

# Event-Triggered $H_\infty$ Control for Continuous-Time Nonlinear System

Dongbin Zhao<sup>1,\*</sup>, Qichao Zhang<sup>1</sup>, Xiangjun Li<sup>2</sup>, and Lingda Kong<sup>2</sup>

<sup>1</sup> State Key Laboratory of Management and Control for Complex Systems, Institute of Automation, Chinese Academy of Sciences, Beijing 100190, China

<sup>2</sup> China Electric Power Research Institute, Beijing 100192, China  
{dongbin.zhao, zhangqichao2014}@ia.ac.cn

**Abstract.** In this paper, the  $H_\infty$  optimal control for a class of continuous-time nonlinear systems is investigated using event-triggered method. First, the  $H_\infty$  optimal control problem is formulated as a two-player zero-sum differential game. Then, an adaptive triggering condition is derived for the closed loop system with an event-triggered control policy and a time-triggered disturbance policy. For implementation purpose, the event-triggered concurrent learning algorithm is proposed, where only one critic neural network is required. Finally, an illustrated example is provided to demonstrate the effectiveness of the proposed scheme.

## 1 Introduction

From the perspective of minmax optimization problem, the  $H_\infty$  control problem can be formulated as a two-player zero-sum differential game [1]. In order to obtain a controller that minimizes a cost function in the presence of worst-case disturbances, ones need to find the Nash equilibrium solution by solving the Hamilton-Jacobi-Isaacs (HJI) equation. Several reinforcement learning (RL) methods [2–4] have been successfully applied to solve the HJI equation for discrete-time systems [5] and continuous-time systems [6, 7].

Due to the capability of computation efficiency, event-triggered control method has been integrated with the RL approach recently [8, 9]. In the event-triggered control method, the controller is updated based on a new sampled state only when an event is triggered at event-triggering instants. This can reduce the communication between the plant and the controller significantly. In [10], an optimal adaptive event-triggered control algorithm was implemented based on an actor-critic structure for continuous-time nonlinear systems. On the other hand, the concurrent learning technique, which can relax the traditional persistency of excitation (PE) condition, was proposed for an uncertain system in [11]. In [12], a related idea called experience replay was adopted in Integral reinforcement learning (IRL) algorithm for constrained-input nonlinear systems.

To the best of our knowledge, there are no results on event-triggered  $H_\infty$  control of nonlinear system via concurrent learning. This is the motivation of our research. In this paper, the  $H_\infty$  control problem is described as a two-player zero-sum differential game and an online event-triggered concurrent learning (ETCL)

---

\* Corresponding author.

algorithm is proposed to approximate the optimal control policy. Simulation results show the effectiveness of the proposed scheme.

## 2 Problem Statement

Consider the following nonlinear system with external disturbance:

$$\dot{x}(t) = f(x) + g(x)u(t) + k(x)w(t), \quad (1)$$

where  $x \in \mathbf{R}^n$  is the state vector,  $u \in \mathbf{R}^m$  is the control input,  $w \in \mathbf{R}^q$  is the nonlinear perturbation with  $w(t) \in L_2(0, \infty)$ .  $f(\cdot) \in \mathbf{R}^n$ ,  $g(\cdot) \in \mathbf{R}^{n \times m}$  and  $k(\cdot) \in \mathbf{R}^{n \times q}$  are smooth nonlinear dynamics. Assume that  $f(x) + g(x)u + k(x)w$  is Lipschitz continuous on a compact set  $\Omega \subseteq \mathbf{R}^n$  with  $f(0) = 0$ . Let  $x(0) = x_0$  be the initial state. Assume that  $w(0) = 0$ , so that  $x = 0$  is an equilibrium of system (1). It is assumed that the system (1) is controllable.

Here, we introduce a sampled-data system that is characterized by a monotonically increasing sequence of event-triggering instants  $\{\lambda_j\}_{j=0}^\infty$ , where  $\lambda_j$  is the  $j$ th consecutive sampling instant with  $\lambda_j < \lambda_{j+1}$ . Define the event-trigger error between the current state  $x(t)$  and the sampled state  $\hat{x}_j$  as follows

$$e_j(t) = \hat{x}_j - x(t), \forall t \in [\lambda_j, \lambda_{j+1}). \quad (2)$$

In the event-triggered control mechanism, the event-triggering condition is determined by the event-trigger error and a state-dependent threshold. When the event-triggering condition is not satisfied at  $t = \lambda_j$ , we say an event is triggered. Then, the system state is sampled that resets the event-trigger error  $e_j(t)$  to zero, and the controller  $v(\hat{x}_j)$  is updated based on the new sampled state. Note that  $v(\hat{x}_j)$  is a function of the event-based state vector. The obtained control sequence  $\{v(\hat{x}_j)\}_{j=0}^\infty$  becomes a continuous-time input signal  $v(t) = \{v(\hat{x}_j, t)\}_{j=0}^\infty$  after using a zero-order hold (ZOH). In order to simplify the expression, we use  $v(\hat{x}_j)$  to represent  $v(\hat{x}_j, t)$  for  $t \in [\lambda_j, \lambda_{j+1})$  in the following presentation.

Similar to the traditional  $H_\infty$  problem, our primary objective is to find a sequence of control inputs  $\{v(\hat{x}_j)\}_{j=0}^\infty$ , which for some prescribed  $\gamma > 0$ , renders

$$J(x_0, v(\hat{x}_j), w) = \sum_{\cup[\lambda_j, \lambda_{j+1})=[0, \infty)} \int_{\lambda_j}^{\lambda_{j+1}} r(x, v(\hat{x}_j), w) dt \quad (3)$$

nonpositive for all  $w(t) \in L_2[0, \infty)$  and  $x(0) = 0$ , where utility  $r(x, v(\hat{x}_j), w) = x^T Q x + v^T(\hat{x}_j) R v(\hat{x}_j) - \gamma^2 \|w(t)\|^2$ ,  $Q$  and  $R$  are symmetric and positive definite matrices, and  $\gamma \geq \gamma^* \geq 0$ . Here,  $\gamma^*$  is the smallest  $\gamma$  such that the system (1) is stabilized. The quantity  $\gamma^*$  is known as the H-infinity gain.

## 3 Event-Triggered Optimal Controller Design

In this section, the  $H_\infty$  control problem is formulated as a two-player zero-sum differential game, where the control input  $u$  is a minimizing player while the

disturbance  $w$  is a maximizing one. It is well known that the solution of  $H_\infty$  control problem is the zero-sum game theoretic saddle point  $(u^*, w^*)$ , where  $u^*$  and  $w^*$  are the optimal control and the worst-case disturbance.

In the time-triggered case, the value function is generally defined as

$$V(u, w) = \int_t^\infty (x^T Q x + u^T R u - \gamma^2 \|w\|^2) d\tau. \quad (4)$$

The corresponding nonlinear zero-sum Bellman equation is

$$r(x, u, w) + (\nabla V)^T (f(x) + g(x)u + k(x)w) = 0, \quad (5)$$

where  $\nabla V = \partial V(x)/\partial x$  is the partial derivative of the value function with respect to the state. Then, the two-player zero-sum game has a unique solution if a saddle point  $(u^*, w^*)$  exists, that is if the Nash condition holds

$$\min_u \max_w V(u, w) = \max_w \min_u V(u, w). \quad (6)$$

Define the Hamiltonian of the time-triggered problem

$$H(x, \nabla V, u, w) = (\nabla V)^T (f + gu + kw) + x^T Q x + u^T R u - \gamma^2 \|w\|^2. \quad (7)$$

Then the associated HJI equation can be written as

$$\min_u \max_w H(x, \nabla V^*, u, w) = 0, \quad (8)$$

where the optimal value function  $V^*$  is the solution to the HJI equation. The associated control and disturbance policies are given as follows:

$$u^*(t) = -\frac{1}{2} R^{-1} g^T(x) \nabla V^*. \quad (9)$$

$$w^*(t) = \frac{1}{2\gamma^2} k^T(x) \nabla V^*. \quad (10)$$

In the event-triggered case, the control input is updated based on the sampled-state information  $\hat{x}_j$  instead of the real state  $x(t)$ . Hence, (9) becomes

$$v^*(\hat{x}_j) = -\frac{1}{2} R^{-1} g^T(\hat{x}_j) \nabla V^*(\hat{x}_j), \forall t \in [\lambda_j, \lambda_{j+1}), \quad (11)$$

where  $\nabla V^*(\hat{x}_j) = \partial V^*(\hat{x}_j)/\partial x(t)$ . By using (10) and (11), the event-triggered HJI equation can be written as

$$\begin{aligned} & (\nabla V^*)^T f(x) + x^T Q x - \frac{1}{2} (\nabla V^*)^T g(x) R^{-1} g^T(\hat{x}_j) \nabla V^*(\hat{x}_j) \\ & + \frac{1}{4} (\nabla V^*(\hat{x}_j))^T g(\hat{x}_j) R^{-1} g^T(\hat{x}_j) \nabla V^*(\hat{x}_j) + \frac{1}{4\gamma^2} (\nabla V^*)^T k(x) k^T(x) \nabla V^* = 0. \end{aligned} \quad (12)$$

**Assumption 1.** The controller  $u(x)$  is Lipschitz continuous with respect to the event-trigger error,

$$\|u(x(t)) - u(\hat{x}_j)\| = \|u(x(t)) - u(x(t) + e_j(t))\| \leq L\|e_j(t)\|, \quad (13)$$

where  $L$  is a positive real constant and  $u(\hat{x}_j) = v(\hat{x}_j)$ .

**Theorem 1.** Suppose that  $V^*(x)$  is the solution of the event-triggered HJI equation (12). For  $\forall t \in [\lambda_j, \lambda_{j+1}), j = 0, \dots, \infty$ , the disturbance policy and control policy are given by (10) and (11), respectively. If the triggering condition is defined as follows

$$\begin{aligned} \|e_j(t)\|^2 \leq e_T &= \frac{(1 - \beta^2)}{L^2 \|s\|^2} \underline{\theta}(Q) \|x\|^2 \\ &+ \frac{1}{L^2} \|v(\hat{x}_j)\|^2 - \frac{\gamma^2}{L^2 \|s\|^2} \|w(t)\|^2, \end{aligned} \quad (14)$$

where  $e_T$  is the threshold,  $\underline{\theta}(Q)$  is the minimal eigenvalue of  $Q$ ,  $\beta \in (0, 1)$  is a designed sample frequency parameter and  $s^T s = R$ . Then the closed-loop system (1) is asymptotically stable.

*Remark 1:* The event-trigger instants  $\{\lambda_j\}_{j=0}^\infty$  is determined by the triggering condition (14). Based on the event-triggered mechanism, an event is generated by the violation of the triggering condition. Note that this method can reduce the communication between the controller and the plant effectively. On the other hand, the sample frequency can be adjusted by the designed parameter  $\beta$  in the triggering condition (14). When  $\beta$  is close to 1 one samples more frequently whereas when  $\beta$  is close to zero, the sampling periods become longer.

## 4 Online Neuro-Optimal Control Scheme

In this section, an online event-triggered concurrent learning (ETCL) algorithm is proposed, where only one critic neural network is required.

According to the Weierstrass high-order approximation theorem, the value function based on NN can be written as

$$V(x) = W_c^T \phi(x) + \varepsilon, \quad (15)$$

where  $W_c \in \mathbf{R}^N$  and  $\phi(x) \in \mathbf{R}^N$  are the critic NN ideal weights and activation function vector, with  $N$  the number of hidden neurons, and  $\varepsilon \in \mathbf{R}$  the critic NN approximation error.

The derivative of (15) with respect to  $x$  can be given by

$$\nabla V(x) = \nabla \phi^T(x) W_c + \nabla \varepsilon. \quad (16)$$

Then, the zero-sum Bellman equation (5) can be rewritten as

$$x^T Q x + v^T(\hat{x}_j) R v(\hat{x}_j) - \gamma^2 \|w(t)\|^2 + W_c^T \nabla \phi(f(x) + g(x)v(\hat{x}_j) + k(x)w(t)) = \varepsilon_H, \quad (17)$$

where the residual error is  $\varepsilon_H = -(\nabla\varepsilon)^T(f(x) + g(x)v(\hat{x}_j) + k(x)w(t))$ . Under the Lipschitz assumption on the system dynamics, the residual error is bounded locally. It is shown in [7] that this error converges uniformly to zero as the number of hidden-layer units increases. That is, there exists  $\varepsilon_{Hmax} > 0$  such that  $\|\varepsilon_H\| \leq \varepsilon_{Hmax}$ .

Let  $\hat{W}_c$  be the estimation of the unknown ideal weight vector  $W_c$ . The actual output of critic NN can be presented as

$$\hat{V}(x) = \hat{W}_c^T \phi(x). \quad (18)$$

Accordingly, the time-triggered disturbance policy (10) and event-triggered control policy (11) can be approximated by

$$\hat{w}(t) = \frac{1}{2\gamma^2} k^T(x) \phi^T(x) \hat{W}_c. \quad (19)$$

$$\hat{v}(\hat{x}_j) = -\frac{1}{2} R^{-1} g^T(\hat{x}_j) \phi^T(\hat{x}_j) \hat{W}_c(\hat{x}_j). \quad (20)$$

where  $\hat{W}_c(\hat{x}_j)$  is the event-based estimation of ideal weight  $W_c$ . Then the closed-loop system dynamics (1) can now be written as

$$\dot{x} = f(x) + g(x)\hat{v}(\hat{x}_j) + k(x)\hat{w}(t), t \geq 0. \quad (21)$$

The approximate Hamilton function is

$$\begin{aligned} & \hat{W}_c^T \nabla \phi(x) f + x^T Q x - \frac{1}{2} \hat{W}_c^T \nabla \phi(x) g(x) R^{-1} g^T(\hat{x}_j) \nabla \phi^T(\hat{x}_j) \hat{W}_c(\hat{x}_j) + \frac{1}{4} \hat{W}_c^T(\hat{x}_j) \times \\ & \nabla \phi(\hat{x}_j) g(\hat{x}_j) R^{-1} g^T(\hat{x}_j) \nabla \phi^T(\hat{x}_j) \hat{W}_c(\hat{x}_j) + \frac{1}{4\gamma^2} \hat{W}_c^T \nabla \phi(x) k(x) k^T(x) \nabla \phi^T(x) \hat{W}_c = e. \end{aligned} \quad (22)$$

where  $e$  is a residual equation error.

Based on concurrent learning, the critic NN' weights can be updated by recorded data concurrently with current data. Define the residual equation error at time  $t_k$  as

$$e(t_k) = r(t_k) + \hat{W}_c^T(t) \sigma_k. \quad (23)$$

where  $r(t_k) = x^T(t_k) Q x(t_k) + \hat{v}^T(\hat{x}_j) R \hat{v}(\hat{x}_j) - \gamma^2 \|\hat{w}(t)\|^2$ ,  $\sigma_k = \nabla \phi(x(t_k)) (f(x(t_k)) + g(x(t_k))\hat{v}(\hat{x}_j) + k(x(t_k))\hat{w}(t))$  are stored data at time  $t_k \in [\lambda_j, \lambda_{j+1})$ ,  $k \in \{1, \dots, p\}$ ,  $j = 0, 1, \dots, \infty$ , and  $p$  is the number of stored samples.

*Condition 1:* Let  $M = [\sigma_1, \dots, \sigma_p]$  be the recorded data corresponding to the critic NN's weights. Then  $M$  contains as many linearly independent elements as the number of corresponding critic NN's hidden neurons, i.e.,  $\text{rank}(M) = N$ .

To derive the minimum value of  $e$ , it is desired to choose  $\hat{W}_c$  to minimize the corresponding squared residual error  $E = \frac{1}{2} e^T e$ . Considering the concurrent learning, we develop a novel weight update law for the critic NN

$$\dot{\hat{W}}_c = -\alpha \sigma \left( \sigma^T \hat{W}_c(t) + r(x, \hat{v}(\hat{x}_j), \hat{w}(t)) \right) - \alpha \sum_{k=1}^p \sigma_k \left( \sigma_k^T \hat{W}_c(t) + r(t_k) \right). \quad (24)$$

where  $\sigma = \nabla \phi(x) (f(x) + g(x)v(\hat{x}_j) + k(x)w)$ ,  $\sigma_k$  is defined in (23),  $k \in \{1, \dots, p\}$  denote the index of a stored data point, and  $\alpha > 0$  denote the learning rate.

*Remark 2:* The online algorithm presented in this paper does not rely on traditional PE condition which is difficult to check online. According to [11], the second term in (24) can be utilised to relax the PE condition with Condition 1.

By defining the weight estimation error of the critic NN as  $\tilde{W}_c = W_c - \hat{W}_c$  and taking the time derivative one has

$$\dot{\tilde{W}}_c = -\alpha \sigma \left( \sigma^T \tilde{W}_c - \varepsilon_H \right) - \alpha \sum_{k=1}^p \sigma_k \left( \sigma_k^T \tilde{W}_c - \varepsilon_H(t_k) \right). \quad (25)$$

**Assumption 2.** *a. The critic NN activation function and its gradient are bounded, i.e.,  $\|\phi(x)\| \leq \phi_M$  and  $\|\nabla\phi(x)\| \leq \nabla\phi_M$ , with  $\phi_M, \nabla\phi_M$  being positive constants.*

*b. The system dynamics  $g(x)$  and  $k(x)$  are upper bounded by positive constants such that  $\|g(x)\| \leq g_M$  and  $\|k(x)\| \leq k_M$ .*

**Theorem 2.** *Consider the nonlinear two-player zero-sum game (1) with the critic neural network (18), the time-triggered disturbance policy (19) and the event-triggered control policy (20). The tuning law based on concurrent learning technique for the continuous-time critic neural network is given by (24). Then the system is asymptotically stable and the critic weight estimation error is guaranteed to be Uniformly Ultimately Bounded (UUB) if the adaptive triggering condition*

$$\begin{aligned} \|e_j(t)\|^2 &\leq \frac{(1-\beta^2)}{L^2\|s\|^2} \underline{\theta}(Q) \|x\|^2 + \frac{1}{4L^2\|R\|^2} \|g^T(\hat{x}_j)\phi^T(\hat{x}_j) \\ &\quad \times \hat{W}_c(\hat{x}_j)\|^2 - \frac{1}{4\gamma^2 L^2\|s\|^2} \|k^T(x)\phi^T(x)\hat{W}_c(t)\|^2 \end{aligned} \quad (26)$$

and the following inequality are satisfied

$$\|\tilde{W}_c\| > \sqrt{\frac{a^2 \sum_{k=1}^{p+1} \varepsilon_{Hmax}^2}{4(a-1)(\underline{\theta}(M) + \sum_{k=1}^p \underline{\theta}(M_k))}} \triangleq B_M \quad (27)$$

for the critic network and  $a > 1$ .

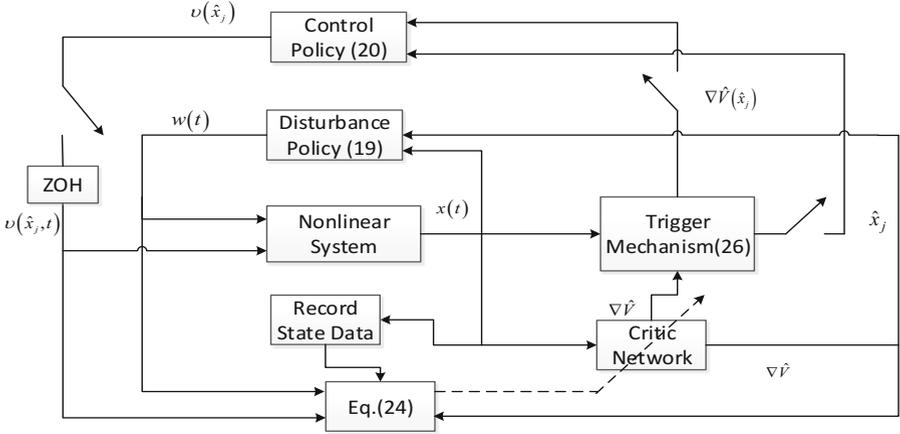
*Remark 3:* Note that the triggering condition (26) is adaptive, because the threshold is designed as function of the system state vector and the critic NN weight estimates. The controller is adjusted with events.

Then we give the structure diagram of the online ETCL algorithm for two-player zero-sum game in Fig. 1.

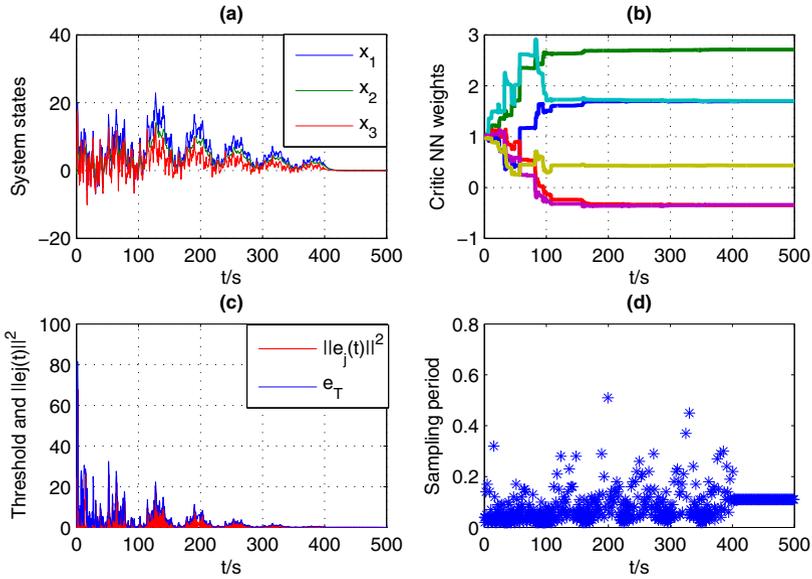
## 5 Simulation

Consider the continuous-time F16 aircraft plant [7]:

$$\dot{x} = \begin{bmatrix} -1.01887 & 0.90506 & -0.00215 \\ 0.82225 & -1.07741 & -0.17555 \\ 0 & 0 & -1 \end{bmatrix} x + \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} u + \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} w$$



**Fig. 1.** Structure diagram of the ETCL algorithm for two-player ZS game



**Fig. 2.** (a) Evolution of system states. (b) Convergence of the critic parameters. (c) Triggering threshold  $e_T$  and  $\|e_j(t)\|^2$ . (d) Sampling period.

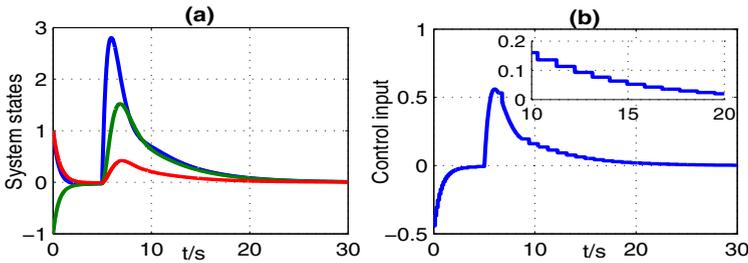
Let  $Q$  and  $R$  be identity matrices with approximate dimensions, and  $\gamma = 5$ . Choose the critic NN activation function as  $\phi(x) = [x_1^2 \ x_1x_2 \ x_1x_3 \ x_2^2 \ x_2x_3 \ x_3^2]^T$ . According to [8], the ideal values of the NN weights are  $W_c = [1.6573 \ 2.7908 \ -0.3322 \ 1.6573 \ -0.3608 \ 0.4370]^T$ . Select the initial state as  $x_0 = [1, -1, 1]^T$ , and  $\alpha = 15$ ,  $p = 10$ ,  $L = 3$ ,  $\beta = 0.8$ . During the learning process, a probing

noise is added to the control input and disturbance for the first 400s. Fig. 2(a) presents the evolution of the system states. Fig. 2(b) shows the convergence of the critic parameters. After 100s the critic parameters converged to  $\bar{W}_c = [1.6563 \ 2.7788 \ -0.3389 \ 1.6490 \ -0.3615 \ 0.4354]^T$  which are nearly the ideal values above. In Fig. 2(c), one can see that the event-trigger error converges to zero as the states converge to zero. The sampling period during the event-triggered learning process for the control policy is provided in Fig. 2(d). In particular, the event-triggered controller uses 1055 samples of the state while the time-triggered controller uses 50000 samples, which means the event-triggered method improved the learning process.

Select a disturbance signal with  $t_0 = 5$  as

$$w(t) = \begin{cases} 8e^{-(t-t_0)} \cos(t-t_0), & t \geq t_0 \\ 0, & t < t_0 \end{cases} \quad (28)$$

Fig. 3 shows the system state trajectories and the event-triggered control input with the  $H_\infty$  event-triggered controller. These simulation results verify the effectiveness of the developed control approach.



**Fig. 3.** (a) Closed-loop system states. (b) Event-triggered control input

## 6 Conclusion

In this paper, we propose an online ETCL algorithm to solve the HJI equation of  $H_\infty$  control problem for nonlinear system. The  $H_\infty$  control problem is described as a two-player zero-sum game, where the control is a minimizing player and the disturbance is a maximizing one. With an event-triggered control policy and a time-triggered disturbance policy, the online ETCL algorithm is presented. For implementation purpose, only one critic NN is used to approximate the value function, the optimal control and disturbance policies. Furthermore, a novel critic tuning law based on concurrent learning technique is given, which can relax the traditional PE condition. In our future work, we will develop an online ETCL algorithm for the unknown two-player zero-sum game system.

**Acknowledgments.** This work is supported in part by National Natural Science Foundation of China (NSFC) under Grant No. 61273136 and No. 61473176, Beijing Nova Program under Grant Z141101001814094, and Science and Technology Foundation of SGCC under Grant DG71-14-032.

## References

1. Basar, T., Olsder, G.J., Cldsder, G.J., et al.: Dynamic noncooperative game theory. Academic Press, London (1995)
2. Zhao, D., Zhu, Y.: MEC—A Near-Optimal Online Reinforcement Learning Algorithm for Continuous Deterministic Systems. *IEEE Transactions on Neural Networks and Learning Systems* 26(2), 346–356 (2015)
3. Zhao, D., Xia, Z., Wang, D.: Model-Free Optimal Control for Affine Nonlinear Systems With Convergence Analysis. *IEEE Transactions on Automation Science and Engineering* (2015), doi:10.1109/TASE.2014.2348991
4. Alippi, C., Ferrero, A., Piuri, V.: Artificial intelligence for instruments and measurement applications. *IEEE Instrumentation & Measurement Magazine* 1(2), 9–17 (1998)
5. Al-Tamimi, A., Abu-Khalaf, M., Lewis, F.L.: Adaptive critic designs for discrete-time zero-sum games with application to control. *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics* 37(1), 240–247 (2007)
6. Abu-Khalaf, M., Lewis, F.L., Huang, J.: Policy iterations on the Hamilton-Jacobi-Isaacs equation for state feedback control with input saturation. *IEEE Transactions on Automatic Control* 51(12), 1989–1995 (2006)
7. Vamvoudakis, K.G., Lewis, F.L.: Online solution of nonlinear two-player zero-sum games using synchronous policy iteration. *International Journal of Robust and Nonlinear Control* 22(13), 1460–1483 (2012)
8. Sahoo, A., Xu, H., Jagannathan, S.: Event-based optimal regulator design for nonlinear networked control systems. In: 2014 IEEE Symposium on Adaptive Dynamic Programming and Reinforcement Learning, pp. 1–8. IEEE Press, Orlando (2014)
9. Zhong, X., Ni, Z., He, H., Xu, X., Zhao, D.: Event-triggered reinforcement learning approach for unknown nonlinear continuous-time system. In: 2014 International Joint Conference on Neural Networks, pp. 3677–3684. IEEE Press, Beijing (2014)
10. Vamvoudakis, K.G.: Event-triggered optimal adaptive control algorithm for continuous-time nonlinear systems. *IEEE/CAA Journal of Automatica Sinica* 1(3), 282–293 (2014)
11. Chowdhary, G., Johnson, E.: Concurrent learning for convergence in adaptive control without persistency of excitation. In: 49th IEEE Conference on Decision and Control (CDC), pp. 3674–3679. IEEE Press, Atlanta (2010)
12. Modares, H., Lewis, F.L., Naghibi-Sistani, M.B.: Integral reinforcement learning and experience replay for adaptive optimal control of partially-unknown constrained-input continuous-time systems. *Automatica* 50(1), 193–202 (2014)