gamma} \in \mathbb{R}_{>0}$ are upper and lower bound constants of the least-squares gain matrix. Furthermore, let the projection operators have lower bounds $\underline{\hat{W}_a} \in \mathbb{R}$ and $\underline{\hat{W}_c} \in \mathbb{R}$ and upper bounds $\overline{\hat{W}_a} \in \mathbb{R}$ and $\overline{\hat{W}_c} \in \mathbb{R}$ for each subsystem j. For brevity and without loss of generality for each subsystem j let $\overline{\Gamma}$, $\underline{\Gamma}$, η_{c1}, η_{c2}, η_{a1}, η_{a2}, λ,and ν have identical gain values, e.g., $\nu = \nu_1 = \nu_2 = \cdots = \nu_j \, \forall j \in \mathbb{S}$. To facilitate sufficient exploration of the state space via simulation of experience and convergence in the Lyapunov-based analysis, the following assumption is made.

Assumption 3. *Over the compact set Ω, a finite set of off-trajectory points $\{y_e : y_e \in \Omega\}_{e=1}^N$ exists such that $0 < \underline{c} \triangleq \frac{1}{N}\inf_{t\in\mathbb{R}_{\ge 0}} \lambda_{\min}\left\{\sum_{e=1}^{N}\frac{\omega_{je}\omega_{je}^T}{1+\nu\omega_{je}^T\Gamma_j\omega_{je}}\right\}$ for all $t \in \mathbb{R}_{\ge 0}$ and $j \in \mathbb{S}$.*

[e] The number of extrapolation points N can vary from subsystem-to-subsystem. However, without loss of generality, let $N = N_j \, \forall j \in \mathbb{S}$.

[f] For brevity, the projection operators are not defined in this paper. See the references for a definition of this projection operator.[25]

E. Stability Analysis

For the dynamical system in (11) and the cost functional in (12), the solution to the ARE is symmetric and PD.[26] Hence, the optimal value function can be bounded as

$$\beta_{1,j}\left\|y\right\|^2 \leq V_j^*\left(y\right) \leq \beta_{2,j}\left\|y\right\|^2, \forall j \in \mathbb{S},\, \beta_{1,j},\, \beta_{2,j} \in \mathbb{R}_{\geq 0}, \tag{24}$$

where $\beta_{1,j}$ and $\beta_{2,j}$ are the minimum and maximum eigenvalues of the solution to the ARE for the j^{th} subsystem.

Existing model-based ADP methods[16] can be applied to an individual subsystem to achieve uniformly ultimately bounded (UUB) stability. Such results cannot be used to conclude stability of the overall switching sequence. Since the critic and actor weight errors $\tilde{W}_{c,j}$ and $\tilde{W}_{a,j}$ are states in each subsystem's Lyapunov function, then there will be an instantaneous increase in the value of the Lyapunov function. The instantaneous jump is due to the difference in $\left\|\tilde{W}_{c,j}\right\|$, $\left\|\tilde{W}_{a,j}\right\|$ and $\left\|\tilde{W}_{c,j+1}\right\|$, $\left\|\tilde{W}_{a,j+1}\right\|$, respectively. To guarantee that the overall switching sequence is stable, each subsystem must remain active sufficiently long so that the instantaneous increase caused by the switching instance is smaller than the increase that precedes it. The sufficient time required to guarantee stability of the overall switching sequence is the dwell-time condition $\tau^* \triangleq \frac{\ln\left(\mu^{N_\sigma}\right)}{\zeta_0-\zeta^*}$, where $\mu \triangleq \frac{\sup_{j\in\mathbb{S}}\{\beta_{2,j}\}\left\|y(0)\right\|^2+\Xi}{\Xi}$, $\Xi \triangleq \underline{\Gamma}^{-1}\left\|\overline{\hat{W}}_c - \underline{\hat{W}}_c\right\|^2 + \left\|\overline{\hat{W}}_a - \underline{\hat{W}}_a\right\|^2$, $\zeta_0 \in \mathbb{R}_{>0}$ is a constant, and $\zeta^* \in (0,\zeta_0)$ is an arbitrarily selected decay rate that satisfies the inequality $\mu^{N_\sigma}e^{-\zeta_0 t} \leq e^{-\zeta^* t}$. Assuming that the dwell-time condition is met, then UUB stability of the overall switching sequence can be proven.[20] To facilitate the subsequent analysis, let $Z_j \triangleq \left[y^T, \tilde{W}_{c,j}^T, \tilde{W}_{a,j}^T\right]$ denote a concatenated state, and let $V_{L,j} : \mathbb{R}^{35} \to \mathbb{R}$ be a candidate Lyapunov function for the j^{th} mode be defined as

$$V_{L,j}\left(Z_j\right) \triangleq \left(y\right) + \frac{1}{2}\tilde{W}_{c,j}^T\Gamma_j^{-1}\tilde{W}_{c,j} + \frac{1}{2}\tilde{W}_{a,j}^T\tilde{W}_{a,j}, \tag{25}$$

where j represents the active subsystem mode. Define the sequence of times instants at which a switching event occurs as $\{t_{N_\sigma}\}$, such that $0 < t_1 < t_2 < \cdots < t_{N_\sigma} < t < t_{N_\sigma+1}$ and $N_\sigma \in \mathbb{N}_{>0}$ denotes the number of switching events. Using the positive definiteness of V_j^* and [27, Lemma 4.3], (25) can generally be bounded as $\underline{v}_{l,j}\left(\left\|Z_j\right\|\right) \leq V_{L,j}\left(Z_j\right) \leq \overline{v}_{l,j}\left(\left\|Z_j\right\|\right)$ using class $\mathcal{K}$ functions $\underline{v}_{l,j}, \overline{v}_{l,j} : \mathbb{R}_{\geq 0} \to \mathbb{R}_{\geq 0}$. Furthermore, let $l \in \mathbb{R}_{>0}$ be a positive bounding constant or the parametric uncertainty terms,[20] and $\Lambda \triangleq \min\left\{\frac{1}{2}\lambda_{\min}\left\{Q\right\}, \frac{1}{16}\left(\eta_{a1}+\eta_{a2}\right), \frac{1}{12}\eta_{c2}\underline{c}\right\}$. Furthermore, define $\mathcal{R} \in \mathbb{R}_{>0}$ as the radius of a ball $\mathcal{B}_\mathcal{R}$ centered at the origin, where $\mathcal{B}_\mathcal{R} \subset \mathbb{R}^n$. Let $\chi \subset \mathbb{R}^{35}$ be a ball centered at the origin with radius $\overline{v}_{l,j}^{-1}\left(\underline{v}_{l,j}\left(\mathcal{R}\right)\right)$.

Theorem 1. *Provided Assumptions 1 and 2 hold, the weight update laws in (21)-(23) are used, and the gain conditions*

$$\eta_{a1} + \eta_{a2} > \frac{1}{\sqrt{\nu\underline{\Gamma}}}\left(\eta_{c1}+\eta_{c2}\right)\overline{WG_\phi}, \tag{26}$$

$$\underline{c} > \frac{3\left(\eta_{c1}+\eta_{c2}\right)^2\overline{W}^2\overline{G_\phi}^2}{16\nu\underline{\Gamma}\left(\eta_{a1}+\eta_{a2}\right)\eta_{a2}} + \frac{3\eta_{a1}}{\eta_{c2}}, \tag{27}$$

$$\max_{j\in\mathbb{S}}\left\{\frac{\beta_{2,j}}{\beta_{1,j}}\right\}\sqrt{\frac{2l}{\Lambda}} < \mathcal{R}, \tag{28}$$

are satisfied, the system state $y\left(\cdot\right)$, the value function weight estimate error $\tilde{W}_{c,j}\left(\cdot\right)$, and the control policy weight estimate error $\tilde{W}_{a,j}\left(\cdot\right)$, are UUB. Hence, the error between the stabilizing control policy for each mode $\hat{u}_j\left(\cdot\right)$ in (19) and its respective optimal control policy $u_j^*\left(\cdot\right)$ in (15) is UUB. The result in [27, Theorem 4.18] can be invoked to show that every trajectory of each subsystem $Z_j\left(t\right)$ that satisfies the initial condition $\left\|Z_j\left(t_0\right)\right\| \leq \overline{v}_{l,j}^{-1}\left(\underline{v}_{l,j}\left(\mathcal{R}\right)\right)$ is bounded for all $t \in \mathbb{R}_{\geq 0}$. That is, $Z_j \in \chi\, \forall t \in \mathbb{R}_{\geq 0}$, $j \in \mathbb{S}$. Since $Z_j \in \chi$ it follows that the individual states of Z_j lie on compact sets; hence, $y \in \Omega$.

Provided a minimum dwell-time τ^* is satisfied, then the switched system consisting of a family of subsystems with the dynamics in (11) with a properly designed dwell-time, $\tau \geq \tau^*$ ensures that the states, critic estimate errors, and actor estimate errors will converge to a neighborhood of the origin for all $t \geq T$, where $V_{L,B} \in \mathbb{R}$ is the maximum ultimate bound for all subsystems, $T \in \mathbb{R}_{\geq 0}$ is the time required to reach the ultimate bound $V_{L,B}$.

Proof. The proof follows from Theorem 1 and 2 of the switched ADP result.[20] □

IV. Simulation

A simulation is performed by switching between 15 different HSV models, which simulate an initial velocity of Mach 7.5 at 85,000 feet.[6] The simulation was executed for 35 seconds to sufficiently cycle through the different HSV temperature profiles.[6] 15 HSV models were used to capture the HSV body temperature profiles at evenly-spaced intervals. Selecting the 15 models is application-specific, and depends on the temperature variation for the HSV forebody and aftbody. The user could select a larger or smaller number of models, which would effect the uncertainty associated with each nominal HSV model and where the switching instances occur. The nominal models are switched every 2.33 seconds, which activates each of the 15 switched subsystems for an equal amount of time. Each linearized model was determined from full nonlinear dynamics[1] and are the identical nominal models from a previous LPV HSV simulation.[6] See the robust HSV control design result[6] for more detail on the HSV, how the disturbances correlate with realistic aerodynamic effects, and how the linearized models were determined. The initial condition of the HSV about its trim condition was

$$x(0) = \begin{bmatrix} -1, & 10^{-3}, & 10^{-8}, & 5, & -5\cdot 10^{-4}, & 0, & 0, & 0, & 0, & 0, & 0 \end{bmatrix}^T. \tag{29}$$

The cost matrices were selected as $Q = 0.1 \cdot I_{5\times 5}$ and $R = 20 \cdot I_{2\times 2}$. The parameters used to facilitate ADP were $\eta_{c1} = 0.01$, $\eta_{c2} = 0.0001$, $\eta_{a1} = 0.01$, $\eta_{a2} = 0.0005$, $\lambda = 0.5$, $\nu = 0.005$, $\overline{\Gamma} = 100$, $\underline{\Gamma} = 0.01$, $\overline{\hat{W}_c} = \overline{\hat{W}_a} = 10\left\|\hat{W}_{c,1}(0)\right\|$, and $\underline{\hat{W}_c} = \underline{\hat{W}_a} = 0.01$. $N = 72$ BE extrapolation points were placed uniformly throughout a subset of the reduced-order model state space. The basis function is selected as

$$\phi(y) = \begin{bmatrix} y_1^2 & y_1y_2 & y_1y_3 & y_1y_4 & y_1y_5 & y_2^2 & y_2y_3 & y_2y_4 & y_2y_5 & y_3^2 & y_3y_4 & y_3y_5 & y_4^2 & y_4y_5 & y_5^2 \end{bmatrix}^T \tag{30}$$

so that $V_j^*(y) = y^T P_j y = W_j^T \phi(y)$, where $P_j \in \mathbb{R}^{5\times 5}$ is a solution to the ARE for the j^{th} subsystem. Initialization of $\hat{W}_{c,1}$ and $\hat{W}_{a,1}$ far from W_1 will cause instability, which is expected due to the initial conditions required to guarantee stability for a switched ADP-based controller.[20] Hence, the weight estimates $\hat{W}_{c,1}(0) = \hat{W}_{a,1}(0)$ are initialized from solving the ARE using the fifteenth nominal state model A_{15} and B_{15}. In this case, the actor and critic weight estimates are initialized for a HSV flight condition that does not match the initial flight condition. Therefore, the weight estimates are not initialized at their ideal values; instead, the weight estimates start within a neighborhood of their ideal values. In practice, one should consider initializing the weights by solving the ARE at the initial trim condition of the HSV. The initial weight estimate values can be based on those from a previous flight-test experiment. The least-squares gain matrix is initialized as $\Gamma_1(0) = 50 \cdot I_{15\times 15}$. When a switching event occurs, let Γ_{j+1} be initialized as the final value from the previously active subsystem Γ_j, e.g., at switching instance t let $\Gamma_j(t^-) = \Gamma_{j+1}(t^+)$.

For simplicity, the weight estimates do not change at each switching instant. For example, at switching instance $t = 2.33$, the new critic weight estimate $\hat{W}_{c,2}(2.33^+) = \hat{W}_{c,1}(2.33^-)$. This is done because the weight estimates are not expected to differ significantly after switching, and the final weight estimates obtained from the previous nominal model should be valid for the next nominal model.

To simulate realistic actuators, a second-order actuator model of the control inputs was used. The commanded control input is determined from (19) such that $u = \hat{u}_j\left(y, \hat{W}_{a,j}\right) = [\delta_{e,cmd}, \delta_{c,cmd}]^T$. The second-order actuator model for the elevator displacement is $\ddot{\delta}_e = -2\zeta_e\omega_e\dot{\delta}_e + \omega_e^2(\delta_{e,cmd} - \delta_e)$, and the second-order actuator model for the canard displacement is $\ddot{\delta}_c = -2\zeta_c\omega_c\dot{\delta}_c + \omega_c^2(\delta_{c,cmd} - \delta_c)$, where $\zeta_e = \zeta_c = 0.7$ is the actuator damping ratio and $\omega_e = \omega_c = 50$ Hz is the actuator natural frequency.[7,28]

For the stability analysis of the developed controller, disturbances were omitted. However, to show robustness of the developed method, bounded nonlinear disturbances were injected into the dynamics as

$$d(t) = \begin{bmatrix} d_{\dot{V}}(t) & 0 & d_{\dot{q}}(t) & 0 & 0 & 0 & d_{\ddot{\eta}_{s,1}}(t) & 0 & d_{\ddot{\eta}_{s,2}}(t) & 0 & d_{\ddot{\eta}_{s,3}}(t) \end{bmatrix}, \tag{31}$$

where $d_{\dot{V}}(t) \in \mathbb{R}$ represents a longitudinal acceleration disturbance, $d_{\dot{\alpha}}(t) \in \mathbb{R}$ represents an angle of attack rate-of-change disturbance, $d_{\dot{q}}(t) \in \mathbb{R}$ represents an angular acceleration disturbance, and $d_{\ddot{\eta}_{s,1}}\, d_{\ddot{\eta}_{s,2}}(t)\, d_{\ddot{\eta}_{s,3}}(t) \in \mathbb{R}$ represent structural mode acceleration disturbances. The magnitude of the disturbances are meant to reflect realistic disturbances of a HSV at Mach 7.5 and 85,000 feet.[6] Figure 1 shows the values of the disturbances in (7). From top-to-bottom, the time-varying disturbances correspond to ΔV, q, $\dot{\eta}_{s1}$, $\dot{\eta}_{s2}$, and $\dot{\eta}_{s3}$, respectively.

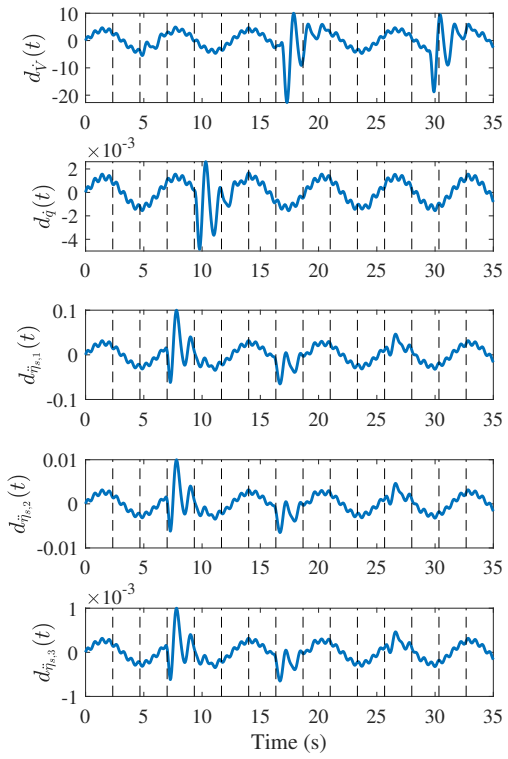

Figure 1. Time-varying disturbances that correspond to the states ΔV, q, $\dot{\eta}_{s1}$, $\dot{\eta}_{s2}$, and $\dot{\eta}_{s3}$, respectively. The effects of these disturbances can be seen in Figures 2-7. The dashed lines designate the time at which a switching instance occurs.

In all of the following figures, the dashed lines designate the time at which a switching instance occurs. Figure 2 shows the forward airspeed velocity relative to the trim condition ΔV of the HSV in the simulation case. From (31), there is a disturbance that is directly being injected into the $\Delta\dot{V}$ state. The disturbance term in (31) prohibits convergence of ΔV to the HSV trim forward velocity. The variation in ΔV small; hence, V is within an acceptable range of variation for a HSV.[29]

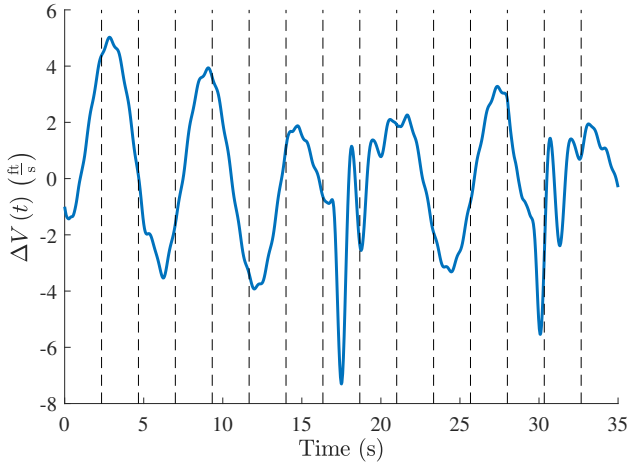

Figure 2. The forward velocity of the HSV measured from its trim condition ΔV. Relative to the trim speed of the HSV (Mach 7.5) the variations in the above graph are small. The injected disturbances prohibit this state from converging. The dashed lines designate the time at which a switching instance occurs.

Figure 3 shows the angle of attack measured from the trim condition $\Delta\alpha$ of the HSV. The values of $\Delta\alpha$ in Figure 3 are relatively small because no disturbance is injected into the dynamics for α. There is, however, a disturbance injected into $\dot{q}$, which is coupled with $\Delta\alpha$. The the angle of attack trim condition is zero; hence,

α lies in an acceptable range for HSVs $\alpha \in (-5, 10)$.[29]

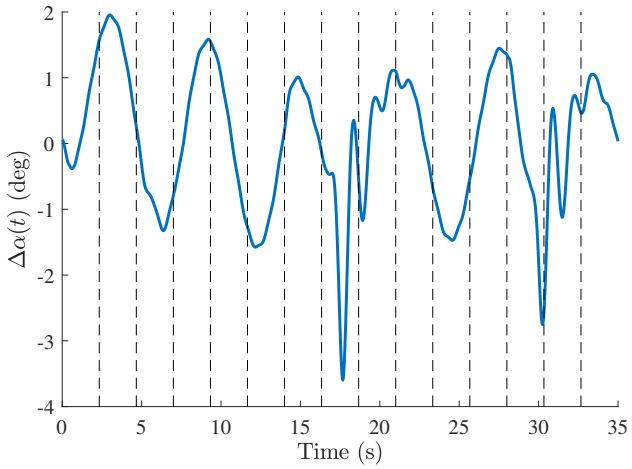

Figure 3. The angle of attack of the HSV measured from its trim condition $\Delta\alpha$. The range of $\Delta\alpha$ is small relative to the bounds of $\Delta\alpha$ required by the trim condition.[6] Similar to Figure 2, the disturbances in $\dot{q}$ prohibit this state from converging. The dashed lines designate the time at which a switching instance occurs.

Figure 4 shows the pitch rate q of the HSV in the simulation case. The trim condition of the pitch rate is zero. From (31), there is a disturbance that is directly being injected into the $\dot{q}$ state. While there are a few instances where the magnitude of q is large (approximately 8 degrees per second), that is the case momentarily and coincides with disturbances injected into other states. The pitch angle of the HSV, which is shown in a subsequent figure, is smaller in magnitude (typically within ± 2 degrees of the trim condition). q is within an acceptable range for HSVs $q \in (-10, 10)$.[29]

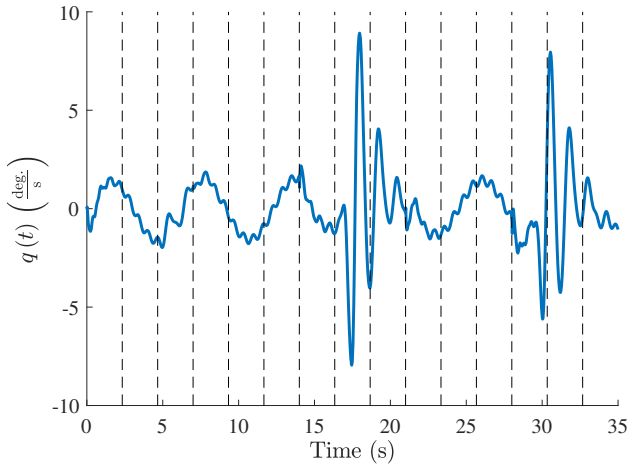

Figure 4. The pitch rate q of the HSV measured from its trim condition. The injected disturbances cause the oscillation in this state. The dashed lines designate the time at which a switching instance occurs.

Figure 5 shows the change in height from the trim condition Δh of the HSV in the simulation case. The altitude of this simulation aircraft is 85,000 feet. Hence, a variation on the order of ± 1 feet is small relative to the overall altitude (approximately 0.001%). The variation in Δh is small; hence, h is within an acceptable range of variation.[29]

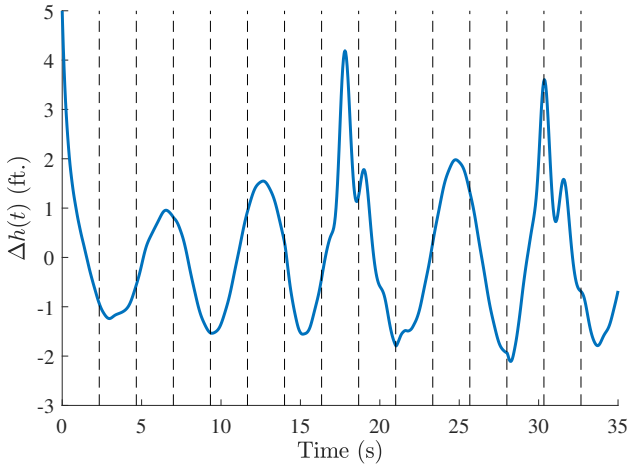

Figure 5. The height of the HSV measured from its trim condition of 85,000 feet Δh. These height variations are very small in comparison to the altitude of the HSV. The dashed lines designate the time at which a switching instance occurs.

Figure 6 shows the pitch angle measured from the HSV trim condition θ for the simulation. There is no disturbance directly injected into this state. However, a disturbance is injected into $\dot{q}$, which is coupled with $\Delta\theta$. Despite the disturbances, the system remains stable. The the pitch angle trim condition is zero; hence θ is within an acceptable range, which is determined from the admissible angle of attack and flight path angle values as $\theta \in (-8, 13)$.[29]

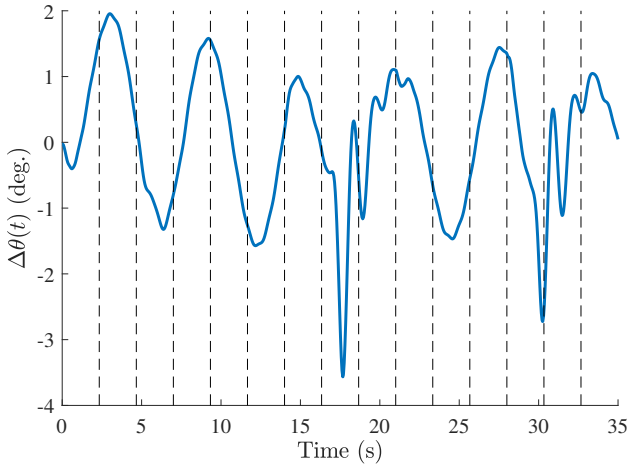

Figure 6. The pitch angle $\Delta\theta$ of the HSV measured from its trim condition. The values of $\Delta\theta$ remain between ± 4 degrees, which is a small range. The dashed lines designate the time at which a switching instance occurs.

Figure 7 shows the actuator deflections δ_e and δ_c for the HSV. While the deflection behavior appears high-frequency, it is achievable by the selected actuators. Note that the displacement of each actuator is within $\pm 15°$ from their respective trim conditions, which are acceptable values for a HSV.[29] The canard is the actuator with the larger amount of displacement. The maximum deflection of the canard δ_c is $-14.92°$. There are large, instantaneous deflections that occur at multiple switching instances. This is due to the discrete change of nominal dynamics and change from W_j to W_{j+1}.

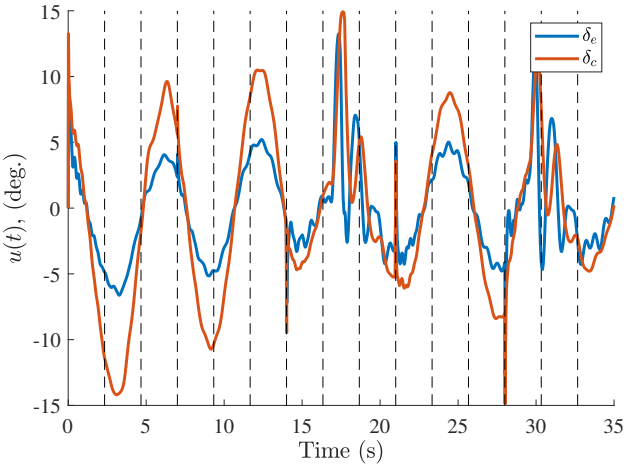

Figure 7. The control inputs of the system. δ_e is the deflection angle of the elevator from the trim condition, and δ_c is the deflection angle of the canard from the trim condition. The switching instances and injected disturbances cause instantaneous jumps in the control inputs. The dashed lines designate the time at which a switching instance occurs.

V. Conclusion

This paper investigated the application of a model-based switched ADP controller to a HSV to account for slowly time-varying aerothermoelastic effects. Simulation results show that the switched ADP controller facilitates state regulation in the presence of disturbances and model uncertainty. Future work will include a Lyapunov-based analysis of the switched ADP problem with model uncertainty, disturbances, and output feedback to reduce the number of output states required to apply ADP-based controllers.

VI. Acknowledgments

This research is supported by Office of Naval Research Grant N00014-13-1-0151, AFOSR award number FA9550-18-1-0109, AFOSR award number FA9550-19-1-0169, NEEC award number N00174-18-1-0003, and AFRL. Any opinions, findings and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the sponsoring agencies or Draper.

References

[1]Bolender, M. A. and Doman, D. B., "Nonlinear Longitudinal Dynamical Model of an Air-Breathing Hypersonic Vehicle," *J. Spacecraft Rockets*, Vol. 44, No. 2, April 2007, pp. 374–387.

[2]Heeg, J., Gilbert, M. G., and Pototzky, A. S., "Active Control of Aerothermoelastic Effects for a Conceptual Hypersonic Aircraft," *J. Aircraft*, Vol. 30, No. 4, 1993, pp. 453–458.

[3]Lind, R., "Linear Parameter-Varying Modeling and Control of Structural Dynamics with Aerothermoelastic Effects," *J. Guid. Control Dynam.*, Vol. 25, No. 4, July 2002, pp. 733–739.

[4]Gao, G. and Wang, J., "Reference command tracking control for an air-breathing hypersonic vehicle with parametric uncertainties," *J. of the Franklin Inst.*, Vol. 350, No. 5, 2013, pp. 1155–1188.

[5]Zhang, L., Nie, L., Cai, B., Yuan, S., and Wang, D., "Switched linear parameter-varying modeling and tracking control for flexible hypersonic vehicle," *Aerosp. Science and Technology*, Vol. 95, 2019, pp. 105445.

[6]Wilcox, Z. D., MacKunis, W., Bhat, S., Lind, R., and Dixon, W. E., "Lyapunov-Based Exponential Tracking Control of a Hypersonic Aircraft with Aerothermoelastic Effects," *AIAA J. Guid. Control Dyn.*, Vol. 33, No. 4, 2010, pp. 1213–1224.

[7]Eugene, L., Kevin, W., and Howe, D., "Robust and adaptive control with aerospace applications," 2013.

[8]Rugh, W. J. and Shamma, J. S., "Research on gain scheduling," *Automatica*, Vol. 36, No. 10, Oct. 2000, pp. 1401–1425.

[9]Shamma, J. and Athans, M., "Gain scheduling: Potential hazards and possible remedies," *IEEE Control System Magazine*, Vol. vol. 12, No. no. 3, 1992, pp. 101–107.

[10]Sutton, R. S. and Barto, A. G., *Reinforcement Learning: An Introduction*, MIT Press, Cambridge, MA, USA, 1998.

[11]Kamalapurkar, R., Walters, P. S., Rosenfeld, J. A., and Dixon, W. E., *Reinforcement learning for optimal feedback control: A Lyapunov-based approach*, Springer, 2018.

[12]Vrabie, D., Vamvoudakis, K. G., and Lewis, F. L., *Optimal Adaptive Control and Differential Games by Reinforcement Learning Principles*, The Institution of Engineering and Technology, 2013.

[13]Lewis, F. L. and Liu, D., *Reinforcement learning and approximate dynamic programming for feedback control*, Vol. 17, John Wiley & Sons, 2013.

[14]Bhasin, S., Kamalapurkar, R., Johnson, M., Vamvoudakis, K. G., Lewis, F. L., and Dixon, W. E., "A novel actor-critic-identifier architecture for approximate optimal control of uncertain nonlinear systems," *Automatica*, Vol. 49, No. 1, Jan. 2013, pp. 89–92.

[15]Kamalapurkar, R., Dinh, H., Bhasin, S., and Dixon, W. E., "Approximate Optimal Trajectory Tracking for Continuous-Time Nonlinear Systems," *Automatica*, Vol. 51, Jan. 2015, pp. 40–48.

[16]Kamalapurkar, R., Walters, P., and Dixon, W. E., "Model-based reinforcement learning for approximate optimal regulation," *Automatica*, Vol. 64, 2016, pp. 94–104.

[17]Kamalapurkar, R., Andrews, L., Walters, P., and Dixon, W. E., "Model-based reinforcement learning for infinite-horizon approximate optimal tracking," *IEEE Trans. Neural Netw. Learn. Syst.*, Vol. 28, No. 3, 2017, pp. 753–758.

[18]Mitrovic, D., Klanke, S., and Vijayakumar, S., "Adaptive Optimal Feedback Control with Learned Internal Dynamics Models," *From Motor Learning to Interaction Learning in Robots*, edited by O. Sigaud and J. Peters, Vol. 264 of *Studies in Computational Intelligence*, Springer Berlin Heidelberg, 2010, pp. 65–84.

[19]Deisenroth, M. P. and Rasmussen, C. E., "PILCO: A model-based and data-efficient approach to policy search," *Int. Conf. Mach. Learn.*, 2011, pp. 465–472.

[20]Greene, M., Abudia, M., Kamalapurkar, R., and Dixon, W. E., "Model-based Reinforcement Learning for Optimal Feedback Control of Switched Systems," *Proc. IEEE Conf. Decis. Control*, 2020, pp. 162–167.

[21]Vamvoudakis, K. G., "Event-triggered optimal adaptive control algorithm for continuous-time nonlinear systems," *IEEE/CAA J. Autom. Sinica*, Vol. 1, No. 3, 2014, pp. 282–293.

[22]Vamvoudakis, K. G., Mojoodi, A., and Ferraz, H., "Event-triggered optimal tracking control of nonlinear systems," *Int. J. Robust Nonlinear Control*, Vol. 27, No. 4, 2017, pp. 598–619.

[23]Williams, T., Bolender, M. A., Doman, D. B., and Morataya, O., "An Aerothermal Flexible Mode Analysis of a Hypersonic Vehicle," *AIAA Paper 2006-6647*, Aug. 2006.

[24]Huang, Y., Wang, D., and Liu, D., "Bounded robust control design for uncertain nonlinear systems using single-network adaptive dynamic programming," *Neurocomputing*, Vol. 266, 2017, pp. 128–140.

[25]Krstic, M., Kanellakopoulos, I., and Kokotovic, P. V., *Nonlinear and Adaptive Control Design*, John Wiley & Sons, New York, NY, USA, 1995.

[26]Anderson, B. D. and Moore, J. B., *Optimal control: linear quadratic methods*, Courier Corp., 1971.

[27]Khalil, H. K., *Nonlinear Systems*, Prentice Hall, Upper Saddle River, NJ, 3rd ed., 2002.

[28]Dickinson, B., Nivison, S., Hart, A., Hung, C., Bialy, B., and Stockbridge, S., "Robust and adaptive control of a rocket boosted missile," *2015 American Control Conference (ACC)*, IEEE, 2015, pp. 2520–2532.

[29]Fiorentini, L., Serrani, A., Bolender, M. A., and Doman, D. B., "Nonlinear Robust Adaptive Control of Flexible Air-Breathing Hypersonic Vehicle," *J. Guid. Control Dynam.*, Vol. 32, No. 2, April 2009, pp. 402–417.