Neurocomputing 377 (2020) 243–255

Contents lists available at ScienceDirect

Neurocomputing

journal homepage: www.elsevier.com/locate/neucom

Integral reinforcement learning-based online adaptive event-triggered control for non-zero-sum games of partially unknown nonlinear systems

Hanguang Su [b], Huaguang Zhang [a,b,*], Shaoxin Sun [b], Yuliang Cai [b]

[a] State Key Laboratory of Synthetical Automation for Process Industries, Northeastern University, Shenyang, Liaoning 110004, PR China
[b] School of Information Science and Engineering, Northeastern University, Shenyang, Liaoning 110004, PR China

ARTICLE INFO

Article history:
Received 30 June 2019
Revised 11 September 2019
Accepted 30 September 2019
Available online 17 October 2019

Communicated by Dr. Derui Ding

Keywords:
Event-triggered control (ETC)
Integral reinforcement learning (IRL)
Adaptive dynamic programming (ADP)
Adaptive critic design
Non-zero-sum (NZS) games

ABSTRACT

This paper develops an integral reinforcement learning (IRL)-based adaptive control method for the multi-player non-zero-sum (NZS) games of the nonlinear continuous-time systems with partially unknown dynamics, in the context of event-triggered mechanism. With the principle of IRL method, the requirement for the system drift dynamics is relaxed in the controller design. Moreover, different from the conventional iteration computation methods, the algorithm developed in this work is implemented in an online adaptive fashion, which provides a new way to combine the IRL algorithm and the event-triggered control framework in solving the NZS game issues. In the event-based algorithm, a state-dependent triggering condition is presented, which not only guarantees the closed-loop system stability, but also reduces the computation and communication loads of the controlled plant. By means of Lyapunov theorem, the uniform ultimate boundedness (UUB) properties of the system states and the critic weight estimation errors have been proved. Finally, two numerical examples are utilized to demonstrate the efficacy of the proposed method.

1. Introduction

In the conventional digital control systems, most of the controller devices are aroused by the periodical sampling signals to execute the computing and updating tasks [1,2]. Generally, a higher sampling frequency of the controlled plant means more information can be collected to design the control inputs and the corresponding control can work on the plants in time, which result in better control performance. However, in some specific applications, such as the networked control systems with geographically distributed sensors, controllers and actuators [3], the transmission bandwidths and computational resources are always constrained. In these instances, higher sampling rate may lead to congestions on the communication networks and more task delays will be caused. Therefore, the tradeoff between the control performance and the communication/computation loads has received intensive attentions by the researchers.

Event-triggered control (ETC) methods [4–6] are deemed as effective not only in reducing the system control loads but also in guaranteeing the achievement of the desired control objectives. Compared with the traditional time-triggered control systems, under the ETC mechanism, the control inputs are recomputed and updated only at the triggering instants, which are determined by some predefined triggering conditions. The core idea of the ETC schemes is that, by selecting proper triggering parameters, less computations and lower transmission frequency can be attained, in the meanwhile the stability of the controlled system is also guaranteed. That is, by applying the event-based control protocol, the system energy consumption is reduced without compromising the control performance, which has been shown qualitatively and quantitatively in the literature [7–9].

The differential game theories [10] have been applied in the fields including military [11,12], industrial manufacturing [13], business decision making [14,15], power system controlling [16,17], etc. In these issues, multiple players are involved to maximize the individual or team-based goals in a cooperative or noncooperative fashion. As for the two-player zero-sum differential game issue [18,19], one's gains are surely the other one's losses, and the decision strategies are made by the players independently of each other. But for the multi-player non-zero-sum (NZS) games [18,20], a balance between attaining the team goals cooperatively and

* Corresponding author at: State Key Laboratory of Synthetical Automation for Process Industries (Northeastern University), Shenyang, Liaoning, 110004, China; School of Information Science and Engineering, Northeastern University, Shenyang, Liaoning, 110004, PR China.
E-mail addresses: suhanguang@sina.com (H. Su), hgzhang@ieee.org (H. Zhang).

https://doi.org/10.1016/j.neucom.2019.09.088
0925-2312/

optimising the individual performance indexes competitively, is the exact objective to be captured. In these problems, the Nash equilibrium can be obtained by solving the coupled Hamilton–Jacobi (HJ) equations [21,22], which are reduced to coupled algebraic Riccati equations in the linear system cases. However, due to the inherent nonlinearity properties and the existence of the coupled terms, the analytic solutions to the coupled HJ equations are nearly impossible to be got.

In order to tackle the aforementioned problem, the adaptive dynamic programming (ADP) methods [23,24], developed from adaptive control designs [25,26] and reinforcement learning (RL) technologies [27], have been utilized to approximate the optimal solutions to the differential games [28–33]. Vamvoudakis and Lewis [21] proposed a novel online adaptive control method to solve the coupled HJ equations by resorting to the policy iteration (PI) algorithm. Then in [32], the authors established an identifier-actor-critic architecture to investigate the NZS games of the unknown nonlinear systems, where a neural network (NN)-based identifier was employed to reconstruct the system dynamics. By means of the Q-learning method [34], the NZS games for a class of deterministic continuous-time linear systems were studied in [31], even though the system dynamics are completely unknown.

As a variant of the RL algorithm, the integral RL (IRL) methods were proposed in [35,36]. In the classical PI-based ADP methods, the approximate optimal solutions of the Hamilton–Jacobi–Bellman (HJB) equations are desired. However, in the IRL schemes, by introducing the integral Bellman equations which involve none of the system dynamic knowledge, the requirement for the system drift dynamics can be relaxed in the controller design process [35]. Considering that in the practical control instances, the system dynamics is hard to be modelled or formulated, the IRL algorithms are more applicable and feasible in dealing with the optimal control problems. And the related studies can be found in [37–40].

Nowadays, the ADP algorithms and the event-triggered mechanism have been combined in addressing the optimal control issues [41,42], the optimal tracking control problems [39,43,44] and the zero-sum differential games [8,45]. In addition, for the systems with unknown states or dynamics, event-based methods have also been presented in [9,46,47], where the observers or identifiers were used to cope with these optimization issues. But to the best of the authors' knowledge, the NZS games under the event-triggered mechanism have not been investigated by the IRL-based algorithms up to now.

In this work, an adaptive control scheme for the NZS games of nonlinear systems subject to unknown drift dynamics is proposed in the context of event-triggered framework. The IRL algorithm is implemented in an online fashion, where the critic NN is built to approximate the optimal value functions corresponding to each of the players. A novel event-triggered condition is developed. In the meanwhile, the adaptive laws of the critic networks are properly designed, which can guarantee the convergence of the critic NN weights and the uniformly ultimately boundedness (UUB) of the closed-loop system state. The contributions of this work are three-fold:

1. For the first time, the NZS differential game issues for the nonlinear systems with partially unknown dynamics are addressed with the aid of the ADP scheme, under the event-triggered mechanism. As the IRL method is used, the requirement for the system drift dynamics is released. Moreover, in contract to the existing event-based ADP algorithms [8,46,47], where the NN-based observers or identifiers were utilized to rebuild the system dynamics, our method avoids introducing the model estimation errors.
2. Compared with the ADP-based methods [35,36], where the offline iteration procedure was applied to approximate the solutions to the optimization problems, the IRL algorithm proposed in this paper is implemented in an online fashion. It provides a new channel to combine the IRL scheme and the ETC mechanism, in consideration that the computing of the triggering conditions and the triggering of the events must be executed in an online control process.
3. In the traditional event-triggered ADP algorithms [45,47,48], the control inputs were involved in the triggering conditions, which means that the communication channels are needed to be established between the triggering derives and the controllers. However, in our design, the triggering condition is state-dependent. That is to say, no controller output signals are transmitted to the triggering determining devices, which will surely save more system computation and communication resources.

The rest of this paper is organized as follows. In Section 2, the unknown nonlinear system is formulated and the knowledge of the multi-player NZS games is introduced. The IRL-based online adaptive control method is proposed in Section 3. The UUB properties of the closed-loop system states and the convergence of the critic NN weight estimation errors are also proved. Two simulation examples are given in Section 4, thereby substantiating the effectiveness and applicability of the developed method. The conclusions are drawn in Section 5, where the future expectation is also presented.

Notations: In this work, $\mathbb{R}$ denotes the set that includes all the real numbers, $\mathbb{R}^n$ represents the n-dimensional *Euclidean* space, and herein we define $\mathbb{R}^{n\times m}$ as the set of all real matrices. $\mathbb{N}^+$ is the set that contains all the positive integers, including a subset of $\mathcal{N}=\{1,\ldots,N\}$. Throughout this paper, t stands for a specific time instant, and any function $f(x(t))$ with regard to t can be simply denoted as f, $f(x)$ or $f(t)$. Matrix $I_{n\times n}$ is the n-dimensional identity matrix and $0_{m\times n}$ is a zero-valued matrix with appropriate dimensions. $\lambda_{\max}(\cdot)$ is the maximal eigenvalue of a matrix, and $\lambda_{\min}(\cdot)$ is the minimal eigenvalue, correspondingly. The function $f(x)\in C^1(\psi)$ means that on the compact set ψ, $f(x)$ is first-order continuously differentiable. $\mathrm{tr}(\cdot)$ is the trace operation of matrices. Moreover, the left-limit operator is defined as $x(t^-)=\lim_{\Delta t\to 0^+}x(t-\Delta t)$.

2. Problem formulation and preliminaries

First of all, the NZS games of the nonlinear systems are formulated in this section, and the time-triggered coupled HJ equations are also derived.

2.1. Time-triggered NZS games

Consider the NZS differential games of the affine nonlinear system with N controllers, which is in the form of

$$\dot{x}=F(x)+\sum_{j=1}^{N}g_j(x)u_j(x) \tag{1}$$

where $x\in\Omega\subset\mathbb{R}^n$ is the system state with Ω a compact set that contains the origin. In the NZS games, the jth controller is also referred to as the jth player with the corresponding control policy $u_j\in\mathbb{R}^n$, $j\in\mathcal{N}$. The smooth function $F(x):\mathbb{R}^n\to\mathbb{R}^n$ and $g_j(x):\mathbb{R}^n\to\mathbb{R}^{n\times n}$ are the drift dynamics and the control input dynamics, respectively. Assume that the function $F(x)$ is unknown. The following assumptions [11,12,41,49,50] are needed throughout this work.

Assumption 1. System (1) is controllable, that is, there exists at least a control policy pair $\{u_1,\ldots,u_N\}$ can asymptotically stabilize system (1), and $x=0$ is the only equilibrium point on the compact set Ω. The system states are all detectable.

Assumption 2. $\forall j \in \mathcal{N}$, the functions $F(x)$ and $g_j(x)$ are all locally Lipschitz on the compact set Ω with $F(0) = 0$.

Assumption 3. For any $x \in \Omega$, the input coefficient matrix $g_j(x)$ has full rank and is norm-bounded, i.e., $\forall j \in \mathcal{N}$, the inverse matrix g_j^{-} always exists and there are positive constants $\underline{g}_j$ and $\bar{g}_j$ such that $0 < \underline{g}_j \le ||g_j(x)|| \le \bar{g}_j$.

With the initial state $x(0)$, $\forall i \in \mathcal{N}$, the infinite-horizon performance index corresponding to the ith player is denoted as

$$J_i(x(0)) = \int_0^\infty \left(x^T M_i x + \sum_{j=1}^N u_j^T R_{ij} u_j\right) ds \equiv \int_0^\infty \rho_i(x, u_1, \ldots, u_N) ds. \tag{2}$$

Here $\rho_i(x, u_1, \ldots, u_N)$ is the utility function with the parameter matrices $M_i \succ 0$ and $R_{ij} \succ 0$ for all $i, j \in \mathcal{N}$. Before proceeding, the definition of the admissible control policies should be provided:

Definition 1 [28,38]**.** For system (1), the control policy pair $u = \{u_1(x), \ldots, u_N(x)\}$ is named as an admissible control policy pair on the compact set Ω, if $\forall i \in \mathcal{N}$, the control policy $u_i(x)$ is continuous on Ω with $u_i(0) = 0$, u can stabilize system (1) and the performance index (2) is finite for any $x(0) \in \Omega$.

As for the NZS game of system (1), the objective is to find the Nash equilibrium point solution, which is defined by

Definition 2 [18,21,38]**.** Consider the addressed nonlinear system (1), an N-tuple admissible control policy pair $u^* = \{u_1^*, \ldots, u_N^*\}$ is said to build a Nash equilibrium point solution for the N-player NZS game of system (1), if $\forall i \in \mathcal{N}$, it holds that

$$J_i^* \equiv J_i(x, u_1^*, \ldots, u_i^*, \ldots, u_N^*) \le J_i(x, u_1^*, \ldots, u_i, \ldots, u_N^*).$$

For any given admissible control policy pair $u = \{u_1, \ldots, u_N\}$, the value function (also named as the cost function) is formulated as

$$V_i(x(t)) = \int_t^\infty \left(x^T M_i x + \sum_{j=1}^N u_j^T R_{ij} u_j\right) ds. \tag{3}$$

Assume that the value function (3) satisfies that $V_i \in C^1(\Omega)$, then $\forall i \in \mathcal{N}$, the differential equivalent form of (3) is converted into the following equation:

$$0 = \rho_i(x, u_1, \ldots, u_N) + (\nabla V_i)^T \left(F(x) + \sum_{j=1}^N g_j u_j\right), \ V_i(0) = 0 \tag{4}$$

where $\nabla V_i = \frac{\partial V_i}{\partial x}$. Here we define the Hamiltonian function as

$$H_i(x, u_1, \ldots, u_N) = \rho_i(x, u_1, \ldots, u_N) + (\nabla V_i)^T \left(F(x) + \sum_{j=1}^N g_j u_j\right). \tag{5}$$

And furthermore, the optimal value function is defined as

$$V_i^* = \min_{u_i} \left\{ \int_t^\infty \left(x^T M_i x + \sum_{j=1}^N u_j^T R_{ij} u_j\right) ds \right\}. \tag{6}$$

Due to the stationarity conditions at the equilibrium point, the optimal control policy of the ith player is obtained as

$$u_i^* = -\frac{1}{2} R_{ii}^{-1} g_i^T(x) \nabla V_i^*, \forall i \in \mathcal{N}. \tag{7}$$

Hence, the time-triggered coupled HJ equations can be got as

$$\begin{aligned} &H_i(x, u_1^*, \ldots, u_N^*, V_i^*) \\ &= x^T M_i x + \sum_{j=1}^N u_j^{*T} R_{ij} u_j^* + (\nabla V_i^*)^T \left(F(x) + \sum_{j=1}^N g_j(x) u_j^*\right) \\ &= x^T M_i x + \frac{1}{4} \sum_{j=1}^N (\nabla V_j^*)^T g_j(x) R_{jj}^{-1} R_{ij} R_{jj}^{-1} g_j^T(x) \nabla V_j^* \\ &\quad + (\nabla V_i^*)^T \left(F(x) - \frac{1}{2} \sum_{j=1}^N g_j(x) R_{jj}^{-1} g_j^T(x) \nabla V_j^*\right) \\ &= 0, \quad V_i^*(0) = 0. \end{aligned} \tag{8}$$

In (8), for the sake of simplification, the notation $(u_j^*)^T$ is rewritten as u_j^{*T}, and the similar terms are also denoted in the same way.

Before ending this section, an important lemma is presented herein, which provides the existence condition of the Nash equilibrium points:

Lemma 1. *Suppose that $\forall i \in \mathcal{N}$, there exists a positive definite function V_i^* that satisfies the coupled HJ equation (8), then the optimal control policies expressed by (7) can asymptotically stabilize system (1), and the control policy pair $\{u_1^*, \ldots, u_N^*\}$ is the Nash equilibrium solution to the NZS game of system (1).*

Lemma 1 has been given in the references of [18,19], and the proof can also be found therein. To avoid the repetition, the proof of Lemma 1 is omitted in this work.

As the drift dynamics $F(x)$ in the Eq. (8) is unknown, it is impossible to solve the coupled HJ equations directly. Therefore, in the subsequent analysis, the adaptive critic designs are employed to get the approximate optimal solutions in an online manner.

3. IRL-based adaptive ETC design for the NZS games

In this section, by integrating the IRL algorithm and the event-triggered framework, a new online adaptive control method is developed to solve the NZS game of system (1), where the critic network is established to approximate the optimal value function. Although the system drift dynamics is unknown, by using our method, the closed-loop system state is proved to be UUB with the help of Lyapunov's theorem.

3.1. Formulation of IRL algorithm

In the classical PI method, the complete system dynamics information is needed in the iteration process. Therefore, to solve the NZS game of nonlinear systems with unknown dynamics, a NN-based identifier was utilized in [32] to reconstruct the system dynamics. Developed from the RL technologies, the IRL algorithm possesses the advantage that the requirement for the knowledge of the system drift dynamics is relaxed in the analysis process. And by the virtue of IRL scheme, the Nash equilibrium solutions were approximated in [37], both in offline and online manners.

As the drift dynamics $F(x)$ in system (1) is unknown, the IRL algorithm is applied in analyzing the NZS game of the addressed system in this work. According to Bellman's principle of optimality [19], for any $T \in (0, t)$, the optimal value function V_i^* satisfies the following equation (also known as the integral Bellman equation):

$$V_i^*(x(t-T)) = V_i^*(x(t)) + \int_{t-T}^t (x^T M_i x + \sum_{j=1}^N u_j^{*T} R_{ij} u_j^*) ds, \ \forall i \in \mathcal{N}. \tag{9}$$

When any admissible control policy pair $u = \{u_1, \ldots, u_N\}$ is considered, the performance index functions satisfy that

$$V_i(x(t-T)) = V_i(x(t)) + \int_{t-T}^{t} (x^T M_i x + \sum_{j=1}^{N} u_j^T R_{ij} u_j) ds, \ \forall i \in \mathcal{N}. \tag{10}$$

And the corresponding Lyapunov equation is defined as

$$LE(V_i(x(t))) = V_i(x(t)) - V_i(x(t-T)) + \int_{t-T}^{t} \left(x^T M_i x + \sum_{j=1}^{N} u_j^T R_{ij} u_j \right) ds, \ \forall i \in \mathcal{N}. \tag{11}$$

This equation will be used in the following adaptive critic design process.

Remark 1. In the conventional PI algorithm, the Eqs. (7) and (8) are used in the iteration process to get the approximate Nash equilibrium point solutions. It has been proved that the IRL algorithm is equivalent to the PI algorithm in solving the coupled HJ equations, and the details can be found in [37]. However, in the IRL algorithm, the integral Bellman equation (9) is solved instead of the coupled HJ equation (8). It should be noticed that the Eq. (9) doesn't involve any explicit terms of the system dynamics. Only the input dynamic function $g_i(x)$ is needed in updating the control policy u_i with the help of Eq. (7). And it's the reason why the requirement for the drift dynamics is obviated in the implementation process of the IRL algorithm.

3.2. Optimal event-triggered controller design

To save the system computation and communication resources, the event-triggered mechanism is introduced in this work, which is characterized by a monotonically increasing sequence $\{t_l\}$. Here t_l represents the lth triggering instant with $l \in \mathbb{N}^+$ and satisfies that $0 = t_0 < t_1 < \cdots < t_l < \cdots$. A predefined triggering condition is employed to determine the triggering instants (that is, the exact instant when the gap between the real-time system state and the system sampled state exceeds a pre-set threshold). In this sampled-data system, the system state is sampled only at the triggering instants, and the sampled state stays the same during the inter-event period, which is formulated as

$$\check{x}_l(t) = x(t_l), \ t_l \le t < t_{l+1}. \tag{12}$$

Furthermore, the control input signals are recomputed only at the triggering instants and hold unchanged until the next triggering:

$$\check{u}_j(t) = u_j(\check{x}_l, t), \ t_l \le t < t_{l+1}. \tag{13}$$

Assuming that there exist no task delays during the ETC process, that is, the recomputation and transmission of the control input signals are executed immediately with no time delays.

To simplify the expression, in the following discussions, the event-triggered error is defined as

$$\pi_l(t) = x(t) - \check{x}_l(t), \ t_l \le t < t_{l+1}. \tag{14}$$

According to (7), when the event-triggered framework is utilized, the optimal control policy is converted into the following form:

$$\check{u}_i^*(t) = -\frac{1}{2} R_{ii}^{-1} g_i^T(\check{x}_l) \nabla \check{V}_i^*, \ t_l \le t < t_{l+1} \tag{15}$$

where $\nabla \check{V}_i^* = \frac{\partial V_i^*}{\partial x}\Big|_{t=t_l}$ with $i \in \mathcal{N}$. And in the meanwhile, the time-triggered HJ equations (8) are transformed into the event-triggered versions:

$$\begin{aligned} &H_i(x, \check{u}_1^*, \ldots, \check{u}_N^*, V_i^*) \\ &= x^T M_i x + \sum_{j=1}^{N} \check{u}_j^{*T} R_{ij} \check{u}_j^* + (\nabla V_i^*)^T \left(F(x) + \sum_{j=1}^{N} g_j(x) \check{u}_j^* \right) \\ &= x^T M_i x + \frac{1}{4} \sum_{j=1}^{N} (\nabla \check{V}_j^*)^T g_j(\check{x}_l) R_{jj}^{-1} R_{ij} R_{jj}^{-1} g_j^T(\check{x}_l) \nabla \check{V}_j^* \\ &\quad + (\nabla V_i^*)^T \left(F(x) - \frac{1}{2} \sum_{j=1}^{N} g_j(x) R_{jj}^{-1} g_j^T(\check{x}_l) \nabla \check{V}_j^* \right), t_l \le t \le t_{l+1}. \end{aligned} \tag{16}$$

Noting that compared with the time-triggered coupled HJ equations (8), as the event-triggered error $\pi_l(t)$ is introduced in (16), the event-triggered Hamiltonian function $H_i(x, \check{u}_1^*, \ldots, \check{u}_N^*, V_i^*)$ is not equal to 0.

Before proceeding on the analysis, the following assumption is necessary, which has been mentioned in the previous works of [8,45] and [48]:

Assumption 4. (Lipschitz continuous condition of the optimal control policies). All the optimal control policies u_i^* are locally Lipschitz with respect to the event-triggered error π_l. That is, $\forall i \in \mathcal{N}$, $l \in \mathbb{N}^+$ and $t_l \le t < t_{l+1}$, there is always a constant $k_i > 0$ satisfying that $||u_i^* - \check{u}_i^*||^2 = ||u_i^*(x) - u_i^*(\check{x}_l)||^2 \le k_i ||x - \check{x}_l||^2 = k_i ||\pi_l||^2$.

The following theorem shows that, when a proper triggering condition is applied, the event-triggered optimal control policies $\check{u}_i^*$ can asymptotically stabilize system (1).

Theorem 1. *Consider the addressed system (1), suppose that Assumptions 1–4 all hold. Assume that for all $i \in \mathcal{N}$, there exists a smooth function V_i^* satisfying Eq.(8) and u_i^* is formulated as (7). When the triggering condition*

$$||\pi_l|| \le \sqrt{\frac{\sigma \lambda_{\min}(M)}{2\, K \lambda_{\max}(\Xi)}} ||x|| \tag{17}$$

is applied, the event-triggered optimal control policies (15) can stabilize system (1) asymptotically. Noting that the definitions of the parameter matrices M, Ξ and K are provided in (20) and (21). In addition, the threshold adjusting parameter $\sigma \in [0, 1)$ is selected by the designer.

Proof. The Lyapunov function is selected as $\mathcal{L} = \sum_{i=1}^{N} V_i^*(x(t))$, here the definition of V_i^* has been given in the expression of Theorem 1.

When the event-based optimal control policies (15) are applied, the orbital derivative of $\mathcal{L}$ along the corresponding closed-loop system is

$$\dot{\mathcal{L}} = \sum_{i=1}^{N} (\nabla V_i^*)^T \times \dot{x} = \sum_{i=1}^{N} (\nabla V_i^*)^T F(x) + \sum_{i=1}^{N} (\nabla V_i^*)^T \sum_{j=1}^{N} g_j(x) \check{u}_j^*. \tag{18}$$

Recalling the coupled HJ equation (8), it holds that

$$\sum_{i=1}^{N} (\nabla V_i^*)^T F(x) = -\sum_{i=1}^{N} \left(x^T M_i x + \sum_{j=1}^{N} u_j^{*T} R_{ij} u_j^* + (\nabla V_i^*)^T \sum_{j=1}^{N} g_j(x) u_j^* \right). \tag{19}$$

Here we denote the augmented optimal control signal vector as $u^* = [u_1^{*T}, \ldots, u_N^{*T}]^T$ and the augmented control error vector as $\tilde{u}^* = [(\check{u}_1^* - u_1^*)^T, \ \ldots, \ (\check{u}_N^* - u_N^*)^T]^T$. By substituting (19) into (18), we have that

$$\dot{\mathcal{L}} = -\sum_{i=1}^{N} x^T M_i x - \sum_{i=1}^{N} \sum_{j=1}^{N} u_j^{*T} R_{ij} u_j^* - \sum_{i=1}^{N} (\nabla V_i^*)^T \sum_{j=1}^{N} g_j(x) \left(u_j^* - \check{u}_j^* \right)$$

$$
\begin{aligned}
&= -x^T Mx - u^{*T} Ru^* - 2u^{*T} Y\bar{u}^* \\
&\le -x^T Mx - u^{*T} Ru^* + u^{*T} Ru^* + \bar{u}^{*T} Y^T R^{-1} Y\bar{u}^* \\
&= -x^T Mx + \bar{u}^{*T} \Xi\bar{u}^* \qquad (20)
\end{aligned}
$$

where $M = \sum_{i=1}^N M_i$, $R = \mathrm{diag}\{\sum_{i=1}^N R_{i1}, \ldots, \sum_{i=1}^N R_{iN}\}$, $Y = \begin{bmatrix} R_{11}g_1^{-1}g_1 & R_{11}g_1^{-1}g_2 & \cdots & R_{11}g_1^{-1}g_N \\ \vdots & & & \vdots \\ R_{NN}g_N^{-1}g_1 & R_{NN}g_N^{-1}g_2 & \cdots & R_{NN}g_N^{-1}g_N \end{bmatrix}$ and $\Xi = Y^T R^{-1} Y$. It's obvious that the matrices M and R are both positive definite. And furthermore, as R^{-1} is positive definite, we can easily get that the maximal eigenvalue of Ξ is positive, considering that the matrix Y is not a zero matrix. That is to say, it holds that $\lambda_{\min}(M) > 0$ and $\lambda_{\max}(\Xi) > 0$. Based on this analysis, we have the following result

$$
\begin{aligned}
\dot{\mathcal{L}} &\le -(1 - \frac{1}{2}\sigma)x^T Mx - \frac{1}{2}\sigma\lambda_{\min}(M)||x||^2 + \lambda_{\max}(\Xi)\sum_{j=1}^N ||\breve{u}_j^* - u_j^*||^2 \\
&\le -(1 - \frac{1}{2}\sigma)x^T Mx - \frac{1}{2}\sigma\lambda_{\min}(M)||x||^2 + \lambda_{\max}(\Xi)K||\pi_l||^2 \qquad (21)
\end{aligned}
$$

where $K = \sum_{j=1}^N k_j$. When the triggering condition (17) is satisfied, it yields that $-\frac{1}{2}\sigma\lambda_{\min}(M)||x||^2 + \lambda_{\max}(\Xi)K||\pi_l||^2 \le 0$. Thus we have that $\dot{\mathcal{L}} \le -(1 - \frac{1}{2}\sigma)x^T Mx < 0$. According to Lyapunov's theorem, it indicates that when the closed-loop control policies (15) are utilized, system (1) is asymptotically stable.

This completes the proof. □

Remark 2. In the analysis of the derivative function $\dot{\mathcal{L}}$, it is the coupled term $2u^{*T} Y\bar{u}^*$ that is intractable to be dealt with, considering that u^* is unknown. By using Young's inequality, the amplification operation is introduced in line 3 of (20), that is, $-2u^{*T} Y\bar{u}^* \le u^{*T} Ru^* + \bar{u}^{*T} Y^T R^{-1} Y\bar{u}^*$. As a result, the unknown term $u^{*T} Ru^*$ is eliminated. And furthermore, by using this operation, the control input information is no more needed in the calculating process of the triggering conditions, and meanwhile the system stability can also be guaranteed. Therefore, in contrast to the existing event-based ADP references of [8,45,48], our method can save more system resources in terms of computation and communication.

3.3. Event-based adaptive critic design

In this section, motivated by the excellent works of [36,37], the IRL algorithm is applied in solving the coupled HJ equations, thus an online adaptive control method is proposed, under the event-triggered mechanism.

First of all, due to the universal approximation of the NNs, on the compact set Ω, the optimal value function V_i^* can be represented in the NN form:

$$
V_i^* = \eta_i^{*T}\varphi_i(x) + \varepsilon_i, \ \ i \in \mathcal{N} \qquad (22)
$$

where $\eta_i^* \in \mathbb{R}^{q_i}$ is the ideal weight vector, $\varphi_i(x) : \mathbb{R}^n \to \mathbb{R}^{q_i}$ is the NN activation function and $\varepsilon_i \in \mathbb{R}$ is the NN approximation error, respectively. As the ideal weight η_i^* is unknown, a critic NN is designed herein to approximate the optimal value function V_i^*:

$$
\hat{V}_i = \hat{\eta}_i^T\varphi_i(x) \qquad (23)
$$

and $\hat{\eta}_i \in \mathbb{R}^{q_i}$ is the estimated critic weight vector. Considering the above Eqs. (7) and (22), the time-based optimal control policies can be formulated as

$$
u_i^* = -\frac{1}{2}R_{ii}^{-1}g_i^T(x)\big((\nabla\varphi_i(x))^T\eta_i^* + \nabla\varepsilon_i\big), \ \ i \in \mathcal{N} \qquad (24)
$$

where $\nabla\varphi_i = \frac{\partial\varphi_i}{\partial x}$ and $\nabla\varepsilon_i = \frac{\partial\varepsilon_i}{\partial x}$. Correspondingly, the event-based optimal control policies (15) are expressed as

$$
\breve{u}_i^* = -\frac{1}{2}R_{ii}^{-1}g_i^T(\breve{x}_l)\big((\nabla\varphi_i(\breve{x}_l))^T\eta_i^* + \nabla\varepsilon_i(\breve{x}_l)\big), \ \ t_l \le t < t_{l+1}. \qquad (25)
$$

Then based on (23), the applied closed-loop control policies under the event-triggered mechanism are formulated as

$$
\breve{u}_i = -\frac{1}{2}R_{ii}^{-1}g_i^T(\breve{x}_l)\big(\nabla\varphi_i(\breve{x}_l)\big)^T\hat{\eta}_i(t_l), \ \ t_l \le t < t_{l+1}. \qquad (26)
$$

When the control policies (26) are utilized, the corresponding Lyapunov equations (11), i.e. the integral versions of the event-triggered Hamiltonian functions over the interval $[t-T, t]$, are referred to as the temporal difference (TD) errors [36] and are presented as

$$
\begin{aligned}
&LE(x, \breve{u}_1, \ldots, \breve{u}_N, \hat{V}_i) \\
&= \hat{V}_i(x(t)) - \hat{V}_i(x(t-T)) + \int_{t-T}^t \left(x^T M_i x + \sum_{j=1}^N \breve{u}_j^T R_{ij}\breve{u}_j\right)ds \\
&= \hat{\eta}_i^T\big(\varphi_i(x(t)) - \varphi_i(x(t-T))\big) + \Theta_i(x, \breve{u}_1, \ldots, \breve{u}_N) \\
&= \hat{\eta}_i^T\vartheta_i(x) + \Theta_i(x, \breve{u}_1, \ldots, \breve{u}_N) \\
&\equiv e_{c,i}, \ \ \forall i \in \mathcal{N} \qquad (27)
\end{aligned}
$$

where $\vartheta_i(x) \equiv \varphi_i(x(t)) - \varphi_i(x(t-T))$ and $\Theta_i(x, \breve{u}_1, \ldots, \breve{u}_N) \equiv \int_{t-T}^t (x^T M_i x + \sum_{j=1}^N \breve{u}_j^T R_{ij}\breve{u}_j)ds$. In addition, an auxiliary term is defined as $e_i \equiv \eta_i^{*T}\vartheta_i + \Theta_i$. Therefore, it holds that $e_{c,i} = e_i - \tilde{\eta}_i^T\vartheta_i$, where $\tilde{\eta}_i = \eta_i^* - \hat{\eta}_i$ is the critic weight estimation error.

The objective of the critic learning phase is to find the weight $\hat{\eta}_i$ to minimize the TD error $e_{c,i}$. Here we define the global error function $E = \sum_{i=1}^N E_i \equiv \frac{1}{2}\sum_{i=1}^N e_{c,i}^2$. By using the gradient descent algorithm [51,52], the tuning laws of the critic weight $\hat{\eta}_i$ are designed as

$$
\begin{aligned}
\dot{\hat{\eta}}_i &= -\alpha_i\frac{1}{(\vartheta_i^T\vartheta_i + 1)^2}\frac{\partial E}{\partial\hat{\eta}_i} = -\alpha_i\frac{1}{(\vartheta_i^T\vartheta_i + 1)^2}\frac{\partial E_i}{\partial\hat{\eta}_i} \\
&= -\alpha_i\frac{\vartheta_i e_{c,i}}{(\vartheta_i^T\vartheta_i + 1)^2} = -\frac{\alpha_i\vartheta_i e_i}{(\vartheta_i^T\vartheta_i + 1)^2} + \frac{\alpha_i\vartheta_i\vartheta_i^T\tilde{\eta}_i}{(\vartheta_i^T\vartheta_i + 1)^2} \qquad (28)
\end{aligned}
$$

with the positive learning rate of α_i. Here we define that $\underline{\vartheta}_i = \frac{\vartheta_i}{\vartheta_i^T\vartheta_i + 1} \in \mathbb{R}^{q_i}$.

As the time derivative of the ideal critic weight satisfies that $\dot{\eta}_i^* = 0$, it leads that

$$
\dot{\tilde{\eta}}_i = -\dot{\hat{\eta}}_i = \alpha_i\frac{\vartheta_i e_{c,i}}{(\vartheta_i^T\vartheta_i + 1)^2} = \frac{\alpha_i\vartheta_i e_i}{(\vartheta_i^T\vartheta_i + 1)^2} - \frac{\alpha_i\vartheta_i\vartheta_i^T\tilde{\eta}_i}{(\vartheta_i^T\vartheta_i + 1)^2}. \qquad (29)
$$

To proceed on, the following assumptions are needed, which have been mentioned by the existing works of [8,21,53]:

Assumption 5. Consider the controlled system (1), for any $i \in \mathcal{N}$ and $t \in [0, +\infty)$, the signal $\underline{\vartheta}_i$ is persistently excited, i.e., there exist positive constants $\zeta_{i,1}$, $\zeta_{i,2}$ and $T_1 \le t$, such that the following inequality

$$
\zeta_{i,1}I_{q_i\times q_i} \le \int_{t-T_1}^t \underline{\vartheta}_i\underline{\vartheta}_i^T ds \le \zeta_{i,2}I_{q_i\times q_i} \qquad (30)
$$

holds for any $i \in \mathcal{N}$.

Assumption 6. For the ith critic NN with $i \in \mathcal{N}$, the ideal weight vector η_i^*, the gradient of the activation function φ_i, the gradient of the NN approaching error ε_i, and the additional term e_i are all bounded. That is, $||\eta_i^*|| \le \bar{\eta}_i$, $||\nabla\varphi_i|| \le \bar{\varphi}_i$, $||\nabla\varepsilon_i|| \le \bar{\varepsilon}_i$ and $|e_i| \le \bar{e}_i$, here the constants $\bar{\eta}_i$, $\bar{\varphi}_i$, $\bar{\varepsilon}_i$ and $\bar{e}_i$ are all positive values.

Remark 3. Assumption 5 implies that $0 < \lambda_{\min}(\underline{\vartheta}_i\underline{\vartheta}_i^T)$. This condition is also known as the persistence of excitation condition (PE condition), which is usually employed in the adaptive control process to guarantee the convergence of the tuning parameters [21,22,31,42]. In general, this condition is satisfied by adding the probing noises into the system dynamics. This point will be further explained in the simulation studies.

As the event-triggered mechanism is utilized in this work, there exist flow dynamics and jump dynamics of the sampled-data system under this framework. Thus the stability analysis on the designed event-based closed-loop control system is made with the help of the impulsive dynamical system. When the augmented system state is set as $X = [x^T, \check{x}_l^T, \tilde{\eta}_1^T, \ldots, \tilde{\eta}_N^T]^T$, the corresponding dynamic model can be formulated as

$$\begin{cases} \dot{X} = \begin{bmatrix} F(x) + \sum_{j=1}^{N} g_j \breve{u}_j \\ \hline 0_{n\times 1} \\ \hline \alpha_1 \dfrac{\vartheta_1 e_{c,1}}{(\vartheta_1^T\vartheta_1+1)^2} \\ \vdots \\ \alpha_N \dfrac{\vartheta_N e_{c,N}}{(\vartheta_N^T\vartheta_N+1)^2} \end{bmatrix}, t \in [t_l, t_{l+1}) \\ X(t) = X(t^-) + \begin{bmatrix} 0_{n\times 1} \\ \hline x(t) - \check{x}_l \\ \hline 0_{q_1\times 1} \\ \vdots \\ 0_{q_N\times 1} \end{bmatrix}, t = t_{l+1} \end{cases} \tag{31}$$

3.4. Stability proof of the online ETC system

Inspired by the works of [31,51], the critic weight $\hat{\eta}_i$ is proved to be convergent to a small neighbourhood of the ideal weight vector in the following lemma:

Lemma 2. *$\forall i \in \mathcal{N}$, let the critic NN weight $\hat{\eta}_i$ be adjusted by the adaptive tuning laws of (28), and the initial weight vector is finite, i.e., $\hat{\eta}_i(0) \in \Omega_{\hat{\eta}_i}$ with $\Omega_{\hat{\eta}_i}$ a bounded compact set. Then there exists an instant $T_c > 0$ such that for all $t > T_c$, it holds that the critic weight estimation error $\tilde{\eta}_i$ is UUB, provided that Assumptions 5 and 6 are both satisfied.*

Proof. In this proof, we select the Lyapunov function as $\mathcal{F} = \sum_{i=1}^{N} \tilde{\eta}_i^T\tilde{\eta}_i$. Based on the system dynamics (31), the time derivative of $\tilde{\eta}_i$ is a flow dynamics. That is, throughout the adaptive control process, there exist no jumps in the value of the parameter $\tilde{\eta}_i$. Moreover, as the signal $\tilde{\eta}_i$ is continuous at the triggering instants, it leads that at $t = t_{l+1}$, the first difference of $\mathcal{F}$ is $\Delta\mathcal{F} = 0$. Therefore in the following discussions, we only focus on the inter-event interval.

When Assumption 5 is satisfied, due to Remark 3, during the inter-event period (i.e., $t_l \le t \le t_{l+1}$), it holds that $\mu_{i,l} \le \min_{t_l\le t<t_{l+1}} \{\lambda_{\min}(\underline{\vartheta}_i(t)\underline{\vartheta}_i^T(t))\}$, here $\mu_{i,l}$ is a positive constant. Then the orbital derivative of $\mathcal{F}$ is formulated as

$$\begin{aligned} \dot{\mathcal{F}} &= 2\sum_{i=1}^{N} \tilde{\eta}_i^T\dot{\tilde{\eta}}_i = 2\sum_{i=1}^{N}\left(\frac{\alpha_i\tilde{\eta}_i^T\vartheta_i e_i}{(\vartheta_i^T\vartheta_i+1)^2} - \frac{\alpha_i\tilde{\eta}_i^T\vartheta_i\vartheta_i^T\tilde{\eta}_i}{(\vartheta_i^T\vartheta_i+1)^2}\right) \\ &\le \sum_{i=1}^{N}\alpha_i\frac{e_i^2 - \tilde{\eta}_i^T\vartheta_i\vartheta_i^T\tilde{\eta}_i}{(\vartheta_i^T\vartheta_i+1)^2} \le -\sum_{i=1}^{N}\alpha_i\left(\tilde{\eta}_i^T\underline{\vartheta}_i\underline{\vartheta}_i^T\tilde{\eta}_i - \bar{e}_i^2\right) \\ &\le -\sum_{i=1}^{N}\alpha_i\lambda_{\min}(\underline{\vartheta}_i\underline{\vartheta}_i^T)||\tilde{\eta}_i||^2 + \Psi \\ &\le -\sum_{i=1}^{N}\alpha_i\mu_{i,l}||\tilde{\eta}_i||^2 + \Psi \end{aligned} \tag{32}$$

where $\Psi = \sum_{i=1}^{N}\alpha_i\bar{e}_i^2$. Hence, when the weight error is on the outside of the bounded set $\{\tilde{\eta}_i | \ \| \tilde{\eta}_i \| \le \sqrt{\frac{\Psi}{\alpha_i\mu_{i,l}}} \equiv \xi_{\tilde{\eta}_i}\}$, it leads that $\dot{\mathcal{F}} < 0$. Then according to the Lyapunov extension theorem, as the critic NN weight vector is initialized in a bounded set, there exists a positive constant T_c, such that as $t > t_l \ge T_c$, the weight error $\tilde{\eta}_i$ is convergent into a small neighbourhood around the origin, which is with the upper bound of $\xi_{\tilde{\eta}_i}$.

Thus the lemma is proved. □

Here comes the closed-loop stability proof of the controlled system with the proposed IRL-based ETC method:

Theorem 2. *Consider the nonlinear system (1) to be addressed, suppose that Assumptions 1–6 all hold. Assume that for all $i \in \mathcal{N}$, there exists a smooth function V_i^* satisfying Eq. (8) and u_i^* is formulated as (7). The augmented impulsive dynamical system is expressed by (31), with the critic NN weights adjusted by the adaptive laws of (28). When the triggering condition*

$$||\pi_l|| \le \sqrt{\frac{\sigma\lambda_{\min}(M)}{2\ K\lambda_{\max}(\Xi)}}||x|| \equiv Z_T(t), \quad t_l \le t < t_{l+1} \tag{33}$$

is utilized, the system state x, the sampled state $\check{x}_l$, and the critic weight estimation errors $\tilde{\eta}_i$ are all UUB.

Proof. The candidate Lyapunov function is selected as

$$W = \nu_1 W_1 + \nu_2 W_2 + \nu_3 W_3 \tag{34}$$

where the function $W_1 = \mathcal{L} = \sum_{i=1}^{N} V_i^*(x)$, $W_2 = \mathcal{F} = \sum_{i=1}^{N}\tilde{\eta}_i^T\tilde{\eta}_i$ has been utilized in Lemma 2, and $W_3 = \sum_{i=1}^{N} V_i^*(\check{x}_l) + \sum_{i=1}^{N}\tilde{\eta}_i^T(t_l)\tilde{\eta}_i(t_l)$ with $t_l \le t < t_{l+1}$. In (34), the parameters ν_1, ν_2 and ν_3 are all adjustable positive constants.

Consider the impulsive system dynamics (31), the analysis is implemented over the following two cases:

Case 1. Between two consecutive triggering instants, i.e., $t \in [t_l, t_{l+1})$. The time derivative of W_1 is formulated as

$$\dot{W}_1 = \sum_{i=1}^{N}(\nabla V_i^*)^T F(x) + \sum_{i=1}^{N}(\nabla V_i^*)^T\sum_{j=1}^{N} g_j(x)\breve{u}_j. \tag{35}$$

Similar to the formula (20), when the practical event-based control policies (26) are employed, it holds that

$$\dot{W}_1 \le -x^T M x + \breve{u}^T\Xi\breve{u} \tag{36}$$

where $\breve{u} = [(\breve{u}_1 - u_1^*)^T, \ldots, (\breve{u}_N - u_N^*)^T]^T$ and the definitions of M and Ξ have been provided by (20). Furthermore, according to Lemma 2, when $t > T_c$ holds, (36) can be rewritten as

$$\begin{aligned} \dot{W}_1 \le & -x^T M x + 2\lambda_{\max}(\Xi)\sum_{j=1}^{N}\left|\left|u_j^* - \breve{u}_j^*\right|\right|^2 \\ & + 2\lambda_{\max}(\Xi)\sum_{j=1}^{N}\left|\left|\breve{u}_j^* - \breve{u}_j\right|\right|^2 \\ \le & -x^T M x + 2\lambda_{\max}(\Xi)K||\pi_l||^2 + 2\lambda_{\max}(\Xi)\sum_{j=1}^{N}\Big|\Big| - \frac{1}{2}R_{jj}^{-1}g_j(\check{x}_l) \\ & \times\left((\nabla\varphi_j(\check{x}_l))^T\tilde{\eta}_j(t_l) + \nabla\varepsilon_j(\check{x}_l)\right)\Big|\Big|^2 \end{aligned}$$

$$\leq -x^T Mx + 2\lambda_{\max}(\Xi)K\left|\left|\pi_l\right|\right|^2 + \lambda_{\max}(\Xi)\sum_{j=1}^{N}\left|\left|R_{jj}^{-1}\right|\right|^2 \bar{g}_j^2\bar{\varphi}_j^2\xi_{\tilde{\eta}_j}^2 + \lambda_{\max}(\Xi)\sum_{j=1}^{N}\left|\left|R_{jj}^{-1}\right|\right|^2 \bar{g}_j^2\bar{\varepsilon}_j^2. \tag{37}$$

Then recalling (32), it yields that

$$\dot{W}_2 \leq -\sum_{i=1}^{N}\alpha_i\mu_{i,l}||\tilde{\eta}_i||^2 + \Psi. \tag{38}$$

What's more, as the function $W_3(t)$ is unchanged during the interval $[t_l, t_{l+1})$, the time derivative of W_3 satisfies that $\dot{W}_3 = 0$ for $t \in [t_l, t_{l+1})$. Combining (37) and (38), the time derivative of W is reduced as

$$\begin{aligned}\dot{W} \leq & \nu_1\left(-x^T Mx + 2\lambda_{\max}(\Xi)K||\pi_l||^2 + \lambda_{\max}(\Xi)\sum_{j=1}^{N}||R_{jj}^{-1}||^2\bar{g}_j^2\bar{\varphi}_j^2\xi_{\tilde{\eta}_j}^2 + \lambda_{\max}(\Xi)\sum_{j=1}^{N}||R_{jj}^{-1}||^2\bar{g}_j^2\bar{\varepsilon}_j^2\right) - \nu_2\left(\sum_{i=1}^{N}\alpha_i\mu_{i,l}||\tilde{\eta}_i||^2 - \Psi\right)\\ \leq & -\nu_1(1-\sigma)\lambda_{\min}(M)||x||^2 - \nu_2 - \sum_{i=1}^{N}\alpha_i\mu_{i,l}||\tilde{\eta}_i||^2 + \nu_1\left(-\sigma\lambda_{\min}(M)||x||^2 + 2\lambda_{\max}(\Xi)K||\pi_l||^2\right) + \Gamma\end{aligned} \tag{39}$$

where $\Gamma = \nu_1\left(\lambda_{\max}(\Xi)\sum_{j=1}^{N}||R_{jj}^{-1}||^2\bar{g}_j^2\bar{\varphi}_j^2\xi_{\tilde{\eta}_j}^2 + \lambda_{\max}(\Xi)\sum_{j=1}^{N}||R_{jj}^{-1}||^2\bar{g}_j^2\bar{\varepsilon}_j^2\right) + \nu_2\Psi > 0$.

As the condition (33) holds, it leads that $-\sigma\lambda_{\min}(M)||x||^2 + 2\lambda_{\max}(\Xi)K||\pi_l||^2 \leq 0$. When at least one of the following conditions is satisfied:

$$||x|| > \sqrt{\frac{\Gamma}{\nu_1(1-\sigma)\lambda_{\min}(M)}} \equiv \varpi_x; \tag{40}$$

$$\|\tilde{\eta}_i\| > \sqrt{\frac{\Gamma}{\nu_2\alpha_i\mu_{i,l}}} \equiv \varpi_{\tilde{\eta}_i} \tag{41}$$

it holds that $\dot{W} < 0$. That is, x and $\tilde{\eta}_i$ are UUB.

Case 2. Then the stability of the controlled system at the triggering instants is analyzed, i.e., $t = t_{l+1}$. As the functions W_1 and W_2 are all time-continuous over the interval $[0, +\infty)$, the first differences of the first two terms in (34) are $\Delta W_1 = \Delta W_2 = 0$. And the first difference of W_3 is formulated as

$$\begin{aligned}\Delta W_3 = & \sum_{i=1}^{N}V_i^*(\check{x}(t_{l+1})) - \sum_{i=1}^{N}V_i^*(\check{x}(t_{l+1}^-)) + \sum_{i=1}^{N}\tilde{\eta}_i^T(t_{l+1})\tilde{\eta}_i(t_{l+1}) - \sum_{i=1}^{N}\tilde{\eta}_i^T(t_l)\tilde{\eta}_i(t_l)\\ = & \left(W_1(t_{l+1}) + W_2(t_{l+1})\right) - \left(W_1(t_l) + W_2(t_l)\right).\end{aligned} \tag{42}$$

According to Case 1, and considering that $W_1(t) + W_2(t)$ is continuous at $t = t_{l+1}$, when the conditions (33), (40) and (41) are satisfied, the function $W_1(t) + W_2(t)$ is non-increasing when $t \in [t_l, t_{l+1}]$. As a result, $W_1(t_{l+1}) + W_2(t_{l+1}) \leq W_1(t_l) + W_2(t_l)$ holds. That is, $\Delta W_3 \leq 0$, and furthermore, it yields that $\Delta W = \nu_3\Delta W_3 \leq 0$ at $t = t_{l+1}$. In other words, the Lyapunov function (34) is still non-increasing at the triggering instants.

Thus the theorem is proved. □

Remark 4. Theorem 2 indicates that, when the proposed event-triggered control policies (26) are employed, the closed-loop stability of the addressed system is guaranteed. The system state x and the critic weight estimation errors $\tilde{\eta}_i$ are all UUB with the corresponding upper bounds of ϖ_x and $\varpi_{\tilde{\eta}_i}$, which are with respect to the parameters σ and α_i. It should be noticed that, by selecting more proper parameters of σ and α_i, smaller ϖ_x and $\varpi_{\tilde{\eta}_i}$ can be obtained.

Based on the above analysis, the implementation procedure of the proposed IRL-based ETC algorithm is presented in Algorithm 1:

Algorithm 1 The IRL-based online adaptive ETC algorithm.

Initialization: For any $i, j \in \mathcal{N}$, selecting the cost function matrixes of M_i, R_{ij};the integral time interval, T;the maximum adaptive control time, $T_{\max}$;the threshold adjusting parameter, σ;the sum of the Lipschitz constants, K;the initial state vector, $x(0)$;the initial critic weight, $\hat{\eta}_i(0)$;the activation functions of the critic NNs, $\varphi_i(x)$;the learning rates of the critic weights, α_i.

1: Set the initial event triggering index as $l = 0$, and the sampled state as $\check{x}_l = x(0)$;
2: Compute the event-based control policy $\check{u}_j$ with (26) and transmit it to the corresponding actuator;
3: If $t_l \geq T_{\max}$, go to Step 6; if else, go on;
4: Adjusting the critic weight $\hat{\eta}_i$ with (28);
5: With the aid of the smart sensors, monitor the real-time system state vector $x(t)$, and compute the triggering condition (33) in the event generator device. If the triggering condition is violated, reset $l \leftarrow l+1$, $t_l \leftarrow t$ and $\check{x}_l \leftarrow x(t)$, then go to Step 2; if not, go to Step 3;
6: **Return** $\hat{\eta}_i$.

Remark 5. Noting that the initial critic weight $\hat{\eta}_i(0)$ should be carefully selected to get an initial admissible control policy pair $\{\hat{u}_1(0), \ldots, \hat{u}_N(0)\}$, which is required in the IRL algorithm [35,36,38]. However, for given nonlinear systems, how to find the admissible control policies is still an open problem. In the following simulation studies, a great deal of exploration experiments have been conducted to find the proper initial weight of $\hat{\eta}_i(0)$ for any $i \in \mathcal{N}$.

4. Simulation studies

Two numerical examples are conducted in this section to demonstrate the theoretical analysis provided above.

Example 1. Consider a two-order nonlinear system [17] with two controllers. The system dynamics is formulated as

$$\begin{aligned}\dot{x} = & \begin{bmatrix}-x_1^3 - 2x_2\\ x_1 + 0.5\cos(x_1^2)\sin(x_2^3)\end{bmatrix} + \begin{bmatrix}0.5 - 0.25\sin(x_1) & 0\\ 0 & -1 + 0.5\sin(x_2)\end{bmatrix}u_1\\ & + \begin{bmatrix}0.45 - 0.15\sin(x_1) & 0\\ 0 & -0.45 + 0.3\sin(x_2)\end{bmatrix}u_2\end{aligned} \tag{43}$$

where $x = [x_1, x_2]^T \in \mathbb{R}^2$ is the system state, with $u_1 \in \mathbb{R}^2$ and $u_2 \in \mathbb{R}^2$ are the outputs of two controllers (i.e., the players). As for the NZS game of system (43), the parameter matrices in the corresponding performance index are selected as $M_1 = 2I_{2\times2}$, $M_2 = I_{2\times2}$, $R_{11} = 2I_{2\times2}$, $R_{12} = R_{21} = I_{2\times2}$ and $R_{22} = 3I_{2\times2}$.

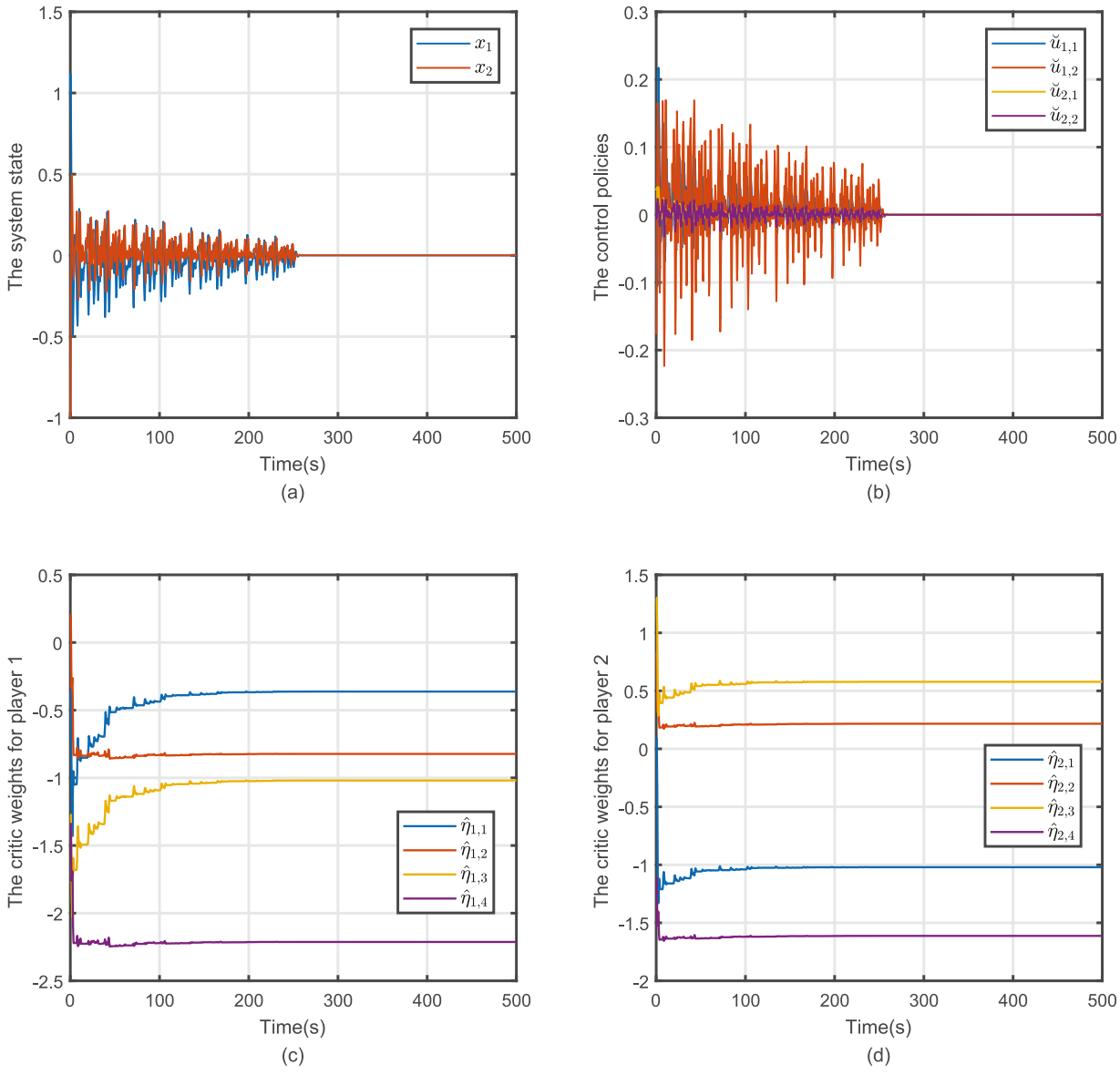

Fig. 1. The evolution of (a) the system state x; (b) the players' policies $\breve{u}_1$ and $\breve{u}_2$; (c) the critic weight $\hat{\eta}_1$; (d) the critic weight $\hat{\eta}_2$.

In the critic learning phase, the integral time interval is set as $T = 1$ s, while the sampling period is selected as 0.01 s. To guarantee that Assumption 6 is satisfied, the activation functions in the critic NNs are selected as $\varphi_i = [\cos(x_1), \cos(x_2), \mathrm{sech}(x_1), \mathrm{sech}(x_2)]^{\mathrm{T}}$ with $i = 1, 2$. To get an initial admissible control policy pair $\{\hat{u}_1(0), \ldots, \hat{u}_N(0)\}$, numerous of exploration experiments have been conducted. And the critic NN weights are initialized as $\hat{\eta}_1(0) = [-0.3587,\ 0.2035,\ -1.2724,\ -1.3381]^T$ and $\hat{\eta}_2(0) = [0.0924,\ 0.8375,\ 1.3020,\ -1.1103]^T$. The learning rates are picked as $\alpha_1 = \alpha_2 = 10$.

Then the triggering-related parameters are given as $\sigma = 0.5$ and $K = 2$. With the initial state of $x(0) = [1, -1]^T$, the IRL-based online adaptive control method is applied on the addressed system for 500 s, during which process the probing noises $b_{e1} = [0.1e^{-0.005t}\big(\sin^2(20t)\cos(20t) + \sin^2(12t)\cos(10t) + \sin^2(28t)\cos(25t) + \sin^5(20t) + \sin^2(12t) + \sin^3(2.5t)\cos(3.5t)\big),\quad 0.2e^{-0.005t}\big(\sin^2(5t)\cos(30t) + \sin^2(1.2t)\cos(t) + \sin^2(1.8t)\cos(2.5t) + \sin^5(2t) + \sin^2(11.2t) + \sin^3(2.4t)\cos(4t)\big)]^T$ are added to the system dynamics to guarantee the PE condition. And the noises are deleted at $t = 250$ s. The corresponding results are presented in Figs. 1 and 2.

From Fig. 1, we can find that the system state x converges to 0 sooner after the noises are removed, which proves that our designed method is effective. The control inputs $\breve{u}_i$ are shown as piecewise continuous signals. The evolution trajectories of the critic weights $\hat{\eta}_1$ and $\hat{\eta}_2$ are given in Fig. 1(c) and (d), which are both convergent. And the final weights are $\hat{\eta}_1 = [-0.3626,\ -0.8233,\ -1.0195,\ -2.2128]^T$, $\hat{\eta}_2 = [-1.0206,\ 0.2163,\ 0.5778,\ -1.6126]^T$. Then the two players' event-based control policies can be formulated by (26).

The triggering results are shown in Fig. 2. The relationship between the event-triggered errors and the triggering thresholds is shown in Fig. 2(a), from which we can find that when the norm of the error π_l exceeds the threshold, the system state is sampled and the error is reset to 0 immediately. The TD errors throughout the control process are presented in Fig. 2(b), which are both convergent to 0. It indicates that the objective of the critic

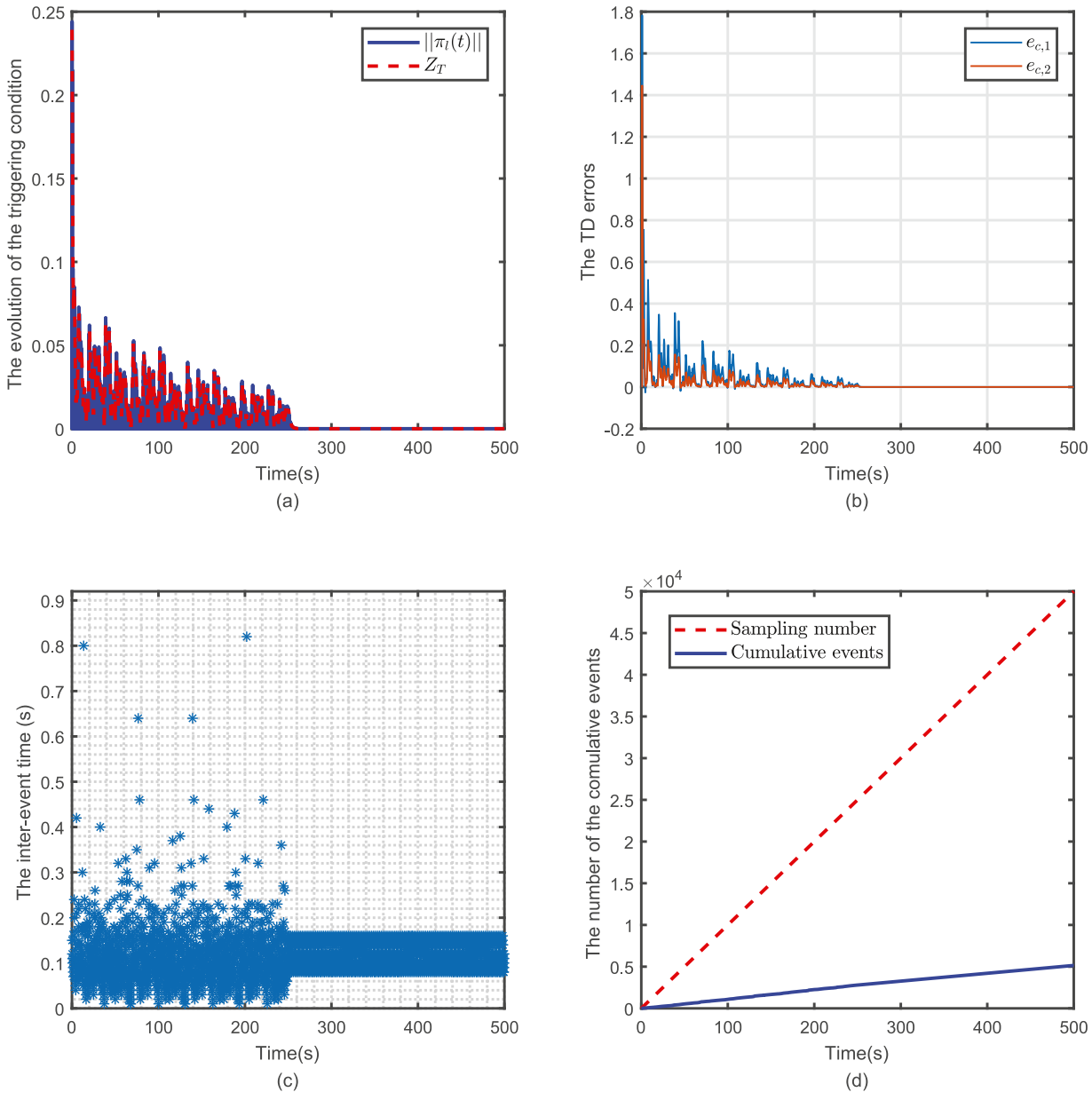

Fig. 2. The evolution of (a) the triggering condition in the adaptive control process; (b) the TD errors; (c) the inter-event times; (d) the cumulative number of the events.

learning has been achieved and the approximate optimal control policies have been attained by our method. Then Fig. 2(c) provides the evolution of the time periods between two consecutive triggering instants. It's emphasized that after $t = 250$ s, the inter-event intervals are all shorter than 0.08 s, which is much larger than the sampling period. That is, a great deal of unnecessary samplings are avoided by using our method and the Zeno behaviour is also eliminated. What's more, Fig. 2(d) indicates that throughout the adaptive control process, 50000 time-based samplings are needed, but the proposed event-based method only needs 5147 events, which means our method reduces the communication and computation burdens significantly.

Example 2. In this example, the proposed event-based control method is tested on the nonlinear system that has been provided in [46]. To verify the effectiveness of our method, five players are considered:

$$\dot{x} = f(x) + g_1(x)u_1 + g_2(x)u_2 + g_3(x)u_3 + g_4(x)u_4 + g_5(x)u_5 \quad (44)$$

where

$$f(x) = \begin{bmatrix} x_2 \\ -4.905\sin(x_1) - 0.5x_2 \end{bmatrix}, g_1 = \begin{bmatrix} 1 & 0 \\ 0 & -0.5 \end{bmatrix},$$
$$g_2 = \begin{bmatrix} 2.5 & 0 \\ 0 & -1 \end{bmatrix}, g_3 = \begin{bmatrix} 0.5 & 0 \\ 0 & -1.5 \end{bmatrix}, g_4 = \begin{bmatrix} 0.5 & 0 \\ 0 & 1 \end{bmatrix},$$
$$g_5 = \begin{bmatrix} 1 & 0 \\ 0 & -0.5 \end{bmatrix}.$$

The five players' policies $u_i \in \mathbb{R}^2$ with $i = 1, \ldots, 5$. As for the NZS game of system (44), the performance index matrices are set as $M_1 = 2I_{2\times2}$, $M_2 = I_{2\times2}$, $M_3 = 1.5I_{2\times2}$, $M_4 = 3I_{2\times2}$, $M_5 = 2.5I_{2\times2}$, $R_{11} = R_{22} = 2I_{2\times2}$, $R_{33} = 3I_{2\times2}$, $R_{44} = 2.5I_{2\times2}$ and $R_{55} = 2I_{2\times2}$. In addition, the other matrices satisfy that $R_{ij} = I_{2\times2}$ with $i, j \in \{1, 2, 3, 4, 5\}$ and $i \neq j$.

In the design of the adaptive controllers, the critic NN weights are initialized as $\hat{\eta}_1(0) = [-0.1517,\ 1.3906,$

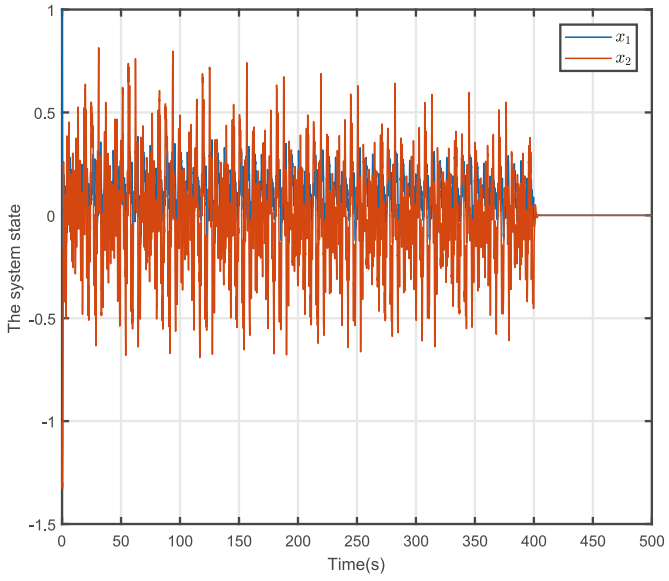

Fig. 3. The evolution of the system state x.

$-1.3731,\ 1.4189]^T$, $\hat{\eta}_2(0) = [-0.9324,\ 0.5014,\ 0.2593,\ 0.5253]^T$, $\hat{\eta}_3(0) = [-0.4169, 0.3608, 0.9335, -1.4422]^T$, $\hat{\eta}_4(0) = [-1.2484, 1.4244, 0.4540,\ -0.8063]^T$, and $\hat{\eta}_5(0) = [-0.2895,\ -1.1339,\ -0.6947,\ -0.7265]^T$. The learning rates in this example are $\alpha_i = 5$ with $i \in \{1, 2, 3, 4, 5\}$. The other parameters are set as $\sigma = 0.75$ and $K = 2$. By selecting the sampling period of 0.01 s, system (44) is controlled by the proposed method for 500 s and the corresponding exploration noises $b_{e2} = [0.5e^{-0.001t}\big(\sin^2(20t)\cos(20t) + \sin^2(12t)\cos(10t) + \sin^2(28t)\cos(25t) + \sin^5(1.2t) + \sin^2(12t) + \sin^3(25t)\cos(35t)\big), 1.2e^{-0.001t}\big(\sin^2(5t)\cos(30t) + \sin^2(37t)\ \cos(t) + \sin^2(18t)\cos(6.5t) + \sin^5(7.2t) + \sin^2(2t) + \sin^3(25t)\ \cos(3t)\big)]^T$ work for the first 400 s. The corresponding results are depicted in Figs. 3–5.

The stability of the controlled system is shown in Fig. 3. It can be found that when the noises are deleted at $t = 400$ s, the system state converges to 0 in a short time, which shows that the obtained control policies are effective in stabilizing the controlled plant. In Fig. 4, the evolutions of the event-based control policies and the critic NN weights are presented, and the weight vectors finally get to $\hat{\eta}_1 = [-0.0305,\ -0.2747,\ -1.2106,\ 0.2621]^T$, $\hat{\eta}_2 = [-0.8947,\ -0.1232,\ 0.3200,\ 0.1179]^T$, $\hat{\eta}_3 = [-0.6845,\ 0.8389,\ 0.7203,\ -0.9069]^T$, $\hat{\eta}_4 = [-1.3579,\ 0.8707,\ 0.4227,\ -0.9699]^T$, and $\hat{\eta}_5 = [-0.2984,\ -0.2112,\ -0.6401,\ 0.1983]^T$.

The evolution of the triggering condition and the TD errors is presented in Fig. 5(a) and (b), respectively. Moreover, from Fig. 5(c), one can find that the lower bound of the inter-event times is 0.03 s after $t = 400$ s. Therefore, the Zeno behaviour is also excluded in this example. In the whole control process, totally 50,000 time-based samplings are needed, but in the meanwhile only 12,262 events are triggered. That is, less recomputations and transmissions of the control input signals are executed in the adaptive control process. Consequently, more system resources can be saved by using our method.

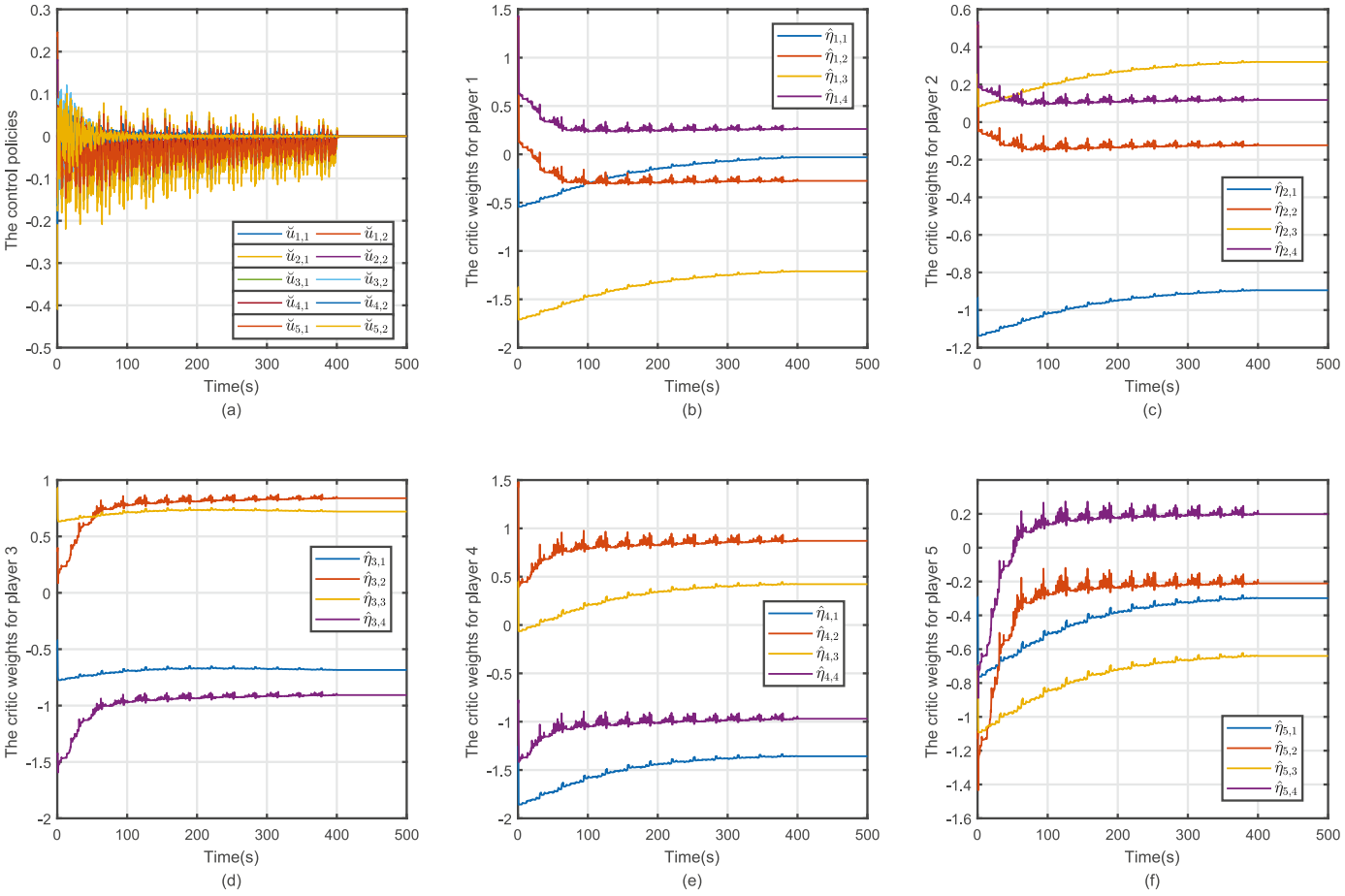

Fig. 4. The evolution of (a) the players' policies $\breve{u}_1$, $\breve{u}_2$, $\breve{u}_3$, $\breve{u}_4$, $\breve{u}_5$; (b) the critic weight $\hat{\eta}_1$; (c) the critic weight $\hat{\eta}_2$; (d) the critic weight $\hat{\eta}_3$; (e) the critic weight $\hat{\eta}_4$; (f) the critic weight $\hat{\eta}_5$.

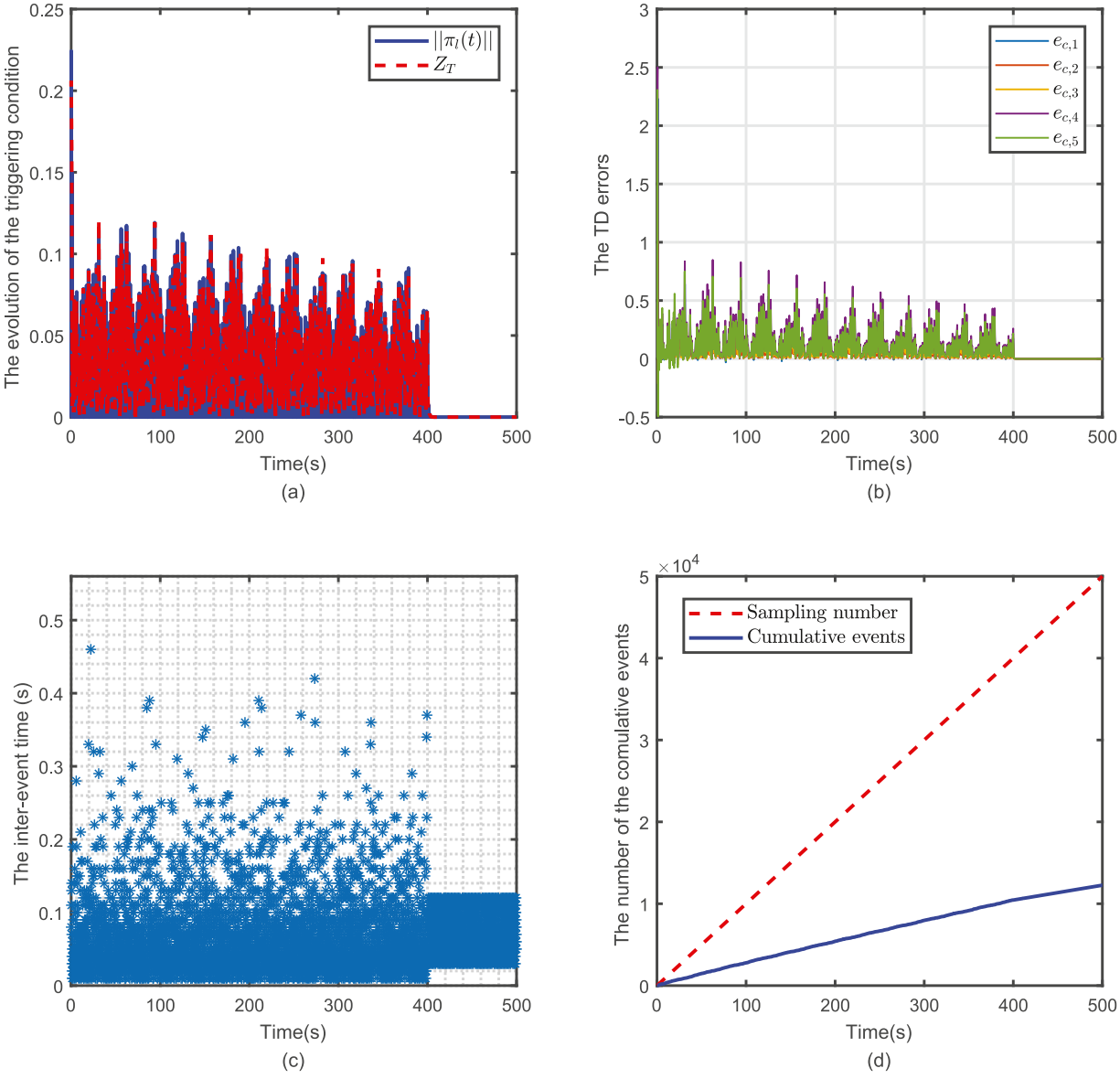

Fig. 5. The evolution of (a) the triggering condition in the adaptive control process; (b) the TD errors; (c) the inter-event times; (d) the cumulative number of the events.

5. Conclusion and future work

In this work, to deal with the NZS games of nonlinear systems subject to unknown system drift dynamics, an IRL-based online adaptive ETC method was proposed. With the aid of the presented algorithm, the requirement for the system drift dynamics is relaxed. In the adaptive control process, the critic NNs were employed to approximate the optimal value functions. The IRL algorithm was implemented in an online manner, and thus made it possible to combine the adaptive critic design method and the ETC mechanism in solving the NZS games. As a novel state-dependent triggering condition was provided, the computation and communication burdens of the whole control process were reduced and the system stability was guaranteed in the meanwhile. Finally, the effectiveness of the proposed method was demonstrated by two numerical examples.

In this work, the affine nonlinear systems with multiple controllers are addressed. But there are still other types of systems to be further investigated. In our future studies, the developed ETC method is expected to be expended to the control fields of more complex plants such as the nonlinear large-scale interconnected systems, the switched systems, and the stochastic systems with unknown dynamics. Moreover, the NZS games of the systems with completely unknown system dynamics, such as the Takagi–Sugeno fuzzy systems [54] and stochastic systems [55], are also interesting issues to be solved.

Declaration of Competing Interest

The authors declare that they have no known competing financial interestsor personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgement

This work was supported by the National Natural Science Foundation of China (No. 61627809, 61433004, 61621004, 61673100), Fundamental Research Funds for Central Universities (N150504011) and Liaoning Revitalization Talents Program (XLYC1801005).

References

[1] Y. Yang, P. Zhang, Study on the influence of inconsistent valve parameters on LCC-HVDC commutation and operation, IEEE Access 7 (2019) 109015–109025.
[2] P. Zhang, D. Lu, A Survey of condition monitoring and fault diagnosis toward integrated O&M for wind turbines, Energies 12 (14) (2019) 2801.
[3] D. Ding, Q. Han, Z. Wang, X. Ge, A survey on model-based distributed control and filtering for industrial cyber-physical systems, IEEE Trans. Ind. Informat. 15 (5) (2019) 2483–2499.
[4] P. Tabuada, Event-triggered real-time scheduling of stabilizing control tasks, IEEE Trans. Autom. Control 52 (9) (2007) 1680–1685.
[5] D. Ding, Z. Wang, Q. Han, A set-membership approach to event-triggered filtering for general nonlinear systems over sensor networks, IEEE Trans. Autom. Control (2019), doi:10.1109/TAC.2019.2934389. In press.
[6] D. Ye, M. Chen, H. Yang, Distributed adaptive event-triggered fault-tolerant consensus of multi-agent systems with general linear dynamics, IEEE Trans. Cybern. 49 (3) (2019) 757–767.
[7] D. Ding, Z. Wang, D.W.C. Ho, G. Wei, Observer-based event-triggering consensus control for multiagent systems with lossy sensors and cyber-attacks, IEEE Trans. Cybern. 47 (8) (2017) 1936–1947.
[8] D. Wang, C. Mu, D. Liu, H. Ma, On mixed data and event driven design for adaptive-critic-based nonlinear H_∞ control, IEEE Trans. Neural Netw. Learn. Syst. 29 (4) (2018) 993–1005.
[9] D. Ding, Z. Wang, Q. Han, G. Wei, Neural-network-based output-feedback control under round-robin scheduling protocols, IEEE Trans. Cybern. 49 (6) (2019) 2372–2384.
[10] A. Friedman, Differential Games, Courier Corporation, Mineola, NY, USA, 2013.
[11] J. Sun, C. Liu, X. Zhao, Backstepping-based zero-sum differential games for missile-target interception systems with input and output constraints, IET Control Theory Appl. 12 (2) (2018) 243–253.
[12] J. Sun, C. Liu, Finite-horizon differential games for missile-target interception system using adaptive dynamic programming with input constraints, Int. J. Syst. Sci. 49 (2) (2018) 264–283.
[13] J. Zhang, Z. Zhang, Y. Han, Research on manufacturability optimization of discrete products with 3d printing involved and lot-size considered, J. Manuf. Syst. 43 (2017) 150–159.
[14] K. Kogan, A. Herbon, Inventory control over a short time horizon under unknown demand distribution, IEEE Trans. Autom. Control 61 (10) (2016) 3058–3063.
[15] S.C. Hsueh, C. Ling, The qualitative properties of symmetric open-loop Nash equilibria in the state-control dynamic system in differential games, Asian J. Control 20 (5) (2018) 1769–1781.
[16] Z. Ni, S. Paul, A multistage game in smart grid security: a reinforcement learning solution, IEEE Trans. Neural Netw. Learn. Syst. 30 (9) (2019) 2684–2695.
[17] D. Wang, H. He, C. Mu, D. Liu, Intelligent critic control with disturbance attenuation for affine dynamics including an application to a microgrid system, IEEE Trans. Ind. Electron. 64 (6) (2017) 4935–4944.
[18] T. Başar, G.J. Olsder, Dynamic Noncooperative Game Theory, SIAM, Philadelphia, PA, 1999.
[19] F.L. Lewis, D.L. Vrabie, V.L. Syrmos, Optimal Control, Wiley, Hoboken, NJ, 2012.
[20] A.W. Starr, Y.C. Ho, Nonzero-sum differential games, J. Optim. Theory Appl. 3 (3) (1969) 184–206.
[21] K.G. Vamvoudakis, F.L. Lewis, Multi-player non-zero-sum games: online adaptive learning solution of coupled hamilton-jacobi equations, Automatica 47 (2011) 1556–1569.
[22] H. Zhang, L. Cui, Y. Luo, Near-optimal control for nonzero-sum differential games of continuous-time nonlinear systems using single-network ADP, IEEE Trans. Cybern. 43 (1) (2013) 206–216.
[23] F. Wang, H. Zhang, D. Liu, Adaptive dynamic programming: an introduction, IEEE Comput. Intell. M. 4 (2) (2009) 39–47.
[24] Q. Wei, D. Liu, F.L. Lewis, Y. Liu, J. Zhang, Mixed iterative adaptive dynamic programming for optimal battery energy control in smart residential microgrids, IEEE Trans. Ind. Electron. 64 (5) (2017) 4110–4120.
[25] X. Xie, Q. Zhou, D. Yue, H. Li, Relaxed control design of discrete-time Takagi–Sugeno fuzzy systems: an event-triggered real-time scheduling approach, IEEE Trans. Syst. Man Cybern. Syst. 48 (12) (2018) 2251–2262.
[26] H. Yang, D. Ye, Distributed fixed-time consensus tracking control of uncertain nonlinear multi-agent systems: a prioritized strategy, IEEE Trans. Cybern. (2019), doi:10.1109/TCYB.2019.2925123. In press.
[27] R.S. Sutton, A.G. Barto, Reinforcement Learning – An Introduction, MIT Press, Cambridge, Massachusetts, 1998.
[28] H. Jiang, H. Zhang, K. Zhang, X. Cui, Data-driven adaptive dynamic programming schemes for non-zero-sum games of unknown discrete-time nonlinear systems, Neurocomputing 275 (2018) 649–658.
[29] M. Johnson, R. Kamalapurkar, S. Bhasin, W.E. Dixon, Approximate n-player nonzero-sum game solution for an uncertain continuous nonlinear system, IEEE Trans. Neural Netw. Learn. Syst. 26 (8) (2015) 1645–1658.
[30] H. Zhang, H. Jiang, C. Luo, G. Xiao, Discrete-time nonzero-sum games for multiplayer using policy-iteration-based adaptive dynamic programming algorithms, IEEE Trans. Cybern. 47 (10) (2017) 3331–3340.
[31] K.G. Vamvoudakis, Non-zero sum nash q-learning for unknown deterministic continuous-time linear systems, Automatica 61 (2015) 274–281.
[32] D. Liu, H. Li, D. Wang, Online synchronous approximate optimal learning algorithm for multiplayer nonzero-sum games with unknown dynamics, IEEE Trans. Syst. Man Cybern. Syst. 44 (8) (2014) 1015–1027.
[33] H. Su, H. Zhang, Y. Liang, Y. Mu, Online event-triggered adaptive critic design for non-zero-sum games of partially unknown networked systems, Neurocomputing 368 (2019) 84–98.
[34] C.J.C.H. Watkins, P. Dayan, Q-learning, Mach. Learn. 8 (3–4) (1992) 279–292.
[35] D. Vrabie, O. Pastravanu, M. Abu-Khalaf, F.L. Lewis, Adaptive optimal control for continuous-time linear systems based on policy iteration, Automatica 45 (2) (2009) 477–484.
[36] D. Vrabie, F.L. Lewis, Neural network approach to continuous-time direct adaptive optimal control for partially unknown nonlinear systems, Neural Netw. 22 (3) (2009) 237–246.
[37] Q. Zhang, D. Zhao, Data-based reinforcement learning for nonzero-sum games with unknown drift dynamics, IEEE Trans. Cybern. 49 (8) (2019) 2874–2885.
[38] R. Song, F.L. Lewis, Q. Wei, H. Zhang, Off-policy actor-critic structure for optimal control of unknown systems with disturbances, IEEE Trans. Cybern. 46 (5) (2016) 1041–1050.
[39] K.G. Vamvoudakis, A. Mojoodi, H. Ferraz, Event-triggered optimal tracking control of nonlinear systems, Int. J. Robust Nonlinear Control 27 (4) (2017) 598–619.
[40] H. Su, H. Zhang, K. Zhang, W. Gao, Online reinforcement learning for a class of partially unknown continuous-time nonlinear systems via value iteration, Optimal Control Appl. Met. 39 (2) (2018) 1011–1028.
[41] D. Wang, H. He, X. Zhong, D. Liu, Event-driven nonlinear discounted optimal regulation involving a power system application, IEEE Trans. Ind. Electron. 64 (10) (2017) 8177–8186.
[42] K.G. Vamvoudakis, H. Ferraz, Model-free event-triggered control algorithm for continuous-time linear systems with optimal performance, Automatica 87 (2018) 412–420.
[43] K. Zhang, H. Zhang, H. Jiang, Y. Wang, Near-optimal output tracking controller design for nonlinear systems using an event-driven ADP approach, Neurocomputing 309 (2018) 168–178.
[44] L. Cui, X. Xie, X. Wang, Y. Luo, J. Liu, Event-triggered single-network ADP method for constrained optimal tracking control of continuous-time non-linear systems, Appl. Math. Comput. 352 (2019) 220–234.
[45] Q. Zhang, D. Zhao, Y. Zhu, Event-triggered H_∞ control for continuous-time nonlinear system via concurrent learning, IEEE Trans. Syst. Man Cybern. Syst. 47 (7) (2017) 1071–1081.
[46] X. Zhong, H. He, D. Wang, Z. Ni, Model-free adaptive control for unknown nonlinear zero-sum differential game, IEEE Trans. Cybern. 47 (3) (2017) 683–694.
[47] Y. Zhu, D. Zhao, H. He, J. Ji, Event-triggered optimal control for partially unknown constrained-input systems via adaptive dynamic programming, IEEE Trans. Ind. Electron. 64 (5) (2017) 4101–4109.
[48] K.G. Vamvoudakis, Event-triggered optimal adaptive control algorithm for continuous-time nonlinear systems, IEEE/CAA J. Autom. Sinica 1 (3) (2014) 282–293.
[49] H. Zhang, H. Su, K. Zhang, Y. Luo, Event-triggered adaptive dynamic programming algorithm for non-zero-sum games of unknown nonlinear systems via generalized fuzzy hyperbolic models, IEEE Trans. Fuzzy Syst. (2019), doi:10.1109/TFUZZ.2019.2896544. In press.
[50] H. Su, H. Zhang, D. Gao, Y. Luo, Adaptive dynamics programming for h_∞ control of continuous-time unknown nonlinear systems via generalized fuzzy hyperbolic models, IEEE Trans. Syst. Man Cybern. Syst. (2019), doi:10.1109/TSMC.2019.2900750. In press
[51] K.G. Vamvoudakis, F.L. Lewis, Online actor-critic algorithm to solve the continuous-time infinite horizon optimal control problem, Automatica 46 (5) (2010) 878–888.
[52] Q. Wei, R. Song, P. Yan, Data-driven zero-sum neuro-optimal control for a class of continuous-time unknown nonlinear systems with disturbance using ADP, IEEE Trans. Neural Netw. Learn. Syst. 27 (2) (2016) 444–458.
[53] D. Ding, Z. Wang, Q. Han, Neural-network-based output-feedback control with stochastic communication protocols, Automatica 106 (2019) 221–229.
[54] Y. Mu, H. Zhang, S. Sun, J. Ren, Robust non-fragile proportional plus derivative state feedback control for a class of uncertain Takagi–Sugeno fuzzy singular systems 356 (12) (2019) 6208–6225.
[55] Y. Wang, W.X. Zheng, H. Zhang, Dynamic event-based control of nonlinear stochastic systems, IEEE Trans. Autom. Control 62 (12) (2017) 6544–6551.

Hanguang Su received the B.S. degree in automation control and the M.S. degree in control engineering from Northeastern University, Shenyang, China, in 2013 and 2015, respectively. He is currently pursuing the Ph.D. degree in control theory and control engineering with the School of Information Science and Engineering, Northeastern University, Shenyang, China. His current research interests include reinforcement learning, optimal control, fuzzy control, adaptive dynamic programming, and their applications.

Huaguang Zhang received the B.S. and M.S. degrees in control engineering from the Northeast Dianli University, Jilin City, China, in 1982 and 1985, respectively, and the Ph.D. degree in thermal power engineering and automation from Southeast University, Nanjing, China, in 1991. He joined the Department of Automatic Control, Northeastern University, Shenyang, China, in 1992, as a Post-Doctoral Fellow for two years. Since 1994, he has been a Professor and the Head of the School of Information Science and Engineering, Institute of Electric Automation, Northeastern University. He has authored or co-authored over 280 journal and conference papers and six monographs and co-invented 90 patents. His current research interests include fuzzy control, stochastic system control, neural networks based control, nonlinear control, and their applications. Dr. Zhang was a recipient of the Outstanding Youth Science Foundation Award from the National Natural Science Foundation Committee of China in 2003. He was named the Cheung Kong Scholar by the Education Ministry of China in 2005. He was also a recipient of the IEEE TRANSACTIONS ON NEURAL NETWORKS 2012 Outstanding Paper Award and the Andrew P. Sage Best Transactions Paper Award 2015. He is the E-Letter Chair of the IEEE CIS Society and the former Chair of the Adaptive Dynamic Programming & Reinforcement Learning Technical Committee on the IEEE Computational Intelligence Society. He was an Associate Editor of the IEEE TRANSACTIONS ON FUZZY SYSTEMS from 2008 to 2013. He is an Associate Editor of Automatica, the IEEE TRANSACTIONS ON NEURAL NETWORKS AND LEARNING SYSTEMS, the IEEE TRANSACTIONS ON CYBERNETICS, and Neurocomputing.

Shaoxin Sun received the B.S. degree in control technology and instrument from Hebei University, Baoding, China, in 2014 and the M.S. degree in control theory and control engineering from Northeastern University, Shenyang, China, in 2016. She is currently pursuing the Ph.D. degree in control theory and control engineering from Northeastern University, Shenyang, China. Her current research interests include fuzzy systems, time-delay systems, fault estimation, fault tolerant control, and stochastic systems.

Yuliang Cai received the B.S. degree in information and computing science from the Ludong University, Yantai, China in 2014, and the M.S. degree in control theory and control engineering from the Dalian University of Technology, Dalian, China in 2017. She is currently working toward the Ph.D. degree in control theory and control engineering at Northeastern University, Shenyang, China. Her research interests include multi-agent system control, fuzzy control, adaptive dynamic programming, etc.