A Self-triggering Control Based on Adaptive Dynamic Programming for Nonzero-sum Game Systems

Yibo Shi and Chaoli Wang
Department of Control Science and Engineering, University of Shanghai for Science and Technology, Shanghai 200093, China

E-mail: clcwang@usst.edu.cn
https://www.usst.edu.cn/main.htm

Abstract

Recently, for the optimal control problem of nonzero-sum game systems, although it is discussed that these methods are event-triggered, it is still necessary to continuously monitor measurement errors during execution, which is difficult to achieve by hardware. In order to avoid continuous detection measurement errors, a self-triggered control based on adaptive dynamic programming is proposed to solve the optimal control problem for continuous-time nonlinear nonzero-sum game systems with unknown drift dynamics. Firstly, the principle of IRL method is used to avoid the requirement of system drift dynamics in the controller design. Then, to approximate the Nash equilibrium solution, a critic neural network is used to estimate the value function. Furthermore, a self-triggered adaptive control scheme is proposed according to Lyapunov theory to ensure the uniform ultimate boundedness (UUB) of the closed-loop system state. The self-triggered control obtained in this paper can calculate the next trigger point by the information of the current trigger moment.

Keywords: Nonzero-sum games, Integral reinforcement learning, Optimal control, Neural network, Self-triggered

1. Introduction

In recent years, the event-triggered control has attracted many researchers because of its ability to reduce transmission load and computation cost. Unlike time-triggered systems, the control inputs of event-triggered systems are updated only at trigger times determined by appropriately designed trigger conditions. In this way, event-triggered control can reduce network bandwidth and computational burden. For multi-player NZS games, event-triggered ADP has become the most common method used to approximate the control input of each player in [1-4]. Mu, Wang, Ni[1] proposed a dynamic event-triggered method for solving the optimal control problem of a fully known continuous-time nonzero-sum game system. Su, Zhang, Liang, etc. [2] used identifier-critic NN to solve continuous time for partially unknown NZS games. Su, Zhang, Sun, etc.[3] used IRL to solve the optimal control for the partially unknown NZS game. Compared with [2], the introduction of identifier NN was avoided, i.e., the identification error was avoided. Zhao, Sun, Wang, etc.[4] proposed an event-triggered ADP is proposed for NZS games of continuous-time nonlinear systems with completely unknown system dynamics. In order to determine whether event-triggered conditions are met, these papers need a general event-triggered scheme to continuously monitor measurement errors. However, it is hard to realize through the hardware. It may weaken its feasibility of physical implementation.

To overcome such an issue, the self-triggered scheme is presented. By the designed triggered condition, the next trigger time instant can be computed by the current one. Moreover, the self-triggered scheme can also save network resources and reduce the communication burden, the same as the event-triggered scheme. To the best of our knowledge, no similar result has been reported in the literature. Based on the above motivation, we focus on a self-triggered nonzero-sum game system based optimal control problem, which avoids continuous monitoring measurement errors.

2. Preliminaries and Problem Formulation

Consider the general N -player NZS differential games[5]

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

where $x \in \mathbb{R}^n$ is the system state, $u_j \in \mathbb{R}^{m_j}$ is the control input for player j. $f(x) \in \mathbb{R}^n$ and $g_j(x) \in \mathbb{R}^{n \times m_j}$ are the unknown drift dynamics and the known input dynamics of the system, respectively.

Assumption 1.[6] $f(x)$ and $g_j(x)$ are Lipschitz continuous on a compact set $\Omega \subset \mathbb{R}^n$ with $f(0) = 0$, $f(\cdot) \le b_f \|x\|$, and $\|g_j(x)\| \le b_{g,j}$, where b_f and $b_{g,j}$ are positive constants.

Define the cost function of system (1) as

$$J_i(x(t), u_i, u_{-i}) = \int_t^{\infty} \left(x^T(\tau) Q_i x(\tau) + \sum_{j=1}^{N} u_i^T(\tau) R_{ij} u_j(\tau) \right) d\tau \tag{2}$$

where $u_{-i} = \{u_1, \dots, u_{i-1}, u_{i+1}, \dots, u_N\}$, $Q_i = Q_i^T \ge 0$, $R_{ii} = R_{ii}^T > 0, R_{ij} = R_{ij}^T \ge 0$.

For a given set of control policies $\{u_i, u_{-i}\}$, the value function for player i can be written

$$V_i(x(t)) = \int_t^{\infty} \left(x^T(\tau) Q_i x(\tau) + \sum_{j=1}^{N} u_i^T R_{ij} u_j \right) d\tau, i \in \mathbb{N}. \tag{3}$$

The optimal control problem is to design a set of control $\{u_1^*, u_2^*, \dots, u_N^*\}$ to stabilize the closed-loop system while minimizing the value function (3). This control combination $\{u_1^*, u_2^*, \dots, u_N^*\}$ corresponds to the Nash equilibrium of NZS games.

The value function $V_i(x(t))$ is assumed to be continuously differentiable. By differentiating V_i along the system trajectories (1), we can write Eq. (3) as:

$$0 = U_i\left(x(t), u_i, u_{-i}\right) + \nabla V_i^T(x(t)) \left(f(x(t)) + \sum_{j=1}^{N} g_j(x(t)) u_j \right) \tag{4}$$

where

$$U_i\left(x(t), u_i, u_{-i}\right) = x^T(t) Q_i x(t) + \sum_{j=1}^{N} u_i^T R_{ij} u_j,\ \nabla V_i = \frac{\partial V_i}{\partial x}.$$

The optimal value function V_i^* can be written as

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

Define the Hamilton-Jacobi-Bellman (HJB) equation as follow:

$$H_i(x, \nabla V_i^*(x(t)), u_i, u_{-i}) =$$
$$U_i(x(t), u_i, u_{-i}) + (\nabla V_i^*(x(t)))^T \left(f(x(t)) + \sum_{j=1}^{N} g_j(x(t)) u_j \right) \tag{5}$$

Using the stationarity conditions $\partial H_i / \partial u_i = 0$ the optimal control input for player i is

$$u_i^*(x(t)) = -\frac{1}{2} R_{ii}^{-1} g_i^T(x(t)) \nabla V_i^*(x(t)) \tag{6}$$

The equivalent transformation of Eq. (6) is

$$V_i(x(t-\Delta t)) = V_i(x(t)) + \int_{t-\Delta t}^{t} U_i(x(\tau), u_i, u_{-i}) d\tau \tag{7}$$

where $\Delta t > 0$ is a time interval.

According to (7), we have

$$V_i^*(x(t-\Delta t)) - V_i^*(x(t)) = \int_{t-\Delta t}^{t} U_i(x(\tau), u_i^*, u_{-i}^*) d\tau \tag{8}$$

It's easy to see from Eq. (8) that there are no more unknown dynamics. Therefore, the identification process of unknown dynamics $f(x)$ is no longer needed, that is, identification error is avoided.

3. Design Self-triggered Control and Stability Analysis

3.1. *Design Self-triggered Control*

From the above time-triggered mechanism, it can be seen that the N-tuple control input $\{u_1, \dots, u_N\}$ is a feedback form of system state updated at each sample time. In self-triggered control, the N-tuple control input $\{u_1, \dots, u_N\}$ is updated only at the trigger time, and the next trigger time is determined by the current trigger time. In this case, a zero-order holder (ZOH) can be used to ensure that the control input is continuous at the trigger time. Define the triggering instant as τ_k, where $\{\tau_k\}_{k=0}^{\infty}$ is a monotonically increasing sequence of time instants with $\tau_0 = 0$. The trigger error is defined as

$$e_k(t) = \breve{x}_k - x(t), t \in [\tau_k, \tau_{k+1}) \tag{9}$$

where $\breve{x}_k = x(\tau_k)$ is the trigger state.

In the framework of self-triggered, the optimal control input (6) can be written as

$$u_i^*(\breve{x}_k) = -\frac{1}{2} R_{ii}^{-1} g_i^T(\breve{x}_k) \nabla V_i^*(\breve{x}_k) \tag{10}$$

where $\nabla V_i^*(\breve{x}_k) = \frac{\partial V_i^*}{\partial x}|_{x=\breve{x}_k}$.

The piecewise continuous control signal can be expressed by a ZOH

$$u_i^*(t)=\begin{cases} u_i^*(\breve{x}_k),\ t\in[\tau_k,\tau_{k+1}) \\ -\frac{1}{2}R_{ii}^{-1}g_i^T(\breve{x}_{k+1})\nabla V_i^*(\breve{x}_{k+1}),\ \ t=\tau_{k+1} \end{cases} \quad (11)$$

In the above analysis, We can get the solution of the optimal control (10) ultimately comes down to the solution of (8), which can be solved by using the critic NN. According to the Weierstrass high-order approximation theorem, we can get

$$\begin{aligned} V_i^*(x) &= \omega_i^{*T}\phi_i(x)+\varepsilon_i(x), \\ \nabla V_i^*(x) &= \nabla\phi_i^T(x)\omega_i^*+\nabla\varepsilon_i(x) \end{aligned} \quad (12)$$

where $\omega_i^*\in\mathbb{R}^{K_i}$ is the unknown ideal weight, $\phi_i:\mathbb{R}^n\to\mathbb{R}^{K_i}$ are linearly independent activation functions, K_i denotes the number of neurons, and ε_i is the approximation error.

Assumption 2.[7] (1) The approximation error $\varepsilon_i(x)$ and its gradient $\nabla\varepsilon_i(x)$ are bounded on Ω, i.e., $\|\varepsilon_i(x)\|\le b_{\varepsilon_i}$, and $\|\nabla\varepsilon_i(x)\|\le b_{\nabla\varepsilon_i}$, with $b_{\varepsilon_i}, b_{\nabla\varepsilon_i}$, being positive constants.

(2) The activation function $\phi_i(x)$ and its gradient $\nabla\phi_i(x)$ are bounded on Ω, i.e., $\|\phi_i(x)\|\le b_{\phi_i}$ and $\|\nabla\phi_i(x)\|\le b_{\nabla\phi_i}$, with $b_{\phi_i}, b_{\nabla\phi_i}$, being positive constants.

According to Eq. (8) and Eq. (12), it can be obtained

$$\begin{aligned} e_i(t) &= \omega_i^{*T}\left[\phi_i(x(t))-\phi_i(x(t-\Delta t))\right] \\ &+\int_{t-\Delta t}^{t}U_i(x(\tau),u_i^*(\breve{x}_k),u_{-i}^*(\breve{x}_k))d\tau \end{aligned} \quad (13)$$

where $e_i(t)=\varepsilon_i(x(t-\Delta t))-\varepsilon_i(x(t))$ is error from the NN approximation error. According to Assumption 2, $e_i(t)$ is bound on Ω, i.e., $\|e_i(t)\|\le b_{e,imax}$, where $b_{e,imax}$ is a positive constant.

Denote $\hat{\omega}_i$ as the estimations of ω_i^*. Then the value function can be approximated as

$$\hat{V}_i(x)=\hat{\omega}_i^T\phi_i(x) \quad (14)$$

Based on (10), the approximate control inputs are

$$\hat{u}_i(\breve{x}_k)=-\frac{1}{2}R_{ii}^{-1}g_i^T(\breve{x}_k)\nabla\phi_i^T(\breve{x}_k)\hat{\omega}_i \quad (15)$$

Using $\hat{V}_i(x)$ to replace $V_i^*(x)$ in Eq. (8). Therefore, the Bellman equation (8) can be written

$$\hat{e}_i(t)=\hat{\omega}_i^T\rho_i(t)+s_i(t) \quad (16)$$

where

$$\begin{aligned} \rho_i(t) &= \phi_i(x(t))-\phi_i(x(t-\Delta t)), \\ s_i(t) &= \int_{t-\Delta t}^{t}U_i(x(\tau),\hat{u}_i(\breve{x}_k),\hat{u}_{-i}(\breve{x}_k))d\tau \end{aligned}$$

As can be seen from Eq. (16), adjusting $\hat{\omega}_i$ can directly affect $\hat{e}_i(t)$. Then the original problem of solving the value function is transformed into minimizing the error $\hat{e}_i(t)$ by adjusting $\hat{\omega}_i$. Consider the objective function

$$E_i(t)=\frac{1}{2}\hat{e}_i^T(t)\hat{e}_i(t) \quad (17)$$

According to the gradient descent method, the update rule of $\hat{\omega}_i$ can be obtained as

$$\dot{\hat{\omega}}_i(t)=-\alpha_i\frac{\rho_i(t)}{\left(1+\rho_i^T(t)\rho_i(t)\right)^2}\left(s_i(t)+\rho_i^T(t)\hat{\omega}_i(t)\right) \quad (18)$$

Denote $\tilde{\omega}_i=\omega_i^*-\hat{\omega}_i$ is the critic weight estimation error and find that $\dot{\tilde{\omega}}_i=-\dot{\hat{\omega}}_i$.

$$\begin{aligned} \dot{\tilde{\omega}}_i(t) &= -\frac{\alpha_i\rho_i(t)\rho_i^T(t)\tilde{\omega}_i(t)}{\left(1+\rho_i^T(t)\rho_i(t)\right)^2} \\ &+\alpha_i\frac{\rho_i(t)}{\left(1+\rho_i^T(t)\rho_i(t)\right)^2}(s_i(t)+\rho_i^T(t)\omega_i^*(t)) \end{aligned} \quad (19)$$

3.2. *Stability Analysis*

Before we discuss the stability of closed-loop systems, we introduce the following assumptions in [5, 6, 8].

Assumption 3. Let the signals $\bar{\rho}_i$ be persistently exciting over the interval $[t, t+T_1]$, i.e., there exist constants $\beta_1>0, \beta_2>0$ such that, for all t,

$$\beta_1 I\le\int_t^{t+T_1}\bar{\rho}_i(\tau)\bar{\rho}_i^T(\tau)d\tau\le\beta_2 I \quad (20)$$

where $\bar{\rho}_i=\rho_i/(\rho_i^T\rho_i+1)$, $i=1,2,\ldots,N$, and I is the identity matrix.

Assumption 4. For $\forall i\in\mathbb{N}$, the control input u_i^* is locally Lipschitz with respect to $e_k(t)$. That is, there exists a constant $L_{u,i}>0$ satisfying that $\|u_i^*(x)-u_i^*(\breve{x}_k)\|^2\le L_{u,i}\|e_k(t)\|^2$.

Theorem 1. Suppose that Assumptions (1)-(4) holds. For the system (1), the critic NN is updating by (18) and the following self-triggered condition

$$\tau_{k+1}=inf\left\{t\,|\,t>\tau_k\cap\left[\frac{\Lambda}{b_f}(e^{b_f(t-\tau_k)}-1)\le\sqrt{\frac{\mathrm{U}(\breve{x}_k)+e^{-\ell t}}{\mathrm{L}}}\right]\right\} \quad (21)$$

is adopted. Then, the close-loop system state and the critic NN weight estimation error $\tilde{\omega}_i$ are all UUB, where $\ell>0$ is the decaying rate,

$$\mathrm{L}=\sum_{i=1}^{N}\left((2\lambda_{max}(R_{ii})+(N-1)b_{g,i}^2)L_{u,i}^2\right),$$

$$U(\breve{x}_k) = \sum_{i=1}^{N} \lambda_{min}(R_{ii}) \| \hat{u}_i(\breve{x}_k) \|^2,$$

$$\Lambda = b_f \| \breve{x}_k \| + \frac{1}{2} \sum_{j=1}^{N} b_{g,j}^{2} \lambda_{max}(R_{jj}^{-1}) b_{\nabla \phi_j} b_{\hat{\omega}_j}.$$

Proof. Due to the limited space, the detailed proof here is omitted. We can send the detailed proof to the readers in need.

4. Conclusion

In this paper, we study the optimal control of non-zero-sum game systems with unknown drift dynamics based on self-triggered control. The IRL method is used to avoid the need of unknown dynamics in the solution process. The solution of Nash equilibrium is obtained by constructing a single layer critic NN. By designing a reasonable self-triggered condition, the calculation and communication burden in the whole control process are reduced and no longer requires continuous detection of measurement errors. The UUB properties of the close-loop state and the critic NN estimation error are proved.

References

1. Mu C X, Wang K, Ni Z. Adaptive learning and sampled-control for nonlinear game systems using dynamic event-triggering strategy[J]. IEEE Transactions on Neural Networks and Learning Systems, 2022.
2. Su H G, Zhang H G, Liang Y L, et al. Online event-triggered adaptive critic design for nonzero-sum games of partially unknown networked systems[J]. Neurocomputing, 2019, 368 (9): 84–98.
3. Su H G, Zhang H G, Sun S X, et al. Integral reinforcement learning-based online adaptive event-triggered control for nonzero-sum games of partially unknown nonlinear systems[J]. Neurocomputing, 2020, 377: 243–255.
4. Zhao Q T, Sun J, Wang G, et al. Event-triggered ADP for nonzero-sum games of unknown nonlinear systems[J]. IEEE Transactions on Neural Networks and Learning Systems, 2021, 33(5): 1905-1913.
5. Vamvoudakis K, Lewis F. Multi-player nonzero-sum games: Online adaptive learning solution of coupled Hamilton-Jacobi equations[J]. Automatica, 2011, 47 (8): 1556–1569.
6. Modares H, Lewis F. Optimal tracking control of nonlinear partially-unknown constrained-input systems using integral reinforcement learning[J]. Automatica. 2014, 50 (7): 1780–1792.
7. Modares H, Lewis F, Sistani N. Integral reinforcement learning and experience replay for adaptive optimal control of partially-unknown constrained-input continuoustime systems[J]. Automatica, 2014, 50 (1): 193–202.
8. Yang X, Wei Q L. Adaptive critic learning for constrained optimal event-triggered control with discounted cost[J]. IEEE Transactions on Neural Networks and Learning Systems, 2021, 32 (1): 91–104.

Authors Introduction

Dr. Yibo Shi

He received the M.Sc. degree in the University of Shanghai for Science and Technology in 2022. He is currently pursuing the Ph.D. degree in control science and engineering at the University of Shanghai for Science and Technology, Shanghai, China. His research interests include adaptive control, optimal control, event-triggered, and game theory.

Dr. Chaoli Wang

He received the B.S. and M.Sc. degrees from Mathematics Department, Lanzhou University, Lanzhou, China, in 1986 and 1992, respectively, and the Ph.D. degree in control theory and engineering from the Beijing University of Aeronautics and Astronautics, Beijing, China, in 1999. He is a Professor with the School of Optical-Electrical and Computer Engineering, University of Shanghai for Science and Technology, Shanghai, China. His current research interests include nonlinear control, robust control, robot dynamic and control, visual servoing feedback control, and pattern identification.