2018 IEEE Conference on Decision and Control (CDC)
Miami Beach, FL, USA, Dec. 17-19, 2018

Dynamic Intermittent Q-Learning for Systems with Reduced Bandwidth

Yongliang Yang *Member, IEEE,* Kyriakos G. Vamvoudakis *Senior Member, IEEE,*
Henrique Ferraz, Hamidreza Modares *Member, IEEE*

***Abstract*— This paper presents a Q-learning based dynamic intermittent mechanism to control linear systems evolving in continuous time. In contrast to existing event-triggered mechanisms, where complete knowledge of the system dynamics is required, the proposed dynamic intermittent control obviates this requirement while providing a quantified level of performance. An internal dynamical system will be introduced to generate the triggering condition. Then, a dynamic intermittent Q-learning is developed to learn the optimal value function and the hybrid controller. A qualitative performance analysis of the dynamic event-triggered control is given in comparison to the continuous-triggered control law to show the degree of suboptimality. The combined closed-loop system is written as an impulsive system, and it is proved to have an asymptotically stable equilibrium point without any Zeno behavior. A numerical simulation of an unknown unstable system is presented to show the efficacy of the proposed approach.**

***Index Terms*— Intermittent Q-learning, dynamic event-triggered control, suboptimal performance.**

I. INTRODUCTION

Feedback is the main principle of control that guarantees several properties of a dynamical plant, including asymptotic stability of equilibrium, optimality and/or disturbance rejection. Recently, several approaches have been formulated, wherein the control is carried out in an open-loop manner between consecutive event instants. Such approaches are classified in event-triggered control [1]–[3] and self-triggered control [4], [5]. Event-triggered control consists of a feedback controller that computes the control input and a triggering mechanism that determines when the control input has to be updated again. Emulation-based approaches have been used to synthesize the intermittent controller by first developing a feedback controller to stabilize the plant without constraints of communication, then subsequently determining the appropriate event-triggering condition to reduce the communication bandwidth, while still ensuring stability [6], [7].

Related Work

In order to relax the above requirements, several papers have focused on the co-design problem, where the design of the controller and the event-triggering condition are carried out simultaneously [8]–[10]. However, these methods are model-based and they require full knowledge of the system dynamics, which can be vulnerable to exhaustive modeling and malicious attacks.

Adaptive dynamic programming (ADP) is a recently developed technique that adopts the idea from reinforcement learning (RL) to approximate the optimal controller for general nonlinear dynamical systems [11]. In contrast to traditional optimal control theory [12], ADP solves the Riccati or Hamilton-Jacobi-Bellman equation in an online manner [13]. Recently, event-triggering is combined with online ADP to develop intermittent control laws [14], [15]. However, complete or partial knowledge of system dynamics is required, which might not be available in many applications. Off-policy RL algorithm [16]–[18] is employed in [19] to obtain the optimal feedback for the event-triggered control in a model-free manner, but the learning process of optimal feedback gain and the event design are separated. Another model-free RL approach, the Q-learning algorithm [20], is adopted in [21] to co-design the optimal feedback gain and the event-triggering condition simultaneously. In this paper, dynamic intermittent feedback proposed in [22] is extended to combine with the Q-learning algorithm to develop a model-free intermittent control to reduce the communication burden even more.

Contributions

The contributions of this paper are threefold. First, an intermittent Q-learning algorithm is combined with an actor-critic structure implemented with a zero-order hold (ZOH) to learn the parameters of optimal Q-function in an online and model-free manner. Specifically, in contrast to existing model-based event-triggered designs, this paper presents a model-free solution to the co-design of both the transmission instants and the control policy in the context of intermittent control. Moreover, an internal dynamical system is introduced to generate the dynamic triggering condition in order to reduce the communication and computation burden even more. The combined closed-loop system is analyzed within the framework of impulsive system approach, and it is proved to have an asymptotically stable equilibrium point without

This work was supported in part by the National Natural Science Foundation of China under Grant No. 61333002, in part by the Fundamental Research Funds for the Central Universities under Grant FRF-TP-18-031A1, in part by NATO under Grant No. SPS G5176, in part by ONR Minerva under Grant No. N00014-18-1-2160 and in part by NSF CAREER CPS-1750789.

Y. Yang is with the School of Automation & Electrical Engineering, University of Science and Technology Beijing, Beijing 100083, China yangyongliang@ieee.org

K. G. Vamvoudakis is with the Daniel Guggenheim School of Aerospace Engineering, Georgia Tech, Atlanta, GA 30332-0150, USA kyriakos@gatech.edu

H. Ferraz is with the Department of Electrical and Computer Engineering, University of California Santa Barbara, Santa Barbara, CA 93106-9560, USA henrique@ece.ucsb.edu

H. Modares is with the the Department of Mechanical Engineering at Michigan State University, East Lansing, MI 48824-1226, USA modaresh@msu.edu

any Zeno behavior. Finally, the performance of the dynamic event-triggered control is compared to the time-triggered control case to show the degree of sub-optimality.

Structure

The remainder of the paper is structured as follows. The optimal stabilization problem of continuous-time linear dynamical systems is formulated in Section II, where both time- and event-triggered control designs are discussed. Section III reviews the static intermittent feedback designs for the cases of model-based and model-free. In Section IV, a novel dynamic intermittent Q-learning-based co-design of event-triggering condition and feedback controller is presented and it is shown that the equilibrium of the closed-loop system is globally asymptotically stable with Zeno-free triggering. A simulation is presented in Section V to verify the proposed algorithm and the conclusions are given in Section VI.

Preliminaries

The notations used in this paper is standard.

$(\cdot)^+$ $\triangleq$ the right-limit operator, $p^+ = \lim_{s\to t^+} p(s)$.

$\mathbb{R}^+$ $\triangleq$ the set of positive real numbers.

t_k $\triangleq$ the k-th consecutive sampling instant

$\{t_k\}_{k=0}^{\infty}$ $\triangleq$ a sequence of monotonically increasing sampling instants satisfying $\lim_{k\to\infty} t_k = \infty$.

$M_{i,j}$ $\triangleq$ (i,j)-th entry of matrix M.

$\bar{\lambda}(M)$ $\triangleq$ the maximum eigenvalue of matrix M.

$\underline{\lambda}(M)$ $\triangleq$ the minimum eigenvalue of matrix M.

$\|M\|_F$ $\triangleq$ $\sqrt{tr(MM^H)}$ (matrix Frobenius norm).

$vech(M)$ $\triangleq$ $\left[\begin{array}{cccccccc} M_{1,1} & \cdots & M_{1,n} & M_{2,2} & \cdots & M_{2,n} & \cdots & M_{n-1,n-1} \quad M_{n-1,n} \quad M_{n,n} \end{array}\right]^T$, $\forall M \in \mathbb{R}^{n\times n}$.

Definition 1. *(Persistent Excitation)* A vector signal $y(t) \in \mathbb{R}^p$ is exciting over the interval $[t, t+T]$ with $T \in \mathbb{R}^+$ if there exists $\beta_1 \in \mathbb{R}^+$ and $\beta_2 \in \mathbb{R}^+$ such that for $\forall t$,

$$\beta_1 I_{p\times p} \leqslant \int_t^{t+T} y(\tau)\, y^{\mathrm{T}}(\tau)\, d\tau \leqslant \beta_2 I_{p\times p} \qquad \square$$

II. Problem Statement

A. Time-Triggered Optimal Control

In this paper, the following continuous-time linear dynamical system is considered

$$\dot{x}(t) = Ax(t) + Bu(t), \qquad x(t_0) = x_0, \tag{1}$$

where $x(t) \in \mathbb{R}^n$ is the state vector, $u(t) \in \mathbb{R}^m$ is the control input, x_0 is the initial state at time $t_0 \geqslant 0$, and $A \in \mathbb{R}^{n\times n}$, $B \in \mathbb{R}^{n\times m}$ are the state and the input matrices.

Let the optimal value function be defined as

$$V^*(x) \coloneqq \min_{u(t)} \int_{t_0}^{\infty} r(x(t), u(t))\, dt, \tag{2}$$

with

$$r(x,u) = \frac{1}{2}\left(x^T H x + u^T R u\right), \tag{3}$$

where $H \geq 0$ and $R > 0$. For a linear system of the form of (1) we can represent the value function as

$$V^*(x) = \frac{1}{2} x^T P x,$$

Assumption 1. The pair (A, B) is stabilizable and $\left(A, \sqrt{H}\right)$ is detectable. $\square$

Under Assumption 1, $P \in \mathbb{R}^{n\times n}$ is the unique positive definite matrix that solves

$$A^T P + PA - PBR^{-1}B^T P + H = 0. \tag{4}$$

Define the Hamiltonian functional as

$$\mathcal{H}\left(u(\cdot); x, \frac{\partial V^*(x)}{\partial x}\right) = \left\langle \frac{\partial V^*(x)}{\partial x}, \dot{x} \right\rangle + r(x,u)$$
$$= \left\langle \frac{\partial V^*(x)}{\partial x}, Ax + Bu \right\rangle + \frac{1}{2}\left(x^T H x + u^T R u\right) \tag{5}$$

Then, based on the Hamiltonian, the optimal control u^* can be obtained as

$$u^*(x) \coloneqq \arg\min_{u(\cdot)} \mathcal{H}\left(u(\cdot); x, \frac{\partial V^*(x)}{\partial x}\right)$$
$$= -R^{-1}B^T \frac{\partial V^*(x)}{\partial x} = -R^{-1}B^T P x \tag{6}$$

Note that in order to solve Eq. (4), complete knowledge of the model of the system is needed, which might not be available in many applications. Also, the optimal control $u^*(t)$ requires continuous update of the control signal, which might be computationally inefficient and increase the communication between controller and sensors/actuators.

In this paper, $u^*(t)$ is referred to as time-triggered optimal control, as opposed to the event-triggered control introduced as follows.

B. Intermittent Feedback Control

In order to increase the computational efficiency and reduce the communication burden, event-triggered controller is obtained by introducing an aperiodic sampling mechanism. Consider the aperiodic state sampling

$$\hat{x}(t) \coloneqq \begin{cases} x(t_k), & \forall t \in [t_k, t_{k+1}) \\ x(t_{k+1}), & t = t_{k+1} \end{cases} \tag{7}$$

The gap between the current state $x(t)$ and the sampled state $\hat{x}(t)$ is denoted as

$$e(t) \coloneqq \hat{x}(t) - x(t) \tag{8}$$

In the sequel, an intermittent control law with aperiodic sampling is introduced based on the time-triggered optimal control (6) as

$$u_e(x) \coloneqq u^*(\hat{x}) = -R^{-1}B^T P \hat{x}. \tag{9}$$

With the above intermittent control law, the continuous dynamics of closed-loop system (1) can be written as

$$\dot{x}(t) = Ax(t) - BR^{-1}B^T P \hat{x}(t) \tag{10}$$

For the event-triggered control policy (9), the following lemma holds.

Lemma 1. *[21] Consider the event-triggered control policy* (9)*. Then, the following facts are true.*

1) *There exists a positive constant* $L \in \mathbb{R}^+$ *such that*

$$\|u^*(x(t)) - u_e(x(t))\| \leqslant L\|e(t)\| \tag{11}$$

2) *The intermittent Hamiltonian from the event-triggered control* (9)*,* $\mathcal{H}\left(u_e(\cdot); x, \frac{\partial V^*(x)}{\partial x}\right)$*, satisfies*

$$\mathcal{H}\left(u_e(\cdot); x, \frac{\partial V^*(x)}{\partial x}\right) \leqslant \frac{L^2\bar{\lambda}(R)}{2}\|e\|^2 \tag{12}$$

□

Lemma 2. *[14] Under Assumption 1, the relationship between the intermittent Hamiltonian and the continuous-triggered Hamiltonian is*

$$\begin{aligned}&\mathcal{H}\left(u_e; x, \frac{\partial V^*(x)}{\partial x}\right) - \mathcal{H}\left(u^*; x, \frac{\partial V^*(x)}{\partial x}\right)\\ &= (u_e - u^*)^T R(u_e - u^*).\end{aligned} \tag{13}$$

□

Remark 1. Note that (9) is a general form of ARE-based event-triggered control policy. Different types of intermittent mechanisms differ in how the state sampling instant sequence $\{t_k\}_{k=0}^{\infty}$ in (7) is determined. As will be shown later, the static event-triggered control $u_s(\cdot)$ and the dynamic event-triggered control $u_d(\cdot)$ have the same form as (9) but the corresponding event-triggering condition is different. □

In the following, it is assumed that finite-time stabilization is not achieved, i.e. $x(t_k) \neq 0$, $\forall k \in \mathbb{N}^+$.

III. Static Intermittent Feedback Design

In this section, static intermittent feedback in both model-based and model-free fashion are briefly reviewed.

A. Static Model-Based Event-Triggered Control

Lemma 3. *[14]* (Static Model-Based Event-triggered Mechanism) *Under Assumptions 1, suppose that the event-triggered controller* $u_s(x) \coloneqq u_e(x)$ *is applied to system* (1) *with the event-triggering condition*

$$\|e\|^2 \leqslant \frac{\left(1-\beta^2\right)\underline{\lambda}(H)}{L^2\bar{\lambda}(R)}\|x\|^2 + \frac{\underline{\lambda}(R)}{L^2\bar{\lambda}(R)}\|u_s\|^2 \tag{14}$$

where $\beta \in (0,1)$ *is a user-defined parameter. Then, the closed-loop system of* (1) *has an asymptotically stable equilibrium point. Moreover, Zeno behavior is guaranteed to be excluded for the event-triggered control* u_s*.* □

According to the event-triggering condition (14) in Lemma 3, the sampling instants can be expressed as

$$\begin{aligned}t_0 &= 0,\\ t_{k+1} &= \inf_{t\in\mathbb{R}^+}\{t > t_k \wedge p \leqslant 0\}.\end{aligned} \tag{15}$$

where

$$p := \left(1-\beta^2\right)\underline{\lambda}(H)\|x\|^2 + \underline{\lambda}(R)\|u_s\|^2 - L^2\bar{\lambda}(R)\|e\|^2. \tag{16}$$

Note that the parameters of the event-triggering condition (14) are time-invariant and $p \geqslant 0$ has to be always satisfied. Therefore, in this paper, the triggering condition (14) is named as static triggering condition, in contrast to the dynamic triggering condition discussed next.

B. Static Model-Free Event-triggered Mechanism

1) Action-Dependent Value Function: In order to develop an algorithm to learn the optimal feedback gain K^* in a model-free manner, the function of a state-action pair, named as action-dependent value function or Q-function, is introduced as

$$\begin{aligned}\mathcal{Q}(x,u_e) \triangleq\ & V^*(x) + \mathcal{H}\left(u_e(\cdot); x, \frac{\partial V^*(x)}{\partial x}\right)\\ & - \mathcal{H}\left(u^*(\cdot); x, \frac{\partial V^*(x)}{\partial x}\right)\end{aligned} \tag{17}$$

This function can be equivalently rewritten in a compact form as

$$\mathcal{Q}(x,u_e) = \frac{1}{2}\bar{x}^T Q\bar{x}, \tag{18}$$

with $\bar{x} \coloneqq \begin{bmatrix} x \\ u_e \end{bmatrix}$ and $Q \coloneqq \begin{bmatrix} Q_{xx} & Q_{xu} \\ Q_{ux} & Q_{uu} \end{bmatrix}$ where $Q_{xx} \coloneqq H + P + PA + A^TP$, $Q_{xu} \coloneqq PB$, $Q_{ux} \coloneqq B^TP$, and $Q_{uu} \coloneqq R$.

Based on the Q-function in (18), the optimal feedback gain K^* can be equivalently expressed as

$$K^* = -Q_{uu}^{-1}Q_{ux} \tag{19}$$

Then, the event-triggered control (9) can be rewritten as

$$u_e = -Q_{uu}^{-1}Q_{ux}\hat{x} \tag{20}$$

In the following, a model-free algorithm is presented to learn the parameters of the Q-function $\mathcal{Q}(x,u_e)$ in (18) and the event-triggered controller $u_e(\cdot)$ in (9) or (20).

2) Actor-Critic Representation: In this subsection, actor-critic structure is employed to parametrize the approximator of the Q-function and the event-triggered controller, i.e., a critic approximator learns $\mathcal{Q}(x,u_e)$ in (18) and an actor approximator with a ZOH learns the event-triggered controller $u_e(\cdot)$ in (9) or (20).

First, the Q-function in (18) can be expressed as

$$\mathcal{Q}(x,u_e) = \frac{1}{2}\bar{x}^T Q\bar{x} = \langle W_c^*, \varphi_c\rangle, \tag{21}$$

where $W_c^* = \frac{1}{2}vech(Q) \in \mathbb{R}^{\frac{(n+m)(n+m+1)}{2}}$ is the ideal critic weight vector with

$$\begin{aligned}vech(Q_{xx}) &= W_c^*\left(1 : \tfrac{n(n+1)}{2}\right)\\ vech(Q_{xu}) &= W_c^*\left(\tfrac{n(n+1)}{2}+1 : \tfrac{n(n+1)}{2}+mn\right)\\ vech(Q_{uu}) &= W_c^*\left(\tfrac{n(n+1)}{2}+mn+1 : \tfrac{(n+m)(n+m+1)}{2}\right)\end{aligned}$$

and $\varphi_c \coloneqq \bar{x}\otimes\bar{x}$ is the critic basis. In order to learn the ideal critic, the following approximator is established

$$\hat{\mathcal{Q}}(x,u_e) = \langle W_c, \varphi_c\rangle,$$

where W_c is approximator of W_c^*.

Similarly, the event-triggered controller $u_e(\cdot)$ in (9) can be equivalently expressed as

$$u_e = -Q_{uu}^{-1}Q_{ux}\hat{x} = \left(W_a^*\right)^T\varphi_a \tag{22}$$

where $W_a^* \in \mathbb{R}^{n\times m}$ is the ideal actor weight and $\varphi_a \coloneqq \hat{x}$ is the actor basis. Then, in order to approximate the event-triggered controller u_e, the following actor approximator is used

$$\hat{u}_e = \left(W_a\right)^T\varphi_a \tag{23}$$

where $\hat{W}_a$ are the approximated weights of W_a^*.

According to linear quadratic optimal control theory [12], the HJB equation can be written as

$$\min_u\left\{\frac{\partial V^*(x)}{\partial x}\left(Ax+Bu\right)+\frac{1}{2}\left(x^THx+u^TRu\right)\right\}=0$$

An equivalent formulation of the above HJB equation, named as integral Bellman equation [23], can be written in terms of the value function $V^*(x)$ as

$$\begin{aligned}&V^*\left(x\left(t-T\right)\right)-V^*\left(x\left(t\right)\right)\\&=\textstyle\int_{t-T}^{t}\frac{1}{2}\left(x^THx+\left(u^*\right)^TRu^*\right)d\tau\end{aligned}$$

By using the action-dependent formulation of Q-function in Section III-B.1, the above integral Bellman equation in terms of the $\mathcal{Q}(x,u_e)$ can be written as

$$\begin{aligned}&\mathcal{Q}\left(x\left(t\right),u_e^*\left(t\right)\right)=\mathcal{Q}\left(x\left(t-T\right),u_e^*\left(t-T\right)\right)\\&-\textstyle\int_{t-T}^{t}\frac{1}{2}\left(x^THx+\left(u_e^*\right)^TRu_e^*\right)d\tau\end{aligned}$$

Define the following critic error $e_c \in \mathbb{R}$ that we would like to eventually drive to zero by picking appropriately W_c,

$$\begin{aligned}&e_c \coloneqq \hat{\mathcal{Q}}\left(x\left(t\right),\hat{u}_e\left(t\right)\right)-\hat{\mathcal{Q}}\left(x\left(t-T\right),\hat{u}_e\left(t-T\right)\right)\\&+\textstyle\int_{t-T}^{t}\frac{1}{2}\left(x^THx+\left(\hat{u}_e\right)^TR\hat{u}_e\right)d\tau\\&=W_c^T\varphi_c\left(t\right)-W_c^T\varphi_c\left(t-T\right)\\&+\textstyle\int_{t-T}^{t}\frac{1}{2}\left(x^THx+\left(\hat{u}_e\right)^TR\hat{u}_e\right)d\tau\end{aligned}$$

Similarly, the actor error e_a can be defined as

$$\begin{aligned}e_a &:= \hat{u}_e\left(\hat{x}\right)-\left(-\hat{Q}_{uu}^{-1}\hat{Q}_{ux}\hat{x}\right)\\&=\left(W_a^T+\hat{Q}_{uu}^{-1}\hat{Q}_{ux}\right)x\left(t_k\right),\forall t\in\left[t_k,t_{k+1}\right)\end{aligned}$$

where $\hat{Q}_{ux}$ and $\hat{Q}_{uu}$ are extracted from the critic weight W_c. The squared-norm of these approximation errors, e_c and e_a, can be expressed as

$$E_c=\frac{1}{2}\|e_c\|^2, E_a=\frac{1}{2}\|e_a\|^2 \tag{24}$$

Based on the above formulations, after applying a gradient descent in (24), the update rule for critic and actor can be determined respectively as

$$\dot{W}_c=-\alpha_c\frac{1}{\left(1+\rho^T\rho\right)^2}\frac{\partial E_c}{\partial W_c}=-\alpha_c\frac{\rho}{\left(1+\rho^T\rho\right)^2}e_c^T \tag{25}$$

$$\begin{cases}\dot{W}_a &= 0, & \forall t\in\left[t_k,t_{k+1}\right)\\ W_a^+ &= W_a-\alpha_a\frac{1}{1+x^Tx}\frac{\partial E_a}{\partial W_a}\\ &= W_a-\alpha_a\frac{x}{1+x^Tx}e_a^T, & t=t_k\end{cases} \tag{26}$$

where $\rho\left(t\right)\coloneqq\varphi_c\left(t\right)-\varphi_c\left(t-T\right)$.

3) Impulsive System Formulation: Define the error for actor and critic weight as

$$\tilde{W}_c \coloneqq W_c^*-W_c, \tag{27}$$

$$\tilde{W}_a \coloneqq W_a^*-W_a, \tag{28}$$

In this subsection, impulsive system formulation [24] of augmented system of x, $\hat{x}$, $\tilde{W}_c$ and $\tilde{W}_a$ is employed for analysis. Considering the actor-critic parametrization in the previous subsection, then the closed systems dynamics in (10) can be rewritten as

$$\dot{x}(t)=Ax(t)+B\left(-Q_{uu}^{-1}Q_{ux}-\tilde{W}_a^T\right)\hat{x}(t)\ \forall t\in\mathbb{R}_0^+ \tag{29}$$

Combining the dynamics in (25), (26) and (29), one can obtain the augmented system with state $\chi \coloneqq \begin{bmatrix}x^T & \hat{x}^T & \tilde{W}_c^T & \tilde{W}_a^T\end{bmatrix}^T$ with the flow ($t\in\left[t_k,t_{k+1}\right)$) and jump ($t=t_{k+1}$) dynamics respectively as in (30), which is shown on top of next page.

4) Static Intermittent Q-Learning: The static intermittent Q-learning design developed in [21] can be summarized as follows.

Lemma 4. *[21]* (Static Intermittent Q-Learning) *Consider the system dynamics given by* (29)*, the Q-function critic approximator given by* (21) *and the actor approximator given by* (22)*. The tuning laws for the weights of the critic and the actor are given by* (25) *and* (26)*, respectively. Then, the origin of the closed-loop impulsive system with state χ for all initial conditions χ_0 is globally asymptotically stable as long as the sampling instants is determined by*

$$\begin{aligned}t_0&=0,\\t_{k+1}&=\inf_{t\in\mathbb{R}^+}\left\{t>t_k\wedge q\leqslant 0\right\}.\end{aligned} \tag{31}$$

with the event-triggering condition

$$q\leqslant 0 \tag{32}$$

$$\begin{aligned}q:=&\left(1-\beta^2\right)\underline{\lambda}\left(H\right)\|x\|^2+\underline{\lambda}\left(R\right)\|u_e\|^2\\&-4\left(L^2+L_1^2\right)\bar{\lambda}\left(R\right)\|e\|^2\end{aligned} \tag{33}$$

where L_1 is a positive constant of unity order, and the following inequalities hold:

$$\frac{\underline{\lambda}\left(H\right)}{\bar{\lambda}\left(R\right)}>\frac{2L_1^2}{\beta^2} \tag{34}$$

$$\alpha_c\gg\alpha_a, 0<\alpha_a<\frac{8\underline{\lambda}\left(R\right)-4}{\underline{\lambda}\left(R\right)+2} \tag{35}$$

□

According to Lemma 4, the parameters of the event-triggering condition (31) are time-invariant, and $q\geqslant 0$ has to be always satisfied. Therefore, in this paper, the triggering condition (31) is named as static intermittent Q-learning, in contrast to the dynamic intermittent Q-learning discussed in the next section.

$$\dot{\chi}=\left[\begin{array}{c} Ax+B\left(-Q_{uu}^{-1}Q_{ux}-\tilde{W}_a^T\right)\hat{x} \\ 0 \\ -\alpha_c\frac{\rho\rho^T}{(1+\rho^T\rho)^2}\tilde{W}_c \\ 0 \end{array}\right], \chi^+=\chi+\left[\begin{array}{c} 0_n \\ e \\ 0_{\frac{(n+m)(n+m+1)}{2}} \\ vec\left(-\alpha_a\frac{xx^T}{1+x^Tx}\tilde{W}_a-\alpha_a\frac{xx^T}{1+x^Tx}\tilde{Q}_{xu}Q_{uu}^{-1}\right) \end{array}\right]. \tag{30}$$

IV. Dynamic Intermittent Q-Learning

In this section, dynamic intermittent Q-learning algorithm is developed. It is shown that the presented dynamic intermittent Q-learning is Zeno-free and has larger inter-event interval compared to the static one. Moreover, the degree of sub-optimality of the dynamic intermittent Q-learning algorithm is discussed.

To formulate the dynamic intermittent feedback control, the following internal dynamical system is required [22]

$$\dot{\varsigma}=-\gamma\varsigma+q,\quad \varsigma(t_0)=\varsigma_0,\ t\in R_0^+ \tag{36}$$

where q is defined in (33) and $\gamma\in\mathbb{R}^+$ is a design parameter. The dynamic intermittent Q-learning, triggers an event when the following condition is satisfied

$$\varsigma(t)+\phi q(t)\leqslant 0, \tag{37}$$

where $\phi\in\mathbb{R}^+$ is a parameter to be designed later. The event-triggering instants sequence can be determined by (37) as

$$\begin{aligned} t_0&=0,\\ t_{k+1}&=\inf_{t\in R^+}\left\{(t>t_k)\wedge(\varsigma(t)+\phi q(t)\leqslant 0)\right\}. \end{aligned} \tag{38}$$

Comparing between (32) and (37), we note that the condition, $q\geqslant 0$, in the static model-free intermittent control can be relaxed to be $\varsigma+\gamma q\geqslant 0$ in the dynamic model-free intermittent control. Consequently, the dynamic event-triggered control can be determined as $u_d(\cdot)\coloneqq u_e(\cdot)$ with the event-triggered condition (37) and the event-triggering instants expressed as (38). The property of the dynamic event-triggered condition (37) can be presented in the following lemma.

Lemma 5. *Let γ be a positive constant, $\varsigma_0,\phi\in\mathbb{R}_0^+$, and q defined as in* (33)*. Then the following conclusions holds.*

1) $\varsigma(t)+\phi q(t)\geqslant 0,\ \forall t\in R_0^+$;
2) $\varsigma\geqslant 0,\ \forall t\in R_0^+$. □

Proof. The proof follows from that of [22, Lemma 2.2]. □

To this end, the dynamic model-free event-triggered co-design based on intermittent Q-learning can be formulated in the next theorem.

Theorem 1. (Dynamic Intermittent Q-Learning) *Consider the system dynamics given by* (30)*, the Q-function critic approximator given by* (21) *and the actor approximator given by* (22)*. Suppose that the signal $\frac{\rho}{1+\rho^T\rho}$ is persistently excited. The tuning laws for the weights of the critic and the actor are given by* (25) *and* (26)*, respectively, along with the dynamic event-triggering condition selected as in* (37)*. Then, the origin of the closed-loop system is globally asymptotically stable.* □

Proof. In order to show the asymptotic stability, the augmented system of (36) and impulsive system (30) and (30) is considered.

Consider the Lyapunov candidate $\mathcal{W}(\chi,\varsigma)=\mathcal{V}(\chi)+\varsigma$. where ς satisfies (36) and $\mathcal{V}(\chi)$ is defined as

$$\mathcal{V}(\chi)=\underbrace{V^*(x)}_{V_1(x)}+\underbrace{V^*(\hat{x})}_{V_2(\hat{x})}+\underbrace{\frac{1}{2}\left\|\tilde{W}_c\right\|^2}_{V_3(\tilde{W}_c)}+\underbrace{\frac{1}{2}tr\left(\tilde{W}_a^T\tilde{W}_a\right)}_{V_4(\tilde{W}_a)}. \tag{39}$$

According to the flow dynamics (30), one can obtain that $\hat{x}$ and $\tilde{W}_a$ are only updated at the event-triggering instants and remain constant during the flows. Therefore, $\dot{V}_2=\dot{V}_4=0$. Then, the time derivative of $\mathcal{V}(\chi)$ is

$$\begin{aligned} \dot{\mathcal{V}}(\chi)&=\dot{V}_1+\dot{V}_3\\ &=\frac{\partial V^*(x)}{\partial x}(Ax+B\hat{u}_d)-\alpha_c\tilde{W}_c^T\frac{\rho\rho^T}{(1+\rho^T\rho)^2}\tilde{W}_c\\ &=\underbrace{\frac{1}{2}(u^*)^TRu^*(x)-\frac{1}{2}x^THx-(u^*)^TR\hat{u}_d}_{\dot{V}_1}\\ &\quad\underbrace{-\alpha_c\tilde{W}_c^T\frac{\rho\rho^T}{(1+\rho^T\rho)^2}\tilde{W}_c}_{\dot{V}_3} \end{aligned} \tag{40}$$

where $(u^*)^TRu^*(x)-(u^*)^TR\hat{u}_d$ in above satisfies

$$\begin{aligned} &\frac{1}{2}(u^*)^TRu^*(x)-(u^*)^TR\hat{u}_d\\ &=\frac{1}{2}(u^*-\hat{u}_d)^TR(u^*-\hat{u}_d)-\frac{1}{2}\hat{u}_d^TR\hat{u}_d\\ &\leqslant\frac{1}{2}\bar{\lambda}(R)\left\|\tilde{W}_a^Tx-W_a^Te\right\|^2-\frac{1}{2}\underline{\lambda}(R)\|\hat{u}_d\|^2\\ &\leqslant 2\left(L^2+L_1^2\right)\bar{\lambda}(R)\|e\|^2+L_1^2\bar{\lambda}(R)\|x\|^2-\frac{1}{2}\underline{\lambda}(R)\|\hat{u}_d\|^2 \end{aligned} \tag{41}$$

where the second inequality results from the facts in (11) and (28). Since $\frac{\rho}{1+\rho^T\rho}$ is persistently excited, then, there exist a positive constant T_{pe} such that

$$\int_t^{t+T_{pe}}\frac{\rho(\tau)\rho^T(\tau)}{\left[1+\rho^T(\tau)\rho(\tau)\right]^2}d\tau\geqslant cI, \tag{42}$$

where $c\in\mathbb{R}^+$ is a positive constant. Suppose also that that there exists $\bar{M}\in\mathbb{R}^+$ such that

$$\max\left\{|M|,\left|\dot{M}\right|\right\}\leqslant\bar{M},\forall t\geqslant t_0 \tag{43}$$

where $M=\frac{\rho}{1+\rho^T\rho}$. Then, one can obtain

$$\dot{V}_3\leqslant-\alpha_c\lambda\left(MM^T\right)\|W_c\|^2. \tag{44}$$

Considering (41), then,

$$\begin{aligned}\dot{V}_1 &\leqslant 2\left(L^2+L_1^2\right)\bar{\lambda}(R)\|e\|^2+L_1^2\bar{\lambda}(R)\|x\|^2-\frac{1}{2}\underline{\lambda}(R)\|\hat{u}_d\|^2\\&\quad-\frac{1}{2}\underline{\lambda}(H)\|x\|^2\\&=-\frac{1}{2}q+L_1^2\bar{\lambda}(R)\|x\|^2-\frac{\beta^2}{2}\underline{\lambda}(H)\|x\|^2\\&\leqslant\left[L_1^2\bar{\lambda}(R)-\frac{\beta^2}{2}\underline{\lambda}(H)\right]\|x\|^2\end{aligned}\tag{45}$$

where q is defined in (33). Based on the fact in (34), $\dot{V}_1<0$ can be guaranteed. Therefore, for the flow dynamics of (30), the derivative of $\mathcal{W}(t)$ satisfies

$$\begin{aligned}&\dot{\mathcal{W}}=\dot{\mathcal{V}}+\dot{\varsigma}\\&=\underbrace{-\frac{1}{2}q+L_1^2\bar{\lambda}(R)\|x\|^2-\frac{\beta^2}{2}\underline{\lambda}(H)\|x\|^2}_{\dot{V}_1}\underbrace{-\gamma\varsigma+\frac{1}{2}q}_{\dot{\varsigma}}+\dot{V}_3\\&\leqslant-\alpha_c\lambda\left(MM^T\right)\|W_c\|^2+\left(L_1^2\bar{\lambda}(R)-\frac{\beta^2}{2}\underline{\lambda}(H)\|x\|^2\right)-\gamma\varsigma\\&<0\end{aligned}\tag{46}$$

Next, consider the jump dynamics given in (30). The difference of the common Lyapunov function (39) can be expressed as

$$\begin{aligned}\Delta\mathcal{V}(\chi)&=\underbrace{V^*\left(x^+\right)-V^*\left(x\left(t_k\right)\right)}_{\Delta V_1}+\underbrace{V^*\left(\hat{x}^+\right)-V^*\left(\hat{x}\left(t_k\right)\right)}_{\Delta V_2}\\&+\underbrace{V_3\left(\tilde{W}_c^+\right)-V_3\left(\tilde{W}_c\left(t_k\right)\right)}_{\Delta V_3}+\underbrace{V_4\left(\tilde{W}_a^+\right)-V_4\left(\tilde{W}_a\left(t_k\right)\right)}_{\Delta V_4}\end{aligned}\tag{47}$$

Note that time evolution of x and $\tilde{W}_c$ are both continuous with no jumps at event-triggering instants, it is evident that $\Delta V_1=\Delta V_3=0$. Based on the fact that $\hat{x}^+=\hat{x}\left(t_{k+1}\right)$, there exists a class $\mathcal{K}$ function $\kappa(\cdot)$ such that[1]

$$\begin{aligned}\Delta V_2&=V^*\left(\hat{x}\left(t_{k+1}\right)\right)-V^*\left(\hat{x}\left(t_k\right)\right)\\&\leqslant\kappa\left(\left\|\hat{x}\left(t_{k+1}\right)-\hat{x}\left(t_k\right)\right\|\right)\end{aligned}\tag{48}$$

holds uniformly for $\forall t_k$. Therefore, one can conclude that $\|\hat{x}(t_k)\|\to 0$, i.e., $\hat{x}(t_k)$ converges to the origin asymptotically. Note that ΔV_4 in (47) satisfies (49) by using using Youngs inequality, Cauchy-Schwarz inequality and the fact in (50) (see top of next page). Therefore, it can be shown that $\Delta V_4<0$ when $\tilde{W}_a$ lies outside the set $\Omega_{\tilde{W}_a}$, which is given in (51) with the actor learning rate α_a satisfying (34). From (49), the set $\Omega_{\tilde{W}_a}$ is forward-invariant. That is, when $\tilde{W}_a$ enters the set $\Omega_{\tilde{W}_a}$, it would stay inside $\Omega_{\tilde{W}_a}$ thereafter. Because the signals in (51) are asymptotically stable, then, the set $\Omega_{\tilde{W}_a}$ vanishes and becomes a single point [25]. Also, for the jump dynamics of the augmented system, note that the variable ς is continuously time-varying, then, the time difference equation of $\varsigma(t)$ is zero. Hence $\Delta\mathcal{W}(t)=\Delta\mathcal{V}(t)$ will converge to the origin asymptotically.

[1]Readers are refereed to [25] for details on class $\mathcal{K}$ functions.

Based on the above analysis, the asymptotic stability of the impulsive augmented system can be guaranteed. This completes the proof. □

Corollary 1. *Let $\{t_k^s\}_{k=1}^{\infty}$ and $\{t_k^d\}_{k=1}^{\infty}$ be the triggering time sequences determined by the static and dynamic intermittent Q-learning as designed in Lemma 4 and Theorem 1, respectively. Assume also that $t_k^s=t_k^d=t_k$ and after writing the next triggering instants by the static and the dynamic intermittent Q-learning as t_{k+1}^s and t_{k+1}^d respectively, one has $t_{k+1}^d\geqslant t_{k+1}^s$.* □

Proof. This will be shown by contradiction. Assume that $t_{k+1}^d<t_{k+1}^s$. Then, based on (38), one has that $q\left(t_{k+1}^d\right)>0$, i.e.,

$$\begin{aligned}&\left(1-\beta^2\right)(H)\left\|x\left(t_{k+1}^d\right)\right\|^2+(R)\left\|u^*\left(t_{k+1}^d\right)\right\|^2\\&>4\left(L^2+L_1^2\right)\bar{\lambda}(R)\left\|e\left(t_{k+1}^d\right)\right\|^2.\end{aligned}\tag{52}$$

Based on (38) and Lemma 5, one has $\varsigma\left(t_{k+1}^d\right)+\theta\mathcal{H}\left(t_{k+1}^d\right)\leqslant 0$, i.e.,

$$\begin{aligned}0&\geqslant\gamma\left(t_{k+1}^d\right)+\phi\Big[\left(1-\beta^2\right)\underline{\lambda}(H)\left\|x\left(t_{k+1}^d\right)\right\|^2\\&+\underline{\lambda}(R)\left\|u^*\left(t_{k+1}^d\right)\right\|^2-4\left(L^2+L_1^2\right)\bar{\lambda}(R)\left\|e\left(t_{k+1}^d\right)\right\|^2\Big]\\&\geqslant\phi\Big[\left(1-\beta^2\right)\underline{\lambda}(H)\left\|x\left(t_{k+1}^d\right)\right\|^2+\underline{\lambda}(R)\left\|u^*\left(t_{k+1}^d\right)\right\|^2\\&-4\left(L^2+L_1^2\right)\bar{\lambda}(R)\left\|e\left(t_{k+1}^d\right)\right\|^2\Big]\\&=\phi q\left(t_{k+1}^d\right).\end{aligned}\tag{53}$$

Note that $\phi\in\left(0,\frac{1}{s}\right]$ is a positive constant, and therefore, (53) yields $q\left(t_{k+1}^d\right)\leqslant 0$, which contradicts the assumption that $q\left(t_{k+1}^d\right)>0$. Therefore, $t_{k+1}^d\geqslant t_{k+1}^s$. □

Remark 2. From Corollary 1, it is shown that the next execution time given by a dynamic event-triggering mechanism is larger than the execution time for static event-triggering mechanism, when starting from the same initial state. Then, Zeno-free property of the dynamic model-free event-triggered co-design by Theorem 1 can be guaranteed. This is because it is shown in Lemma 4 that the static event-triggering mechanism excludes Zeno-behavior. □

The the degree of sub-optimality about the dynamic event-triggered Q-learning algorithm is discussed as follows.

Corollary 2. *Consider the dynamic model-free event-triggered co-design in Theorem 1. Then, the cost of $u_d(\cdot)$ is*

$$\begin{aligned}&J\left(u_d(\cdot);x_0\right)=J\left(u^*(\cdot);x_0\right)\\&+\int_{t_0}^{\infty}\left\|u_d\left(x(\tau)\right)-u^*\left(x(\tau)\right)\right\|_R d\tau.\end{aligned}\tag{54}$$

□

Proof. Applying now the intermittent control policy $u_s(\cdot)$ to

$$
\begin{aligned}
&\Delta V_4 = V_4\left(\tilde{W}_a^+\right) - V_4\left(\tilde{W}_a\left(t_k\right)\right)\\
&= \frac{1}{2\alpha_a} tr\left[-\alpha_a \tilde{W}_a^T \frac{x\left(t_k\right) x(t_k)^T}{1+x(t_k)^T x\left(t_k\right)} \tilde{W}_a\right] + \frac{1}{2\alpha_a} tr\left[-\alpha_a \tilde{W}_a^T \frac{x\left(t_k\right) x(t_k)^T}{1+x(t_k)^T x\left(t_k\right)} \tilde{Q}_{xu} Q_{uu}^{-1}\right]\\
&+ \frac{1}{2\alpha_a} tr\left[-\alpha_a \tilde{W}_a^T \frac{x\left(t_k\right) x(t_k)^T}{1+x(t_k)^T x\left(t_k\right)} \tilde{W}_a\right] + \frac{1}{2\alpha_a} tr\left[\alpha_a^2 \tilde{W}_a^T \frac{x\left(t_k\right) x(t_k)^T}{1+x(t_k)^T x\left(t_k\right)} \frac{x\left(t_k\right) x(t_k)^T}{1+x(t_k)^T x\left(t_k\right)} \tilde{W}_a\right]\\
&+ \frac{1}{2\alpha_a} tr\left[\alpha_a^2 \tilde{W}_a^T \frac{x\left(t_k\right) x(t_k)^T}{1+x(t_k)^T x\left(t_k\right)} \frac{x\left(t_k\right) x(t_k)^T}{1+x(t_k)^T x\left(t_k\right)} \tilde{Q}_{xu} Q_{uu}^{-1}\right] + \frac{1}{2\alpha_a} tr\left[-\alpha_a Q_{uu}^{-1} \tilde{Q}_{xu}^T \frac{x\left(t_k\right) x(t_k)^T}{1+x(t_k)^T x\left(t_k\right)} \tilde{W}_a\right]\\
&+ \frac{1}{2\alpha_a} tr\left[\alpha_a^2 Q_{uu}^{-1} \tilde{Q}_{xu}^T \frac{x\left(t_k\right) x(t_k)^T}{1+x(t_k)^T x\left(t_k\right)} \frac{x\left(t_k\right) x(t_k)^T}{1+x(t_k)^T x\left(t_k\right)} \tilde{W}_a\right]\\
&+ \frac{1}{2\alpha_a} tr\left[\alpha_a^2 Q_{uu}^{-1} \tilde{Q}_{xu}^T \frac{x\left(t_k\right) x(t_k)^T}{1+x(t_k)^T x\left(t_k\right)} \frac{x\left(t_k\right) x(t_k)^T}{1+x(t_k)^T x\left(t_k\right)} \tilde{Q}_{xu} Q_{uu}^{-1}\right]\\
&\leqslant -\left\|\tilde{W}_a^T x\left(t_k\right)\right\|^2 + \bar{\lambda}\left(Q_{uu}^{-1}\right)\left\|\tilde{W}_a^T x\left(t_k\right)\right\| \left\|\frac{x\left(t_k\right)}{1+x(t_k)^T x\left(t_k\right)}\right\| \left\|\tilde{Q}_{xu}\right\| + \frac{\alpha_a}{2}\left\|\tilde{W}_a^T x\left(t_k\right)\right\|^2 \left\|\frac{x\left(t_k\right) x(t_k)^T}{\left[1+x(t_k)^T x\left(t_k\right)\right]^2}\right\|\\
&+ \alpha_a \bar{\lambda}\left(Q_{uu}^{-1}\right)\left\|\tilde{W}_a^T x\left(t_k\right)\right\| \left\|\frac{x\left(t_k\right)}{1+x(t_k)^T x\left(t_k\right)}\right\| \left\|\frac{x\left(t_k\right) x(t_k)^T}{\left[1+x(t_k)^T x\left(t_k\right)\right]^2}\right\| \left\|\tilde{Q}_{xu}\right\| + \frac{\alpha_a \bar{\lambda}\left(Q_{uu}^{-1}\right)}{2}\left\|\tilde{Q}_{xu}\right\|^2 \left\|\frac{x\left(t_k\right) x(t_k)^T}{\left[1+x(t_k)^T x\left(t_k\right)\right]^2}\right\|^2\\
&\leqslant \left\|\tilde{W}_a^T x\left(t_k\right)\right\|^2 \left(1 - \frac{\bar{\lambda}\left(Q_{uu}^{-1}\right)}{4} - \frac{\alpha_a}{8} - \frac{\alpha_a \bar{\lambda}\left(Q_{uu}^{-1}\right)}{4}\right) + \frac{\bar{\lambda}\left(Q_{uu}^{-1}\right)}{4}\left\|\tilde{Q}_{xu}\right\|^2 + \frac{\alpha_a \bar{\lambda}\left(Q_{uu}^{-1}\right)}{4}\left\|\tilde{Q}_{xu}\right\|^2 + \frac{\alpha_a\left[\bar{\lambda}\left(Q_{uu}^{-1}\right)\right]^2}{2}. \qquad (49)
\end{aligned}
$$

$$
\left\|\frac{1}{1+x(t_k)^T x\left(t_k\right)}\right\| \leqslant 1, \left\|\frac{x\left(t_k\right) x(t_k)^T}{1+x(t_k)^T x\left(t_k\right)}\right\| \leqslant 1, \left\|\frac{x\left(t_k\right)}{1+x(t_k)^T x\left(t_k\right)}\right\| \leqslant \frac{1}{2}, \left\|\frac{x\left(t_k\right) x(t_k)^T}{\left[1+x(t_k)^T x\left(t_k\right)\right]^2}\right\| \leqslant \frac{1}{4}. \qquad (50)
$$

$$
\Omega_{\tilde{W}_a} = \left\{\tilde{W}_a \left| \left\|\tilde{W}_a\right\| \leqslant \sqrt{\frac{\frac{\bar{\lambda}\left(Q_{uu}^{-1}\right)^2}{4} + \frac{\alpha_a \bar{\lambda}\left(Q_{uu}^{-1}\right)}{4} + \frac{\alpha_a\left[\bar{\lambda}\left(Q_{uu}^{-1}\right)\right]^2}{2}}{1 - \frac{\bar{\lambda}\left(Q_{uu}^{-1}\right)}{4} - \frac{\alpha_a}{8} - \frac{\alpha_a \bar{\lambda}\left(Q_{uu}^{-1}\right)}{4}}} \left\|\tilde{Q}_{xu}^{-1}\right\| \right.\right\}. \qquad (51)
$$

the system (1), then (2) yields,

$$
\begin{aligned}
&J\left(u_d(\cdot); x_0\right) = \int_{t_0}^{\infty}\left[x^T(t) H x(t) + u_d^T(t) R u_d(t)\right] dt\\
&= V^*\left(x_0\right) + \int_{t_0}^{\infty}\left[x^T(t) H x(t) + u_d^T(t) R u_d(t)\right] dt\\
&+ \int_{t_0}^{\infty}\left[\frac{\partial V^*(x(t))}{\partial x(t)}\right]^T \left[A x + B u^*(\hat{x})\right] dt. \qquad (55)
\end{aligned}
$$

Using now Lemma 2 the proof completes. □

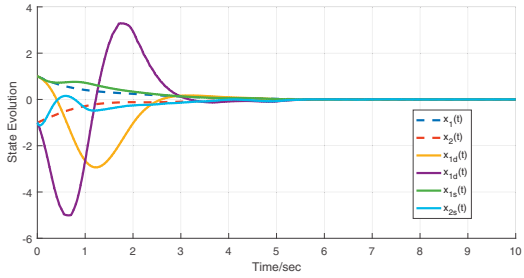

Fig. 1. The evolution of the state for continuous, static and dynamic intermittent feedback. The i-th component of state for continuous, static and dynamic cases are denoted as x_i, x_{is} and x_{id}, for $i = 1, 2$.

V. Simulation Study

In this section, the example in [1] is employed to show the effectiveness of the proposed model-free dynamic intermittent control policy. Consider the linear system $\dot{x} = Ax + Bu$, where $A = \begin{bmatrix} 0 & 1 \\ -2 & 3 \end{bmatrix}$ and $B = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$ are unknown matrices to the designer. The parameters for the utility function in (3) are selected as $H = 0.01 I_2$ and $R = 0.01$. Both the static and dynamic model-free co-design approach in this paper are used to develop the event-triggering condition and optimal feedback gain simultaneously. The static model-free triggering parameter in (33) is selected as $\beta = 0.5$, $L = 17$ and $L_1 = 2.7$. The length of interval for integral Q-learning algorithm is selected as $T = 0.05$. The learning rate for the critic and actor is $\alpha_c = 10$ and $\alpha_a = 0.001$, respectively.

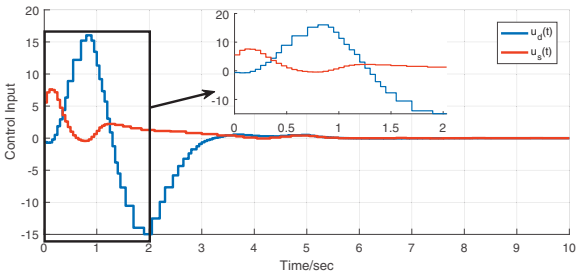

Fig. 2. The evolution of the control input. The control input for static and dynamic intermittent Q-learning are denoted as u_s and u_d, respectively.

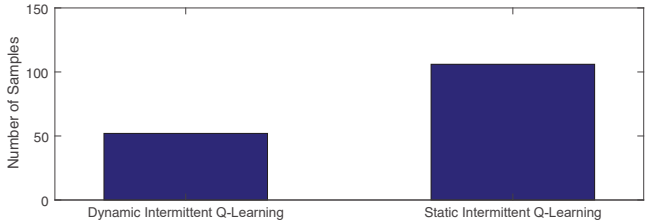

Fig. 3. Number of state samples used in static and dynamic intermittent Q-learning.

For the dynamic model-free triggering parameters, γ in (36) is selected as $\gamma = 1$ and $\phi = 0.1$. The results of continuous-triggered control, static and dynamic model-free triggered control are shown in Figures 1 – 3. From Figure 3, one can observe that the dynamic triggering approach can further decrease the number of triggering instants. Therefore, the dynamic intermittent Q-learning outperforms the static one in terms of communication bandwidth.

VI. Conclusions

This paper presents a Q-learning based dynamic intermittent feedback for continuous-time linear systems. In contrast to existing event-triggered designs, where complete knowledge of the system dynamics is required, the proposed method is able to obviate this requirement by using the intermittent Q-learning algorithm. The actor-critic approximator structure is employed to co-design the event-triggering condition and controller. The combined closed-loop system can be written as an impulsive system, which is proved to have an asymptotically stable equilibrium point without any Zeno behavior. A qualitative performance analysis of the dynamic Q-learning is given in comparison to the continuous optimal feedback and the degree of sub-optimality is established. A numerical simulation of an unknown unstable system is presented to show the efficacy of the proposed approach. Future work will be focused on extending the proposed dynamic intermittent Q-learning algorithm to a distributed synchronization problem of model-free multi-agent systems.

References

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

[2] J. Lunze and D. Lehmann, "A state-feedback approach to event-based control," *Automatica*, vol. 46, no. 1, pp. 211 – 215, 2010.

[3] W. P. M. H. Heemels, M. C. F. Donkers, and A. R. Teel, "Periodic event-triggered control for linear systems," *IEEE Transactions on Automatic Control*, vol. 58, no. 4, pp. 847–861, April 2013.

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

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

[6] D. Nesic and A. R. Teel, "Input-output stability properties of networked control systems," *IEEE Transactions on Automatic Control*, vol. 49, no. 10, pp. 1650–1667, Oct 2004.

[7] R. Postoyan, P. Tabuada, D. Nesic, and A. Anta, "A framework for the event-triggered stabilization of nonlinear systems," *IEEE Transactions on Automatic Control*, vol. 60, no. 4, pp. 982–996, April 2015.

[8] M. Abdelrahim, R. Postoyan, J. Daafouz, D. Nei, and M. Heemels, "Co-design of output feedback laws and event-triggering conditions for the l2-stabilization of linear systems," *Automatica*, vol. 87, pp. 337 – 344, 2018.

[9] D. Antunes, W. P. M. H. Heemels, and P. Tabuada, "Dynamic programming formulation of periodic event-triggered control: Performance guarantees and co-design," in *2012 IEEE 51st IEEE Conference on Decision and Control (CDC)*, Dec 2012, pp. 7212–7217.

[10] T. Gommans, D. Antunes, T. Donkers, P. Tabuada, and M. Heemels, "Self-triggered linear quadratic control," *Automatica*, vol. 50, no. 4, pp. 1279 – 1287, 2014.

[11] Y. Yang, D. Wunsch, and Y. Yin, "Hamiltonian-driven adaptive dynamic programming for continuous nonlinear dynamical systems," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 28, no. 8, pp. 1929–1940, Aug 2017.

[12] F. L. Lewis, D. Vrabie, and V. L. Syrmos, *Optimal control*. John Wiley & Sons, 2012.

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

[14] K. G. Vamvoudakis, A. Mojoodi, and H. Ferraz, "Event-triggered optimal tracking control of nonlinear systems," *International Journal of Robust and Nonlinear Control*, vol. 27, no. 4, pp. 598–619, 2017.

[15] L. Dong, X. Zhong, C. Sun, and H. He, "Event-triggered adaptive dynamic programming for continuous-time systems with control constraints," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 28, no. 8, pp. 1941–1952, Aug 2017.

[16] 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, October 2012.

[17] H. Modares, F. L. Lewis, and Z. P. Jiang, "H_∞ tracking control of completely unknown continuous-time systems via off-policy reinforcement learning," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 26, no. 10, pp. 2550–2562, Oct 2015.

[18] B. Kiumarsi, F. L. Lewis, and Z.-P. Jiang, "H_∞ control of linear discrete-time systems: Off-policy reinforcement learning," *Automatica*, vol. 78, no. Supplement C, pp. 144 – 152, April 2017.

[19] Y. Yang, H. Modares, K. G. Vamvoudakis, Y. Yin, and D. C. Wunsch, "Model-free event-triggered containment control of multi-agent systems," in *2018 American Control Conference (ACC)*, June 2018, to appear.

[20] K. G. Vamvoudakis, "Q-learning for continuous-time linear systems: A model-free infinite horizon optimal control approach," *Systems & Control Letters*, vol. 100, pp. 14 – 20, 2017.

[21] 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, 2018.

[22] A. Girard, "Dynamic triggering mechanisms for event-triggered control," *IEEE Transactions on Automatic Control*, vol. 60, no. 7, pp. 1992–1997, July 2015.

[23] F. L. Lewis, D. Vrabie, and K. G. Vamvoudakis, "Reinforcement learning and feedback control: Using natural decision methods to design optimal adaptive controllers," *IEEE Control Systems*, vol. 32, no. 6, pp. 76–105, Dec 2012.

[24] W. M. Haddad, V. Chellaboina, and S. G. Nersesov, *Impulsive and hybrid dynamical systems*. Princeton, NJ, USA: Princeton University Press, 2006.

[25] H. K. Khalil, *Nonlinear Systems*, 3rd ed. Prentice Hall, 2002.