Hamiltonian-Driven Hybrid Adaptive Dynamic Programming

Yongliang Yang, *Member, IEEE*, Kyriakos G. Vamvoudakis, *Senior Member, IEEE*, Hamidreza Modares, *Member, IEEE*, Yixin Yin, *Member, IEEE*, and Donald C. Wunsch, *Fellow, IEEE*

***Abstract*—This article presents a model-based hybrid adaptive dynamic programming (ADP) framework consisting of continuous feedback-based policy evaluation and policy improvement steps as well as an intermittent policy implementation procedure. This results in an intermittent ADP with a quantifiable performance and guaranteed closed-loop stability of the equilibrium point. To investigate the effect of aperiodic sampling on the communication bandwidth and the control performance of the intermittent ADP algorithms, we use a Hamiltonian-driven unified framework. With such a framework, it is shown that there is a tradeoff between the communication burden and the control performance. We finally show that the developed policies exhibit Zeno-free behaviors. Simulation examples show the efficiency of the proposed framework along with quantifiable comparisons of the policies with different intermittent information.**

***Index Terms*—Adaptive dynamic programming (ADP), communication bandwidth, control performance, Hamiltonian-driven framework, intermittent control, tradeoff.**

I. Introduction

IN THE context of cyber-physical systems (CPS) and the Internet of Things (IoT), where the resources, namely, the actuators, sensors, and controllers are remotely located and communicate through a shared medium or cloud, it is critical for such resources to be used efficiently and effectively to assure stability and optimality. Recently, emerging control scheduling strategies, including event-triggered control [1], [2] and self-triggered control [3], [4], have been used to reduce the communication burden. This is achieved mostly by operating the system in an open-loop manner, and closing the loop only when a predetermined condition that guarantees stability and/or performance is violated.

The threshold in the event-triggering condition should be properly designed to avoid the Zeno-behavior. An exponentially decaying threshold is used for consensus control of multiagent systems in [5]. However, the exponentially decaying threshold may lead to infinitely fast sampling [6]. Liu and Jiang [6] introduced an additional dynamical system to overcome the limitation of the exponentially decreasing threshold signal. *Girard* extended the event-triggered control [1] with a dynamic mechanism that may reduce even further the communication and computation burden [7]. Yang *et al.* [8] employed the dynamic triggering mechanism for the containment control problem of multiagent systems. However, due to the dependency of the performance on the update of the control input, reducing the communication load may have an impact on the efficiency. In this article, we quantify the tradeoff between the bandwidth and the performance of the intermittent adaptive dynamic programming (ADP) within a Hamiltonian-driven framework.

Reinforcement learning (RL) and/or ADP, which evolves from off-line iteration [9]–[12] to online learning [13], [14], have been successfully developed to provide an efficient way to obtain optimal controllers in an adaptive manner online [15], [16]. Extensions of ADP algorithms have been made for continuous-time [17] and discrete-time systems [18], single-agent [19] and multiagent systems [20]–[22]. Recently, intermittent ADP has been developed to reduce the communication load and the computation burden [23]–[27]. Existing intermittent ADP methods aim to design an event-triggering condition to guarantee the closed-loop stability of the equilibrium point during the learning phase. However, there are no results to quantify the performance. In order to investigate the effect of aperiodic state sampling on the learning phase of the ADP, this article extends the Hamiltonian-driven continuous

Manuscript received April 9, 2019; revised August 28, 2019; accepted December 12, 2019. This work was supported in part by the National Natural Science Foundation of China under Grant 61903028 and Grant 61333002, in part by the Fundamental Research Funds for the China Central Universities of USTB under Grant FRF-TP-18-031A1 and Grant FRF-BD-19-002A, in part by the China Postdoctoral Science Foundation under Grant 2018M641197, in part by the National Science Foundation under Grant CAREER CPS-1851588 and Grant S&AS 1849198, in part by ONR Minerva under Grant N00014-18-1-2160, in part by the Lifelong Learning Machines Program from DARPA/Microsystems Technology Office, and in part by the Army Research Laboratory under Agreement W911NF-18-2-0260. This article was recommended by Associate Editor D. Yue. *(Corresponding author: Yongliang Yang.)*

Yongliang Yang and Yixin Yin are with the Key Laboratory of Knowledge Automation for Industrial Processes, Ministry of Education, School of Automation and Electrical Engineering, University of Science and Technology Beijing, Beijing 100083, China (e-mail: yangyongliang@ieee.org; yyx@ies.ustb.edu.cn).

Kyriakos G. Vamvoudakis is with the Daniel Guggenheim School of Aerospace Engineering, Georgia Institute of Technology, Atlanta, GA 30332 USA (e-mail: kyriakos@gatech.edu).

Hamidreza Modares is with the Department of Mechanical Engineering, Michigan State University, East Lansing, MI 48824 USA (e-mail: modaresh@msu.edu).

Donald C. Wunsch is with the Department of Electrical and Computer Engineering, Missouri University of Science and Technology, Rolla, MO 65401 USA (e-mail: dwunsch@mst.edu).

Color versions of one or more of the figures in this article are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TSMC.2019.2962103

ADP [28] to the intermittent case with stability, bandwidth, and performance analysis.

This article presents a more comprehensive framework, referred to as Hamiltonian-driven hybrid ADP. First, the Hamiltonian-driven continuous ADP [28] is used for the learning phase, including policy evaluation and policy improvement steps. Then, Hamiltonian-driven intermittent ADPs, including both static and dynamic intermittent ADPs, are developed for the policy implementation such that the control is updated in an intermittent fashion. The bandwidth and performance analysis of the continuous and intermittent ADPs are performed using the Hamiltonian-driven framework.

Contributions: The contributions of this article are twofold. We provide the theoretical analysis to compare the communication bandwidth for the static and dynamic intermittent feedback in terms of the interevent interval and its lower bound. Moreover, compared to existing event-triggered control design, the control performance improvement for the intermittent feedback is considered in this article. A unified framework for the Hamiltonian-driven ADP is presented, which consists of a learning phase using continuous feedback and an implementation phase using intermittent feedback.

Structure: The remainder of this article is structured as follows. Section II briefly summarizes preliminary results on optimal control problem and the Hamiltonian-driven continuous ADP. Hamiltonian-driven intermittent ADP, including the static and dynamic intermittent cases, are developed in Section III. In Section IV, we provide comparisons in terms of the communication bandwidth, and the performance between the static, and the dynamic intermittent ADP. A simulation example is conducted in Section V to demonstrate the results. Finally, Section VI concludes and talks about future work.

II. Preliminaries and Problem Formulation

A. Optimal Control Problem

Consider the input-affine nonlinear dynamical system

$$\dot{x}(t) = f(x) + g(x)u(t), x(t_0) = x_0,\ t \geq 0 \tag{1}$$

where $x \in \mathbb{R}^n$ is the state, $u \in \mathbb{R}^m$ is the control input, and $f(\cdot) : \mathbb{R}^n \to \mathbb{R}^n$ and $g(\cdot) : \mathbb{R}^n \to \mathbb{R}^{n\times m}$ are nonlinear functions.

Assumption 1: It is assumed that $f(0) = 0$ and $g(0) = 0$ and that the dynamics $f(x)$ and $g(x)$ are locally Lipschitz continuous on a compact set $\Omega \subseteq \mathbb{R}^n$ containing the origin, and the system (1) is stabilizable and zero-state observable on Ω.

It is desired to minimize the following infinite horizon cost functional:

$$J(u(\cdot); x_0) = \int_{t_0}^{\infty} U(x(t), u(t))\mathrm{d}t \tag{2}$$

where $U(x, u) := \|x\|_Q + \|u\|_R$ with $\|x\|_Q = x^T Qx$, $\|u\|_R = u^T Ru$, $Q \succ 0$, and $R \succ 0$. To begin with, the following definition of Hamiltonian is needed, which will play a critical role in our Hamiltonian-driven ADP framework.

Definition 1 [29]: A policy u is defined to be admissible with respect to (2) on Ω, denoted by $u \in \Psi(\Omega)$, if:

1) u is continuous on Ω;
2) $u(0) = 0$;
3) u stabilizes (1) on Ω;
4) for $\forall x_0 \in \Omega$ $V(x_0)$ is finite.

Definition 2: For a given admissible policy u, the Hamiltonian functional of is defined as

$$\begin{aligned}&\mathcal{H}(u; x(t), V(x(t); u))\\ &\quad = U(x(t), u) + \left\langle \frac{\partial V(x(t))}{\partial x}, \dot{x}(t) \right\rangle\end{aligned} \tag{3}$$

where $\langle \cdot, \cdot \rangle$ denotes the inner product between two vectors of appropriate dimensions, $x(t)$ is the continuously differentiable solution to the system (1) with initial state x_0, and $V(\cdot)$ is a continuously differentiable positive definite function.

Note that (3) is slightly different from the Hamiltonian found in [30]–[33]. In our case, x will stand for an arbitrary state of the state space rather than the solution of (1). Both x and $V(x)$ will be viewed as parameters in this article.

Definition 3: Let $x \in \mathbb{R}^n$ be an arbitrary state of the state space and u be an admissible control input. A continuous differentiable and positive definite function $V(x)$ is called a value function if it satisfies, for $\forall x$

$$\mathcal{H}(u; x, V(x; u)) = 0 \tag{4}$$

with the boundary condition $V(0; u) = 0$.

The functional $J(\cdot)$ in (2) is termed as the cost evaluation of the given policy $u(\cdot)$. On the other side, the value function $V(\cdot)$ is termed as the cost evaluation of an arbitrary state in the state space. Under some certain conditions, that will be described later, the value function is equivalent to the cost functional when a given admissible policy u is applied to system (1) from a predetermined state x. The value function (4) is also referred to, as the generalized HJB (GHJB) equation [34].

The cost functional $J(\cdot)$ in (2) depends on the orbital trajectory of the system (1). In contrast, the value function $V(\cdot)$ is independent of the system trajectory because it is derived by solving the nonlinear Lyapunov equation (4) for $\forall x$. By following [28], the value function $V(\cdot)$, obtained by solving (4), is equivalent to the cost functional $J(\cdot)$ for a given admissible policy u.

Based on (3), for the optimal policy u^*, the necessary and sufficient condition for optimality can be written as [28], [35]

$$\begin{aligned}0 &= \mathcal{H}\big(u^*(t); x(t), V^*(x(t))\big)\\ &= \min_{u(t)} \mathcal{H}\big(u(t); x(t), V^*(x(t))\big)\ \forall x\end{aligned} \tag{5}$$

where V^* is the optimal value function satisfying $\mathcal{H}(u^*; x, V^*(x)) = 0$. Assuming that the minimum on the right-hand side exists and is unique, then, one has

$$u^*(x) = -\frac{1}{2}R^{-1}g^T(x)\frac{\partial V^*(x)}{\partial x} \quad \forall x. \tag{6}$$

The HJB equation (5) can be interpreted within the framework of Hamiltonian-driven ADP by following [28]. Given that the Hamiltonian functional is parametrized by the optimal value function $V^*(\cdot)$, then the control policy that attains the minimum of the Hamiltonian is termed as the optimal control $u^*(\cdot)$ and is unique. This is illustrated in Fig. 1.

The existence and uniqueness of the optimal value function are guaranteed in [36] under stabilizability and observability

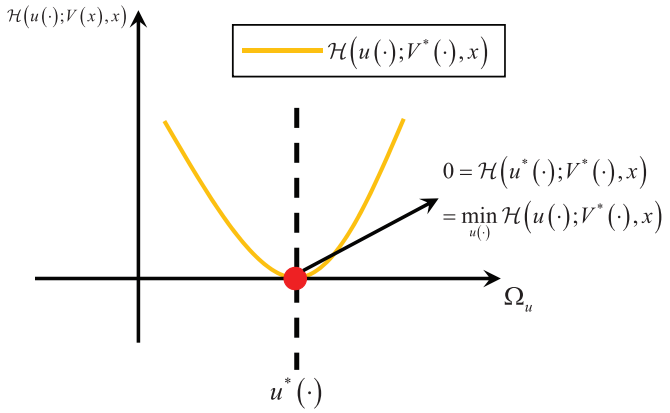


Fig. 1. Hamiltonian-based interpretation of the HJB equation (5). The solid curve represents the Hamiltonian functional parameterized by the optimal value function $V^*(\cdot)$. Once can observe that: 1) the minimum of the Hamiltonian functional is 0 and 2) the control policy which attains the minimum of the Hamiltonian is $u^*(\cdot)$, which is the solution to the HJB equation (5) and solves the continuous-time optimal control problem.

assumptions and the quadratic form of the performance (2). However, solving the HJB equation (5) is a challenging problem since it is a nonlinear partial differential equation and does not have an analytical solution. In the following, we shall use a Hamiltonian-driven ADP framework to approximate such a solution.

B. Hamiltonian-Driven Continuous ADP

In [28], Hamiltonian-driven ADP framework, which contains three fundamental steps for solving the optimal control problem with respect to the performance (2) given the system (1), is presented as follows.

1) *Policy Evaluation:* To build a criterion that evaluates an arbitrary admissible control $u(\cdot)$, i.e., calculate the corresponding cost $J(u(\cdot))$.
2) *Policy Comparison:* To establish a rule that compares two different admissible policies $u(\cdot)$ and $v(\cdot)$.
3) *Policy Improvement:* Based on the current admissible control $u_k(\cdot)$, $k \in \mathbb{Z}$, design a successive control $u_{k+1}(\cdot)$ with an improved cost $J(u_{k+1}(\cdot))$.[1]

In this article, the Hamiltonian-driven ADP framework using continuous-time feedback is named as the Hamiltonian-driven continuous ADP.

In the Hamiltonian $\mathcal{H}(u; x, V)$, the argument is the policy u and the value function V is the parameter. To investigate the property of the Hamiltonian $\mathcal{H}(u; x, V)$ parameterized by a fixed value function V, by completing the squares, the minimum of the Hamiltonian functional can be expressed as

$$\begin{aligned} h(V) &= \min_u \mathcal{H}(u; x, V) \\ &= \left[\frac{\partial V(x)}{\partial x}\right]^T f(x) - \frac{1}{4}\left[\frac{\partial V}{\partial x}\right]^T g(x) R^{-1} g^T(x) \frac{\partial V}{\partial x} \\ &\quad + x^T Q x \end{aligned} \tag{7}$$

[1]Since we are considering a minimization problem, "improved" is achieved given that the cost is monotonically decreasing, i.e., $J(u_{k+1}(\cdot)) < J(u_k(\cdot))$.

with

$$\bar{u}(x; V) = \arg\min_u \mathcal{H}(u; x, V) = -\frac{1}{2} R^{-1} g^T \frac{\partial V(x)}{\partial x}. \tag{8}$$

Note that $\bar{u}(\cdot)$ is different from $u^*(\cdot)$, since it is not parametrized by the optimal value function $V^*(x)$. The distance between $u(\cdot)$, and $\bar{u}(\cdot)$ is expressed as

$$d(\bar{u}, u) = \|\bar{u}(x; V) - u(x)\|_R \quad \forall x. \tag{9}$$

The Hamiltonian-driven ADP framework is shown in Fig. 2. The policy evaluation step of the Hamiltonian-driven framework is shown in Fig. 2(a), where the solid and dashed lines correspond to the Hamiltonian parameterized by the value functions $V(\cdot)$ and $V'(\cdot)$, respectively. Based on the GHJB equation (4), the value function $V(\cdot)$ is the one makes the Hamiltonian $\mathcal{H}(\cdot; V, x)$ to be zero given the admissible $u(\cdot)$.

The policy comparison step is illustrated in Fig. 2(b). In Fig. 2(b), the Hamiltonian functionals parameterized by the value function $V_1(\cdot)$ and $V_2(\cdot)$ correspond to the dashed and solid line, respectively. The distance between the prescribed control policy, $u_i(\cdot)$, and the policy that attains the minimum of the corresponding Hamiltonian, $\bar{u}_i(\cdot) = \arg\min_u \mathcal{H}(u; x, V_i)$, is denoted as d_i, for $i = 1, 2$. The minimum of the Hamiltonian functional parameterized by the value function $V_i(\cdot)$ is denoted as h_i, for $i = 1, 2$. In [28], it is shown that h_i and d_i indicate the comparison between V_i: 1) $d_2 < d_1 \Rightarrow V_2(x) < V_1(x)$ and 2) $h_2 > h_1 \Rightarrow V_2(x) < V_1(x)$, for $\forall x$.

The successive process of policy improvement in the Hamiltonian-driven framework is shown in Fig. 2(c), which can be viewed as a special case of policy comparison. In Fig. 2(c), $V_i(\cdot)$ is the value function corresponding to u_i, i.e., $H(u_i; x, V_i(\cdot)) = 0$. u_{i+1} is obtained by minimizing $H(u_i; x, V_i(\cdot))$, i.e., $u_{i+1} = \arg\min_u H(u; x, V_i(\cdot))$. This iterative process is guaranteed to yield a policy sequence $\{u_i\}_{i+1}^{\infty}$, which would approach to the optimal policy $u^* = \arg\min_u H(u; x, V^*(\cdot))$ [28].

Remark 1: Compared to the policy comparison step, the policy improvement step provides an explicit method to obtain the policy $u_{i+1}(\cdot)$ with an improved performance compared to the $u_i(\cdot)$. Then, the policy iteration algorithm can be formulated as a successive minimization of the iterative Hamiltonian $H(u_i(\cdot); V_i(\cdot), x)$ [28].

Corollary 1: Suppose that the sequence of control policies $\{u_i(\cdot)\}_{i=0}^{\infty}$ and the value functions $\{V_i(\cdot)\}_{i=0}^{\infty}$ are generated by the Hamiltonian-driven continuous ADP. Then, for $i \in \mathbb{Z}^+$

$$h(V_i) + d(u_i, u_{i+1}) = 0. \tag{10}$$

Proof: First, based on (8) and [28, Th. 1], one has

$$u_{i+1}(x) = \bar{u}(x; V_i). \tag{11}$$

From (7) and (9), we can write

$$\begin{aligned} h(V_i) + d(u_i, u_{i+1}) &= \left[\frac{\partial V_i(x)}{\partial x}\right]^T f(x) + \left[\frac{\partial V_i(x)}{\partial x}\right]^T g(x) u_i(x) \\ &\quad + x^T Q x + u_i^T(x) R u_i(x) = 0. \end{aligned}$$

This completes the proof. ■

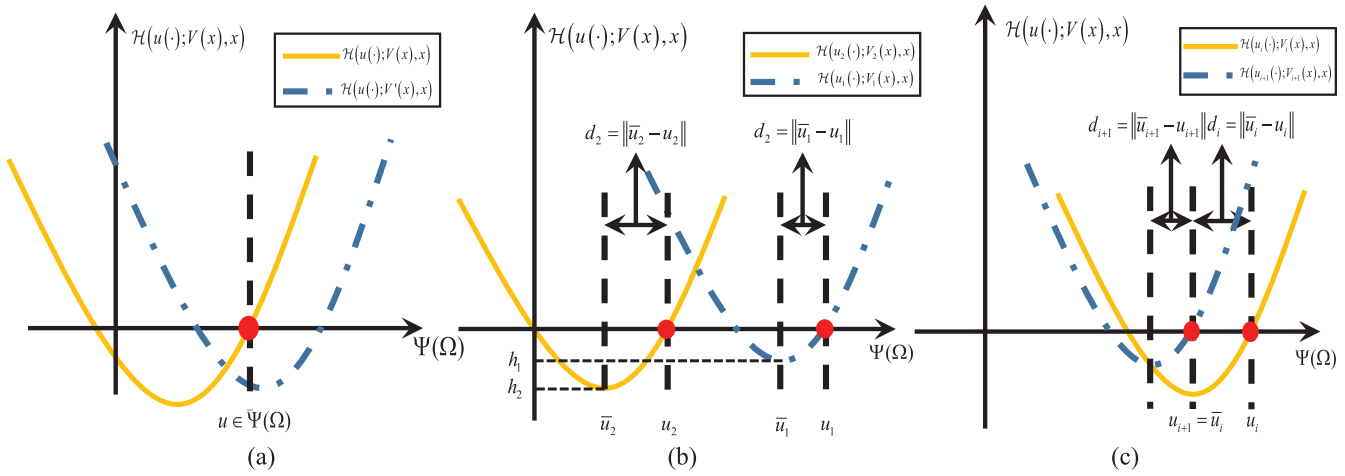


Fig. 2. Hamiltonian-driven ADP framework. (a) Policy evaluation step. (b) Policy comparison step. (c) Policy improvement step (successive minimization of Hamiltonian).

III. Hamiltonian-Driven Intermittent ADP

In the Hamiltonian-driven continuous ADP, the iterative policy has to be calculated continuously for implementation since the continuously varying signal $x(t)$ is fed into the policy $u_i(\cdot)$. In this section, we develop an intermittent ADP design to reduce the computation burden of the policy update.

Consider the aperiodic state sampling, implemented through a data-sampled component, as follows:

$$\hat{x}(t) = x(t_k) \ \forall t \in [t_k, t_{k+1}) \tag{12}$$

with the gap between x and $\hat{x}$ denoted as

$$e(t) = \hat{x}(t) - x(t) \ \forall t \in [t_k, t_{k+1}). \tag{13}$$

Given the sequence of the iterative policy $\{u_i(\cdot)\}_{i=0}^{\infty}$ and its corresponding value function sequence $\{V_i(\cdot)\}_{i=0}^{\infty}$ obtained by the Hamiltonian-driven continuous ADP, the iterative intermittent ADP policy sequence $\{u_i^e(\cdot)\}_{i=0}^{\infty}$ can be determined as

$$\begin{cases} u_0^e(x) = u_0(\hat{x}) \\ u_{i+1}^e(x) = u_{i+1}(\hat{x}) = -\frac{1}{2}R^{-1}g^T(\hat{x})\frac{\partial V_i(\hat{x})}{\partial \hat{x}}, i \in \mathbb{Z}^+. \end{cases} \tag{14}$$

Assumption 2: There exists $L \in \mathbb{R}^+$ such that

$$\left\| u_i^e(x(t)) - u_i(x(t)) \right\| \le L\|e(t)\| \ \forall i \in \mathbb{Z}^+.$$

Assumption 2 is common in event-triggered control, [1], [2], [7], which can be satisfied when the controller is affine with respect to the sampling error. Moreover, it is sufficient to the existence of a solution and is required for the triggering condition design and stability discussion [1]. As shown in this article, the Lipschitz constant of the control mapping affects the event-sampling frequency through the threshold design.

Corollary 2 [1]: Suppose that Assumptions 1 and 2 holds. Then, applying the intermittent control (14) to the system (1) yields the Lipschitz continuity of $f(x) + g(x)u_i^e(x)$ on Ω, i.e., there exists a positive constant A_e such that

$$\left\| f(x) + g(x)u_i^e(x) \right\| \le A_e\|x\| + A_e\|e\|.$$

The following lemma provides a Hamiltonian-based relationship between $u_i(x)$ and $u_i(\hat{x})$.

Lemma 1: Suppose that Assumption 2 holds. Assume that the policy sequence $\{u_i(\cdot)\}_{i=0}^{\infty}$ and the value function sequence $\{V_i(\cdot)\}_{i=0}^{\infty}$ are generated by the Hamiltonian-driven continuous ADP. Then, the iterative intermittent feedback policy satisfies

$$\mathcal{H}(u_i^e, V_i, x) - \mathcal{H}(u_i, V_i, x) \le d(u_{i+1}, u_i) + L\bar{\lambda}(R)\|e\|.$$

Proof: Considering the Hamiltonian of the iterative continuous feedback $u_i(\cdot)$, and using lemma [28, Lemma 2], one can obtain

$$\begin{aligned} 0 &= \mathcal{H}(u_i, V_i, x) \\ &= \left[\frac{\partial V_i(x)}{\partial x}\right]^T \left[f(x) + g(x)u_i\right] + x^TQx + u_i^TRu_i. \end{aligned} \tag{15}$$

From (7), and [28, Th. 1], one has

$$\begin{aligned} h(V_i) &= \min_u \mathcal{H}(u; V_i, x) = \mathcal{H}(u_{i+1}, V_i, x) \\ &= \left[\frac{\partial V_i(x)}{\partial x}\right]^T f(x) - u_{i+1}^TRu_{i+1} + x^TQx \le 0. \end{aligned} \tag{16}$$

The Hamiltonian of the intermittent policy $u_i^e(\cdot)$ can be parametrized by $V_i(\cdot)$ to write

$$\begin{aligned} \mathcal{H}(u_i^e(\cdot); V_i(\cdot), x) &= \left[\frac{\partial V_i(x)}{\partial x}\right]^T f(x) \\ &\quad + \left[\frac{\partial V_i(x)}{\partial x}\right]^T g(\hat{x})\hat{u}_i + x^TQx + \hat{u}_i^TR\hat{u}_i. \end{aligned} \tag{17}$$

Based on (16), one has

$$\begin{aligned} \mathcal{H}(\hat{u}_i, V_i, x) &= h(V_i) + (u_{i+1} - \hat{u}_i)^TR(u_{i+1} - \hat{u}_i) \\ &\le \left\| u_{i+1}(x) - u_i(x) + u_i(x) - u_i(\hat{x}) \right\|_R \\ &\le \|u_{i+1}(x) - u_i(x)\|_R + \left\| u_i(x) - u_i(\hat{x}) \right\|_R \\ &= d(u_{i+1}, u_i) + \bar{\lambda}(R)L\|e\|. \end{aligned} \tag{18}$$

This completes the proof. ■

In the Hamiltonian-driven intermittent ADP design, the sampled state $\hat{x}(t)$ is plugged into the iterative policy $u_i(\cdot)$,

obtained by the Hamiltonian-driven continuous ADP. We will now present two different types of intermittent ADP designs.

A. Static Intermittent ADP

The following theorem provides the static intermittent ADP design.

Theorem 1: Suppose that the policy and value function sequences $\{u_i(\cdot)\}_{i=0}^{\infty}$ and $\{V_i(\cdot)\}_{i=0}^{\infty}$ are generated by the Hamiltonian-driven continuous ADP. Suppose also that the static iterative intermittent ADP policy $u_i^e(\cdot)$ is determined as in (14) with the triggering instant determined by the condition

$$\|e\|^2 \leq \frac{\left(1-\sigma^2\right)\underline{\lambda}(Q)\|x\|^2 + \bar{\lambda}(R)\left\|u_i\left(\hat{x}\right)\right\|^2}{\bar{\lambda}(R)L^2} \tag{19}$$

where $\sigma \in (\max\{0, 1-([\bar{\lambda}(R)L^2]/[\underline{\lambda}(Q)])\}, 1)$. Then:

1) the origin of the closed-loop system is asymptotically stable;
2) the static intermittent ADP is Zeno-free in the sense that the interevent interval $\tau_s = \min_k(t_{k+1}-t_k)$ determined by (19) is lower-bounded by

$$\underline{\tau}_s = \int_0^{L_x} \frac{1}{A_e + 2A_e s + A_e s^2}\mathrm{d}s \tag{20}$$

with $L_x = \sqrt{(b/a)}$, $a = \bar{\lambda}_R L^2$, $b = \underline{\lambda}_Q(1-\sigma^2)$.

Proof:

1) We shall start by using the iterative value function $V_i(\cdot)$ as a Lyapunov candidate. Applying now the intermittent control to (1) yields

$$\dot{V}_i = \left[\frac{\partial V_i(x)}{\partial x}\right]^T \left[f(x) + g(x)u_i\left(\hat{x}\right)\right]. \tag{21}$$

Based on (16), the following holds:

$$\left[\frac{\partial V_i(x)}{\partial x}\right]^T f(x) = h(V_i) - Q(x) - u_i^{\mathrm{T}}\left(\hat{x}\right)Ru_i\left(\hat{x}\right). \tag{22}$$

Inserting now (22) into (21), yields

$$\dot{V}_i = h(V_i) - Q(x) - \left\|u_i\left(\hat{x}\right)\right\|_R + d\left(u_{i+1}(x), u_i\left(\hat{x}\right)\right).$$

Based on (18), the following holds:

$$d\left(u_i\left(\hat{x}\right), u_{i+1}(x)\right) \leq d(u_i(x), u_{i+1}(x)) + L\bar{\lambda}(R)\|e\|.$$

Then, $\dot{V}_i$ satisfies

$$\begin{aligned}\dot{V}_i \leq\ & h(V_i) + d(u_i, u_{i+1}) \\ & - Q(x) - u_i^{\mathrm{T}}\left(\hat{x}\right)Ru_i\left(\hat{x}\right) + L\bar{\lambda}(R)\|e\| \\ =\ & - Q(x) - u_i^{\mathrm{T}}\left(\hat{x}\right)Ru_i\left(\hat{x}\right) + L\bar{\lambda}(R)\|e\|\end{aligned} \tag{23}$$

where the equality results from Corollary 1. By adding and subtracting $\sigma^2\underline{\lambda}(Q)\|x\|^2$ to (23), one obtains

$$\begin{aligned}\dot{V}_i \leq\ & \bar{\lambda}(R)L^2\|e\|^2 - \underline{\lambda}(Q)\left(1-\sigma^2\right)\|x\|^2 - \bar{\lambda}(R)\left\|u_i\left(\hat{x}\right)\right\|^2 \\ & - \sigma^2\underline{\lambda}(Q)\|x\|^2.\end{aligned} \tag{24}$$

Therefore, $\dot{V}_i \leq -\underline{\lambda}(Q)\sigma^2\|x\|^2 < 0$ can be guaranteed by the triggering condition (19).

2) At the triggering instant t_k, one has

$$\|e\|^2 \geq \frac{b\|x\|^2 + \bar{\lambda}(R)\left\|u_i\left(\hat{x}\right)\right\|^2}{a} \geq \frac{b\|x\|^2}{a} = L_x^2\|x\|^2. \tag{25}$$

Therefore, in an interval between two consecutive triggering instants, $y := (\|e\|/\|x\|)$ evolves from 0 to L_x. According to [1] and Corollary 2, one has

$$\dot{y} \leq A_e + 2A_e y + A_e y^2, \quad y(t_0) = y_0. \tag{26}$$

Denote $\phi(t, \phi_0)$ as the solution of the differential equation

$$\dot{\phi} = A_e + 2A_e\phi + A_e\phi^2, \quad \phi(t_0) = y_0. \tag{27}$$

Then, we can conclude that $y(t) \leq \phi(t, \phi_0)$ by using the comparison principle [37]. Therefore, the interevent interval is lower bounded. This completes the proof. ■

Note that the triggering condition (19) is equivalent to $p \geq 0$ with

$$p = b\|x\|^2 - a\|e\|^2 + \bar{\lambda}(R)\left\|u_i\left(\hat{x}\right)\right\|^2. \tag{28}$$

The sequence of the triggering instants that is determined by the condition (19) can be rewritten as

$$\begin{aligned}t_0 &= 0 \\ t_{k+1} &= \inf_{t\in\mathbb{R}^+}\{t > t_k \wedge p \leq 0\}.\end{aligned} \tag{29}$$

That is, the inequality $p \geq 0$ has to be satisfied $\forall t$.

B. Dynamic Intermittent ADP

In order to further reduce the communication load, the requirement that $p \geq 0$ has to be satisfied $\forall t$ is relaxed. We will now introduce a dynamic intermittent ADP. To begin with, an internal dynamical system is required

$$\dot{\eta} = -\mu\eta + p, \ \eta(t_0) = \eta_0 \tag{30}$$

where p is given by (28) and $\eta_0, \mu \in \mathbb{R}^+$.

We are now ready to present the dynamic intermittent ADP, i.e., an event is triggered when the following condition is satisfied:

$$\eta(t) + \theta p(t) \leq 0 \tag{31}$$

where $\theta \in \mathbb{R}^+$ is a parameter to be designed later.

The triggering instants sequence can be determined by (31) as

$$\begin{aligned}t_0 &= 0 \\ t_{k+1} &= \inf_{t\in\mathbb{R}^+}\{(t > t_k) \wedge (\eta(t) + \theta p(t) \leq 0)\}.\end{aligned} \tag{32}$$

Then, $p \geq 0$ in static intermittent ADP to is relaxed as $\eta(t) + \theta p(t) \geq 0$ in dynamic intermittent ADP. Consequently, the dynamic intermittent policy takes the form of (14) with the triggering instants determined by (32).[2]

Lemma 2 [7]: Let $\mu, \eta_0, \theta \in \mathbb{R}_0^+$, and p defined as in (28). Then the following conditions hold:

1) $\eta(t) + \theta p(t) \geq 0\ \forall t \in \mathbb{R}^+$;
2) $\eta \geq 0\ \forall t \in \mathbb{R}^+$.

[2]Note that the dynamic intermittent policy and the static intermittent policy share the same form, but they are different because of the event-triggering conditions (28) and (31), respectively.

The following theorem guarantees the stability of the equilibrium point of the closed-loop system with a dynamic intermittent mechanism given by (30) and (31).

Theorem 2: Suppose that the sequence of continuous feedback policies $\{u_i(\cdot)\}_{i=0}^{\infty}$ and value functions $\{V_i(\cdot)\}_{i=0}^{\infty}$ are generated by Hamiltonian-driven continuous ADP. Let σ be selected as in Theorem 1 and $\mu, \eta_0, \theta \in \mathbb{R}^+$. The iterative intermittent feedback policy sequence $\{u_i^e(\cdot)\}_{i=0}^{\infty}$ is determined as in (14) with the triggering instant determined by (32). Then:

1) the origin of the closed-loop system is asymptotically stable;
2) the dynamic intermittent ADP is Zeno-free. Moreover, let $q = 2A_e - \mu$, $\mu \in (0, 2A_e)$, $\theta \in ((1/2q), (1/q)]$, the interevent interval $\underline{\tau}_d = \min_k(t_{k+1} - t_k)$ determined implicitly by (32) is lower-bounded by a positive constant $\underline{\tau}_d$, which is given by

$$\underline{\tau}_d = \int_0^1 \frac{1}{A_e\sqrt{\frac{a}{b}} + \left(A_e + \frac{\mu}{2}\right)s + A_e\sqrt{\frac{b}{a}}s^2 + \frac{1}{2\theta}s^3}\mathrm{d}s. \quad (33)$$

Proof:

1) For the augmented system of (16) and (30), consider the candidate Lyapunov function $W_i : \mathbb{R}^n \times \mathbb{R}_0^+ \to \mathbb{R}_0^+$ defined as $W_i(x, \eta) = V_i(x) + \eta$, which is a positive definite and radially unbounded scalar-valued function. Based on Theorem 1, for $\forall t \in \mathbb{R}_0^+$, the orbital derivative of $W_i(x, \eta)$ yields

$$\begin{aligned}\dot{W}_i(x,\eta) &= \dot{V}_i(x) + \dot{\eta} \le \left(-\underline{\lambda}_Q\sigma^2\|x\|^2 - p\right) + (-\mu\eta + p)\\ &= -\underline{\lambda}_Q\sigma^2\|x\|^2 - \mu\eta. \end{aligned} \quad (34)$$

According to Lemma 2 and Theorem 1, one can conclude that $\dot{W}_i(x, \eta) < 0$. Therefore, $W_i(x, \eta)$ decreases, and both $x(t)$ and $\eta(t)$ converge to the origin asymptotically.

2) From Theorem 1, $(\|e\|/\|x\|)$ evolves from 0 to L_x during the interval $[t_k, t_{k+1})$. Equivalently, the term $[(\sqrt{a}\|e\|)/(\sqrt{b}\|x\|)]$ evolves from 0 to 1. For the dynamic intermittent mechanism, at the triggering instant, according to (31), one has

$$\eta(t) + \theta\left[b\|x(t)\|^2 + \bar{\lambda}_R\|u_d(t)\|^2 - a\|e(t)\|^2\right] \le 0 \quad (35)$$

which can further yield

$$\begin{aligned}a\theta\|e(t)\|^2 &\ge \eta(t) + b\theta\|x(t)\|^2 + \theta\bar{\lambda}_R\|u_d(t)\|^2\\ &\ge \eta(t) + b\theta\|x(t)\|^2. \end{aligned} \quad (36)$$

Therefore, when $\theta > 0$, in the interval of $[t_k, t_{k+1})$, $\xi(t) := [(\sqrt{a\theta}\|e(t)\|)/(\sqrt{\eta(t) + b\theta\|x(t)\|^2})]$, evolves from 0 to 1. The interevent interval is investigated, by examining the dynamics of $\xi(t)$ for $t \in [t_k, t_{k+1})$

$$\begin{aligned}\dot{\xi} = &-\frac{\sqrt{a\theta}\|e\|}{2\left(\eta + b\theta\|x\|^2\right)^{\frac{3}{2}}}\left(\dot{\eta} + 2b\theta x^T\dot{x}\right)\\ &+ \frac{\sqrt{a\theta}e^T\dot{e}}{\sqrt{\eta + b\theta}\cdot\|e\|\cdot\|x\|}\end{aligned}$$

with $\xi(t_k) = 0$. Considering Corollary 2 and the facts

$$\begin{aligned}\dot{e} &= -\dot{x}\\ \dot{\eta} &= -\mu\eta + b\|x\|^2 + \bar{\lambda}_R\left\|u_i(\hat{x})\right\|^2 - a\|e\|^2\\ &\ge -\mu\eta + b\|x\|^2 - a\|e\|^2 \end{aligned} \quad (37)$$

then ξ satisfies (38), as shown at the bottom of this page. From $\theta \in ((1/2q), (1/q)]$ and $q = 2A_e - \mu$, one obtains $\theta \in (0, [1/(2A_e - \mu)]]$. Assume that $\psi(t, \psi_0)$ is the solution of the following differential equation:

$$\dot{\psi} = A_e\sqrt{\frac{a}{b}} + \left(A_e + \frac{\mu}{2}\right)\psi + A_e\sqrt{\frac{b}{a}}\psi^2 + \frac{1}{2\theta}\psi^3, \psi_0 = \xi_0.$$

Using the comparison principle [37] and (38), then $\xi(t)$ satisfies $\xi(t) \le \psi(t, \psi_0)$. Moreover, the time needed by $\xi(t)$ to evolve from 0 to 1 is lower bounded by the positive constant $\underline{\tau}_d$ in Theorem 2. Therefore, the intermittent condition (31) is Zeno-free. ■

IV. Hamiltonian-Driven Unified Framework for Hybrid ADP

As mentioned in the introduction, there is a tradeoff between the bandwidth and the performance for the intermittent feedback designs. In this section, we quantify this tradeoff to show that the static intermittent ADP has better performance, whereas the dynamic intermittent ADP has a more efficient bandwidth.

A. Bandwidth Discussion

In this section, the static and dynamic intermittent ADP with respect to the bandwidth size is compared. In the following, let $\{t_j^s\}_{j=1}^{\infty}$ and $\{t_k^d\}_{k=1}^{\infty}$ be the triggering instants determined by Theorems 1 and 2, respectively. Denote the system trajectories

$$\begin{aligned}\dot{\xi} &\le \frac{\sqrt{a\theta}A_e}{\sqrt{\eta + b\theta\|x\|^2}}(\|x\| + \|e\|) + \frac{\sqrt{a\theta}\|e\|}{2\left(\eta + b\theta\|x\|^2\right)^{\frac{3}{2}}}\left(\mu\eta - b\|x\|^2 + a\|e\|^2 + 2b\theta A_e\|x\|^2 + 2b\theta A_e\|x\|\|e\|\right)\\ &\le A_e\sqrt{\frac{a}{b}} + A_e\xi + A_e\sqrt{\frac{b}{a}}\xi^2 + \frac{1}{2\theta}\xi^3 + \frac{\sqrt{a\theta}\|e\|}{2\left(\eta + b\theta\|x\|^2\right)^{\frac{3}{2}}}\left(\mu\eta - b\|x\|^2 + 2b\theta A_e\|x\|^2\right)\\ &\le A_e\sqrt{\frac{a}{b}} + A_e\xi + A_e\sqrt{\frac{b}{a}}\xi^2 + \frac{1}{2\theta}\xi^3 + \frac{\mu}{2}\xi + \frac{b\theta\|x\|^2}{2\left(\eta + b\theta\|x\|^2\right)}\left(-\mu - \frac{1}{\theta} + 2A_e\right)\xi\\ &\le A_e\sqrt{\frac{a}{b}} + \left(A_e + \frac{\mu}{2}\right)\xi + A_e\sqrt{\frac{b}{a}}\xi^2 + \frac{1}{2\theta}\xi^3 \end{aligned} \quad (38)$$

as $x_s(t)$ and $x_d(t)$, respectively. Then, the sampled states using the static and dynamic triggering condition can be expressed as $\hat{x}_s(t) = x(t_j^s)$ for $t \in [t_j^s, t_{j+1}^s)$ and $\hat{x}_d(t) = x(t_k^d)$ for $t \in [t_k^d, t_{k+1}^d)$, respectively.

From Theorems 1 and 2, the local stability can be guaranteed by the static and dynamic triggering condition. Therefore, $x_s(t)$ and $x_d(t)$ are remain bounded on compact sets. Also, there exists Lipschitz constant such that

$$\begin{aligned}\|e_d\|^2 - \|e_d + \Delta - \Delta_0\|^2 &\le \rho_1\|\Delta - \Delta_0\| \\ \|x_d - \Delta\|^2 - \|x_d\|^2 &\le \rho_2\|\Delta\|. \end{aligned} \tag{39}$$

In addition, the control input is also bounded for both static and dynamic intermittent feedback, i.e., $\|u_i(\hat{x}_s)\| \le \bar{u}$, $\|u_i(\hat{x}_s)\| \le \bar{u}$ for $\bar{u} > 0$. Then, one has

$$\begin{aligned}\big\|u_i\big(\hat{x}_s\big)\big\| - \big\|u_i\big(\hat{x}_d\big)\big\| &\le \big|\big\|u_i\big(\hat{x}_s\big)\big\| - \big\|u_i\big(\hat{x}_d\big)\big\|\big| \\ &\le \big\|u_i\big(\hat{x}_s\big) - u_i\big(\hat{x}_d\big)\big\| \le L\big\|\hat{x}_s - \hat{x}_d\big\|. \end{aligned} \tag{40}$$

Theorem 3: Let the lower bound of the interevent interval for static and dynamic intermittent ADP be $\underline{\tau}_s$ and $\underline{\tau}_d$, respectively. Suppose that $t_j^s \equiv t_k^d$ and write the next triggering instants using the static and the dynamic intermittent mechanisms as t_{j+1}^s and t_{k+1}^d, respectively. Then:

1) $\underline{\tau}_s < \underline{\tau}_d$;
2) if $x_s(t_j^s) = x_d(t_k^d)$, then, $t_{k+1}^d \ge t_{j+1}^s$;
3) there exists $\delta > 0$ such that if $\|x_s(t_j^s) - x_d(t_k^d)\| < \delta$, then $t_{k+1}^d \ge t_{j+1}^s$.

Proof:

1) Based on Theorem 1, by letting $s := (\sqrt{a}\|e\|/\sqrt{b}\|x\|)$, s evolves from 0 to 1 between two successive events for the static intermittent ADP. Then, $\underline{\tau}_s$ in (20) can be equivalently expressed as

$$\underline{\tau}_s = \int_0^1 \frac{1}{A_e\sqrt{\frac{a}{b}} + 2A_e s + A_e\sqrt{\frac{a}{b}}s^2}\mathrm{d}s. \tag{41}$$

It remains now to compare the denominators of $\underline{\tau}_s$ and $\underline{\tau}_d$ for $s \in (0, 1)$. From Theorems 1 and 2, the denominators of $\underline{\tau}_s$ and $\underline{\tau}_d$ can be written as

$$\begin{aligned} D_s &= \int_0^1 \left[A_e\sqrt{\frac{a}{b}} + 2A_e s + A_e\sqrt{\frac{a}{b}}s^2\right]\mathrm{d}s \\ D_d &= \int_0^1 \left[A_e\sqrt{\frac{a}{b}} + \Big(A_e + \frac{\mu}{2}\Big)s + A_e\sqrt{\frac{b}{a}}s^2 + \frac{1}{2\theta}s^3\right]\mathrm{d}s. \end{aligned}$$

Note that we have $\sigma \in (\max\{0, 1 - [(\bar{\lambda}_R L^2)/(\underline{\lambda}_Q)]\}, 1)$. From (42), after subtracting D_s from D_d, one has

$$\begin{aligned} D_d - D_s &= \Big(\frac{\mu}{2} - A_e\Big)\int_0^1 s\mathrm{d}s + \frac{1}{2\theta}\int_0^1 s^3\mathrm{d}s \\ &= \frac{1}{4\theta}\left(\frac{1}{2} - \theta q\right). \end{aligned} \tag{42}$$

Consider the parameters design in Theorems 1 and 2 that satisfy $\mu \in (0, 2A_e), \theta \in ((1/2q), (1/q)], q = 2A_e - \mu$. Then, $D_d < D_s$. Therefore, $\underline{\tau}_s - \underline{\tau}_d = [(D_d - D_s)/(D_s D_d)]$, where both D_s and D_d are positive. One finally obtains $\underline{\tau}_s < \underline{\tau}_d$.

2) This will be shown by contradiction. Assume that $t_{k+1}^d < t_{j+1}^s$. Then, based on (29), one has that $p(t_{k+1}^d) > 0$. Based on (32) and Lemma 2, one has $\eta(t_{k+1}^d) + \theta p(t_{k+1}^d) \le 0$, i.e.,

$$0 \ge \eta\Big(t_{k+1}^d\Big) + \theta p\Big(t_{k+1}^d\Big) \ge \theta p\Big(t_{k+1}^d\Big).$$

Note that $\theta \in ((1/2q), (1/q)]$ is a positive constant, therefore, the above equation yields $p(t_{k+1}^d) \le 0$, which contradicts the fact that $p(t_{k+1}^d) > 0$. Therefore, $t_{k+1}^d < t_{j+1}^s$.

3) First, the static and dynamic triggering condition yields the sampling error $e_s(t) = x_s(t) - x_s(t_j^s)$ and $e_d(t) = x_d(t) - x_d(t_k^d)$. Define $\Delta(t) = x_d(t) - x_s(t)$ and $\Delta_0 = x_d(t_k^d) - x_s(t_j^s)$. Then, the dynamics of $\Delta(t)$ is

$$\dot{\Delta} = \big[f(x_d) + g(x_d)u_i\big(\hat{x}_d\big)\big] \tag{43}$$

$$- \big[f(x_d - \Delta) + g(x_d - \Delta)u_i(x_d + e_d - \Delta_0)\big]. \tag{44}$$

From Corollary 2, one has $\|\dot{\Delta}\| \le A_e\|\Delta\| + A_e\|\Delta_0\|$. Therefore, using the comparison lemma yields

$$\|\Delta(t)\| \le \|\Delta_0\|\left[2e^{A_e\left(t - t_j^s\right)} - 1\right]. \tag{45}$$

The static triggering condition guarantees

$$\|e_s\|^2 + \varepsilon_1 \le \frac{b}{a}\|x_s\|^2 + \frac{1}{L^2}\big\|u_i\big(\hat{x}_s\big)\big\|^2, t \in \Big[t_j^s, t_{j+1}^s\Big) \tag{46}$$

for some $\varepsilon_1 > 0$. Using the definitions of $\Delta(t)$ and Δ_0, (46) can be rewritten as

$$\|e_d + \Delta_0 - \Delta\|^2 + \varepsilon_1 \le \frac{b}{a}\|x_d - \Delta\|^2 + \frac{1}{L^2}\big\|u_i\big(\hat{x}_s\big)\big\|^2. \tag{47}$$

Considering the facts in (39), one has

$$\begin{aligned}\|e_d\|^2 - \rho_1\|\Delta - \Delta_0\| &\le \|e_d + \Delta - \Delta_0\|^2 \\ \|x_d - \Delta\|^2 &\le \rho_2\|\Delta\| + \|x_d\|^2. \end{aligned}$$

Then, (47) yields

$$\begin{aligned}\|e_d\|^2 + \varepsilon_1 &\le \frac{b}{a}\Big(\rho_2\|\Delta\| + \|x_d\|^2\Big) + \frac{1}{L^2}\big\|u_i\big(\hat{x}_s\big)\big\|^2 \\ &\quad + \rho_1\|\Delta - \Delta_0\| \\ &\le C_1\|\Delta_0\| + \frac{b}{a}\|x_d\|^2 + \frac{1}{L^2}\big\|u_i\big(\hat{x}_s\big)\big\|^2 \end{aligned} \tag{48}$$

where $C_1 = ((b/a)\rho_2 + \rho_1)(2e^{A_e(t - t_j^s)} - 1) + \rho_1$. From (40), one has

$$-2L\bar{u}\|\Delta_0\| \le \big\|u_i\big(\hat{x}_s\big)\big\|^2 - \big\|u_i\big(\hat{x}_d\big)\big\|^2. \tag{49}$$

Therefore, for $\|\Delta_0\| < [(\varepsilon_1)/(2L\bar{u} + C_1)] := \delta$, there exists $\varepsilon_2 = \varepsilon_1 - (2L\bar{u} + C_1)\|\Delta_0\| > 0$ such that

$$\begin{aligned}\|e_d\|^2 + \varepsilon_2 &\le \frac{b}{a}\|x_d\|^2 + \frac{1}{L^2}\big\|u_i\big(\hat{x}_d\big)\big\|^2 \\ &< \frac{b}{a}\|x_d\|^2 + \frac{1}{L^2}\big\|u_i\big(\hat{x}_d\big)\big\|^2 + \frac{\eta(t)}{\theta}. \end{aligned} \tag{50}$$

Therefore, $t_{j+1}^s - t_j^s \le t_{k+1}^d - t_k^d$. This completes the proof. ■

Remark 2: By proposition 3) of Theorem 3, it is shown that the state trajectories generated by the static and dynamic intermittent feedback approach a small neighborhood of each other in finite time and stay close to each other from then on. In this scenario, the interevent interval of the dynamic intermittent feedback is no smaller than the static case. This fact also contributes to the control performance discussions in the next section.

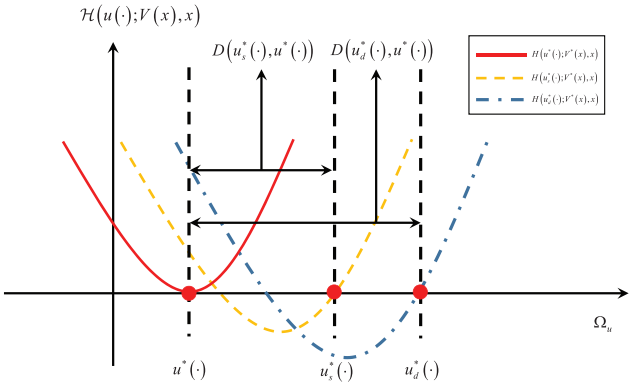


Fig. 3. Policy comparison of $u_d^*(\cdot)$ and $u_s^*(\cdot)$ within the Hamiltonian-driven framework in terms of performance. Closer distance between the static intermittent policy $u_s^*(\cdot)$ and the time-triggered policy $u^*(\cdot)$ corresponds to higher control input update frequency and superior performance, whereas further distance between the static intermittent policy $u_d^*(\cdot)$ and the time-triggered policy $u^*(\cdot)$ corresponds to lower control input update frequency and inferior performance.

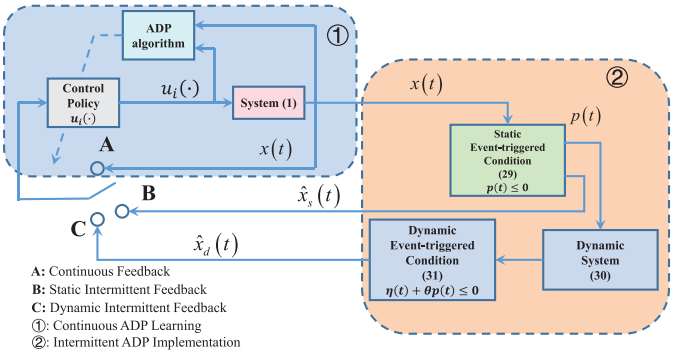


Fig. 4. Hamiltonian-driven unified framework for hybrid ADP, with $\hat{x}_s$, $\hat{x}_d$ denoting the state sampling using triggering conditions of the static and dynamic intermittent ADP, respectively, and $u_i(\cdot)$ denoting the iterative policy obtained by Hamiltonian-driven continuous ADP.

B. Performance Quantification

The following result discusses the performance of the intermittent ADP compared to the continuous ADP in terms of cost defined in (2).

Theorem 4: Denote $u_j^*(\cdot)$ with $j = s, d$ as the optimal policy $u^*(\cdot)$ with the static and dynamic triggering conditions in Theorems 1 and 2, respectively. Then, the cost of the intermittent feedback policy is

$$J\left(u_j^*; x_0\right) = J\left(u^*; x_0\right) + \int_{t_0}^{\infty} \left\| u_j^*(x) - u^*(x) \right\|_R \mathrm{d}\tau, \ \text{for} \, j = s, d. \quad (51)$$

Proof: The proof follows from [23]. ■

The aforementioned result holds for both the static and dynamic intermittent ADP. Note that the performance of the intermittent ADP policy is closely related to the distance between the optimal policy $u^*(\cdot)$ and the intermittent policy $u_j^*(\cdot)$ defined as

$$D\left(u_j^*(\cdot), u^*(\cdot)\right) = \int_{t_0}^{\infty} \left\| u_j^*(x(\tau)) - u^*(x(\tau)) \right\|_R \mathrm{d}\tau. \quad (52)$$

From Theorem 3, the interevent interval of the static intermittent ADP is larger than the dynamic case. Then, it can be inferred that

$$D\left(u_s^*(\cdot), u^*(\cdot)\right) \le D\left(u_d^*(\cdot), u^*(\cdot)\right). \quad (53)$$

In addition, from Theorem 4, one can obtain

$$J\left(u_s^*; x_0\right) < J\left(u_d^*; x_0\right). \quad (54)$$

Similar to the policy comparison in the Hamiltonian-driven continuous ADP, comparison between static and dynamic intermittent ADP in terms of the performance can be made accordingly, as shown in Fig. 3. From Section II-B, the policy evaluation step in the Hamiltonian-driven ADP framework can be used to compare u_s^* and u_d^* in terms of performance.

Remark 3: For the static (or dynamic) triggering condition in Theorem 1 (or Theorem 2), when the parameter σ approaches 1, which implies that the event is triggered more often and $u_s^* \to u^*$ (or $u_d^* \to u^*$), the term $D(u_s^*(\cdot), u^*(\cdot))$ or $D(u_d^*(\cdot), u^*(\cdot))$ approaches zero. On the other hand, selecting σ close to 0 increases the interevent interval, $t_{k+1}^s - t_k^s$ (or $t_{k+1}^d - t_k^d$), and the performance difference between u_s^* (or u_d^*) and u^*. This means that the intermittent feedback will be far from the continuous feedback.

C. Hamiltonian-Driven Hybrid ADP

Fig. 4 shows the Hamiltonian-driven unified framework consists of continuous learning and intermittent implementation. The policy learning phase employs the continuous ADP, which consists of policy evaluation and policy improvement. The policy implementation is based on the intermittent ADP, including static and dynamic intermittent ADP as designed by Theorems 1 and 2. When Option *A* is activated, continuous ADP learning is used to evaluate and improve the iterative policy $u_i(\cdot)$. When Option *B* or *C* is activated, intermittent ADP is selected to implement the iterative policy u_i. The continuous ADP learning guarantees the convergence of the iterative policy $u_i(\cdot)$ to the optimal policy $u^*(\cdot)$ [28]. Also, Theorems 1 and 2 show that the iterative policy obtained by continuous ADP guarantees the closed-loop stability and Zeno-free property.

Remark 4: The Hamiltonian-driven hybrid ADP framework can be related to existing ADP variants as follows.

1) Compared to [28], two types of intermittent feedback are presented to guarantee the closed-loop stability and avoid the Zeno-phenomenon. In addition, the communication bandwidth and control performance comparison between the continuous and intermittent feedback policies are discussed in Theorems 3 and 4, respectively. It is shown that there is a tradeoff between the bandwidth and performance.
2) In the continuous learning phase of the Hamiltonian-driven hybrid ADP, model-based ADP methods [9]–[12] can be used for policy evaluation and policy improvement.
3) The learning process can be offline implemented with iteratively updated in the subset of the state space asynchronously, as shown in [12].

4) When the system dynamics $f(x)$ and $g(x)$ are unknown, neural network-based approximation method can be employed to compensate the system uncertainty [38], [39].
5) To obviate the requirement of model knowledge, action-dependent value function approximation can be used, such as Q-learning-based ADP methods [24], [40].
6) Off-policy RL, another type of model-free RL variant, can be employed to merge the policy evaluation and policy improvement in the continuous learning phase [21]. This will be discussed in detail in the next section.

D. Model-Free Extension

As mentioned in Section IV-C, the learning phase in hybrid ADP is based on system dynamics $f(\cdot)$ and $g(\cdot)$. This section briefly gives the model-free extension of the learning phase using off-policy RL approaches [13], [41], [42].

Assume that the continuous feedback $\mu(x(t))$, called behavior policy, is applied to the system (1), then, the dynamics of the system can be written as

$$\dot{x} = f(x) + g(x)u_i(x) + g(x)[\mu(x) - u_i(x)].$$

Denote the value function corresponding to the iterative policy $u_i(x)$ as $V_i(x)$, then

$$\|x\|_Q + \|u_i(x)\|_R + \left\langle \frac{\partial V_i(x)}{\partial x}, f(x) + g(x)\mu(x) \right\rangle - 2\left[-\frac{1}{2}\left(\frac{\partial V_i(x)}{\partial x} \right)^{\mathrm{T}} g(x)R^{-1} \right] R[u_i(x) - \mu(x)] = 0.$$

Note that for an strictly increasing sequence of instants $\{t_l\}_{l=0}^{\infty}$

$$\int_{t_l}^{t_{l+1}} \left\langle \frac{\partial V_i(x)}{\partial x}, \dot{x} \right\rangle \mathrm{d}\tau = V_i(x(t_{l+1})) - V_i(x(t_l)).$$

Integrating the above equation on $[t_l, t_{l+1}]$ yields

$$V_i(x(t_{l+1})) - V_i(x(t_l)) + \int_{t_l}^{t_{l+1}} \left(\|x\|_Q + \|u_i(x)\|_R \right) \mathrm{d}\tau - 2\int_{t_l}^{t_{l+1}} \left[u_{i+1}(x) \right]^T R[u_i(x) - u(x)] \mathrm{d}\tau = 0 \tag{55}$$

where $u_{i+1}(x) = -(1/2)R^{-1}g(x)^{\mathrm{T}}[(\partial V_i(x))/(\partial x)]$. Therefore, the model-free learning begins from the initial admissible policy mapping $\hat{u}_i(\cdot)$ and tries find $\hat{u}_{i+1}(x)$ and $\hat{V}_i(x)$ to minimize the norm of Bellman residual, which is defined as

$$e = \hat{V}_i(x(t_{l+1})) - \hat{V}_i(x(t_l)) + \int_{t_l}^{t_{l+1}} \left(\|x\|_Q + \left\| \hat{u}_i(x) \right\|_R \right) \mathrm{d}\tau - 2\int_{t_l}^{t_{l+1}} \left[\hat{u}_{i+1}(x) \right]^T R\left[\hat{u}_i(x) - \mu(x) \right] \mathrm{d}\tau \tag{56}$$

where $\hat{u}_i(\cdot)$ and $\hat{V}_i(\cdot)$ are the approximated policy and value function using function approximators such as neural networks. Then, actor-critic off-policy RL approach can be used to evaluate and improve the iterative policy $u_i(\cdot)$ [13], [41]–[43].

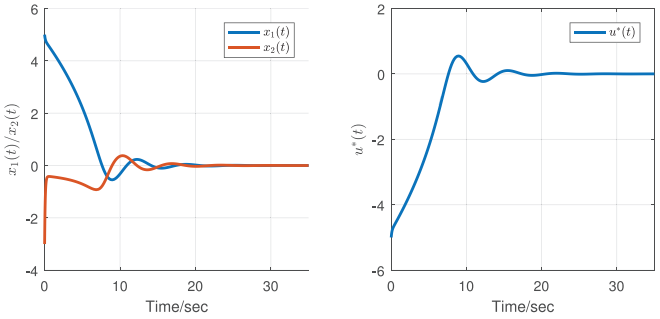


Fig. 5. Evolution of the continuous-triggered control policy $u^*(x)$ with the method in [44].

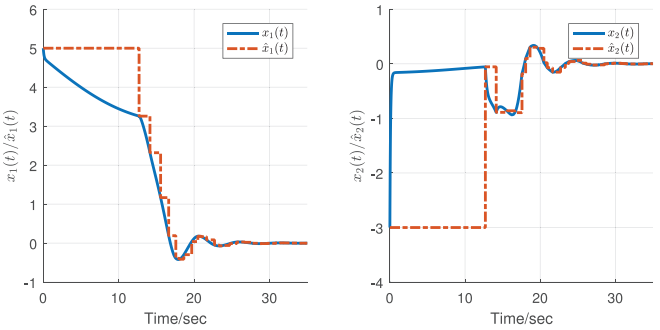


Fig. 6. Evolution $x(t)$ for static intermittent feedback (29).

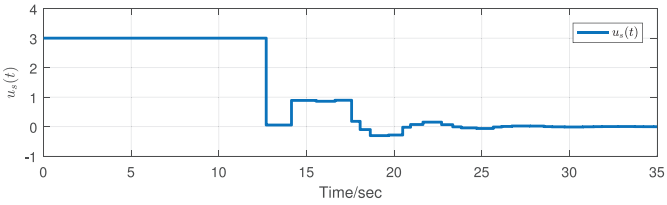


Fig. 7. Evolution of $u_s(t)$ for static intermittent feedback (29).

V. Simulation

In order to validate the effectiveness of the presented Hamiltonian intermittent control policies the example adopted from [44] will be used.

Consider the controlled Van der Pol oscillator

$$\dot{x} = \begin{bmatrix} x_2 \\ -x_1 + 0.5\left(1 - x_2^2\right)x_2 \end{bmatrix} + \begin{bmatrix} 0 \\ 1 \end{bmatrix} u$$

with user-defined matrices $Q = I$ and $R = 1$. The optimal value function for this system is $V^*(x) = x_1^2 + x_2^2$ and the optimal controller is $u^*(x) = -x_2$. Fig. 5 shows the evolution of the optimal continuous-triggered control policy $u^*(x)$ with the method in [44].

Select now $\sigma = 0.8$ of Theorem 1. The evolution of the states and the control signal are shown in Figs. 6 and 7. Select the parameters of Theorem 2, as $\mu = 0.8$, $\theta = 0.3$. The evolution of the states and the control signal are shown in Figs. 8 and 9. In order to compare the communication burden of the three Hamiltonian driven ADP techniques, we show the total number of sampling updates in Fig. 10. The triggering instants for both static and dynamic intermittent mechanisms are given in Fig. 11. From Fig. 10, one can observe that the static intermittent mechanism uses 50 samples of the state as opposed

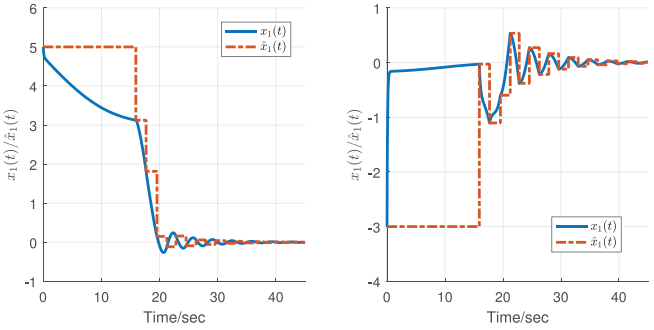


Fig. 8. Evolution of the state for dynamic intermittent feedback (32).

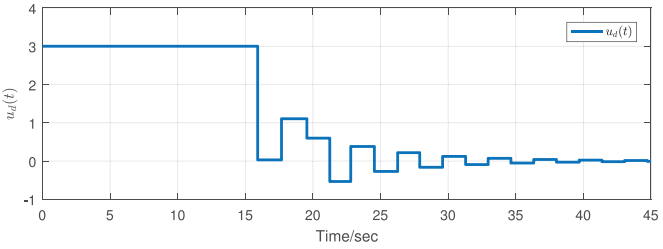


Fig. 9. Evolution of the $u_d(t)$ for dynamic intermittent feedback (32).

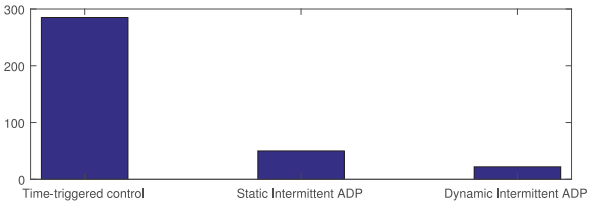


Fig. 10. Number of state samples used in continuous-triggered policy from [44], static (29) and dynamic (32) intermittent control.

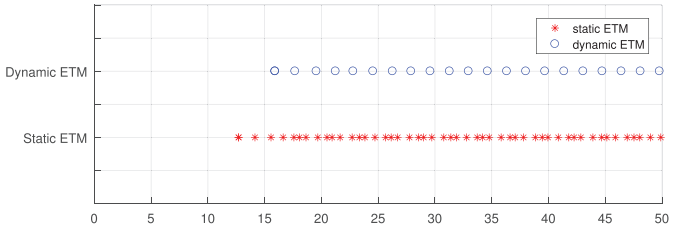


Fig. 11. Triggering instants for static (29) and dynamic (32) intermittent feedback mechanisms.

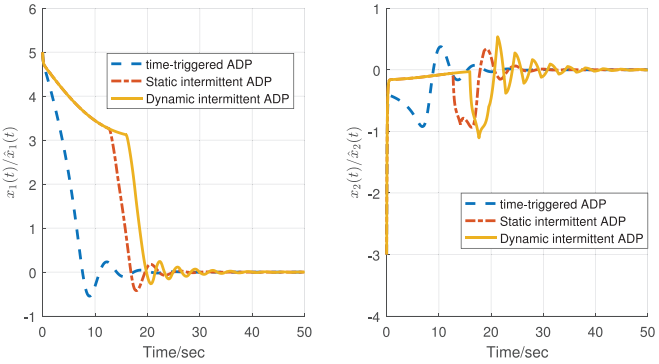


Fig. 12. Comparison of state evolution for continuous-triggered [44], static (29) and dynamic (32) intermittent cases.

to 285 of the continuous-triggered controller. The dynamic intermittent mechanism uses 22 only samples of the state. Fig. 11 also support this point. Figs. 12 and 13 illustrate the

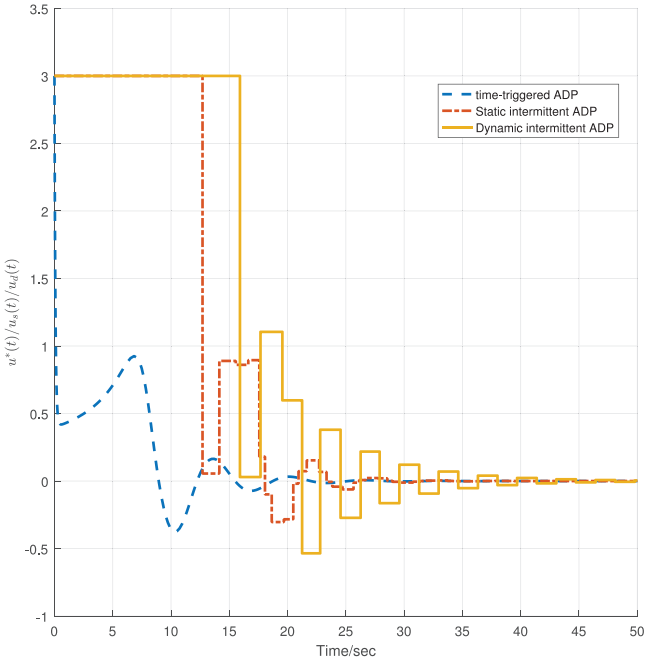


Fig. 13. Comparison of control input evolution for continuous-triggered [44], static (29) and dynamic (32) intermittent cases.

performance, where the state trajectories and control input signals of three mechanisms are given. One can observe that as expected, the static intermittent mechanism outperforms the dynamic one in terms of communication bandwidth.

VI. Conclusion

This article presented a Hamiltonian-driven unified framework for hybrid ADP, which consists of continuous feedback ADP learning phase and intermittent feedback implementation phase. Both the closed-loop stability and Zeno-free property are guaranteed for the intermittent ADP implementation phase. Moreover, the quantifiable tradeoff between the communication cost and the control performance for static and dynamic intermittent control policies is discussed. A simulation example is conducted to verify the efficacy of the presented Hamiltonian-driven unified framework. Future work is to integrate model-free RL-based algorithms with the developed hybrid framework to approximate optimal policy without the requirement of complete knowledge of system dynamics.

Acknowledgment

The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the Army Research Laboratory or the U.S. Government.

References

[1] P. Tabuada, "Event-triggered real-time scheduling of stabilizing control tasks," *IEEE Trans. Autom. Control*, vol. 52, no. 9, pp. 1680–1685, Sep. 2007.

[2] D. P. Borgers and W. P. M. H. Heemels, "Event-separation properties of event-triggered control systems," *IEEE Trans. Autom. Control*, vol. 59, no. 10, pp. 2644–2656, Oct. 2014.

[3] M. Mazo, A. Anta, and P. Tabuada, "An ISS self-triggered implementation of linear controllers," *Automatica*, vol. 46, no. 8, pp. 1310–1314, 2010.

[4] X. Wang and M. D. Lemmon, "Self-triggered feedback control systems with finite-gain $\mathcal{L}_2$ stability," *IEEE Trans. Autom. Control*, vol. 54, no. 3, pp. 452–467, Mar. 2009.

[5] G. S. Seyboth, D. V. Dimarogonas, and K. H. Johansson, "Event-based broadcasting for multi-agent average consensus," *Automatica*, vol. 49, no. 1, pp. 245–252, 2013.

[6] T. Liu and Z.-P. Jiang, "Event-based control of nonlinear systems with partial state and output feedback," *Automatica*, vol. 53, pp. 10–22, Mar. 2015.

[7] A. Girard, "Dynamic triggering mechanisms for event-triggered control," *IEEE Trans. Autom. Control*, vol. 60, no. 7, pp. 1992–1997, Jul. 2015.

[8] Y. Yang, H. Modares, K. G. Vamvoudakis, Y. Yin, and D. C. Wunsch, "Dynamic intermittent feedback design for H_∞ containment control on a directed graph," *IEEE Trans. Cybern.*, to be published, doi: 10.1109/TCYB.2019.2933736.

[9] D. Liu and Q. Wei, "Policy iteration adaptive dynamic programming algorithm for discrete-time nonlinear systems," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 25, no. 3, pp. 621–634, Mar. 2014.

[10] Q. Wei, D. Liu, and H. Lin, "Value iteration adaptive dynamic programming for optimal control of discrete-time nonlinear systems," *IEEE Trans. Cybern.*, vol. 46, no. 3, pp. 840–853, Mar. 2016.

[11] Q. Wei, D. Liu, Q. Lin, and R. Song, "Discrete-time optimal control via local policy iteration adaptive dynamic programming," *IEEE Trans. Cybern.*, vol. 47, no. 10, pp. 3367–3379, Oct. 2017.

[12] Q. Wei, F. L. Lewis, D. Liu, R. Song, and H. Lin, "Discrete-time local value iteration adaptive dynamic programming: Convergence analysis," *IEEE Trans. Syst., Man, Cybern., Syst.*, vol. 48, no. 6, pp. 875–891, Jun. 2018.

[13] Y. Jiang and Z.-P. Jiang, "Computational adaptive optimal control for continuous-time linear systems with completely unknown dynamics," *Automatica*, vol. 48, no. 10, pp. 2699–2704, 2012.

[14] K. G. Vamvoudakis and F. L. Lewis, "Online actor–critic algorithm to solve the continuous-time infinite horizon optimal control problem," *Automatica*, vol. 46, no. 5, pp. 878–888, 2010.

[15] F. Lewis and D. Liu, *Reinforcement Learning and Approximate Dynamic Programming for Feedback Control*. Hoboken, NJ, USA: Wiley, 2013.

[16] D. Liu, H. Li, and D. Wang, "Online synchronous approximate optimal learning algorithm for multi-player non-zero-sum games with unknown dynamics," *IEEE Trans. Syst., Man, Cybern., Syst.*, vol. 44, no. 8, pp. 1015–1027, Aug. 2014.

[17] D. Wang, D. Liu, Q. Zhang, and D. Zhao, "Data-based adaptive critic designs for nonlinear robust optimal control with uncertain dynamics," *IEEE Trans. Syst., Man, Cybern., Syst.*, vol. 46, no. 11, pp. 1544–1555, Nov. 2016.

[18] Y. Yang, Z. Guo, H. Xiong, D.-X. Ding, Y. Yin, and D. C. Wunsch, "Data-driven robust control of discrete-time uncertain linear systems via off-policy reinforcement learning," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 30, no. 12, pp. 3735–3747, Dec. 2019.

[19] H. Modares and F. L. Lewis, "Linear quadratic tracking control of partially-unknown continuous-time systems using reinforcement learning," *IEEE Trans. Autom. Control*, vol. 59, no. 11, pp. 3051–3056, Nov. 2014.

[20] H. Modares, S. P. Nageshrao, G. A. D. Lopes, R. Babuška, and F. L. Lewis, "Optimal model-free output synchronization of heterogeneous systems using off-policy reinforcement learning," *Automatica*, vol. 71, pp. 334–341, Sep. 2016.

[21] Y. Yang, H. Modares, D. C. Wunsch, and Y. Yin, "Leader–follower output synchronization of linear heterogeneous systems with active leader using reinforcement learning," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 29, no. 6, pp. 2139–2153, Jun. 2018.

[22] Y. Yang, H. Modares, D. C. Wunsch, and Y. Yin, "Optimal containment control of unknown heterogeneous systems with active leaders," *IEEE Trans. Control Syst. Technol.*, vol. 27, no. 3, pp. 1228–1236, May 2019.

[23] K. G. Vamvoudakis, "Event-triggered optimal adaptive control algorithm for continuous-time nonlinear systems," *IEEE/CAA J. Automatica Sinica*, vol. 1, no. 3, pp. 282–293, Jul. 2014.

[24] K. G. Vamvoudakis and H. Ferraz, "Model-free event-triggered control algorithm for continuous-time linear systems with optimal performance," *Automatica*, vol. 87, pp. 412–420, Jan. 2018.

[25] Y. Yang, K. G. Vamvoudakis, H. Ferraz, and H. Modares, "Dynamic intermittent Q-learning-based model-free suboptimal co-design of $\mathcal{L}_2$-stabilization," *Int. J. Robust Nonlin. Control*, vol. 29, no. 9, pp. 2673–2694, 2019.

[26] X. Zhong and H. He, "An event-triggered ADP control approach for continuous-time system with unknown internal states," *IEEE Trans. Cybern.*, vol. 47, no. 3, pp. 683–694, Mar. 2017.

[27] M. Ha, D. Wang, and D. Liu, "Event-triggered adaptive critic control design for discrete-time constrained nonlinear systems," *IEEE Trans. Syst., Man, Cybern., Syst.*, to be published, doi: 10.1109/TSMC.2018.2868510.

[28] Y. Yang, D. Wunsch, and Y. Yin, "Hamiltonian-driven adaptive dynamic programming for continuous nonlinear dynamical systems," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 28, no. 8, pp. 1929–1940, Aug. 2017.

[29] M. Abu-Khalaf and F. L. Lewis, "Nearly optimal control laws for nonlinear systems with saturating actuators using a neural network HJB approach," *Automatica*, vol. 41, no. 5, pp. 779–791, 2005.

[30] F. L. Lewis and D. Vrabie, "Reinforcement learning and adaptive dynamic programming for feedback control," *IEEE Circuits Syst. Mag.*, vol. 9, no. 3, pp. 32–50, 3rd Quart., 2009.

[31] 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.

[32] D. P. Bertsekas, *Dynamic Programming and Optimal Control*, 1st ed. Belmont, MA, USA: Athena Sci., 1995.

[33] D. Liberzon, *Calculus of Variations and Optimal Control Theory: A Concise Introduction*. Princeton, NJ, USA: Princeton Univ. Press, 2012.

[34] R. W. Beard, G. N. Saridis, and J. T. Wen, "Galerkin approximations of the generalized Hamilton–Jacobi–Bellman equation," *Automatica*, vol. 33, no. 12, pp. 2159–2177, 1997.

[35] F. L. Lewis, D. Vrabie, and V. L. Syrmos, *Optimal Control*. New York, NY, USA: Wiley, 2012.

[36] A. J. van der Schaft, "$\mathcal{L}_2$-gain analysis of nonlinear systems and nonlinear state-feedback H_∞ control," *IEEE Trans. Autom. Control*, vol. 37, no. 6, pp. 770–784, Jul. 1992.

[37] H. K. Khalil, *Nonlinear Systems*. Upper Saddle River, NJ, USA: Prentice-Hall, 2002.

[38] S. Bhasin, R. Kamalapurkar, M. Johnson, K. G. Vamvoudakis, F. L. Lewis, and W. E. Dixon, "A novel actor–critic–identifier architecture for approximate optimal control of uncertain nonlinear systems," *Automatica*, vol. 49, no. 1, pp. 82–92, 2013.

[39] R. Kamalapurkar, H. Dinh, S. Bhasin, and W. E. Dixon, "Approximate optimal trajectory tracking for continuous-time nonlinear systems," *Automatica*, vol. 51, pp. 40–48, Jan. 2015.

[40] Q. Wei, F. L. Lewis, Q. Sun, P. Yan, and R. Song, "Discrete-time deterministic Q-learning: A novel convergence analysis," *IEEE Trans. Cybern.*, vol. 47, no. 5, pp. 1224–1237, May 2017.

[41] R. Song, F. L. Lewis, Q. Wei, and H. Zhang, "Off-policy actor–critic structure for optimal control of unknown systems with disturbances," *IEEE Trans. Cybern.*, vol. 46, no. 5, pp. 1041–1050, May 2016.

[42] R. Song, F. L. Lewis, and Q. Wei, "Off-policy integral reinforcement learning method to solve nonlinear continuous-time multiplayer nonzero-sum games," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 28, no. 3, pp. 704–713, Mar. 2017.

[43] H. Modares, F. L. Lewis, and Z.-P. Jiang, "H_∞ tracking control of completely unknown continuous-time systems via off-policy reinforcement learning," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 26, no. 10, pp. 2550–2562, Oct. 2015.

[44] L. Rodrigues, D. Henrion, and M. A. Fallah, "An inverse optimality method to solve a class of optimal control problems," *arXiv preprint arXiv:1002.2900*, 2010.

Yongliang Yang (Member, IEEE) received the B.S. degree in electrical engineering from Hebei University, Baoding, China, in 2011, and the Ph.D. degree in electrical engineering from the University of Science and Technology Beijing (USTB), Beijing, China, in 2017.

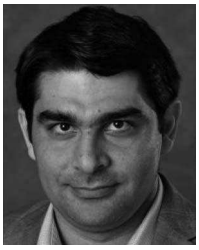
He was a Visiting Scholar with the Missouri University of Science and Technology, Rolla, MO, USA, from 2015 to 2017, sponsored by China Scholarship Council. He is currently an Assistant Professor with the School of Automation and Electrical Engineering, USTB. His research interests include adaptive optimal control, distributed optimization and control, and CPSs.

Dr. Yang was a recipient of the Best Ph.D. Dissertation of China Association of Artificial Intelligence, the Best Ph.D. Dissertation of USTB, the Chancellors Scholarship in USTB, and the Excellent Graduates Awards in Beijing. He also serves as the Reviewer of several international journals and conferences, including *Automatica*, the IEEE TRANSACTIONS ON AUTOMATIC CONTROL, IEEE TRANSACTIONS ON CYBERNETICS and IEEE TRANSACTIONS ON NEURAL NETWORKS AND LEARNING SYSTEMS.

Kyriakos G. Vamvoudakis (Senior Member, IEEE) was born in Athens, Greece. He received the Diploma (a 5 year degree, equivalent to a Master of Science) degree (Highest Hons.) in electronic and computer engineering from the Technical University of Crete, Chania, Greece, in 2006, and the M.S. and Ph.D. degrees in electrical engineering from the University of Texas at Arlington, Arlington, TX, USA, in 2008 and 2011, respectively, under the supervision of F. L. Lewis.

From May 2011 to January 2012, he was working as an Adjunct Professor and a Faculty Research Associate with the University of Texas at Arlington and the Automation and Robotics Research Institute, Fort Worth, TX, USA. From 2012 to 2016 he was a Project Research Scientist with the Center for Control, Dynamical Systems and Computation, University of California, at Santa Barbara, Santa Barbara, CA, USA. He was an Assistant Professor with the Kevin T. Crofton Department of Aerospace and Ocean Engineering, Virginia Tech, Blacksburg, VA, USA, until 2018. He currently serves as an Assistant Professor with the Daniel Guggenheim School of Aerospace Engineering, Georgia Institute of Technology, Atlanta, GA, USA. His research interests include approximate dynamic programming, game theory, cyber-physical security, networked control, smart grid, and safe autonomy.

Dr. Vamvoudakis is a recipient of the 2019 ARO YIP Award, the 2018 NSF CAREER Award, and of several international awards, including the 2016 International Neural Network Society Young Investigator Award, the Best Paper Award for Autonomous/Unmanned Vehicles at the 27th Army Science Conference in 2010, the Best Presentation Award at the World Congress of Computational Intelligence in 2010, and the Best Researcher Award from the Automation and Robotics Research Institute in 2011. He is a member of Tau Beta Pi, Eta Kappa Nu, and Golden Key Honor Societies. He is listed in Who's Who in the World, Who's Who in Science and Engineering, and Who's Who in America. He has also served on various international program committees and has organized special sessions, workshops, and tutorials for several international conferences. He currently is a member of the Technical Committee on Intelligent Control of the IEEE Control Systems Society, the Technical Committee on Adaptive Dynamic Programming and Reinforcement Learning of the IEEE Computational Intelligence Society, a member of the IEEE Control Systems Society Conference Editorial Board, an Associate Editor of *Automatica*, *IEEE Computational Intelligence Magazine*, the *Journal of Optimization Theory and Applications*, IEEE CONTROL SYSTEMS LETTERS, and a Registered Electrical/Computer Engineer (PE), and a member of the Technical Chamber of Greece. He is a Senior Member of AIAA.



Hamidreza Modares (Member, IEEE) received the B.S. degree in electrical engineering from the University of Tehran, Tehran, Iran, in 2004, the M.S. degree in electrical engineering from the Shahrood University of Technology, Shahroud, Iran, in 2006, and the Ph.D. degree in electrical engineering from the University of Texas at Arlington, Arlington, TX, USA, in 2015.

He was a Faculty Research Associate with the University of Texas at Arlington from 2015 to 2016, and an Assistant Professor with the Missouri University of Science and Technology, Rolla, MO, USA, from 2016 to 2018. He is currently an Assistant Professor with the Mechanical Engineering Department, Michigan State University, East Lansing, MI, USA. His current research interests include cyber-physical systems, reinforcement learning, distributed control, robotics, and machine learning.

Dr. Modares was a recipient of the Best Paper Award from the 2015 IEEE International Symposium on Resilient Control Systems. He is an Associate Editor of the IEEE TRANSACTIONS ON NEURAL NETWORKS AND LEARNING SYSTEMS.

Yixin Yin (Member, IEEE) received the B.S., M.S., and Ph.D. degrees in electrical engineering from the University of Science and Technology Beijing (USTB), Beijing, China, in 1982, 1984, and 2002, respectively.

He was a Visiting Scholar with several universities in Japan, including the University of Tokyo, Tokyo, Japan, the Kyushu Institute of Technology, Kitakyushu, Japan, Kanagawa University, Yokohama, Japan, Chiba University, Chiba, Japan, and the Muroran Institute of Technology, Muroran, Japan. He served as the Dean of the School of the Information Engineering, USTB from 2000 to 2011, where he was the Dean of the School of Automation and Electrical Engineering, from 2011 to 2017, and he is currently a Professor with the School of Automation and Electrical Engineering. His major research interests include modeling and control of complex industrial processes, computer aided design of control system, intelligent control, and artificial life.

Prof. Yin is a recipient of several national awards, including the Outstanding Young Educator in 1993, the Award of Science and Technology Progress in Education in 1994, the Award of National Science and Technology Progress in 1995, the Special Allowance of the State Council in 1994, the Best Paper of Japanese Acoustical Society in 1999, and the Award of Metallurgical Science and Technology in 2014. He is a fellow of Chinese Society for Artificial Intelligence and a member of Chinese Society for Metals and Chinese Association of Automation.

Donald C. Wunsch (Fellow, IEEE) received the B.S. degree in applied mathematics from the University of New Mexico, Albuquerque, NM, USA, in 1984, the M.S. degree in applied mathematics and the Ph.D. degree in electrical engineering from the University of Washington, Seattle, WA, USA, in 1987 and 1991, respectively, and the Executive M.B.A. degree in business administration from Washington University in St. Louis, St. Louis, MO, USA, in 2006.

He was with Texas Tech University, Lubbock, TX, USA; Boeing, Seattle, WA, USA; Rockwell International, Albuquerque, NM, USA; and International Laser Systems, Albuquerque, NM, USA. He is currently the Mary K. Finley Missouri Distinguished Professor with the Missouri University of Science and Technology (Missouri S&T), Rolla, MO, USA, where he is also the Director of the Applied Computational Intelligence Laboratory, a multidisciplinary research group. His current research interests include clustering/unsupervised learning; biclustering; adaptive resonance and adaptive dynamic programming architectures, hardware, and applications; neurofuzzy regression; autonomous agents; games; and bioinformatics.

Dr. Wunsch was a recipient of the NSF CAREER Award, the 2015 International Neural Networks Society (INNS) Gabor Award, and the 2019 Ada Lovelace Service Award. He has produced 22 Ph.D. recipients in computer engineering, electrical engineering, systems engineering, and computer science. He served as the IJCNN General Chair, and on several boards, including the St. Patricks School Board, the IEEE Neural Networks Council, the INNS, and the University of Missouri Bioinformatics Consortium, and the Chair of the Missouri S&T Information Technology and Computing Committee as well as the Student Design and Experiential Learning Center Board. He is an INNS Fellow and the former INNS President.