Neural Networks 124 (2020) 95–108

Contents lists available at ScienceDirect

Neural Networks

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

Differential-game for resource aware approximate optimal control of large-scale nonlinear systems with multiple players

Avimanyu Sahoo [a,*], Vignesh Narayanan [b]

[a] 555 Engineering North, Division of Engineering Technology, Oklahoma State University, Stillwater, OK 74078, United States of America
[b] Washington University, St. Louis, MO, United States of America

ARTICLE INFO

Article history:
Received 5 April 2019
Received in revised form 8 December 2019
Accepted 30 December 2019
Available online 14 January 2020

Keywords:
Approximate dynamic programming
Event-driven control
Neural network control
Nonzero sum game
Optimal control

ABSTRACT

In this paper, we propose a novel differential-game based neural network (NN) control architecture to solve an optimal control problem for a class of large-scale nonlinear systems involving N-players. We focus on optimizing the usage of the computational resources along with the system performance simultaneously. In particular, the N-players' control policies are desired to be designed such that they cooperatively optimize the large-scale system performance, and the sampling intervals for each player are desired to reduce the frequency of feedback execution. To develop a unified design framework that achieves both these objectives, we propose an optimal control problem by integrating both the design requirements, which leads to a multi-player differential-game. A solution to this problem is numerically obtained by solving the associated Hamilton-Jacobi (HJ) equation using event-driven approximate dynamic programming (E-ADP) and artificial NNs online and forward-in-time. We employ the critic neural networks to approximate the solution to the HJ equation, i.e., the optimal value function, with aperiodically available feedback information. Using the NN approximated value function, we design the control policies and the sampling schemes. Finally, the event-driven N-player system is remodeled as a hybrid dynamical system with impulsive weight update rules for analyzing its stability and convergence properties. The closed-loop practical stability of the system and Zeno free behavior of the sampling scheme are demonstrated using the Lyapunov method. Simulation results using a numerical example are also included to substantiate the analytical results.

1. Introduction

Event-driven sampling and control (Tabuada, 2007; Wang & Lemmon, 2011) is a resource aware control scheme to reduce the computational burden on the embedded controllers without significant loss of performance. This control scheme is most suitable for large-scale systems, such as multi-player systems (Johnson, Kamalapurkar, Bhasin, & Dixon, 2015; Vamvoudakis & Lewis, 2011), multi-agent systems (Ma et al., 2016), interconnected systems (Narayanan & Jagannathan, 2018), and networked control systems (NCS) (Sahoo & Jagannathan, 2017), which require significant amount of communication and computational resources. Event-driven sampling schemes use smart sensors with additional electronics to evaluate the system stability and performance to determine the sampling instants, and in the resulting event-driven control approach the system is allowed to run as an open-loop system, i.e., without any feedback based control update, until a significant change in the stability or the performance of the system is observed. In other words, the control policy is only updated on the verge of instability or loss of performance.

The most important aspect of this resource aware control approach is the design of the event-driven sampling condition for determining the sampling and the control execution instants. Various approaches have been reported in the literature to design these conditions for both single (Tabuada, 2007; Wang & Lemmon, 2011) and multi-agent systems (Ma et al., 2016). For example, in the state-based sampling conditions, the state error (introduced by the aperiodic sampling scheme) is compared with a state dependent threshold (Tabuada, 2007; Tallapragada & Chopra, 2014), or a constant threshold (Albattat, Gruenwald, & Yucelen, 2016), and an event is triggered when the state error crosses the threshold. Since the state dependent threshold is designed such that it converges to zero as the system state converges to the desired equilibrium point, asymptotic stability results for the event-driven system can be achieved (Tabuada, 2007). However, the constant threshold based conditions lead to practical stability (Albattat et al., 2016). Alternatively, the event-driven sampling conditions are also designed using a dynamic triggering threshold, which incorporates a dynamic variable in the threshold function such that this variable is adaptively

* Corresponding author.
E-mail addresses: avimanyu.sahoo@okstate.edu (A. Sahoo), vnxv4@mst.edu (V. Narayanan).

https://doi.org/10.1016/j.neunet.2019.12.031
0893-6080/

updated based on some stability/performance requirements (Girard, 2015).

On the other hand, to deal with system uncertainties, artificial neural networks (NNs) are employed in tandem with event-triggered learning to develop NN control schemes (Narayanan, Jagannathan, & Ramkumar, 2018; Sahoo, Xu, & Jagannathan, 2013, 2016). Although these approaches (Narayanan et al., 2018; Sahoo et al., 2013, 2016) relaxed the requirement of system knowledge, their primary objective was to ensure stability of the closed-loop system with reduced computation. In addition to the reduction in resource utilization, optimization of a desired performance index, such as minimizing fuel cost, following the shortest path, and navigating a path in minimum time, etc., is equally important for control applications. To this end, optimal control policies for linear and nonlinear systems are designed by solving the associated Riccati and Hamilton-Jacobi-Bellman (HJB) equations, respectively (Lewis, Vrabie, & Syrmos, 2012). Neural networks have been instrumental in approximating the solution to the HJB equation using Bellman or temporal difference errors (Bellman, 2013; Bertsekas, 1995) since it is almost impossible to obtain an analytical solution to the HJB equation for nonlinear systems, and these approaches are commonly referred to as approximate dynamic programming (ADP) (Bertsekas, 1995; Werbos, 2007) and reinforcement learning (RL) (Sutton & Barto, 1998) based control schemes. A major challenge associated with the ADP-based control approaches, e.g., policy iteration (Liu & Wei, 2014) and value iteration (Heydari, 2017), is the requirement of a large number of iterative updates for convergence (Bertsekas, 1995) of the learning algorithm, limiting their real-time implementation. Therefore, various online and real-time ADP schemes were also reported in the literature, for e.g., Dierks and Jagannathan (2012).

To further reduce the computational burden of the ADP schemes and NN learning, event-driven ADP (E-ADP) schemes have also been introduced to obtain approximate optimal solutions online and forward-in-time using aperiodic feedback information (for e.g., Dong, Zhong, Sun, & He, 2017; Liu & Jiang, 2015; Sahoo, Narayanan, Vignesh & Jagannathan, 2017; Wang, He, Zhong, & Liu, 2017). It was demonstrated in the works presented in Dong et al. (2017), Liu and Jiang (2015), Sahoo, Narayanan et al. (2017) and Wang et al. (2017) that the E-ADP schemes can reduce the computational burden while guaranteeing near optimal performance. Moreover, the event-driven sampling conditions were adaptively (Sahoo, Xu & Jagannathan, 2017) updated to ensure the approximation accuracy of the value function along with optimality and stability. One commonality in these schemes (Dong et al., 2017; Liu & Jiang, 2015; Narayanan & Jagannathan, 2016; Sahoo, Narayanan et al., 2017; Wang et al., 2017) is that the sampling intervals were not optimized based on any performance criterion, and in general, the sampling instants were selected such that the stability of the system was retained while executing the near optimal control policies at these sampling instants with the sampled states.

For large-scale systems (Johnson et al., 2015; Vamvoudakis & Lewis, 2011) with multiple players, each player's performance is influenced by the others. Therefore, optimization of the system performance requires a cooperative solution that accounts for the action of each player. Differential game theory-based approaches (Friedman, 2013) are employed to solve such optimization problems to reach an equilibrium solution, known as Nash solution (Nash, 1951; Starr & Ho, 1969). The Nash equilibrium does not allow any single player to improve its own performance by unilaterally modifying its policy (Nash, 1951). A special case of the N-player differential game is the two-player non-cooperative zero-sum game where the Nash equilibrium solution is a saddle point solution (Başar & Bernhard, 1995).

An ample amount of research results on the application of multi-player differential game theory (Friedman, 2013; Starr & Ho, 1969) for control design of N-player systems are available in the literature (Johnson et al., 2015; Liu, Li, & Wang, 2014; Vamvoudakis & Lewis, 2011; Zhao, Zhang, Wang, & Zhu, 2016). For systems with uncertain nonlinear dynamics ADP/RL based approaches using NNs were proposed in Johnson et al. (2015), Liu et al. (2014), Vamvoudakis and Lewis (2011) and Zhao et al. (2016) to obtain the non-zero sum (NZS) feedback Nash equilibrium. The work presented in Vamvoudakis and Lewis (2011) reported an online ADP approach for both linear and nonlinear systems using policy iteration with actor–critic network architecture. Various other works that study approximate Nash solution for systems with uncertain dynamics were also reported in the literature (for e.g., Johnson et al., 2015; Liu et al., 2014; Song, Lewis, & Wei, 2017; Zhao et al., 2016). With the requirement of continuous availability of the state information, the aforementioned multi-player game based designs (Case, 1969; Johnson et al., 2015; Kamalapurkar, Klotz, & Dixon, 2014; Liu et al., 2014; Song et al., 2017; Vamvoudakis & Hespanha, 2018; Vamvoudakis & Lewis, 2011; Zhao et al., 2016) lead to a significant amount of computation and communication cost.

To maximize the advantages of event-driven sampling and control for multi-player systems, the sampling intervals must also be optimized along with the control policies. In our previous work (Sahoo, Narayanan et al., 2017), a unified design scheme using a zero-sum game based approach was proposed for designing the control and the sampling scheme. However, the work in Sahoo, Narayanan et al. (2017) cannot be trivially extended to the multi-player domain due to the fact that the objective function must take into account both the non-cooperative (player level) and cooperative (system level) optimization, which is a challenging open problem.

Therefore, in this work, we present a resource aware differential game-based event-driven sampling and near optimal NN control scheme for a nonlinear N-player system. The optimal control problem is formulated as a cooperative non-zero sum game to obtain an equilibrium solution for all the players, whereas the optimization of sampling intervals and policy of each player is treated as non-cooperative min–max problem to optimize the use of computational resources. To make the problem tractable, a single novel performance index is introduced for each player by combining both these objectives such that the solution to this optimization problem when employed to design the controllers and the sampling mechanisms for the large-scale system leads to a desired system performance with limited computational effort.

A single critic NN is designed at each player and updated using event-based Bellman error to obtain an approximate solution for the associated Hamilton–Jacobi (HJ) equation. Using the NN approximated solution, the control policy for each player is designed, and the worst-case threshold value for the control policy error due to aperiodic event-driven sampling is also computed, which is then used to design the sampling conditions. The critic NN weights at each player are updated both at the sampling instants and during the inter-sample times resulting in an impulsive weight update scheme. The practical stability of the event-driven N-player system under aperiodic sampling is guaranteed using the extension of the Lyapunov stability theory for hybrid systems. Further, the Zeno free behavior of the sampling scheme is ensured by enforcing a lower bound on the adaptive sampling threshold. The lower bound rules out the situation for the adaptive threshold to become zero during the learning period of the NN. An analytical formula is also presented showing that the inter-sample times are lower bounded by a non-zero positive constant. Some preliminary results related to this work was reported in Sahoo, Narayanan, and Jagannathan (2018). In contrast, we reformulate the performance index to accommodate the design requirements, and, therefore, solve an

entirely different optimization problem in this paper. Moreover, in this paper, we present both exact and approximate solutions for the optimization problem along with rigorous stability proofs and numerical validation.

Contributions. To the authors' best knowledge, a unified differential game based event-driven control scheme for N-player systems is not reported in the literature. The proposed approach enables the co-design of the sampling intervals and the optimal control policy of each player by using the non-cooperative game while cooperatively optimizing the overall system performance. The primary contributions of the paper are: (i) Definition of a novel performance index for optimizing the system performance and the sampling intervals for an N-player system; (ii) Design of the NN parameter update laws using event-driven Bellman error; (iii) Development of an improved event-driven sampling condition design to elongate sampling intervals for a multi-player large-scale system; and (iv) derivation of the stability and convergence results using the extension of Lyapunov stability theory and demonstration of Zeno freeness of the sampling schemes.

The rest of the paper is organized as follows. In Section 2, we present a brief background on N-player optimal control problem, and then, we define the problem addressed in this paper. In Section 3, we derive the exact solution for the proposed problem, and in Section 4, we present an approximate solution that is practically computable. In Sections 5 and 6, we present the simulation results and conclusions, respectively. The proofs for lemmas, theorems, and corollaries are detailed in the Appendix.

Notations. The n-dimensional Euclidean space is denoted by $\Re^n$, $\Re^+$ is the set of positive real numbers, and $\mathbb{N}$ is the set of natural numbers. Superscript * denotes the optimal value of the quantity. The norm operator $\|.\|$ denotes the Frobenius norm for matrices and the Euclidean norm for vectors. The notations $\lambda_{min}(\cdot)$ and $\lambda_{max}(\cdot)$ are the minimum and the maximum eigenvalues. The transpose operator is denoted by $(\cdot)^T$, the first difference by Δ, and the gradient by ∇.

2. Background and problem statement

2.1. Background

Consider the nonlinear dynamics of an N-player system in input affine form, represented by

$$\dot{x} = F(x) + \sum_{i=1}^{N} G_i(x)u_i, \tag{1}$$

where the system state vector is denoted by $x \in \Re^n$ with initial state $x(0) = x_0$ and the ith player's control policy by $u_i \in \Re^{m_i}$, $i = 1, 2, \ldots, N$. The vector function $F : \Re^n \to \Re^n$ with $F(0) = 0$ is the internal dynamics and matrix function $G_i : \Re^n \to \Re^{n\times m_i}$, $\forall i = 1, \ldots, N$ are the control coefficient function. The functions $F(\cdot)$ and $G(\cdot)$ are known and satisfy the following assumptions.

Assumption 1. The N-player system in (1) is stabilizable, and zero state observable. Further, all the state vectors are available for measurement (Vamvoudakis, Modares, Kiumarsi, & Lewis, 2017).

Assumption 2. The functions $F(x)$, and $G_i(x)$ are locally Lipschitz continuous $\forall x \in \Omega_x \subset \Re^n$ where Ω_x is a compact set.

Assumption 3. The matrix function $G_i(x)$ is uniformly bounded $\forall x \in \Omega_x$ and satisfies $0 < G_{im} < \|G_i(x)\| \leq G_{iM}$, $\forall i = 1, 2, \ldots, N$ where $G_{im} \in \Re^+$ and $G_{iM} \in \Re^+$ are constants.

Note that Assumptions 1 and 2 ensure the existence of the solution of the HJ equation formulated in Section 2 and are standard assumptions in the ADP/RL literature (Kamalapurkar, Dinh, Bhasin, & Dixon, 2015; Vamvoudakis et al., 2017).

The performance index for the ith-player in an N-player NZS differential game (Johnson et al., 2015) can be defined as

$$J_i(x_0; u_1, \ldots, u_N) = \int_0^\infty (r_i(x(\tau), u_1(\tau), \ldots, u_N(\tau)))d\tau, \tag{2}$$

where the utility function is denoted by $r_i(x, u_1, \ldots, u_N) = x^T Q_i x + u_i^T R_{ii} u_i + \sum_{j=1, j\neq i}^{N} u_j^T R_{ij} u_j$. The user defined penalty matrices $Q_i \in \Re^{n\times n} > 0$, $R_{ii} \in \Re^{m_i\times m_i} > 0$, and $R_{ij} > 0 \in \Re^{m_j\times m_j}$, $i = 1, 2, \ldots, N$ are symmetric matrices. Further, in this context, the control objective is to design a set of Nash equilibrium policies $(u_1^*, u_2^*, \ldots, u_N^*)$ by minimizing the performance index (2), given by

$$\min_{u_i} \quad J_i(x; u_1, \ldots, u_N), \quad \text{s.t. } \dot{x} = F(x) + \sum_{i=1}^{N} G_i(x)u_i.$$

The set of Nash equilibrium policies can be obtained by solving the corresponding HJ equations (Johnson et al., 2015). An analytical solution of the coupled partial differential HJ equation is, in general, impossible to obtain, and the solution to the HJ equation is approximated by using the ADP/RL based approach with NNs as approximators. Typically, the continuous system state vector $x(t)$ is used as the input to the NN. However, in an event-driven sampling and control framework for the N-player system, $x(t)$ is not available continuously due to adaptive aperiodical sampling. This results in a sparse input, in the time domain, to the NN approximator, which may affect the approximation accuracy and, in turn, the cost. Therefore, the performance optimization of N-player event-triggered system requires- (1) optimal sampling schemes to reduce approximation error and save computational resources by reaching a trade-off between system performance and resource utilization, and (2) optimal control policies. Next, we formalize these requirements and present the resulting control problem.

2.2. Problem formulation

In the event-driven sampling approach, a sampling mechanism orchestrates the aperiodic sampling instants. For an N-player system, each player requires a sampling mechanism to determine its controller execution instants with the sampled states and these sampling mechanisms are required to be asynchronous. Let the sampling instances for ith player sampling mechanism be defined by a sequence of time instants, $\{t_i^k\}_{k=0}^{\infty}$, with $t_i^k < t_i^{k+1}$ and $t_i^0 = 0$. The players will execute their policies only at time instants t_i^k, $\forall k \in \{0, \mathbb{N}\}$, $i = 1, 2, \ldots, N$ with the sampled state vector $x(t_i^k)$.

The state information at the ith player between two sampling instants can be defined as

$$x_i^e(t) = x(t_i^k), \quad t_i^k \leq t < t_i^{k+1}, \quad \forall k \in \{0, \mathbb{N}\}, \tag{3}$$

where $x_i^e(t)$ is the sampled state vector at the ith player. This implies that the previous sampled state information is retained until the next sample arrives.

The control policy with sampled state information x_i^e of the ith player is given by

$$u_i^e(t) = \zeta_i(x_i^e(t)), \quad t_i^k \leq t < t_i^{k+1}, \quad \forall k \in \{0, \mathbb{N}\}, \tag{4}$$

where $\zeta_i : \Re^n \to \mathbb{U} \subset \Re^{m_i}$ is a vector function where $\mathbb{U}$ is the set of all admissible control policies. The dynamics of the N-player system in (1) with sampled control input (4) is given as

$$\dot{x} = F(x) + \sum_{i=1}^{N} G_i(x)u_i^e, \quad i = 1, 2, \ldots, N. \tag{5}$$

Note that we dropped the argument t for brevity, i.e., $x(t)$ is written as x.

The sampled control policy, u_i^e, introduces an error in the system dynamics due to the aperiodic control execution when compared to (1). In this case, the dynamics in (5) can be rewritten as

$$\dot{x} = F(x) + \sum_{i=1}^{N} G_i(x)u_i + \sum_{i=1}^{N} G_i(x)e_i^u \tag{6}$$

where e_i^u is the control policy error due to the aperiodic sampling, and it is defined as the error between the continuous control policy u_i and sampled control policy u_i^e, i.e.,

$$e_i^u(t) = u_i^e(t) - u_i(t),\ i = 1, 2, \ldots, N,\quad t_i^k \le t < t_i^{k+1}. \tag{7}$$

Note that the control policy error, e_i^u, due to the aperiodic sampling depends on the length of the intervals $\delta_i^k = t_i^{k+1} - t_i^k, \forall k \in \{0, \mathbb{N}\}, i = 1, 2, \ldots, N$ and affects the system stability and performance cost. Therefore, an optimal control problem in the event-driven sampling and control context is to reach a trade-off between the system performance and the frequency of feedback. In a single player system, this can be formulated as a zero-sum game problem (Sahoo, Narayanan, & Jagannathan, 2019) and a saddle point solution can be obtained. However, for an N-player system, this problem leads to a combination of a zero-sum game at the player level and an NZS game at the system level.

In view of the above, the problem addressed in this paper can be defined as an event-driven hybrid optimization problem. Our objective is to develop a set of optimal event-driven control policies for the N-player large-scale system in the presence of worst-case control policy errors $e_i^u, i = 1, 2, \ldots, N$ such that the resulting control policies optimize the system performance cooperatively while reducing the required computations. An exact solution to the above stated problem is presented next.

3. Exact solution to the game

In this section, we formally define the optimization problem to ensure a desired system performance with reduced frequency of feedback and derive an exact solution to this problem.

The performance index in (2) can be redefined to formulate the proposed optimization problem, which can be decomposed into a zero-sum game at the player level and NZS game at the overall system level, and it is given as

$$J_i(x_0; u_1, \ldots, u_N, \bar{e}_i^u) = \int_0^\infty (r_i(x(\tau), u_1(\tau), \ldots, u_N(\tau), \bar{e}_i^u(\tau)))d\tau, \tag{8}$$

where $r_i(x, u_1, \ldots, u_N, \bar{e}_i^u) = x^T Q_i x + (u_i^T R_{ii} u_i - \gamma_{ii}^2 \bar{e}_i^{uT} \bar{e}_i^u) + \sum_{j=1, j\neq i}^{N} u_j^T R_{ij} u_j$. The signal $\bar{e}_i^u \in \Re^{m_i}, \forall i = 1, 2 \cdots, N$ is an exogenous signal, which will be used as the threshold for control policy error, e_i^u, to determine the sampling intervals. The penalty factors are given by $\gamma_{ii} > \gamma^* > 0$ for all $i, j = 1, \ldots, N$, where γ^* is the minimum value of γ such that the cost function is finite for all admissible control policies in the set $\mathbb{U}$.

Remark 1. The rationale behind such a definition of the performance index in (8) is that the minimization of $J_i, \forall i = 1, 2, \ldots, N$ will lead to an optimal control policy, u_i^*, which will attenuate the worst-case exogenous signal e_i^{u*} (optimal value of $\bar{e}_i^u$). The performance index (8) also uses other players' inputs as part of the minimization problem, which leads to a cooperative solution to reach Nash equilibrium (Starr & Ho, 1969). Alternatively, this performance index results in a zero-sum game for the ith player (when the other players are absent), and it leads to a non-zero-sum game for the overall system (when the feedback is continuous), which is an improved version of (2). As proved by the authors in Başar and Bernhard (1995) for zero-sum games, the worst-case exogenous signal e_i^{u*} is the limiting value of the error with guaranteed stability and can be used for maximizing the sampling intervals between two consecutive sampling instants for controller update.

By definition (Johnson et al., 2015), the set of optimal N-tuple $\{u_1^*, \ldots, u_i^*, \ldots, u_N^*, e_i^{u*}\}$ is the Nash equilibrium solution if the optimal value V_i^* at each player satisfies the inequality $V_i^*(x, u_1^*, \ldots, u_i^*, \ldots, u_N^*, e_i^{u*}) \le V_i^*(x, u_1^*, \ldots, u_i, \ldots, u_N^*, e_i^{u*}), i = 1, 2, \ldots, N$. The Hamiltonian $\mathcal{H}_i$ with dynamic state constraint in (5) can be defined as (Lewis et al., 2012)

$$\begin{aligned}\mathcal{H}_i(x, u_1, \ldots, u_N, \bar{e}_i^u, V_{ix}^*) &= x^T Q_i x + u_i^T R_{ii} u_i - \gamma_{ii}^2 \bar{e}_i^{uT} \bar{e}_i^u + \sum_{j=1, j\neq i}^{N} u_j^T R_{ij} u_j \\ &+ V_{ix}^{*T}[F(x) + \sum_{i=1}^{N} G_i(x)u_i + \sum_{i=1}^{N} G_i(x)e_i^u],\end{aligned} \tag{9}$$

where $V_{ix}^* = \partial V_i^*/\partial x,\ i = 1, 2, \ldots, N$. The following standard assumption is necessary to proceed further.

Assumption 4. The set of optimal value functions $V_i^*(x),\ i = 1, 2, \ldots, N$, which is the solution of the HJ equation, exists and is continuously differentiable. Further, the gradient of the value function V_{ix}^* is Lipschitz continuous $\forall x \in \Omega_x$ and satisfies $\|V_{ix}^*\| \le L_V \|x\|$ where $L_V \in \Re^+$ is a constant.

Remark 2. The assumption is standard in the ADP literature (Johnson et al., 2015; Kamalapurkar, Andrews, Walters, & Dixon, 2017; Vamvoudakis & Lewis, 2011) and valid for systems with local Lipschitz continuous internal dynamics, and which satisfy local stabilizability and zero state observability conditions (Vamvoudakis et al., 2017).

The optimal control policy $u_i^* = \arg\min_{u_i}\ \mathcal{H}_i(x, u_1, \ldots, u_N, \bar{e}_i^u, V_{ix}^*)$, by using the stationarity condition (Başar & Bernhard, 1995; Lewis et al., 2012), $\frac{\partial \mathcal{H}_i}{\partial u_i} = 0$, can be computed as

$$u_i^* = -\frac{1}{2} R_{ii}^{-1} G_i^T(x) V_{ix}^*,\ i = 1, 2, \ldots, N. \tag{10}$$

Similarly, the worst-case threshold $e_i^{u*} = \arg\max_{\bar{e}_i^u}\ \mathcal{H}_i(x, u_1, \ldots, u_N, \bar{e}_i^u)$ of the exogenous signal $\bar{e}_i^u$, can be computed using $\frac{\partial \mathcal{H}_i}{\partial \bar{e}_i^u} = 0$, and is given by

$$e_i^{u*} = \frac{1}{2\gamma_{ii}^2} G_i^T(x) V_{ix}^*,\ i = 1, 2, \ldots, N. \tag{11}$$

The HJ equation, after inserting the expression for the optimal control policies in (10) and the worst-case threshold values in (11), becomes

$$\begin{aligned}\mathcal{H}_i(x, u_1^*, \ldots, u_N^*, e_i^{u*}, V_{ix}^*) &= x^T Q_i x + u_i^{*T} R_{ii} u_i^* - \gamma_{ii}^2 e_i^{u*T} e_i^{u*} + \sum_{j=1, j\neq i}^{N} u_j^{*T} R_{ij} u_j^* \\ &+ V_{ix}^{*T}[F(x) + \sum_{i=1}^{N} G_i(x)u_i^* + \sum_{i=1}^{N} G_i(x)e_i^{u*}].\end{aligned} \tag{12}$$

The optimal control policy (10) with the sampled state x_i^e at ith controller can be expressed as

$$u_i^{e*} = -\frac{1}{2} R_{ii}^{-1} G_i^T(x_i^e) V_{ix_i^e}^*,\quad i = 1, 2, \ldots, N,\quad t_i^k \le t < t_i^{k+1} \tag{13}$$

where $V_{ix_i^e}^* = (\partial V_i^*(x)/\partial x)\big|_{x_i^e}$.

Note that the control policy error e_i^u in the dynamics (6) may drive the system to instability. Updating the control policy at the instant when e_i^u approaches the limiting value e_i^{u*} will elongate the sampling intervals while retaining the stability (proved in Theorem 1 later). To maximize the sampling intervals, we will use the worst-case exogenous signal (11) as the threshold for the control error policy (7), referred to as event-driven sampling condition, and it is given by

$$\|e_i^u(t)\| \leq \max\{r^2, \|e_i^{u*}(t)\|\}, \quad i = 1, \ldots, N \tag{14}$$

where $r \in \Re^+$ is a constant and $\|e_i^{u*}(t)\| = \|\frac{1}{2\gamma_{ii}^2} G_i^T(x) V_{ix}^*\|$. Note that each player will have its own sampling-mechanism to evaluate the sampling condition (14) to determine the sampling and the control execution instants. The next theorem claims the stability of the closed-loop event-driven system.

Theorem 1. *Consider the N-player event-sampled system in (5), the performance index (8), and the sampled optimal control policy (13). Let Assumptions 1–4 hold, the optimal value function $V_i^*(x)$ satisfies the HJ equation (12), and there exist constants $\gamma_{ii} \in \Re^+$ such that $\lambda_{min}(R_{ii}^{-1}) > \frac{1}{\gamma_{ii}^2}, \forall i = 1, \ldots, N$. Then, the closed-loop system states x is uniformly ultimately bounded (UUB) when the sampling instants are selected and control policies are updated at the violation of the inequality (14) and the gain condition satisfies $\frac{q_{i,min}}{2} > \frac{1}{2} N L_V^2 G_{iM}^2 (1 + \frac{1}{\gamma^{*2}})$, where $q_{i,min} = \lambda_{min}(Q_i)$.*

Proof. See Appendix.

Remark 3. From the proof of Theorem 1, with sampling condition defined in Case I of the proof, i.e., $\|e_i^u(t)\| \leq \|e_i^{u*}(t)\|$, the system states converge to zero asymptotically. However, enforcing a lower bound r^2 on the threshold (RHS) in sampling condition (14) leads to bounded stability as shown in Case II of the proof. The lower bound r^2 of the threshold ensures Zeno free behavior of the system as presented in the following corollaries.

Corollary 1. *Let the hypothesis of Theorem 1 hold. The sampled optimal control policies in (13) also render the event-driven system (5) UUB stable if the sampling instants are selected at the violation of the following inequality:*

$$L_\zeta \|e_i^x(t)\| \leq \max\{r^2, \|e_i^{u*}(t)\|\}, \forall i = 1, \ldots, N \tag{15}$$

where $e_i^x(t) = x(t) - x_i^e(t), \quad t_i^k \leq t < t_i^{k+1}, \quad \forall k \in \{0, \mathbb{N}\}, i = 1, \ldots, N$ is the state error at the ith player due to aperiodic sampling.

Proof. Refer to Appendix.

Remark 4. The event-driven sampling condition in (15) is a conservative condition when compared to the condition (14) and may lead to a higher number of events. Therefore, guaranteeing a lower bound on the inter-sample times generated by the condition (15) will also guarantee a lower bound on the inter-sample times generated by (14), i.e., the Zeno free behavior of the closed-loop sampled system.

In the next corollary, the inter-sample times are analyzed to rule out the Zeno behavior.

Corollary 2. *Let the hypotheses of Theorem 1 and Corollary 1 hold. Then, the sampled system (5) with sampled control policy (13) and sampling condition (14) is Zeno free, i.e., the inter-sample times $\delta_i^k = t_i^{k+1} - t_i^k, \forall k \in \{0, \mathbb{N}\}$ are lower bounded by a non zero positive number.*

Proof. Refer to Appendix.

Since a closed form solution to the HJ partial differential equation (12) is almost impossible to obtain Bellman (2013), an approximate solution using critic NNs for each player is presented next.

4. Approximate Nash equilibrium solution

In this section, an approximate solution to the HJ equation is presented using critic NNs in an event-driven sampling and control framework.

4.1. Critic neural network and control policy design

The block diagram of the proposed E-ADP based solution is shown in Fig. 1, where each player possesses a critic NN to approximate the value function and to compute its control policy while the sampling mechanism for each player evaluates its own sampling condition (defined latter) asynchronously. Note that the critic neural networks also receive other players' sampled control policies at the asynchronous sampling instants for cooperative policy design. In particular, each player uses the latest received control policy information while computing its own policy at the event-driven sampling instants.

Provided that a solution to the HJ equation, i.e., the optimal value function exists and it is continuously differentiable (Assumption 4), the universal approximation property of NN (Lewis, Jagannathan, & Yesildirak, 1998) can be invoked to approximate the solution to the HJ equation (12) in the compact set Ω_x. The selection of NNs as an approximator for the optimal value function is based on the fact that NNs can potentially learn a sparse representation when compared to spectral approximation type methods as discussed in Abu-Khalaf and Lewis (2005).

Therefore, the optimal value function in a NN based parametric form can be expressed as

$$V_i^*(x) = W_i^{*T} \phi_i(x) + \varepsilon_i(x) \tag{16}$$

where $W_i^* \in \Re^{l_{w_i}}$ is the unknown constant target weight vector defined as $W_i^* = arg \min_{W^*} \left(\sup_{x \in \Omega_x} \|W_i^{*T} \phi_i(x) - V_i^*(x)\| \right)$ and $\phi_i(x) \in \Re^{l_{w_i}}$ is the activation function, and $\varepsilon_i(x) \in \Re$ is the approximation error for the ith critic NN. Note that the approximation error $\varepsilon_i(x)$ is a function of system states and increases with the number of states. This requires the number of neurons in the NN to be increased to attain an arbitrary small approximation error (Lewis et al., 1998). Before we proceed further, the following standard assumption is necessary (Kamalapurkar et al., 2017, 2015, 2014; Lewis et al., 1998).

Assumption 5. The critic NNs' target weights, activation functions, gradient of activation functions, reconstruction errors, and gradient of reconstruction errors are bounded in compact sets and satisfy $\|W^*\| \leq W_{iM}, \sup_{x \in \Omega_x} \|\phi_i(x)\| \leq \phi_{iM}, \sup_{x \in \Omega_x} \|\nabla_x(\phi_i(x))\| \leq \phi'_{iM}, \sup_{x \in \Omega_x} \|\varepsilon_i(x)\| \leq \varepsilon_{iM}, \sup_{x \in \Omega_x} \|\nabla_x(\varepsilon_i(x))\| \leq \varepsilon'_{iM}$ for $i = 1, 2, \ldots, N$ where $\nabla_x(\cdot) = \partial(\cdot)/\partial x$, $W_{iM}, \phi_{iM}, \phi'_{iM}, \varepsilon_{iM}, \varepsilon'_{iM} \in \Re^+$ are constants.

Note that Assumption 5 can be realized with the proper selection of the activation function, weight parameters, and the number of neurons (Lewis et al., 1998). This is a standard assumption in the neural network literature (Kamalapurkar et al., 2017, 2015, 2014; Lewis et al., 1998).

Moving on, the optimal control policy (10) using the gradient of the value function (16) can be expressed as

$$u_i^* = -\frac{1}{2} R_{ii}^{-1} G_i^T [\nabla_x^T(\phi_i(x)) W_i^* + \nabla_x^T(\varepsilon_i)], \tag{17}$$

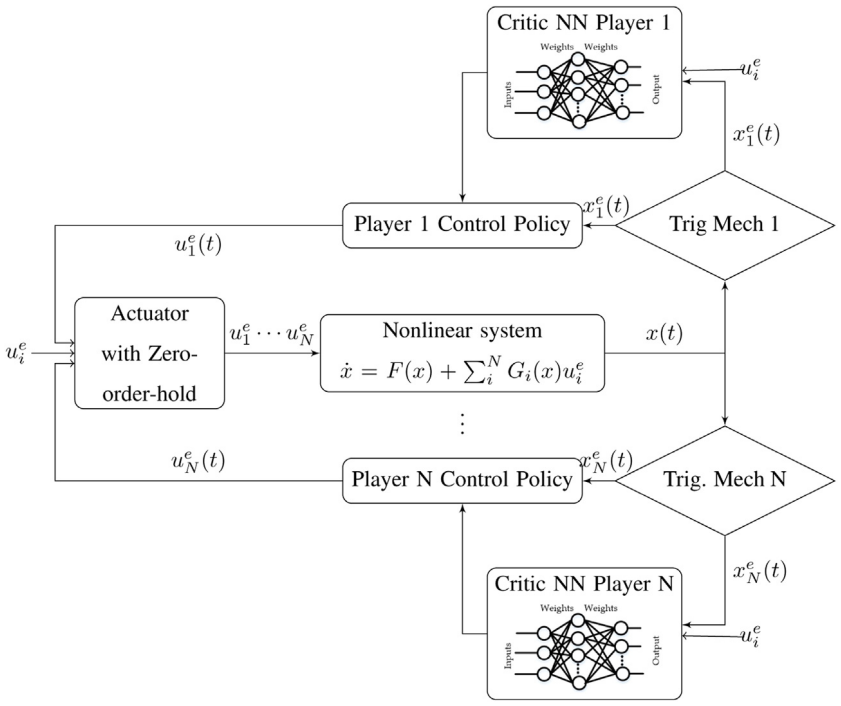

Fig. 1. Block diagram of the proposed E-ADP scheme for the N-player Nash solution.

for $i = 1, 2, \ldots, N$. Similarly, the worst-case threshold value in (11), in a parametric form, using the gradient of the value function (16) can be expressed as

$$e_i^{u*} = \frac{1}{2\gamma_{ii}^2} G_i^T [\nabla_x^T(\phi_i(x)) W_i^* + \nabla_x^T(\varepsilon_i)]. \tag{18}$$

The value function with estimated weight $\hat{W}_i \in \Re^{l_{w_i}}$ and sampled state information x_i^e can be expressed as

$$\hat{V}_i(x) = \hat{W}_i^T(t)\phi_i(x_i^e), \quad i = 1, 2, \ldots, N. \tag{19}$$

From (19), the estimated sampled control policy $\hat{u}_i^e$, $i = 1, 2, \ldots, N$, can be rewritten as

$$\hat{u}_i^e = -\frac{R_{ii}^{-1}}{2} G_{i,e}^T \hat{V}_{ix_i^e} = -\frac{R_{ii}^{-1}}{2} G_{i,e}^T \nabla_{x_i^e}^T(\phi_i(x_i^e))\hat{W}_i, \ \forall \ i = 1, 2, \ldots, N \tag{20}$$

where $G_{i,e} = G_i(x_i^e)$ and $\nabla_{x_i^e}(\cdot) = \partial(\cdot)/\partial x \mid_{x_i^e}$. Similarly, the estimated threshold $\hat{\bar{e}}_i^u$, $i = 1, 2, \ldots, N$ can be computed as

$$\hat{\bar{e}}_i^u = \frac{1}{2\gamma_{ii}^2} G_i^T \hat{V}_{ix_i^e} = \frac{1}{2\gamma_{ii}^2} G_i^T \nabla_{x_i^e}^T(\phi_i(x_i^e))\hat{W}_i, \ i = 1, 2, \ldots, N. \tag{21}$$

In the next step, we will update the critic NN weights using event-driven Bellman errors such that the weights converge close to their respective target values. Since the sampled control policies are executed aperiodically with time-varying sampling intervals $\delta_i^k = t_i^{k+1} - t_i^k$, $k \in \{0, \mathbb{N}\}$, the performance index (8) can be rewritten as

$$J_i(x_0; u_1, \ldots, u_N, \bar{e}_i^u) = \sum_{k=0}^{\infty} \int_{t_i^k}^{t_i^{k+1}} (r_i(x, u_1, \ldots, u_N, \bar{e}_i^u))d\tau. \tag{22}$$

By Bellman principle of optimality, the Bellman equation can be expressed as

$$0 = \int_{t_i^k}^{t_i^{k+1}} (r_i(x, u_1, \ldots, u_N, \bar{e}_i^u))d\tau + V_i^*(x(t_i^{k+1})) - V_i^*(x(t_i^k)). \tag{23}$$

Inserting the NN approximation of the optimal value function (16), the Bellman equation (23) leads to

$$0 = \int_{t_i^k}^{t_i^{k+1}} (r_i(x, u_1, \ldots, u_N, \bar{e}_i^u))d\tau + W_i^{*T} \Delta\phi_{i,k} + \Delta\varepsilon_{i,k} \tag{24}$$

where $\Delta\phi_{i,k} = \phi_i(x(t_i^{k+1})) - \phi_i(x(t_i^k))$ and $\Delta\varepsilon_{i,k} = \varepsilon_i(x(t_i^{k+1})) - \varepsilon_i(x(t_i^k))$. Since the estimated value function (19) does not satisfy the Bellman equation (23), the resulting error $E_{i,k}^e$ can be expressed as

$$E_{i,k}^e = \int_{t_i^{k-1}}^{t_i^k} (x^T Q_i x + u_i^T R_{ii} u_i - \gamma_{ii}^2 \bar{e}_i^{uT} \bar{e}_i^u + \sum_{j=1, j\neq i}^{N} u_j^T R_{ij} u_j)d\tau + \hat{V}_i(x(t_i^k)) - \hat{V}_i(x(t_i^{k-1})) \tag{25}$$

where $E_{i,k}^e$ is also referred to as the Bellman error or the temporal difference error. Recalling the estimated value of the critic NN (19), the Bellman error can be expressed as

$$E_{i,k}^e = \int_{t_i^{k-1}}^{t_i^k} (x^T Q_i x + u_i^T R_{ii} u_i - \gamma_{ii}^2 \bar{e}_i^{uT} \bar{e}_i^u + \sum_{j=1, j\neq i}^{N} u_j^T R_{ij} u_j)d\tau + \hat{W}_i \Delta\phi_{i,k-1} \tag{26}$$

where $\Delta\phi_{i,k-1} = \phi(x(t_i^k)) - \phi(x(t_i^{k-1}))$.

Remark 5. Note that the Bellman error in (26) is in a computable form and needs state and control policies information for the current and previous sampling instants, i.e., $x(t_i^k)$ and $x(t_i^{k-1})$ to evaluate it.

Now the objective is to update the estimated critic NN (19) such that it converges to a neighborhood of optimal critic (16) to ensure that the Bellman error (26) is reduced arbitrarily close to zero. To do this, the critic NN weights are updated using a gradient based update law with event-driven Bellman error (26) in two stages, given by

$$\dot{\hat{W}}_i(t) = -\alpha_{i1} \frac{\Delta\phi_{i,k-1}}{\rho_{i,e}^2} E_{i,k}^{eT}, \ t_i^k < t < t_i^{k+1}, \tag{27}$$

$$\hat{W}_i^+(t) = \hat{W}_i(t) - \alpha_{i2}\frac{\Delta\phi_{i,k-1}}{\rho_{i,e}^2}E_{i,k}^{eT}, \quad t = t_i^k, \tag{28}$$

for $i = 1, 2, \ldots, N, \forall k \in \{0, \mathbb{N}\}$ where $\alpha_{i1}, \alpha_{i2} > 0$ are the learning gains and $\rho_{i,e} = 1+\Delta\phi_{i,k-1}^T\Delta\phi_{i,k-1}$ is the normalization term. The notation $\hat{W}_i^+(t) = \hat{W}_i(t^+)$ and defined as $\lim_{s\to t}\hat{W}_i(s)$.

Remark 6. The two stage update law can be viewed as the flow and the jump dynamics of the impulsive weight update rule, where (27) is referred to as flow dynamics, i.e., update during the sampling intervals $t_i^k < t < t_i^{k+1}$ and (28) is referred to as jump dynamics, i.e., the update at the sampling instants $t = t_i^k, \forall k \in \{0, \mathbb{N}\}$. Since the state information is received by the controller at the sampling instants only, the event-driven Bellman error (26) is evaluated at the sampling instants only and used to update the weights both at jumps and during flow.

Similar to the sampling conditions in Section 3, the event-driven sampling conditions can be designed using the estimated thresholds (21) and it is given by

$$\|e_i^u(t)\| \leq \max\{r^2, \|\hat{\bar{e}}_i^u(t)\|\}, \quad i = 1, \ldots, N. \tag{29}$$

where $\|\hat{\bar{e}}_i^u\| = \|\frac{1}{2\gamma_{ii}^2}G_i^T\nabla_{x_i^e}^T(\phi_i(x_i^e))\hat{W}_i\|$.

To analyze the stability and the convergence properties of the controlled system, define the augmented states $\xi_i = [x^T\ \tilde{W}_i^T]^T$ where $\tilde{W}_i = W_i^* - \hat{W}_i$ is the critic NN weight estimation error. The event-driven system can be formulated as a nonlinear impulsive hybrid dynamical system (Haddad, Chellaboina, & Nersesov, 2014), using (6), (27), and (28), and is given by

$$\dot{\xi}_i(t) = \begin{bmatrix} F(x) + \sum_{i=1}^N G_i(x)u_i + \sum_{i=1}^N G_i(x)e_i^u \\ \alpha_{i1}\frac{\Delta\phi_{i,k-1}}{\rho_{i,e}^2}E_{i,k}^{eT} \end{bmatrix}, \quad t_i^k < t < t_i^{k+1}, \ \forall k, i, \tag{30}$$

and

$$\xi_i^+(t) = \begin{bmatrix} x(t) \\ \tilde{W}_i + \alpha_{i2}\frac{\Delta\phi_{i,k-1}}{\rho_{i,e}^2}E_{i,k}^{eT} \end{bmatrix}, \ t = t_i^k, \ \forall k, i \tag{31}$$

where (30) and (31) are referred to as the flow dynamics and the jump dynamics of the impulsive nonlinear system, respectively.

To prove the stability of the impulsive system (30) and (31), we will use the extension of the Lyapunov theorem for hybrid dynamical systems (Haddad et al., 2014). Before presenting the stability results in the next theorem, the definition of persistency of excitation (PE) condition (Ioannou & Fidan, 2006) is presented for completeness, which is used to complete the proof of Theorem 2.

Definition 1. A regressor vector $\frac{\Delta\phi_{i,k}}{1+\Delta\phi_{i,k}^T\Delta\phi_{i,k}}$ is said to be persistently exciting over an interval $[t-T, t], \forall t \in \Re^+$, if there exists a $T > 0, \tau_1 > 0, \tau_2 > 0$ such that $\tau_1 I \leq \int_{t-T}^t \frac{\Delta\phi_{i,k}\Delta\phi_{i,k}^T}{(1+\Delta\phi_{i,k}^T\Delta\phi_{i,k})^2}d\tau \leq \tau_2 I$ where I is the identity matrix of appropriate dimension.

Theorem 2. *Given the N player event-driven system (5), the sampled control policies (13), and the critic NN weight update laws (27) and (28), the augmented system is represented as an impulsive dynamical system given by (30) and (31). Let Assumptions 1 and 5 hold, and the critic NN weights are initialized in a compact set. Suppose the initial sampling instant is set to be $t_0^i = 0$, there exists a positive minimum inter-sample time $\delta_i^k > 0, \forall i$, the regressor vector satisfies the PE condition, and the initial control policy $u_i(0)$ is admissible. Then the closed-loop state ξ is locally ultimately bounded (UB) in the set $\mathcal{B}_{x\tilde{W}}$ (defined in the proof) provided the sampling instants are selected and the control policies are updated at the violation of the sampling conditions (29) and for constants $\alpha_{i1} > 0$, $\alpha_{i2} < \frac{1}{2}$ and $\gamma_{ii} > 0, i = 1, \ldots, N$, the gain conditions satisfy $\lambda_{min}(R_{ii}^{-1}) > \frac{1}{\gamma_{ii}^2}$ and $\frac{1}{2}q_{i,min} > \frac{3N}{2\gamma^{*2}}L_V^2G_{iM}^2$ where $\beta_i = \frac{\gamma^{*2}}{2\phi_{iM}'^2}\frac{\alpha_{i1}}{\rho_{im}}$.*

Proof. Refer to Appendix.

Corollary 3. *Let the hypothesis of Theorem 2 hold. Then the estimated value function (19) and control policy (20), respectively, converge arbitrarily close to the neighborhood of the optimal value function (16) and control policy (13).*

Proof. The proof is a direct consequence of the boundedness of the critic NN weight estimation error proved in Theorem 2 and, hence, omitted.

The following corollary analyzes the Zeno free behavior of the closed-loop system to relax the assumption of non-zero minimum inter-sample time in Theorem 2.

Corollary 4. *Given the N player event-driven system (5), the sampled control policies (13), and the critic NN weight update laws (27) and (28), the augmented system is represented as an impulsive dynamical system given by (30) and (31). Let Assumptions 1 and 5 hold, and the critic NN weights are initialized in a compact set. Suppose the initial sampling instant is set to be $t_0^i = 0$, the regressor vector satisfies the PE condition and the initial control policy $u_i(0)$ is admissible. Then the closed-loop event-driven system is locally UB if the sampling instants are selected and control policy is updated at the violation of the following inequality*

$$L_\zeta\|e_i^x(t)\| \leq \max\{r^2, \|\hat{\bar{e}}_i^u(t)\|\}, \tag{32}$$

for $i = 1, \ldots, N$. Further, the inter-sample times are lower bounded by a non-zero positive number and given by

$$\delta_i^k = t_i^{k+1} - t_i^k > \frac{1}{\nu_1}\ln(1 + \frac{\nu_1 r^2}{\nu_2 L_\zeta}) > 0 \tag{33}$$

where the parameters $\nu_1 > 0$ and $\nu_2 > 0$.

Proof. The proof follows the steps of Corollaries 1 and 2.

Remark 7. The ultimate bound $\mathcal{B}_{x\tilde{W}}$ for the event-triggered system state ξ, defined in (60) and (65), is a function of the approximation error bound ε_{M_i} defined in Assumption 5. Since the approximation error is a function of the number of states n, the ultimate bound is a function of system states. It is shown that with an increase in the number of neurons the approximation error can be reduced arbitrarily close to zero (Lewis et al., 1998), which in turn will reduce the radius of the ultimate bound $\mathcal{B}_{x\tilde{W}}$ arbitrarily close to zero.

5. Simulation results

In this section, an academic example is considered for numerical simulation. We use a nonlinear system with three players whose dynamics are given by $\dot{x} = F(x) + G_1(x)u_1 + G_2(x)u_2 + G_3(x)u_3$, where $x \in \Re^2$ is the state vector and $u_i \in \Re, i = 1, 2, 3$ are control inputs and $F(x) = \begin{bmatrix} x_2 - 2x_1 \\ -\frac{1}{2}x_1 - x_2 + \frac{1}{4}x_2(cos(2x_1)+2)^2 \\ +\frac{1}{4}x_2(sin(4x_1^2)+2)^2 \end{bmatrix}$, $G_1(x) = \begin{bmatrix} 0 \\ cos(2x_1)+2 \end{bmatrix}$, $g_2(x) = \begin{bmatrix} 0 \\ sin(4x_1^2)+2 \end{bmatrix}$ and $g_3(x) = \begin{bmatrix} 0 \\ cos(4x_1^2)+2 \end{bmatrix}$. The performance indices for the 3-player system

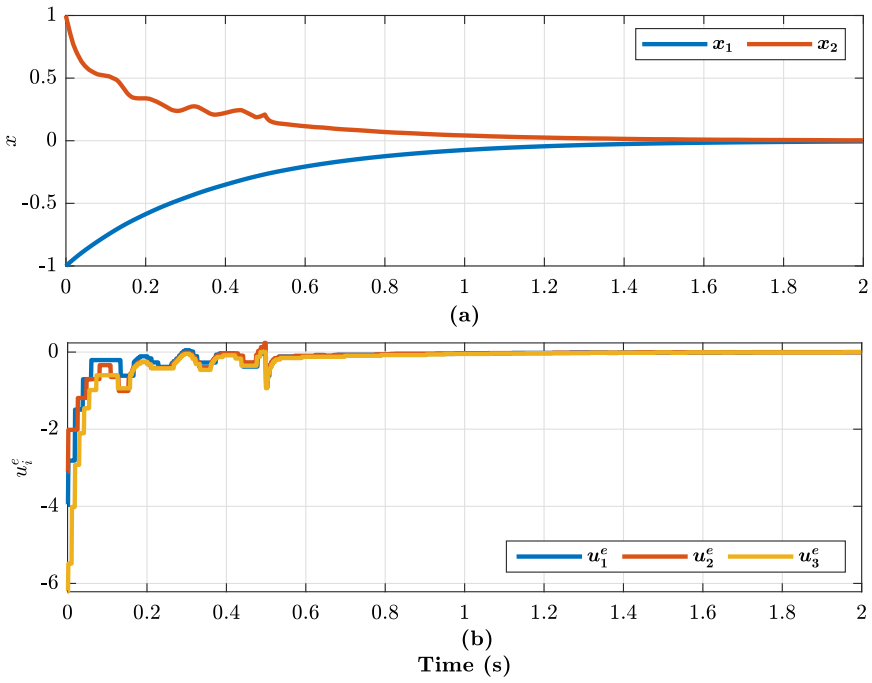

Fig. 2. Convergence of (a) system states and (b) player 1, 2, and 3's event-based control policies.

are defined as

$$J_1 = \int_t^\infty (x^T Q_1 x + u_1^T R_{11} u_1 - \gamma_{11}^2 \bar{e}_1^{uT} \bar{e}_1^u + u_2^T R_{12} u_2 + u_3^T R_{13} u_3) d\tau, \tag{34}$$

$$J_2 = \int_t^\infty (x^T Q_2 x + u_2^T R_{22} u_2 - \gamma_{22}^2 \bar{e}_2^{uT} \bar{e}_2^u + u_1^T R_{21} u_1 + u_3^T R_{23} u_3) d\tau, \tag{35}$$

and

$$J_3 = \int_t^\infty (x^T Q_3 x + u_3^T R_{33} u_3 - \gamma_{33}^2 \bar{e}_3^{uT} \bar{e}_3^u + u_1^T R_{31} u_1 + u_2^T R_{32} u_2) d\tau. \tag{36}$$

The simulation parameters were selected as follows. The penalty matrices for the performance indexes (34), (35), and (36) were $Q_1 = Q_2 = Q_3 = I_2$, $R_{11} = 0.2, R_{12} = 0.1, R_{13} = 0.1, R_{21} = 0.2, R_{22} = 0.2, R_{23} = 0.2, R_{31} = 0.2, R_{32} = 0.2$, and $R_{33} = 0.2$ where I_2 is 2×2 identity matrix and $\gamma_{11} = 0.8$, $\gamma_{22} = 1$, and $\gamma_{33} = 0.7$. The learning gains were selected as $\alpha_{11} = 1.5$, $\alpha_{12} = 0.4$, $\alpha_{21} = 5$, $\alpha_{22} = 0.3$, $\alpha_{31} = 5$, and $\alpha_{32} = 0.3$, and the initial critic NN weight $W = [1,\ 1,\ 1]^T$, activation function $\phi_i = [x_1^2, x_1x_2, x_2^2]^T$, initial state $x_0 = [-1, 1]^T$ and $r = 0.1$. Note that the selection of learning gains $\alpha_{(\cdot)}$ and $\gamma_{(\cdot)}$ are as per the gain conditions obtained in Theorem 2. The activation functions ϕ_i for $i = 1, 2, \ldots, N$ are selected based on the simulation example. The activation function and its derivative are bounded due to the fact that the initial control policy is admissible. A random noise, generated from the uniform distribution in the interval of [0, 1] is added to the control input for 1 s to satisfy the PE condition of the regressor.

The time history of the system states are plotted in Fig. 2(a) and the control policies are shown in Fig. 2(b). Both the plots show the boundedness of the parameters and convergence to a close neighborhood of zero. The Bellman errors, shown in Fig. 3(a)–(c), converge close to zero earlier than the state convergence. This implies the Nash equilibrium is attained with the convergence of the approximated value function in (19) to the optimal value function (16). The estimated weights of critic NNs for each player are shown in Fig. 4(a)–(c). Note that the performance index for each player is different due to different penalty matrices and, hence, the target weights are different. The cumulative cost at each player is plotted in Fig. 5.

Table 1
Resource saving performance of the players for the chosen γ_{ii}, $i = 1, 2, 3$.

Player	Number of triggering	Minimum inter-sample time (s)	Average inter-sample time (s)
Player 1	49	0.002	0.0359
Player 2	62	0.002	0.0235
Player 3	63	0.002	0.0316

From the resource saving perspective, Figs. 6 and 7 show the evolution of the estimated sampling threshold along with control policy error and inter-sample times, respectively, for all the players. It is clear from Fig. 6 that the control policy error is allowed to increase until it reaches the estimated worst-case threshold. This elongates the sampling intervals while guaranteeing near optimality in the sense of Nash. Fig. 7 shows the time intervals between two consecutive sampling instants. It is clear from the plots that the sampling intervals are different at each of the players implying an asynchronous sampling and control execution. The number of samplings, minimum, and average sampling intervals are listed in Table 1. Note that the inclusion of the exploration noise leads to a higher number of samplings during the learning phase as shown in Fig. 7. Further, a change in penalty factor γ_{ii} will result in a different number of samplings. The selection γ_{ii} depends on the gain conditions and the user defined performance requirement

6. Conclusions

We presented a differential game based event-driven sampling and approximate optimal NN control design for an N-player

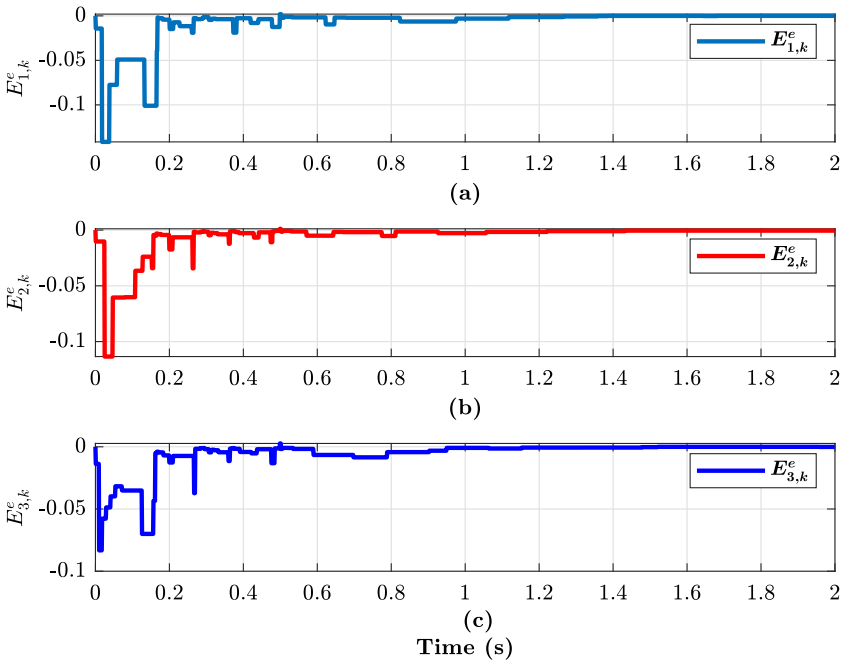

Fig. 3. Convergence of Bellman error (a) player 1, (b) player 2, and (c) player 3.

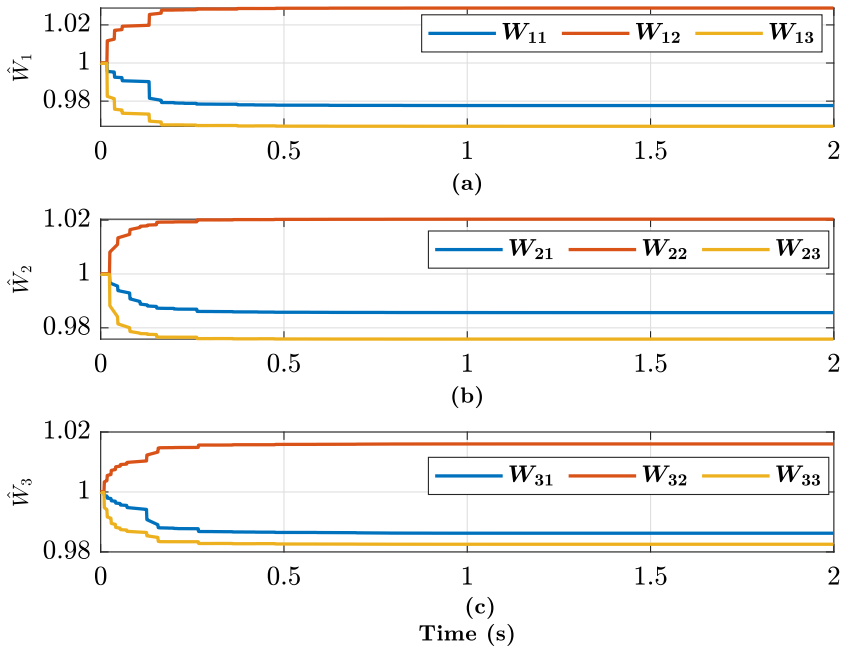

Fig. 4. Convergence of critic neural network wights for all players.

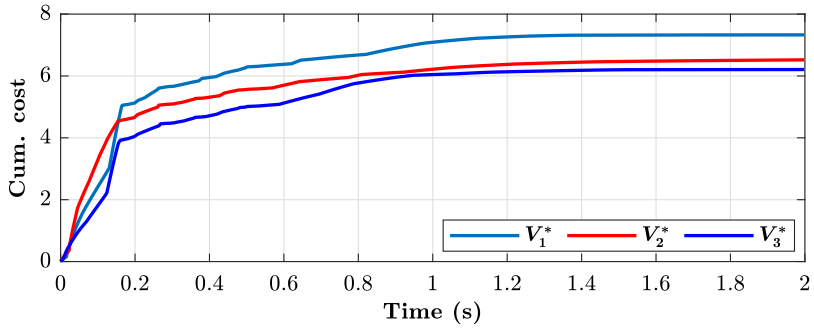

Fig. 5. Cumulative cost for each player.

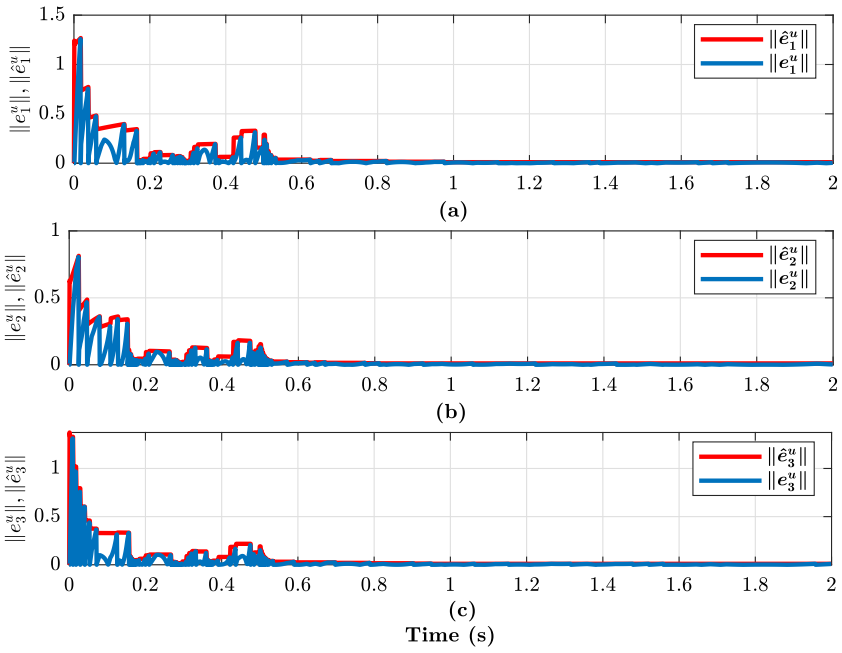

Fig. 6. Evolution of threshold and control policy error (a) player 1, (b) player 2, and (c) player 3.

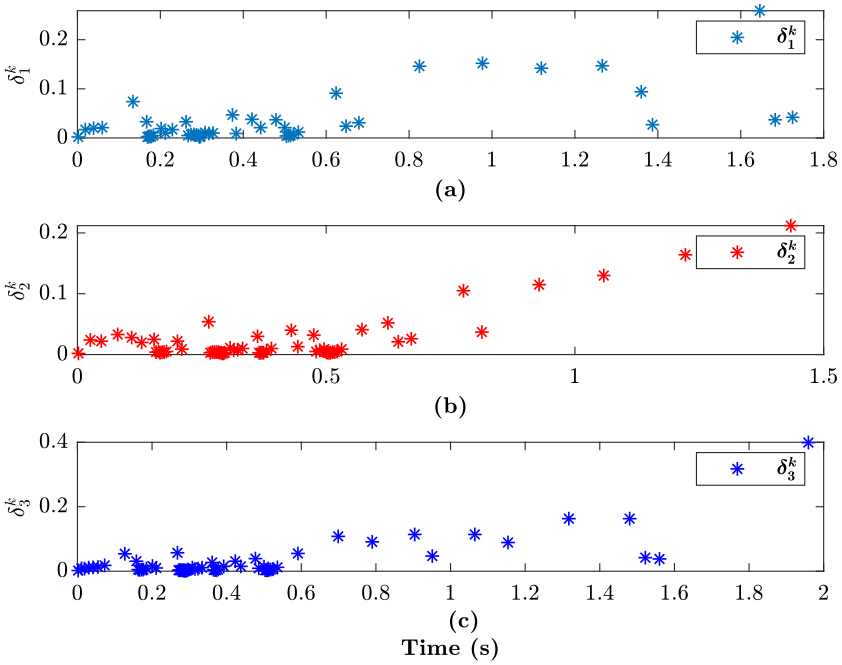

Fig. 7. Inter-sample times for (a) player 1, (b) player 2, and (c) player 3.

system. Improved event-driven sampling conditions with asynchronous sampling among all players are presented. The sampling intervals are determined using the estimated worst-case threshold value for control policy error. The Zeno free behavior for the sampling scheme is enforced by lower bounding the threshold. The near optimal Nash solution is obtained using critic NNs for the proposed NZS differential game for the large-scale system. With aperiodically available state information, the critic NNs are able to approximate the solution of the HJ equation for each player. From the simulation results, it is observed that the Bellman errors converged to a small-neighborhood of zero before the convergence of the system states implying convergence to a local optimal solution. A higher number of sampling instants was observed during the learning phase of the NN, as expected, and then sampling intervals elongated with the convergence of the NN weights. The proposed event-driven design considered the apriori knowledge of the control co-efficient matrix function to compute the control policies. Relaxing the knowledge of control co-efficient matrix function will be interesting and included in our future research.

Appendix

Proof of Theorem 1. Consider a positive definite radially unbounded function $L(x)$ as the candidate Lyapunov function where $L(x) = V_i^*(x)$ is the optimal value function for the ith player.

The time derivative of the Lyapunov function candidate can be expressed as

$$\dot{L}_i(x) = V_{ix}^{*T}\dot{x} = V_{ix}^{*T}[F(x) + \sum_{i=1}^{N} G_i(x_i)u_i^e]. \tag{37}$$

Recalling the definition of the control policy error (7) and optimal control policy u_i^*, $i = 1, \ldots, N$, the first time-derivative is

$$\dot{L}_i(x) = V_{ix}^{*T}[F(x) + \sum_{i=1}^{N} G_i(x)u_i^* + \sum_{i=1}^{N} G_i(x)e_i^{u*}]$$
$$- V_{ix}^{*T}\sum_{i=1}^{N} G_i(x)(e_i^{u*} - e_i^u). \tag{38}$$

From the HJ equation (12) we have

$$V_{ix}^{*T}[F(x) + \sum_{i=1}^{N} G_i(x)u_i^* + \sum_{i=1}^{N} G_i(x)e_i^{u*}]$$
$$= -x^T Q_i x - (u_i^{*T}R_{ii}u_i^* - \gamma_{ii}^2 e_i^{u*T}e_i^{u*}) - \sum_{\substack{j=1\\ j\neq i}}^{N} u_j^{*T}R_{ij}u_j^*. \tag{39}$$

The first derivative in (38) with (39), optimal control policy (10), and worst-case threshold from (11), along with Cauchy-Schwarz (C-S) can be expressed as

$$\dot{L}_i(x) \leq -x^T Q_i x - \frac{1}{4}V_{ix}^{*T}G_i(R_{ii}^{-1} - \frac{1}{\gamma_{ii}^2}I)G_i^T V_{ix}^*$$
$$- \sum_{\substack{j=1\\ j\neq i}}^{N}\frac{1}{4}V_{jx}^{*T}G_j R_{jj}^{-1}R_{ij}R_{jj}^{-1}G_j^T V_{jx}^* + \|V_{ix}^*\|\sum_{i=1}^{N}\|G_i\|(\|e_i^{u*}\| + \|e_i^u\|). \tag{40}$$

Recall the sampling condition (14), there are two cases that arise to complete the proof.

Case I: For sampling condition $\|e_i^u(t)\| \leq max\{r^2, \|e_i^{u*}(t)\|\} = \|e_i^{u*}(t)\|$, $i = 1, \ldots, N$

Substituting the sampling condition and recalling worst-case threshold in (11) and Assumption 4, the first derivative leads to

$$\dot{L}_i(x) \leq -\frac{q_{i,min}}{2}\|x\|^2 - \frac{1}{4}(V_{ix}^{*T}G_i(R_{ii}^{-1} - \frac{1}{\gamma_{ii}^2}I)G_i^T V_{ix}^*)$$
$$- \sum_{\substack{j=1\\ j\neq i}}^{N}\frac{1}{4}V_{jx}^{*T}G_j R_{jj}^{-1}R_{ij}R_{jj}^{-1}G_j^T V_{jx}^* - (\frac{q_{i,min}}{2} - \frac{N}{\gamma^{*2}}L_V^2 G_{iM}^2)\|x\|^2. \tag{41}$$

where $\gamma_{ii} > \gamma^* > 0, \forall i = 1, 2, \ldots, N$ and $q_{i,min} = \lambda_{min}(Q)$. By proper selection of $\gamma_{ii}, i, = 1, 2, \ldots, N$ based on the penalty matrices Q and R_{ii}, the condition $\lambda_{min}(R_{ii}^{-1}) > \frac{1}{\gamma_{ii}^2}I$ can be satisfied. Hence, the first derivative of the Lyapunov function candidate is negative. By the theorem, the system states converge to zero asymptotically.

It remains to show the second case where the sampling condition is bounded by a constant threshold.

Case II: For sampling condition $\|e_i^u(t)\| \leq max\{r^2, \|e_i^{u*}(t)\|\} = r^2$, for all $i = 1, 2, \ldots, N$.

The first derivative in (40), with the sampling condition in case II and the worst-case control policy error (11), can be expressed as

$$\dot{L}_i(x) \leq -q_{i,min}\|x\|^2 - \frac{1}{4}V_{ix}^{*T}G_i(R_{ii}^{-1} - \frac{1}{\gamma_{ii}^2}I)G_i^T V_{ix}^* - \sum_{\substack{j=1\\ j\neq i}}^{N}\frac{1}{4}V_{jx}^{*T}G_j R_{jj}^{-1}$$
$$\times R_{ij}R_{jj}^{-1}G_j^T V_{jx}^* + \frac{N}{2\gamma^{*2}}G_{iM}^2 L_V^2\|x\|^2 + \frac{N}{2\gamma^{*2}}G_{iM}L_V\|x\|r^2. \tag{42}$$

By using Cauchy-Schwarz inequality to separate the cross terms and combining similar terms, the first derivative leads to

$$\dot{L}_i(x) \leq -\frac{q_{i,min}}{2}\|x\|^2 - \frac{1}{4}V_{ix}^{*T}G_i(R_{ii}^{-1} - \frac{1}{\gamma_{ii}^2}I)G_i^T V_{ix}^*$$
$$- \sum_{\substack{j=1\\ j\neq i}}^{N}\frac{1}{4}V_{jx}^{*T}G_j R_{jj}^{-1}R_{ij}R_{jj}^{-1}G_j^T V_{jx}^*$$
$$- (\frac{q_{i,min}}{2} - \frac{N}{\gamma^{*2}}L_V^2 G_{iM}^2)\|x\|^2 + \frac{N}{8\gamma^{*2}}r^4. \tag{43}$$

By proper selection of $\gamma_{ii}, i, = 1, 2, \ldots, N$ based on the penalty matrices Q and R_{ii}, the conditions $\lambda_{min}(R_{ii}^{-1}) > \frac{1}{\gamma_{ii}^2}$ and $\frac{q_{i,min}}{2} > \frac{N}{\gamma^{*2}}L_V^2 G_{iM}^2$ can be satisfied. Consequently, the first derivative of the Lyapunov function $\dot{L}_i(x)$ is negative outside the set $\mathcal{B}_x \triangleq \{x \in \Omega_x \mid \|x\| < \frac{Nr^4}{8q_{i,min}\gamma^{*2}}\}$. By Lyapunov Theorem (Khalil, 1996) the system state states are UUB.

From Cases I and II, the system states are UUB. This completes the proof.

Proof of Corollary 1. By Assumptions 2 and 4, the optimal control policies are Lipschitz continuous in the compact set Ω_x. Alternatively, there exists a computable constant $L_\zeta \in \Re^+$ such that $\|u_i^e - u_i\| = \|\zeta_i(x_i^e) - \zeta_i(x)\| \leq L_\zeta\|x_i^e - x\| = L_\zeta\|e_i^x\|$.

Therefore, the control policy error $\|e_i^u\| = \|u_i^e - u_i\| \leq L_\zeta\|e_i^x\|$, $i = 1, \ldots, N$. From the sampling condition (15), it holds that $\|e_i^u\| \leq L_\zeta\|e_i^x\| \leq \max\{r^2, \|e_i^{u*}(t)\|\}$. This implies the inequality in (14) is also satisfied if the sampling instants are determined using (15). By invoking Theorem 1, the event-driven system is UUB. This completes the proof.

Proof of Corollary 2. From Corollary 1, the sampling condition (15) implies the condition (14). Therefore, it suffices to show that the inter-sample times generated by the condition in (15) are lower bounded by a non-zero positive number, to show the Zeno freeness of the closed-loop system.

Further, from the sampling condition (15), the lower bound of the threshold is r^2. For computing the inter-sample times, we compute the evolution time of the sampled state error e_i^x from 0 to the minimum threshold r^2, which must be lower bounded by a non-zero positive number.

The sampling condition (15) can equivalently be expressed as $\|e_i^x\| \leq \frac{r^2}{L_\zeta}$, $i = 1, \ldots, N$. The dynamics of $\|e_i^x\|$ can be written as

$$\frac{d}{dt}\|e_i^x\| \leq \|\dot{e}_i^x\| = \|\dot{x}\| = \|F(x) + \sum_{i=1}^{N} G_i(x)u_i^* + \sum_{i=1}^{N} G_i(x)e_i^u\|. \tag{44}$$

From Assumptions 2 and 4, the inequality $\|F(x) + \sum_{i=1}^{N} G_i(x)u_i^*\| \leq \kappa\|x\|$ holds. Substituting the above results and using triangle inequality, it further holds that

$$\frac{d}{dt}\|e_i^x\| \leq \|F(x) + \sum_{i=1}^{N} G_i(x)u_i^*\| + \|\sum_{i=1}^{N} G_i(x)e_i^u\|$$
$$\leq \kappa\|x\| + NG_{iM}L_\zeta\|e_i^x\| \leq (\kappa + NG_{iM}L_\zeta)\|e_i^x\| + \kappa\|x_i^e\|$$
$$= \nu_1\|e_i^x\| + \nu_2,$$

for all $i = 1, 2, \ldots, N$ with $\nu_1 = \kappa + NG_{iM}L_\zeta$ and $\nu_2 = \kappa\|x_i^e\|$. Note that ν_2 is a constant during the inter-sample times $t_i^k \leq t < t_i^{k+1}$, $\forall i, k$. The solution to the differential inequality with initial condition $e_i^x(t_i^k) = 0$ is upper bounded (Khalil, 1996) by

$$\|e_i^x(t)\| \leq \nu_2\int_{t_i^k}^{t} e^{\nu_1(t-\tau)}d\tau, \quad \forall i, k. \tag{45}$$

At the next sampling instants t_i^{k+1} with sampling threshold r^2/L_ζ, it holds

$$\frac{r^2}{L_\zeta} = \|e_i^x(t_i^{k+1})\| \leq \frac{\nu_2}{\nu_1}(e^{\nu_1(t_i^{k+1}-t_i^k)} - 1), \forall i, k. \tag{46}$$

Solving the above inequality, the inter-sample times can be obtained as

$$\delta_i^k = t_i^{k+1} - t_i^k > \frac{1}{\nu_1}\ln(1 + \frac{\nu_1 r^2}{\nu_2 L_\zeta}) > 0, \forall i, k. \tag{47}$$

Therefore, the minimum inter-sample time $\delta = \inf_{\forall k}(t_i^{k+1} - t_i^k) > 0$, $\forall i$ and the event-sampled system does not show Zeno behavior. This completes the proof.

Proof of Theorem 2. To demonstrate the local UB of the system, we evaluate a single Lyapunov candidate function at both inter-sample times, i.e., flow duration, and at the sampling instants, i.e., jump instants.

Flow duration: Consider a continuously differentiable positive definite candidate Lyapunov function, L and given as

$$L = \beta_i L_i(x) + \sum_{i=1}^{N} L_i(\tilde{W}_i) \tag{48}$$

where $L_i(x) = V_i^*(x)$, $L_i(\tilde{W}_i) = (1/2)\tilde{W}_i^T\tilde{W}_i$, and $\beta_i > 0$ is a constant defined later in the proof.

Consider the first term of the Lyapunov candidate $L_i(x) = V_i^*(x)$. The time derivative is the same as in (40) in proof of Theorem 1, given by

$$\dot{L}_i(x) \leq -x^T Q_i x - \frac{1}{4}V_{ix}^{*T}G_i(R_{ii}^{-1} - \frac{1}{\gamma_{ii}^2}I)G_i^T V_{ix}^* - \sum_{\substack{j=1\\j\neq i}}^{N}\frac{1}{4}V_{jx}^{*T}G_jR_{jj}^{-1}$$
$$\times R_{ij}R_{jj}^{-1}G_j^T V_{jx}^* + \|V_{ix}^*\|\sum_{i=1}^{N}\|G_i\|(\|e_i^{u*}\| + \|e_i^u\|). \tag{49}$$

From the sampling condition (29), two cases arise to complete the proof.

Case I: For sampling condition $\|e_i^u(t)\| \leq max\{r^2, \|\hat{e}_i^u(t)\|\} = r^2$, $i = 1, \ldots, N$

Substituting the sampling condition in (49), the first derivative of the Lyapunov function candidate is

$$\dot{L}_i(x) \leq -\frac{q_{i,min}}{2}\|x\|^2 - \frac{1}{4}V_{ix}^{*T}G_i(R_{ii}^{-1} - \frac{1}{\gamma_{ii}^2}I)G_i^T V_{ix}^* - \sum_{\substack{j=1\\j\neq i}}^{N}\frac{1}{4}V_{jx}^{*T}G_jR_{jj}^{-1}$$
$$\times R_{ij}R_{jj}^{-1}G_j^T V_{jx}^* - (\frac{q_{i,min}}{2} - \frac{N}{\gamma^{*2}}L_V^2G_{iM}^2)\|x\|^2 + \frac{N}{8\gamma^{*2}}r^4. \tag{50}$$

By selecting learning parameters and the penalty matrices, the gain conditions $\lambda_{min}(R_{ii}^{-1}) > \frac{\beta_i}{\gamma_{ii}^2}$ and $\frac{1}{2}q_{i,min} > \frac{N}{\gamma^{*2}}L_V^2G_{iM}^2$ can be satisfied. This ensures that the fourth term is also negative.

Case II: For sampling condition $\|e_i^u(t)\| \leq max\{r^2, \|\hat{e}_i^u(t)\|\} = \|\hat{e}_i^u(t)\|$, $i = 1, \ldots, N$

Substituting the above sampling condition in (49), using Assumption 4, and Young's inequality, the first derivative is upper bounded as

$$\dot{L}_i(x) \leq -x^T Q_i x - \frac{1}{4}V_{ix}^{*T}G_i(R_{ii}^{-1} - \frac{1}{2\gamma_{ii}^2}I)G_i^T V_{ix}^* - \sum_{\substack{j=1\\j\neq i}}^{N}\frac{1}{4}V_{jx}^{*T}G_jR_{jj}^{-1}$$
$$\times R_{ij}R_{jj}^{-1}G_j^T V_{jx}^* + \frac{N}{2\gamma^{*2}}L_V^2G_{iM}^2\|x\|^2 + L_V\|x\|\sum_{i=1}^{N}\frac{1}{2\gamma^{*2}}G_{iM}\phi_{iM}'\|\hat{W}_i\|. \tag{51}$$

By using the definition of the critic NN weight estimation error $\tilde{W}_i = W_i^* - \hat{W}_i$, the first derivative is upper bounded as

$$\dot{L}_i(x) \leq -x^T Q_i x - \frac{1}{4}V_{ix}^{*T}G_i(R_{ii}^{-1} - \frac{1}{\gamma_{ii}^2}I)G_i^T V_{ix}^* - \sum_{\substack{j=1\\j\neq i}}^{N}\frac{1}{4}V_{jx}^{*T}G_jR_{jj}^{-1}$$
$$\times R_{ij}R_{jj}^{-1}G_j^T V_{jx}^* + \frac{N}{2\gamma^{*2}}L_V^2G_{iM}^2\|x\|^2$$
$$+ \frac{1}{2\gamma^{*2}}L_V\|x\|G_{iM}\phi_{iM}'\sum_{i=1}^{N}\|W_i^* - \tilde{W}_i\|. \tag{52}$$

By Assumption 5, the target weights are bounded as $\|W_i^*\| \leq W_{iM}$. Using the inequalities $\|a - b\| \leq \|a\| + \|b\|$, and $ab \leq \frac{\alpha}{2}a^2 + \frac{1}{2\alpha}b^2$ to separate the cross terms, the first derivative is bound as

$$\dot{L}_i(x) \leq -x^T Q_i x - \frac{1}{4}V_{ix}^{*T}G_i(R_{ii}^{-1} - \frac{1}{\gamma_{ii}^2}I)G_i^T V_{ix}^* - \sum_{\substack{j=1\\j\neq i}}^{N}\frac{1}{4}V_{jx}^{*T}G_jR_{jj}^{-1}$$
$$\times R_{ij}R_{jj}^{-1}G_j^T V_{jx}^* + \frac{N}{2\gamma^{*2}}L_V^2G_{iM}^2\|x\|^2 + \frac{N}{2\gamma^{*2}}L_V^2G_{iM}^2\|x\|^2$$
$$+ \frac{N}{8\gamma^{*2}}\phi_{iM}'^2W_{iM}^2$$
$$+ \frac{N}{2\gamma^{*2}}L_V^2G_{iM}^2\|x\|^2 + \frac{1}{8N\gamma^{*2}}\phi_{iM}'^2\sum_{i=1}^{N}\|\tilde{W}_i\|^2. \tag{53}$$

Combining similar terms, we can express the first derivative as

$$\dot{L}_i(x) \leq -\frac{1}{2}q_{i,min}\|x\|^2 - \frac{1}{4}V_{ix}^{*T}G_i(R_{ii}^{-1} - \frac{1}{\gamma_{ii}^2}I)G_i^T V_{ix}^*$$
$$- \sum_{\substack{j=1\\j\neq i}}^{N}\frac{1}{4}V_{jx}^{*T}G_jR_{jj}^{-1}$$
$$\times R_{ij}R_{jj}^{-1}G_j^T V_{jx}^* - (\frac{1}{2}q_{i,min} - \frac{3N}{2\gamma^{*2}}L_V^2G_{iM}^2)\|x\|^2$$
$$+ \frac{1}{2\gamma^{*2}}\phi_{iM}'^2\sum_{i=1}^{N}\|\tilde{W}_i\|^2 + B_i^x \tag{54}$$

where $B_i^x = \frac{N}{8\gamma^{*2}}\phi_{iM}'^2W_{iM}^2$.

Next, we will consider the second term of the Lyapunov function. The, first derivative of the second term $\dot{L}_i(\tilde{W}_i) = \tilde{W}_i^T\dot{\tilde{W}}_i$. Using the weight tuning rule, $\dot{L}_i(\tilde{W}_i) = \alpha_{i1}\tilde{W}_i^T\frac{\Delta\phi_{i,k-1}}{\rho_{i,e}^2}E_{i,k}^{eT}$.

Subtracting (24) from (26), the Bellman error can be expressed in terms of the NN weight estimation error as

$$E_{i,k}^e = -\tilde{W}_i^T\Delta\phi_{i,k-1} - \Delta\varepsilon_{i,k-1}, \ \forall i \tag{55}$$

where $\Delta\varepsilon_{i,k-1} = \varepsilon_i(x(t_i^k)) - \varepsilon_i(x(t_i^{k-1}))$.

Substituting (55) in the first derivative $\dot{L}_i(\tilde{W}_i)$, we have

$$\dot{L}_i(\tilde{W}_i) = -\frac{\alpha_{i1}}{\rho_{i,e}^2}\tilde{W}_i^T\Delta\phi_{i,k-1}\Delta\phi_{i,k-1}^T\tilde{W}_i$$
$$+ \frac{\alpha_{i1}}{\rho_{i,e}^2}\tilde{W}_i^T\Delta\phi_{i,k-1}\Delta\varepsilon_{i,k-1}^T, \ \forall i. \tag{56}$$

Applying Young's inequality to separate the cross terms, the first derivative is upper bounded as

$$\dot{L}_i(\tilde{W}_i) \leq -\frac{\alpha_{i1}}{2\rho_{im}}\|\tilde{W}_i\|^2 + \alpha_{i1}\Delta\varepsilon_{iM}^2, \ \ i = 1, \ldots, N. \tag{57}$$

where $0 < \rho_{i,m} \leq \frac{\Delta\phi_{i,k-1}\Delta\phi_{i,k-1}^T}{\rho_{i,e}^2}$ by ensuring the PE condition (Ioannou & Fidan, 2006) of the regressor vector and $\|\Delta\varepsilon_{i,k-1}\| \leq \Delta\varepsilon_{iM}$ with $\Delta\varepsilon_{iM} >$ is a constant.

Combining the Lyapunov time-derivatives (54) and (57), the first derivative of the overall Lyapunov function

$$\begin{aligned}\dot{L}_i(\zeta_i) \leq & -\frac{\beta_i}{2}q_{i,min}\|x\|^2 - \frac{\beta_i}{4}V_{ix}^{*T}G_i(R_{ii}^{-1} - \frac{1}{\gamma_{ii}^2}I)G_i^TV_{ix}^* \\ & - \beta_i\sum_{\substack{j=1\\ j\neq i}}^{N}\frac{1}{4}V_{jx}^{*T}G_jR_{jj}^{-1} \\ & \times R_{ij}R_{jj}^{-1}G_j^TV_{jx}^* - \beta_i(\frac{1}{2}q_{i,min} - \frac{3N}{2\gamma^{*2}}L_V^2G_{iM}^2)\|x\|^2 \\ & + \frac{\beta_i}{2\gamma^{*2}}\phi_{iM}^{'2}\sum_{i=1}^{N}\|\tilde{W}_i\|^2 + \bar{B}_i^x \\ & - \frac{\alpha_{i1}}{2\rho_{im}}\sum_{i=1}^{N}\|\tilde{W}_i\|^2 + N\alpha_{i1}\Delta\varepsilon_{iM}^2 \end{aligned} \tag{58}$$

where $\bar{B}_i^x = max\{\frac{N}{8\gamma^{*2}}r^4, B_i^x\}$. Defining $\beta_i = \frac{\gamma^{*2}}{2\phi_{iM}^{'2}}\frac{\alpha_{i1}}{\rho_{im}} > 0$, the first derivative is upper bounded by

$$\begin{aligned}\dot{L}_i \leq & -\frac{\beta_i}{2}q_{i,min}\|x\|^2 - \frac{\beta_i}{4}(V_{ix}^{*T}G_i(R_{ii}^{-1} - \frac{1}{\gamma_{ii}^2}I)G_i^TV_{ix}^*) \\ & - \beta_i\sum_{\substack{j=1\\ j\neq i}}^{N}\frac{1}{4}V_{jx}^{*T}G_jR_{jj}^{-1}R_{ij}R_{jj}^{-1}G_j^TV_{jx}^* \\ & - \beta_i(\frac{1}{2}q_{i,min} - \frac{3N}{2\gamma^{*2}}L_V^2G_{iM}^2)\|x\|^2 - \frac{\alpha_{i1}}{4\rho_{im}}\sum_{i=1}^{N}\|\tilde{W}_i\|^2 + \Upsilon_i^c \end{aligned} \tag{59}$$

where $\Upsilon_i^c = \bar{B}_i^x + N\alpha_{i1}\Delta\varepsilon_{iM}^2$. By selecting the gain conditions $\lambda_{min}(R_{ii}^{-1}) > \frac{1}{\gamma_{ii}^2}$, $\frac{1}{2}q_{i,min} > \frac{3N}{2\gamma^{*2}}L_V^2G_{iM}^2$, the fourth term is also negative.

Defining $B_{i,m} \triangleq \min\{\frac{\beta_i}{2}q_{i,min}, \beta_i(\frac{1}{2}q_{i,min} - \frac{3N}{2\gamma^{*2}}L_V^2G_{iM}^2), \frac{\alpha_{i1}}{4\rho_{im}}\}$, the first derivative of the Lyapunov function is

$$\dot{L}_i \leq -2B_{i,m}\|x\|^2 - B_{i,m}\|\tilde{W}_i\|^2 + \Upsilon_i^c, i = 1, 2, \ldots, N \tag{60}$$

Therefore, it can be concluded that for $\alpha_{i1} > 0$, the first derivative $\dot{L}_i$ is negative outside the balls $\mathcal{B}_x = \{x \in \mathcal{R}^n | \|x\| < \sqrt{\frac{\Upsilon_i^c}{B_{i,m}}} \triangleq B_i^{xw}\}$ or $\mathcal{B}_{\tilde{W}_{fl}} = \{\tilde{W}_i \in \mathcal{R}^{l_{w_i}} | \|\tilde{W}\| < B_i^{xw}\}$. By Lyapunov theorem, the system states and NN weight estimation errors will converge to the set $\{\mathcal{B}_x \bigcup \mathcal{B}_{\tilde{W}_{fl}}\}$ and are locally UB in flow period.

Jump dynamics: Consider the Lyapunov function from the flow period. Along the jump dynamics, the system state $x^+ = x$. Therefore, the first difference

$$\Delta L_i(x_i) = V_i^*(x_i^+) - V_i^*(x_i) = 0. \tag{61}$$

The first difference of the second term can be expressed as $\Delta L_i(\tilde{W}_i) = \frac{1}{2}\tilde{W}_i^{+T}\tilde{W}_i^+ - \frac{1}{2}\tilde{W}_i^T\tilde{W}_i$. Along the NN weight estimation error dynamics for the jump instants,

$$\begin{aligned}\Delta L_i(\tilde{W}_i) = & \frac{1}{2}\left[\tilde{W}_i - \alpha_{i2}\frac{\Delta\phi_{i,k-1}\Delta\phi_{i,k-1}^T\tilde{W}_i}{\rho_{i,e}^2} - \alpha_{i2}\frac{\Delta\phi_{i,k-1}\Delta\varepsilon_{i,k-1}^T}{\rho_{i,e}^2}\right]^T \\ & \left[\tilde{W}_i - \alpha_{i2}\frac{\Delta\phi_{i,k-1}\Delta\phi_{i,k-1}^T\tilde{W}_i}{\rho_{i,e}^2} - \alpha_{i2}\frac{\Delta\phi_{i,k-1}\Delta\varepsilon_{i,k-1}^T}{\rho_{i,e}^2}\right] - \frac{1}{2}\tilde{W}_i^T\tilde{W}_i. \end{aligned} \tag{62}$$

Expanding the first difference and separating the cross terms by applying Young's inequality, the first difference leads to

$$\begin{aligned}&\Delta L_i(\tilde{W}_i) \\ &\leq \frac{1}{2}\left[-\alpha_{i2}\frac{\tilde{W}_i^T\Delta\phi_{i,k-1}\Delta\phi_{i,k-1}^T\tilde{W}_i}{\rho_{i,e}^2} + 2\alpha_{i2}^2\frac{\tilde{W}_i^T\Delta\phi_{i,k-1}\Delta\phi_{i,k-1}^T\Delta\phi_{i,k-1}\Delta\phi_{i,k-1}^T\tilde{W}}{\rho_{i,e}^4}\right. \\ &\left. + \alpha_{i2}\frac{\Delta\varepsilon_{i,k-1}^T\Delta\varepsilon_{i,k-1}}{\rho_{i,e}^2} + 2\alpha_{i2}^2\frac{\Delta\varepsilon_{i,k-1}^T\Delta\phi_{i,k-1}^T\Delta\phi_{i,k-1}\Delta\varepsilon_{i,k-1}}{\rho_{i,e}^4}\right]. \end{aligned} \tag{63}$$

From the following facts $\|\frac{\Delta\phi_{i,k-1}\Delta\phi_{i,k-1}^T}{\rho_{i,e}^2}\| \leq 1$, $\|\frac{1}{\rho_{i,e}^2}\| \leq 1$, and $\|\frac{1}{\rho_{i,e}^4}\| \leq 1$, the first difference is upper bounded by

$$\Delta L_i(\tilde{W}_i) \leq -\alpha_{i2}\rho_{im}(\frac{1}{2} - \alpha_{i2})\|\tilde{W}_i\|^2 + \Upsilon_i^d \tag{64}$$

where $\Upsilon_i^d = \alpha_{i2}^2\Delta\varepsilon_{iM}^2 + \frac{1}{2}\alpha_{i2}\Delta\varepsilon_{iM}^2$.

Finally, combining both the first differences (61) and (64), the overall first difference at the jump instants is upper bounded as

$$\Delta L_i \leq -\alpha_{i2}\rho_{im}(\frac{1}{2} - \alpha_{i2})\sum_{i=1}^{N}\|\tilde{W}\|^2 + \bar{\Upsilon}_i^d \tag{65}$$

where $\bar{\Upsilon}_i^d = \sum_{i=1}^{N}\Upsilon_i^d$ and the learning gain $0 < \alpha_{i2} < \frac{1}{2}$. From (65), the first difference of the Lyapunov function is negative outside the ball $\mathcal{B}_{\tilde{W}_{jp}} = \{\tilde{W}_i \in \mathcal{R}^{l_{w_i}} | \|\tilde{W}\| < \sqrt{\frac{\bar{\Upsilon}_i^d}{\alpha_{i2}(\frac{1}{2}\rho_{im} - \alpha_{i2})}} \triangleq B_i^w\}$. Therefore, there exists an integer $\bar{N} > 0$ such that for all $t_i^k > t_i^{\bar{N}}$, the NN weight estimation error converges to the ultimate bound (Haddad et al., 2014).

Consequently, from both the flow and jump dynamics, the system states and the weight estimation errors are locally UB in the ball defined by $\mathcal{B}_{x\tilde{W}} = \{\mathcal{B}_x \bigcup \mathcal{B}_{\tilde{W}_{fl}} \bigcup \mathcal{B}_{\tilde{W}_{jp}}\}$ for all time. This completes the proof.

References

Abu-Khalaf, M., & Lewis, F. L. (2005). Nearly optimal control laws for nonlinear systems with saturating actuators using a neural network HJB approach. *Automatica*, *41*(5), 779–791.

Albattat, A., Gruenwald, B., & Yucelen, T. (2016). On event-triggered adaptive architectures for decentralized and distributed control of large-scale modular systems. *Sensors*, *16*(8), 1297.

Başar, T., & Bernhard, P. (1995). *H_∞- optimal control and related minimax design problems*. Boston: Birkhäuser.

Bellman, R. (2013). *Dynamic programming*. Courier Corporation.

Bertsekas, D. P. (1995). *Dynamic programming and optimal control, Vol. 1*. Belmont, MA: Athena scientific.

Case, J. H. (1969). Toward a theory of many player differential games. *SIAM Journal on Control*, *7*(2), 179–197.

Dierks, T., & Jagannathan, S. (2012). A self-tuning optimal controller for affine nonlinear continuous-time systems with unknown internal dynamics. In *IEEE conference on decision and control* (pp. 5392–5397).

Dong, L., Zhong, X., Sun, C., & He, H. (2017). Event-triggered adaptive dynamic programming for continuous-time systems with control constraints. *IEEE Transactions on Neural Networks and Learning Systems*, *28*(8), 1941–1952.

Friedman, A. (2013). *Differential games*. Courier Corporation.

Girard, A. (2015). Dynamic triggering mechanisms for event-triggered control. *IEEE Transactions on Automatic Control*, *60*(7), 1992–1997.

Haddad, W. M., Chellaboina, V., & Nersesov, S. G. (2014). *Impulsive and hybrid dynamical systems: stability, dissipativity, and control, Vol. 49*. Princeton University Press.

Heydari, A. (2017). Stability analysis of optimal adaptive control under value iteration using a stabilizing initial policy. *IEEE Transactions on Neural Networks and Learning Systems*.

Ioannou, P., & Fidan, B. (2006). *Adaptive control tutorial (Advances in design and control), Vol. 1*. PA: SIAM.

Johnson, M., Kamalapurkar, R., Bhasin, S., & Dixon, W. E. (2015). Approximate *N*-player nonzero-sum game solution for an uncertain continuous nonlinear system. *IEEE Transactions on Neural Networks and Learning Systems*, *26*(8), 1645–1658.

Kamalapurkar, R., Andrews, L., Walters, P., & Dixon, W. E. (2017). Model-based reinforcement learning for infinite-horizon approximate optimal tracking. *IEEE Transactions on Neural Networks and Learning Systems*, *28*(3), 753–758.

Kamalapurkar, R., Dinh, H., Bhasin, S., & Dixon, W. E. (2015). Approximate optimal trajectory tracking for continuous-time nonlinear systems. *Automatica*, *51*, 40–48.

Kamalapurkar, R., Klotz, J. R., & Dixon, W. E. (2014). Concurrent learning-based approximate feedback-Nash equilibrium solution of *N*-player nonzero-sum differential games. *IEEE/CAA journal of Automatica Sinica*, *1*(3), 239–247.

Khalil, H. K. (1996). *Noninear systems, Vol. 2(5)*. New Jersey: Prentice-Hall, 5–1.

Lewis, F., Jagannathan, S., & Yesildirak, A. (1998). *Neural network control of robot manipulators and non-linear systems*. CRC Press.

Lewis, F. L., Vrabie, D., & Syrmos, V. L. (2012). *Optimal control*. John Wiley & Sons.

Liu, T., & Jiang, Z.-P. (2015). A small-gain approach to robust event-triggered control of nonlinear systems. *IEEE Transactions on Automatic Control*, *60*(8), 2072–2085.

Liu, D., Li, H., & Wang, D. (2014). Online synchronous approximate optimal learning algorithm for multi-player non-zero-sum games with unknown dynamics. *IEEE Transactions on Systems, Man, and Cybernetics A*, *44*(8), 1015–1027.

Liu, D., & Wei, Q. (2014). Policy iteration adaptive dynamic programming algorithm for discrete-time nonlinear systems. *IEEE Transactions on Neural Networks and Learning Systems*, *25*(3), 621–634.

Ma, H., Wang, Z., Wang, D., Liu, D., Yan, P., & Wei, Q. (2016). Neural-network-based distributed adaptive robust control for a class of nonlinear multiagent systems with time delays and external noises. *IEEE Transactions on Systems, Man, and Cybernetics A*, *46*(6), 750–758.

Narayanan, V., & Jagannathan, S. (2016). Approximate optimal distributed control of uncertain nonlinear interconnected systems with event-sampled feedback. In *Proceedings of IEEE 55th conference on decision and control* (pp. 5827–5832).

Narayanan, V., & Jagannathan, S. (2018). Event-triggered distributed approximate optimal state and output control of affine nonlinear interconnected systems. *IEEE Transactions on Neural Networks and Learning Systems*, *29*(7), 2846–2856.

Narayanan, V., Jagannathan, S., & Ramkumar, K. (2018). Event-sampled output feedback control of robot manipulators using neural networks. *IEEE Transactions on Neural Networks and Learning Systems*, *30*(6), 1651–1658.

Nash, J. (1951). Non-cooperative games. *Annals of Mathematics*, 286–295.

Sahoo, A., & Jagannathan, S. (2017). Stochastic optimal regulation of nonlinear networked control systems by using event-driven adaptive dynamic programming. *IEEE Transactions on Cybernetics*, *47*(2), 425–438.

Sahoo, A., Narayanan, V., & Jagannathan, S. (2017). Optimal sampling and regulation of uncertain interconnected linear continuous time systems. In *Proceedings of IEEE symposium series on computational intelligence* (pp. 1–6).

Sahoo, A., Narayanan, V., & Jagannathan, S. (2018). Event-triggered control of N-player nonlinear systems using nonzero-sum games. In *Symposium series on computational intelligence* (pp. 1447–1452). IEEE.

Sahoo, A., Narayanan, V., & Jagannathan, S. (2019). A min–max approach to event-and self-triggered sampling and regulation of linear systems. *IEEE Transactions on Industrial Electronics*, *66*(7), 5433–5440.

Sahoo, A., Xu, H., & Jagannathan, S. (2013). Adaptive event-triggered control of a uncertain linear discrete time system using measured input and output data. In *American control conference* (pp. 5672–5677).

Sahoo, A., Xu, H., & Jagannathan, S. (2016). Neural network-based event-triggered state feedback control of nonlinear continuous-time systems. *Transactions on Neural Networks and Learning Systems*, *27*(3), 497–509.

Sahoo, A., Xu, H., & Jagannathan, S. (2017). Approximate optimal control of affine nonlinear continuous-time systems using event-sampled neurodynamic programming. *IEEE Transactions on Neural Networks and Learning Systems*, *28*(3), 639–652.

Song, R., Lewis, F. L., & Wei, Q. (2017). Off-policy integral reinforcement learning method to solve nonlinear continuous-time multiplayer nonzero-sum games. *IEEE Transactions on Neural Networks and Learning Systems*, *28*(3), 704–713.

Starr, A. W., & Ho, Y.-C. (1969). Nonzero-sum differential games. *Journal of Optimization Theory and Applications*, *3*(3), 184–206.

Sutton, R. S., & Barto, A. G. (1998). *Reinforcement learning: An introduction, Vol. 1*. Cambridge: MIT Press.

Tabuada, P. (2007). Event-triggered real-time scheduling of stabilizing control tasks. *IEEE Transactions on Automatic Control*, *52*(9), 1680–1685.

Tallapragada, P., & Chopra, N. (2014). Decentralized event-triggering for control of nonlinear systems. *IEEE Transactions on Automatic Control*, *59*(12), 3312–3324.

Vamvoudakis, K. G., & Hespanha, J. P. (2018). Cooperative Q-learning for rejection of persistent adversarial inputs in networked linear quadratic systems. *IEEE Transactions on Automatic Control*, *63*(4), 1018–1031.

Vamvoudakis, K. G., & Lewis, F. L. (2011). Multi-player non-zero-sum games: Online adaptive learning solution of coupled Hamilton–Jacobi equations. *Automatica*, *47*(8), 1556–1569.

Vamvoudakis, K. G., Modares, H., Kiumarsi, B., & Lewis, F. L. (2017). Game theory-based control system algorithms with real-time reinforcement learning: How to solve multiplayer games online. *IEEE Control Systems Magazine*, *37*(1), 33–52.

Wang, D., He, H., Zhong, X., & Liu, D. (2017). Event-driven nonlinear discounted optimal regulation involving a power system application. *IEEE Transactions on Industrial Electronics*, *64*(10), 8177–8186.

Wang, X., & Lemmon, M. D. (2011). Event-triggering in distributed networked control systems. *IEEE Transactions on Automatic Control*, *56*(3), 586–601.

Werbos, P. J. (2007). Using ADP to understand and replicate brain intelligence: The next level design? In *Neurodynamics of cognition and consciousness* (pp. 109–123). Springer.

Zhao, D., Zhang, Q., Wang, D., & Zhu, Y. (2016). Experience replay for optimal control of nonzero-sum game systems with unknown dynamics. *IEEE Transactions on Cybernetics*, *46*(3), 854–865.