

# Event-Triggered Distributed Approximate Optimal State and Output Control of Affine Nonlinear Interconnected Systems

Vignesh Narayanan, *Student Member, IEEE*, and Sarangapani Jagannathan, *Fellow, IEEE*

**Abstract**—This paper presents an approximate optimal distributed control scheme for a known interconnected system composed of input affine nonlinear subsystems using event-triggered state and output feedback via novel hybrid learning scheme. First, the cost function for the overall system is redefined as the sum of cost functions of individual subsystems. A distributed optimal control policy for the interconnected system is developed using the optimal value function of each subsystem. To generate the optimal control policy, forward-in-time, neural networks (NNs) are employed to reconstruct the unknown optimal value function at each subsystem online. In order to retain the advantages of event triggered feedback for an adaptive optimal controller, a novel hybrid learning scheme is proposed to reduce the convergence time for the learning algorithm. The development is based on the observation that, in the event triggered feedback, the sampling instants are dynamic and results in variable inter event time. To relax the requirement of entire state measurements, an extended nonlinear observer is designed at each subsystem to recover the system internal states from the measurable feedback. Using a Lyapunov-based analysis it is demonstrated that the system states, observer errors remain locally uniformly ultimately bounded (UUB) and the control policy converge to a neighborhood of the optimal policy. Simulation results are presented to demonstrate the performance of the developed controller.

## I. INTRODUCTION

CONTROL of complex interconnected systems is one of the actively pursued areas of research in the control community [1]–[6]. The composition of interacting subsystems presents a unique challenge in designing control algorithms for such interconnected systems. Various control schemes for such interconnected system have been developed which can be broadly categorized into strictly decentralized design, adaptive controllers, robust controllers and distributed controllers. The interactions between subsystems are assumed to be weak and decentralized controllers are designed [1] while in robust control approach, in addition to the decentralized control policy, a compensation term for the interconnections is added [2], [3].

In the design approach which uses adaptive controllers, the additional compensation term is adaptive and it is designed to learn the interconnection terms to cancel their effects [5], [6]. In summary, the controllers in [1]–[6] are designed at

This research is funded in part by the intelligent systems center, Rolla, NSF ECCS #1406533 and CMMI #1547042. N. Vignesh and S. Jagannathan are with the Department of Electrical and Computer Engineering, Missouri University of Science and Technology, Rolla, MO, 65409, USA. e-mail: (vnvx4@mst.edu, sarangap@mst.edu).

Digital Object Identifier: 10.1109/TNNLS.2017.2693205

each subsystem as function of local states. By communicating the states to other subsystems and using the states of the neighboring subsystems, it was demonstrated that the transient performance of each subsystem could be improved [6]; further, various distributed control schemes are given in [7]–[9] and the references therein.

One of the impediments for implementing distributed control algorithms is the communication cost involved due to sharing of states among subsystems. To mitigate these costs, event-triggered controllers were proposed [7], [9]–[15]. Initially, the focus of the event triggered control research was to design an event triggering mechanism to reduce the frequency of control implementation using latest sensor measurements without compromising the system stability. However, in addition to stability, optimality is desired.

When the system dynamics are linear and known, optimal control problem can be solved to obtain a backward-in-time solution using Riccati equation [16]. When the system dynamics are nonlinear, solution to the Hamilton-Jacobi-Bellman (HJB) equation is required for optimal policy. Since the HJB equation does not have a closed-form solution [17], inspired by the reinforcement learning (RL) techniques [18], a suite of learning algorithms based on dynamic programming were proposed. These learning algorithms generate an approximation of the optimal value function and an approximate optimal control policy [1]–[3], [14], [15], [17], [19]–[27], [31] and are broadly classified as approximate dynamic programming (ADP) schemes.

The optimal value function is approximated by using an artificial neural network (NN) without solving the HJB equation directly. In order to learn the NN weights which minimize the approximation error, HJB residual error, the continuous time equivalent of Bellman error, is used. Starting with the policy/value iteration (PI/VI) based techniques proposed in [3], [15], [21], several improvements were suggested to implement the algorithms online. For PI/VI algorithms to converge, sufficiently large number of iterations are needed within each sampling interval [20], [23]. In contrast, several online ADP schemes are proposed in [1], [14], [20], [21], [23], [26], [28] which are suitable for online implementation.

These results of ADP based learning controllers were used to develop decentralized control schemes for interconnected systems using continuous/periodic feedback which guarantee stability and optimality [1], [3]. The RL based online ADP methods [1], [3] applied to interconnected systems typically requires extensive computations and exchange of feedback

information among subsystems through a communication network. Comparing with the traditional ADP design, the event based method samples the state and updates the controller only when it is necessary. Therefore, the computation and transmission costs are reduced.

The authors in [14], [15] developed near optimal controllers using event-triggered feedback when the system dynamics are nonlinear and uncertain by using one step temporal difference learning (TD ADP) [14] and PI based control scheme [15]. The event triggering condition introduced in [14] facilitated learning during the initial learning period. The event-triggering mechanism used estimated NN weights to determine the sampling instants and hence, required a mirror estimator. Moreover, the design [14], [15] requires an initial stabilizing control policy while TD learning demands longer convergence time, PI algorithm demands larger inter-event time.

In RL methods, to relax the need for accurate knowledge of the state transition probability and reward distribution, generalized policy iteration (GPI) [18] algorithm based on the classical dynamic programming was proposed. In the GPI algorithm, policy evaluation and improvement are two iterative steps. Depending on the number of iterations in each of these steps, several RL schemes are developed to generate a sequence of control actions which maximize certain reward function. The policy evaluation step learns the optimal value function and the policy improvement step learns the greedy action. For online control algorithms, the temporal difference learning (TDL) based RL schemes with one-step policy evaluation are more suitable. In TDL methods, using the one step feedback and the estimated future cost (bootstrapping), the value function parameters are updated [18].

Inspired by the TD ADP design in [23], this paper presents an online learning framework for interconnected systems by using event triggered state and output information. Several NNs [29] will be designed for estimating the optimal value functions by minimizing the HJB error [18]. Using the event-based control framework, the communication and computational resource utilization are significantly reduced.

To overcome the requirement of larger inter-event time as demanded by the event based PI algorithm and to reduce the convergence time of the event based TD learning algorithm, a TD ADP scheme combined with iterative learning between two event sampling instants is developed. As the event triggering instants are decided based on a dynamic condition, the time between any consecutive events is not fixed. Therefore, embedding finite number of iterations to tune the NN weights while assuring stable operation is non-trivial; especially due to the fact that the initial NN parameters and the initial control policy play a vital part in determining the stability during the learning phase. The net result is the development of a novel hybrid learning scheme using RL approach for approximate optimal regulation of interconnected dynamical systems with event triggered feedback information. First, the state vector of each subsystem is communicated to others.

Next, to relax the requirement of measuring the entire state vector, nonlinear observers are designed at each subsystem to estimate overall system state vector using outputs that are communicated only at event based sampling instants from

the other subsystems. The hybrid learning scheme with the observers is analyzed using Lyapunov technique. It is shown that closed-loop system is stable with both event triggered state and output feedback. Finally, two simulation examples are used to evaluate the effectiveness of the analytical design presented in the paper.

The contributions of this paper include: 1) development of an approximately optimal controller for the interconnected system using state and output feedback with event-triggered ADP approach in the presence of communication; 2) design of a novel hybrid learning scheme, with full state measurements and for the case when only the outputs are available, to reduce the convergence time of the TD learning algorithm ; 3) design of an adaptive event-triggering mechanism using locally available information; 4) design of extended nonlinear observers that utilizes the event-triggered output vector at each subsystem to relax the need for the entire state-vector to be measured and broadcasted, and 5) demonstration of local uniform ultimate boundedness (UUB) of the closed-loop system using Lyapunov analysis.

In the following presentation,  $\mathbb{N}$  is used to denote the set of natural numbers,  $\mathbb{R}$  is used to denote the set of real numbers. The norm operator  $\|\cdot\|$  for a vector denotes its Euclidean norm and for a matrix, its Frobenius norm;  $\cup$  denotes the set union operation,  $A \subseteq B$  implies  $A$  is a subset of  $B$  and  $A \in B$  denotes  $A$  is a member of the set  $B$ ,  $\exists a \in \mathbb{R}$  implies there exists a real number  $a$ .

## II. BACKGROUND

### A. System dynamics

Consider a nonlinear input affine system composed of  $N$  interconnected subsystems. Let the dynamics of each subsystem be represented as

$$\begin{aligned} \dot{x}_i(t) &= f_i(x_i) + g_i(x_i)u_i(t) + \sum_{j=1, j \neq i}^N \Delta_{ij}(x_i, x_j), \\ x_i(0) &= x_{i0}, \quad y_i = C_i x_i \end{aligned} \quad (1)$$

where  $x_i \in S_i \subseteq \mathbb{R}^{n_i \times 1}$  represents the state vector,  $\dot{x}_i \in \mathbb{R}^{n_i \times 1}$  represents the state derivative with respect to time for the  $i^{th}$  subsystem,  $u_i \in \mathbb{R}^{m_i}$  represents the control action,  $f_i(x_i) : \mathbb{R}^{n_i} \rightarrow \mathbb{R}^{n_i}$ ,  $g_i(x_i) : \mathbb{R}^{n_i} \rightarrow \mathbb{R}^{n_i \times m_i}$ ,  $\Delta_{ij} : \mathbb{R}^{n_i \times n_j} \rightarrow \mathbb{R}^{n_i}$ , represents the nonlinear dynamics, input gain function and the interconnection map between the  $i^{th}$  and  $j^{th}$  subsystems, respectively;  $y_i \in \mathbb{R}^{p_i}$  is the output vector with  $C_i \in \mathbb{R}^{p_i \times n_i}$ , a constant matrix and  $S_i$  is a compact set. The dynamics of the augmented system are expressed as

$$\dot{X}(t) = F(X) + G(X)U(t), \quad X(0) = X_0 \quad (2)$$

where  $X \in S \subseteq \mathbb{R}^{n \times 1}$ ,  $U \in \mathbb{R}^m$ ,  $G(X) : \mathbb{R}^n \rightarrow \mathbb{R}^{n \times m}$ ,  $F(X) : \mathbb{R}^n \rightarrow \mathbb{R}^n$ ,  $U = [u_1^T, \dots, u_N^T]^T$ ,  $m = \sum_{i=1}^N m_i$ ,  $n = \sum_{i=1}^N n_i$ ,  $\dot{X} = [\dot{x}_1^T, \dots, \dot{x}_N^T]^T$ ,  $G(X) = \text{diag}(g_1(x_1), \dots, g_N(x_N))$ ,  $F(X) = [(f_1(x_1) + \sum_{j=2}^N \Delta_{1j})^T, \dots, (f_N(x_N) + \sum_{j=1}^{N-1} \Delta_{Nj})^T]^T$  and  $S$  is a compact set obtained as a result of finite union of  $S_i$ . The following assumptions on the system dynamics will be made in the analysis presented in this paper.

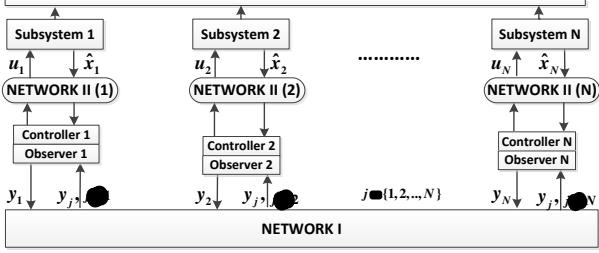


Fig. 1. Interconnected system

*Assumption 1:* Each subsystem described by (1) and the interconnected system (2) are controllable.

*Assumption 2:* The nonlinear maps  $F(X), G(X)$  are Lipschitz functions [30] in the compact set  $S$ .

*Assumption 3:* There exists  $g_{im}, g_{iM} > 0 : g_{im} < \|g_i(x_i)\| \leq g_{iM}, \forall i \in \{1, \dots, N\}$ .

*Assumption 4:* The feedback controller using event-triggered states assumes that the states are measurable. This will be relaxed in the subsequent design using outputs and extended nonlinear observers. Delay and packet loss in the communication network are assumed to be absent.

Note that the dynamics of the system (2) and the component subsystems (1) do not explicitly describe how the subsystems are interconnected. Interconnected systems with state interactions will be the focus of this paper and the systems with output/control interactions are not dealt in this paper. The control scheme for the augmented system (2) is represented in the form of a block diagram in Fig. 1.

As seen in Fig. 1, at each subsystem an event-sampling mechanism monitors the subsystem states/outputs to determine the feedback/broadcast instants. For the case of output feedback, only the outputs from each subsystem are broadcast and the output vector is used to reconstruct the states of all the subsystems to be used in the controller. Due to the flexibility offered by the networked control architecture, the interconnected system represented in Fig. 1 consisting of two networks is preferred. The Network I enables information exchange between subsystems, while Network II is a local communication network which closes the feedback control loop of each subsystem. The communication resources involved in the control design of such systems motivated the use of event based feedback.

Define a subsequence  $\{t_k\}_{k \in \mathbb{N}} \subset t$  to represent the event triggering instants. The state of the  $i^{th}$  subsystem at the sampling instant  $t_k^i$  is denoted as  $x_i(t_k^i)$ . During the inter-event period, latest sensor measurements are not updated at the controller. The difference between the actual state and the states available at the controller results in an event-sampling error given by

$$e_i(t) = x_i(t) - x_i(t_k^i), \quad t_k^i \leq t < t_{k+1}^i. \quad (3)$$

This error is reset to zero at the sampling instants due to the feedback update. A brief background on the design of

optimal controller using aperiodic, event-triggered feedback is presented in the next subsection.

### B. Event based optimal control policy

Let the performance of the interconnected system (2), be evaluated using the following function

$$V(X, U) = \int_t^\infty [Q(X) + U^T(\tau)RU(\tau)]d\tau \quad (4)$$

where  $R \in \mathbb{R}^{m \times m}$ ,  $Q(X) : \mathbb{R}^n \rightarrow \mathbb{R}$  with  $Q(0) = 0$ , represent positive definite functions which penalize the states and control action, respectively. Define a compact set  $B$ . Use the integral in (4) to denote the infinite horizon value function  $V(X(t))$  defined in  $B$ . If  $V(X(t))$  and its derivative are continuous in its domain, the time derivative of the  $V(X(t))$  (4) is given by [16], [23]

$$\dot{V}(X(t)) = -[Q(X) + U^T(t)RU(t)] = \frac{\partial V^T}{\partial X} \dot{X}. \quad (5)$$

Assuming that a minimum of the value function exists and it is unique [21], the optimal control policy can be obtained as

$$U^* = -\frac{R^{-1}}{2} G^T(X) \frac{\partial V^*}{\partial X}. \quad (6)$$

Substituting (6) in (5), the HJB equation [23] is obtained as

$$H = Q(X) + \frac{\partial V^*}{\partial X} F(X) - \frac{1}{4} \frac{\partial V^*}{\partial X} G(X) R^{-1} G^T(X) \frac{\partial V^*}{\partial X}. \quad (7)$$

When the feedback is aperiodic and event-based, the Hamiltonian in (7) between events can be represented using a piecewise continuous control input as

$$H = \left[ Q(X) + U^{*T}(t_k) R U^*(t_k) \right] + \frac{\partial V^*}{\partial X} \dot{X}. \quad (8)$$

The piecewise continuous control policy which minimizes the Hamiltonian in (8) is defined as

$$U^*(t_k) = -\frac{R^{-1}}{2} G^T(X_e) \frac{\partial V^*}{\partial X_e} \quad (9)$$

with  $X_e = X(t_k)$ , the state held at the actuator using a zero order hold (ZOH) circuit between  $t_k, t_{k+1}$ , for all  $k \in \mathbb{N}$ .

*Remark 1:* The control policy will be piecewise continuous due to the limited feedback availability and ZOH. The system dynamics can be considered to be driven by the event-sampling error (3) which is nonzero between events.

The function approximation property of NNs with event-triggered feedback is presented next.

### C. Neural network approximation using event based feedback

With the following standard assumption, the effect of the aperiodic event based feedback on the approximation property of the NN observed in [14] is stated next.

*Assumption 5:* The NN reconstruction error and its derivative,  $\varepsilon_i(x), \nabla_x \varepsilon_i(x)$ , the constant target weights  $\theta_i^*$  and the activation function  $\phi(x)$ , which satisfies  $\phi(0) = 0$ , are bounded in the compact set  $S$ .

Given,  $\chi : A \subseteq \mathbb{R}^n \rightarrow \mathbb{R}$ , a smooth function in a compact set  $A$  and  $\varepsilon_M > 0, \exists \theta^* \in \mathbb{R}^{P \times 1} : \chi(x) = \theta^{*T} \phi(x_e) + \varepsilon_e, t_k \leq$

$t < t_{k+1}$ , with  $\|\varepsilon_e\| < \varepsilon_M$ ,  $\forall x_e \in A$ , where,  $\phi(x_e)$  is a basis function driven by the inputs  $x(t), e(t)$  and  $\varepsilon_e = \theta^{*T}(\phi(x_e + e) - \phi(x_e)) + \varepsilon$ , the reconstruction error driven by  $e(t)$ . The error  $e(t)$  is due to the difference between  $x(t_k), x(t)$ .

*Remark 2:* The NN approximation with state vector sampled at event triggering instants as input is a function of event-sampling error. Since the reconstruction error  $\varepsilon_e$  depends on the error due to event sampling, a direct relationship between approximation accuracy and the frequency of events is revealed.

*Remark 3:* One of the motivations behind the proposed learning algorithm is to decouple the relationship between the accuracy of approximation and the sampling frequency. In [14], this trade-off is handled by designing the event triggering condition based on the estimated weights and the states of the system. This resulted in a inverse relationship between the inter-event time and the weight estimation error, thereby forcing more events when the difference between the estimated NN weight and the target weights is large.

In the next section, the control scheme is introduced and the stability results are presented in section IV.

### III. DISTRIBUTED CONTROLLER DESIGN

In this section, firstly, a novel hybrid learning scheme is used in the design of distributed approximately optimal controller with state feedback. Using a NN based online approximator, optimal value function is approximated at each subsystem. Taking into account the interactions between subsystems, the distributed control law is desired to be a function of  $X(t)$ . Later, nonlinear observers are introduced to relax the requirement of full state measurements. In order to avoid redundancy, only the important results are presented for the output feedback controllers.

With the following assumption, the design of distributed control policy is introduced.

*Assumption 6:*  $V^*(X) \in C^1(S)$  is a unique solution to the HJB equation, where  $C^1(S)$  represents the class of continuous functions defined in  $S$  and have continuous derivatives in  $S$ .

*Proposition [26]:* Consider the augmented system dynamics in (2) with the individual subsystems (1),  $\forall i \in \{1, 2, \dots, N\}$ ,  $\exists u_i^*$  which is a function of  $X(t)$ , such that the cost function (4) is minimized.

*Proof:* First, consider the infinite horizon value function defined by (4) for the augmented system in (2). Define  $R = \text{diag}(R_1, R_2, \dots, R_N)$ ,  $Q(X) = \sum_{i=1}^N Q_i(X)$ ,  $U(X(t)) = [u_1^T, \dots, u_N^T]^T$ ,  $\frac{\partial V^T}{\partial X} = \left[ \frac{\partial V_1^T}{\partial x_1}, \frac{\partial V_2^T}{\partial x_2}, \dots, \frac{\partial V_N^T}{\partial x_N} \right]$ ,  $V(X(t)) = \int_t^\infty \left( \sum_{i=1}^N [Q_i(X) + u_i^T R_i u_i] \right) d\tau = \int_t^\infty \sum_{i=1}^N V_i(X(\tau)) d\tau$ . The Hamiltonian (8) becomes  $H(X) = \left[ \frac{\partial V_1^T}{\partial x_1}, \dots, \frac{\partial V_N^T}{\partial x_N} \right] [\dot{x}_1^T, \dots, \dot{x}_N^T]^T + \left( \sum_{i=1}^N [Q_i(X) + u_i^T R_i u_i] \right) = \sum_{i=1}^N H_i(X, u_i)$ :  $H_i(X, u_i) = \left( \frac{\partial V_i^T}{\partial x_i} \dot{x}_i + Q_i(X) + u_i^T R_i u_i \right)$ .

For optimality, each subsystem should generate a control policy from  $H_i(x, u_i)$  as

$$u_i^* = -\frac{1}{2} R_i^{-1} g_i^T(x_i) \frac{\partial V_i^*}{\partial x_i}, \forall i \in 1, 2, \dots, N. \quad (10)$$

By designing controllers at each subsystem to generate (10), cost function (4) of the augmented system is minimized.

*Remark 4:* A strictly decentralized controller can be realized by designing (10) as a function of  $x_i(t)$ . Despite the simplicity of such controller, the efforts in [6] highlighted the unacceptable performance observed, especially in the transient period, as a result of such design approach. Therefore, (10) is desired to be a function of  $X(t)$  and it can be considered as

$$u_i^* = -\frac{1}{2} R_i^{-1} g_i^T(x_i) \frac{\partial V_{i,i}^*}{\partial x_i} - \frac{1}{2} R_i^{-1} g_i^T(x_i) \sum_{\substack{j=1 \\ j \neq i}}^N \frac{\partial V_{i,j}^*}{\partial x_i} \quad (11)$$

where  $V_{i,j}^*$  is the cost due to interconnections and  $V_{i,i}^*$  is the optimal cost of the  $i^{th}$  subsystem when the interconnections are absent and  $V_i^* = V_{i,i}^* + V_{i,j}^*$ . The control policy as expressed in (11) is composed of two parts. The first part denotes the optimal control policy for a decoupled subsystem wherein the interconnections are absent while the second part compensates for the interconnections.

*Remark 5:* Note that the control policy (10) is considered to be distributed and it is equivalent to (11). In the decentralized control policy, the second term in (11) is zero [1]. This term explicitly takes into account the interconnection terms in the subsystem dynamics and it is expected to compensate for the interconnections. An equivalent control policy for the linear interconnected system using Riccati solution can be obtained as  $u_i^* = -k_{ii}x_i - \sum_{\substack{i=1 \\ i \neq j}}^N k_{ij}x_j$ . Here,  $k_{ii}, k_{ij}$  are the diagonal and off-diagonal entries of the Kalman gain matrix corresponding to the optimal controller for the interconnected system (2), with linear dynamics.

The design using state feedback is presented next.

#### A. State feedback controller design

We use the artificial NNs [29] to represent the optimal value function in a parametric form using NN weights and a set of basis function with a bounded approximation error. Using the parameterized representation, the value function is represented as  $V(X) = \theta^T \phi(X) + \varepsilon(X)$ , where  $\phi(X)$  is a basis function and  $\varepsilon(X)$  is the bounded approximation/reconstruction error. Let the target NN weights be  $\theta_i^*$  and the estimated NN weights be  $\hat{\theta}_i$  at the  $i^{th}$  subsystem. The parameterized HJB equation with approximate optimal value function can be obtained as

$$Q_i(X) + \theta_i^{*T} \nabla_x \phi(x) \bar{f}_i(x) - \frac{1}{4} \theta_i^{*T} \nabla_x \phi(x) D_i \nabla_x^T \phi(x) \theta_i^* + \varepsilon_{i_{HJB}} = 0 \quad (12)$$

where  $\varepsilon_{i_{HJB}} = \nabla_x \varepsilon_i^T (\bar{f}_i(x) - \frac{D_i}{2} (\nabla_x^T \phi(x) \theta_i^* + \nabla_x \varepsilon_i)) + \frac{1}{4} \nabla_x \varepsilon_i^T D_i \nabla_x \varepsilon_i$ ,  $\bar{f}_i(x) = f_i(x_i) + \sum_{j=1}^N \Delta_{ij}(x_i, x_j)$ , the partial derivative of the optimal cost function  $V_i^{*T}$  with respect to  $x_i$  is  $\nabla_x^T \phi(x) \theta_i^*$  and  $D_i = D_i(x_i) = g_i(x_i) R_i^{-1} g_i^T(x_i)$ . Let  $\|\nabla_x^T \phi(x) \theta_i^*\| \leq V_{xiM}$ ,  $\|D_i\| \leq D_{iM}$ . Now, using the estimated weights  $\hat{\theta}_i$ , the control input (10), can be written as

$\hat{u}_i = -0.5R_i^{-1}g_i^T\hat{\theta}_i^T\nabla_x\phi(x)$  and the parameterized Hamiltonian equation is derived as

$$\hat{H}_i = Q_i(X) + \hat{\theta}_i^T\nabla_x\phi(x)\bar{f}_i(x) - \frac{1}{4}\hat{\theta}_i^T\nabla_x\phi(x)D_i\nabla^T_x\phi(x)\hat{\theta}_i. \quad (13)$$

In the GPI literature, equation (7) is used to evaluate the value function for the given policy. Since it is a consistency condition, if the estimated value function is the true optimal value function for the control policy (10), then  $\hat{H}_i = 0$ . Due to the estimated quantity  $\hat{\theta}_i$ , the value function calculated using the estimated weights is not equal to the optimal value function. This will result in a HJB residual error and  $\hat{H}_i = 0$  is no longer true. The estimates  $\hat{\theta}_i$  are now updated such that the HJB residual error is minimized. Levenberg-Marquardt algorithm [21] can be used as a weight update rule and the weight estimates evolve based on the dynamic equation given by  $\dot{\hat{\theta}}_i = \frac{-\alpha_{i1}\sigma_i\hat{H}_i}{(\sigma_i^T\sigma_i+1)^2}$ , where  $\alpha_{i1}$  is the learning step and  $\sigma_i = \nabla_x\phi(x)\bar{f}_i(x) - \frac{1}{2}\nabla_x\phi(x)D_i\nabla_x^T\phi(x)\hat{\theta}_i$ . This weight tuning rule ensures the HJB residual error convergence while stability of the closed-loop system when the estimated weights are used in the control policy is not a given, especially, if the initial control policy is not stabilizing. Therefore, to relax the dependence on the initial control policy in dictating the stability of the closed-loop system, a conditional stabilizing term was appended in the weight update rule proposed in [23]. Here, we propose the following weight update rule

$$\dot{\hat{\theta}}_i(t) = -\frac{\alpha_{i1}}{(\sigma_i^T\sigma_i+1)^2}\sigma_i\hat{H}_i + \frac{1}{2}\beta_i\nabla_x\phi(x)D_iL_{ix}(x_i) - \kappa_i\hat{\theta}_i \quad (14)$$

where  $\kappa_i, \beta_i$  are positive design parameters,  $L_{ix}(x_i)$  is the partial derivative of a positive definite Lyapunov function for the  $i^{th}$  subsystem with respect to the state. Since the controller has access to the feedback information only when an event is triggered, (14) will have to be slightly modified and this will be presented in the next subsection.

*Remark 6:* By utilizing the nonlinear maps  $g_i$ , the stabilizing term in (14) is appended to the NN weight tuning rule to relax the requirement of initial stabilizing control [23]. In the event-triggered implementation of the controller presented in this paper, the stabilizing term in the update rule ensures stability of the closed-loop system at the event based sampling instants and the sigma-modification term ensures that the weights are bounded.

The event triggered state feedback controller design is introduced next.

### B. Event triggered state feedback controller

For the near optimal distributed control design with event-triggered state feedback, the error (3) introduced due to aperiodic feedback will drive the control policy between two event based sampling instants. With the estimated optimal value function and the estimated optimal control policy, the Hamiltonian is represented as

$$\hat{H}_i(X, \hat{u}_{i,e}) = \frac{\partial\hat{V}^T}{\partial x_i}\dot{x}_{i,e} + [Q_i(X_e) + \hat{u}_{i,e}^T(t)R_i\hat{u}_{i,e}(t)] \quad (15)$$

where  $(.)_e$  denotes the influence of (3) due to event-based feedback and this notation will be followed henceforth. Using the parameterized representation of the approximate value function, we get

$$\begin{aligned} \hat{H}_i &= Q_i(X_e) + \hat{\theta}_i^T\nabla_x\phi(x_e)\bar{f}_i(x_e) \\ &\quad - \frac{1}{4}\hat{\theta}_i^T\nabla_x\phi(x_e)D_{i,\varepsilon}\nabla^T_x\phi(x_e)\hat{\theta}_i \end{aligned} \quad (16)$$

where  $D_{i,\varepsilon} = D_i(x_{i,e})$ . Finally, we propose the NN weight tuning rule which minimizes the HJB residual error, with  $\rho = (\sigma_{i,e}^T\sigma_{i,e} + 1)$ , as

$$\dot{\hat{\theta}}_i(t) = \begin{cases} -\frac{\alpha_{i1}}{\rho^2}\sigma_i\hat{H}_i + \frac{1}{2}\beta_i\nabla_x\phi(x)D_iL_{ix}(x_i) - \kappa_i\hat{\theta}_i, & t = t_k^i \\ 0, & t \in (t_k^i, t_{k+1}^i). \end{cases} \quad (17)$$

The estimated NN weights,  $\hat{\theta}_i$ , at each subsystem are not updated between events. To determine the time instants  $t_k$ , a decentralized event-triggering condition is required. Define a locally Lipschitz Lyapunov candidate function  $L_i(x_i)$ , for the  $i^{th}$  subsystem such that  $L_i(x_i) > 0, \forall x_i \in S \setminus \{\vec{0}\}$ . Events are generated such that the following condition is satisfied

$$L_i(x_i(t)) \leq (1 + t_k - t)\Gamma_i L_i(x_i(t_k)), \quad t_k \leq t < t_{k+1} \quad (18)$$

with  $0 < \Gamma_i < 1$ .

*Remark 7:* Note that the event-triggering condition (18) requires only the local states. Also note that the  $k^{th}$  event sampling instant at any two subsystems need not be the same and  $t_k^i$  used in the equations above represents the time instant of the occurrence of the  $k^{th}$  event at the  $i^{th}$  subsystem. Since the estimated weights are not used in (18) a mirror estimator is not required [14].

Next, the nonlinear observer which utilizes the output from the subsystems obtained at event-based sampling instants to reconstruct the internal state information is presented which requires the following standard assumption.

*Assumption 7:* The subsystems are assumed to be observable. This is required to enable reconstruction of the states from the measured outputs.

### C. Event triggered output feedback controller

Output feedback controllers use the measured quantity to estimate the internal system states using observers. The estimated states are then utilized to design the controllers. Since, it is desired that the outputs be communicated among subsystems, the observers at each subsystem are designed so that they estimate the state vector of all the subsystems using the event-triggered outputs.

To avoid redundancy, all the equations for the controller are not explicitly presented for output feedback based design. For the implementation of output feedback controller, estimated states will replace the actual states in the design equations presented in the previous subsection. However, the stability analysis for output feedback controller is presented in detail. In order to develop an event-triggering condition, we could substitute the outputs in place of the states in (18). In the analysis, the event-triggering condition can be represented in terms of the state vector using the linear map  $C_i$ .

Next, the observer which estimates the state vector using the measured output with measurement error is presented.

In order to estimate the system state vector using the output information obtained at the event-based sampling instants, consider the observer at  $i^{th}$  subsystem with dynamics

$$\dot{\hat{X}}_i(t) = F(\hat{X}_i) + G(\hat{X}_i)U_{i,e}(t) + \mu_i[Y_{i,e}(t) - C\hat{X}_i(t)] \quad (19)$$

where  $\hat{X}_i, \mu_i, Y_{i,e}$  represent the overall estimated state vector, observer gain matrix and event-triggered output vector of the overall system, respectively, at the  $i^{th}$  subsystem,  $C$  is the augmented matrix composed of  $C_i$ , each with appropriate dimensions. The output vector is a function of the measurement error since the output from each subsystem is shared only when an event is triggered.

Defining the difference between the actual state and the estimated state vectors at the  $i^{th}$  subsystem as the state estimation error,  $\tilde{X}_i(t) = X_i(t) - \hat{X}_i(t)$ , the evolution of the state estimation error is described by the differential equation

$$\begin{aligned} \dot{\tilde{X}}_i(t) &= F(X_i) + G(X_i)U_{i,e}(t) - [F(\hat{X}_i) + G(\hat{X}_i)U_{i,e}(t)] \\ &\quad - \mu_i[Y_{i,e}(t) - C\hat{X}_i(t)]. \end{aligned} \quad (20)$$

Next, the boundedness of the state estimation error with event-triggered output feedback is presented assuming the distributed control policy is admissible.

*Lemma 1:* For the augmented system given in (2) composed of interconnected subsystems given in (1), consider the proposed observer (19) at each subsystem with the error dynamics (20) and let the measurement error (3) be bounded. The observer estimation error is locally UUB, provided the control policy is optimal and the observer gains are chosen such that  $\eta_{i,o1}, \eta_{i,o2} > 0$ , where the design variables  $\eta_{i,o1}, \eta_{i,o2}$  are defined in the proof.

*Proof:* See appendix.

Since the separation principle does not hold for nonlinear systems, the stability of the controllers together with the observers, operating online, should be analyzed.

Note that the convergence of the NN weights is coupled with the number of events when the weight update rule (17) is used. This significantly reduces the convergence time [14]. To decouple this relationship between the number of events and the learning time, a new NN weight adaption rule is introduced in the next subsection.

#### D. Hybrid learning algorithm

The results of event based function approximation [14] shows that the approximation error in the optimal value function and the optimal control action generated will depend on the frequency of events. The TD ADP scheme in [14] presents a NN approximator wherein the NN weight updates occur only at the event triggering instants  $t_k^i$ . In contrast, the ADP scheme in [15] performs iterative learning, assuming significant iterations could be carried out during the inter-event period resulting in a greedy policy at every event-triggered update of the control action. It should be noted that the iterative updates can be related to the GPI in RL wherein the finite

iterations using the past values reduce the HJB error and aid the estimated weights move towards their target weights.

Thus, the learning scheme proposed here is inspired by the GPI and the NN weights are tuned using the weight tuning rule

$$\dot{\hat{\theta}}_i(t) = \begin{cases} -\frac{\alpha_{i1}}{\hat{\rho}^2} \hat{\sigma}_i \hat{H}_i + \frac{1}{2} \beta_i \nabla_x \phi(\hat{x}) \hat{D}_i L_{ix}(\hat{x}_i) - \kappa_i \hat{\theta}_i, & t = t_k^i \\ -\frac{\alpha_{i1}}{\hat{\rho}^2(t_k^i)} \hat{\sigma}_{i,e}(t_k^i) \hat{H}_{i,e}(\hat{\theta}_i(t)) - \kappa_i \hat{\theta}_i(t), & t_k^i < t < t_{k+1}^i. \end{cases} \quad (21)$$

To denote the use of estimated states from the observer,  $(\cdot)$  notation is used for the functions  $D_i, \rho_i, \sigma_i$ . Whenever an event occurs, new feedback information is updated at the controller and broadcast to the neighboring subsystems. The weights are tuned with the new feedback information and the updated weights are used to generate the control action which is applied at the actuator. In the inter-event period, past feedback values are used to evaluate the value function and the policy using the HJB equation. This is done by adjusting the estimated weights in the inter-event period according to (21) so that  $\hat{\theta}_i$  moves towards  $\theta_i^*$ . The stability of the system is preserved as a consequence of the additional stabilizing term in (21). Using the actual states in place of the estimated states, the update rules for the hybrid learning scheme can be derived for the state feedback controller.

*Remark 8:* As the time between two successive events increases, more time is available for the iterative weight updates. Therefore, HJB residual error is reduced considerably resulting in an approximately optimal control action at every event triggering instant [15].

*Remark 9:* The event-sampling condition [10] was demonstrated to have large average inter-event period than the existing event sampling schemes. It should be noted that the proposed learning algorithm can be implemented with any event-triggering condition.

*Remark 10:* In the traditional RL literature, the GPI is used and a family of TD algorithms are presented, such as TD(0),  $n$ -TD, TD( $\lambda$ ) [18]. All these learning algorithms [18] have a fixed number of iterative weight updates for policy evaluation. In contrast, the event-triggered control framework cannot ensure fixed inter-event time and hence, the proposed hybrid algorithm is most relevant and applicable in the event based online learning control framework.

For the stability analysis, first, using the fact that the optimal control policy results in a stable closed-loop system, a time-varying bound on the closed-loop dynamics are defined [23] as  $\|F(X) + G(X)U^*\| \leq \psi \|X\|$ , with  $\psi > 0$ . It was also shown in [23] that there exists positive constant  $\zeta_1$  such that,  $\|L_x(X)\| \|f(X) + g(X)U^*\| \leq -\zeta_1 \|L_x(X)\|^2$ , with the Lyapunov function  $L(X)$ , its gradient  $L_x(X)$ , with respect to the state vector. For example,  $L(X) = 0.5(X^T X)$ , we get  $\|X^T\| \|f(X) + g(X)U^*\| \leq -\zeta_1 \|X\|^2$ , which will be used to analyze the proposed controller. With these results, the stability analysis of the proposed state-feedback controller, output feedback controller with event-triggered feedback will be presented in the next section.

#### IV. STABILITY ANALYSIS

In this section, Lyapunov stability theory [30] is used to analyze the closed loop stability of the nonlinear interconnected system with the proposed event-triggered distributed controller using state and output feedback. For the analysis of the event-triggered controller, first, we prove that the proposed distributed controller admits a Lyapunov function for the closed loop system which satisfies local input-to-state stability like conditions, resulting in local UUB of all the states, weight estimation error and state estimation error. Further, the stability during the inter-event time and sampling instants are analyzed. This ensures that the event based implementation of the controller will result in stable operation of the closed-loop system.

With the following equations, the stability results are presented next. Let the error in the NN weight estimate be defined as  $\tilde{\theta}_i = \theta_i^* - \hat{\theta}_i$ , and the target weights be constants and bounded by  $\theta_{iM}$ . Consider the Hamiltonian (16), and the ideal HJB equation given in (12), adding and subtracting  $Q_i(X)$  in (16) and rewriting the Hamiltonian in terms of  $\tilde{\theta}_i$ , we get the following equations:

$$\begin{aligned} \hat{H}_{i,e} &= -\tilde{\theta}_i^T \sigma_{i,e} + \frac{1}{4} \tilde{\theta}_i^T \nabla_x \phi(x_e) D_{i,\varepsilon} \nabla^T x \phi(x_e) \tilde{\theta}_i + Q_i(x_e) \\ &+ \theta_i^{*T} [\nabla_x \phi(x_e) \bar{f}_i(x_e) - \nabla_x \phi(x) \bar{f}_i(x)] - \varepsilon_{i_{HJB}} - Q_i(X) \\ &+ \frac{1}{4} \theta_i^{*T} [\nabla_x \phi(x) D_i \nabla^T x \phi(x) - \nabla_x \phi(x_e) D_{i,\varepsilon} \nabla^T x \phi(x_e)] \theta_i^*. \end{aligned} \quad (22)$$

Similarly, for the case of output feedback, we have

$$\begin{aligned} \hat{H}_i &= -\tilde{\theta}_i^T \hat{\sigma}_{i,e} + \frac{1}{4} \tilde{\theta}_i^T \nabla_x \phi(\hat{x}_e) \hat{D}_{i,\varepsilon} \nabla^T x \phi(\hat{x}_e) \tilde{\theta}_i + Q_i(\hat{X}_e) \\ &+ \frac{1}{4} \theta_i^{*T} [\nabla_x \phi(x) D_i \nabla^T x \phi(x) - \nabla_x \phi(\hat{x}_e) \hat{D}_{i,\varepsilon} \nabla^T x \phi(\hat{x}_e)] \theta_i^* \\ &- \varepsilon_{i_{HJB}} + \theta_i^{*T} (\nabla_x \phi(\hat{x}_e) \bar{f}_i(\hat{x}_e) - \nabla_x \phi(x) \bar{f}_i(x)) - Q_i(X). \end{aligned} \quad (23)$$

First, the stability results of the output feedback control scheme are presented in detail.

*Theorem 1:* Consider the nonlinear dynamics of the augmented system (2) with the equilibrium point at origin. Let the initial states  $x_{i0}, \hat{X}_{i0} \in S$  and let  $\hat{\theta}_i(0)$  be defined in a compact set  $\Omega_{i\theta}$ . Use the update rule defined in (14), with the estimated states, to tune the NN weights. With the estimated states evolving according to the observer dynamics given by (19), there exists  $\eta_{i's} > 0$  such that  $\tilde{\theta}_i, X(t)$  and the observer error dynamics are locally uniformly ultimately bounded by  $\xi_{icl}$  in the presence of a bounded external input. The constants,  $\eta_{i's}$  and the bound,  $\xi_{icl}$ , are defined in the proof.

*Proof:* Refer to the Appendix.

*Remark 11:* This analytical result in Theorem 1 is equivalent to the local ISS condition [30]. The reconstruction and the measurement errors can be considered as external inputs to the system. However, the boundedness of the event-based measurement error will be established in the next theorem using the decentralized event-triggering condition.

*Theorem 2:* Consider the nonlinear interconnected system described by (2) wherein the initial states  $x_{i0}, \hat{X}_{i0} \in S$ , and let the NN weights be initialized in a compact set  $\Omega_{i\theta}$ . Consider the weight tuning rule defined in (21) using the estimated states

and the event-triggering mechanism satisfying (18), with the measured outputs at each subsystem. With the estimated states evolving according to the observer dynamics given by (19), there exists  $\eta_{i's} > 0$  such that  $\tilde{\theta}_i, X(t)$  and the observer error dynamics are locally uniformly ultimately bounded by  $\xi_{icl}$  wherein the bound is obtained independent of the measurement error. The constants,  $\eta_{i's}$  and the bound,  $\xi_{icl}$ , are defined in the proof.

*Proof:* Refer to the Appendix.

*Corollary:* 1) Consider the nonlinear interconnected system given by (2) with origin being the equilibrium point and the initial states  $x_{i0}, \hat{X}_{i0} \in S$ , and let  $\hat{\theta}_i(0)$  be defined in a compact set  $\Omega_{i\theta}$ . Use the update rule defined in (14) to tune the NN weights at each subsystem. Then, there exists computable positive constants  $\alpha_{i1}, \beta_i, \kappa_i$  such that  $\tilde{\theta}_i$  and  $X(t)$  are locally uniformly ultimately bounded with the bounds  $\xi_\theta, \xi_x$  respectively, when there is a non-zero bounded measurement error. 2) Using the event-sampling condition (18), it can be shown the closed-loop system is locally UUB when the NN weights are tuned using (17) and (21).

*Proof:* Since the stability results for the state feedback controller can be directly obtained from Theorem 1 and Theorem 2 by setting the observer estimation error to zero, detailed derivations are not provided for the corollary. Refer to the Appendix for the main results.

*Remark 12:* Results from Theorems 1 and 2 can be used along with Assumption 2 to establish the non-zero minimum inter-event time [7], [9], [10]. However, since the inter-event time is dynamically changing, ensuring sufficient time availability to carry out significantly large number of weight updates between any successive events is not feasible. Therefore, algorithms like policy iteration or value iteration are restrictive for event based control implementation.

*Remark 13:* Redundant events can be prevented by using a dead-zone operator as soon as the states of each subsystem converge to their respective bounds.

*Remark 14:* The learning algorithm and the corresponding stability results derived for the closed-loop nonlinear system can be easily extended for linear interconnected system.

*Remark 15:* The event-sampling mechanism at each subsystem operates asynchronously, resulting in lesser network congestion. However, suitable communication protocol is required to be utilized along with the proposed controller to minimize the packet losses due to collision and other undesired network performance [7], [9]. Further, it is shown that the event triggering condition ensures continuity of the Lyapunov function for states at the sampling instants [10].

*Remark 16:* The weight tuning rules for the online approximator in (21) are used for event-triggered implementation of state and output feedback controllers. The bounds  $\xi_{icl}$  can be made arbitrarily small by appropriate choice of  $\alpha_{i1}, \beta_i, \kappa_i$  in the weight update rule satisfying the Lyapunov stability results.

*Remark 17:* The iterative learning, presented in [3], [15], [21], results in the value function approximate that yield approximately optimal, hence, stabilizing control input at each time step. This yields  $\tilde{\theta}_i = 0$  in each of the algorithms [3], [15], [21] which reduces the complexity of analysis. In this paper, the stabilizing term  $\frac{1}{2} \beta_i \nabla_x \phi(\hat{x}) \hat{D}_i L_{ix}(\hat{x}_i)$  in the weight

tuning rule (21) is used to ensure stability of the closed loop system in the presence of non-zero  $\hat{\theta}_i$ .

*Remark 18:* In the adaptive control theory, the sigma/epsilon modification [29] terms in the adaptation rule ensures that the actual weights are bounded in the presence of bounded disturbances. It also helps in avoiding the parameter drift and also relaxes the PE condition. In all the ADP designs [21], the PE condition is required for convergence of the weight estimation errors and it is achieved by adding random signal to the control policy [14], [20], [23]. This also had an additional benefit of being an exploratory signal. In RL literature, the dilemma of exploration versus exploitation is greatly discussed [18]. For a learning problem, the exploratory noise signal helps the learning mechanism to explore the search space to find the exact solution and ensures observability conditions while learning [21]. However, for the online control problem, stability is more important and is given priority. Therefore, explicitly adding random exploratory signal to the control policy is undesirable.

*Remark 19:* In the RL literature, the one step TD algorithm is proven to have convergence issues [18] due to bootstrapping. This occurs as the parameter values that approximate the value function grow unbounded as the approximation is based on 'guesses' [18]. However, convergence results for online one-step TD algorithms are presented in [14], [20], [23] under certain conditions. These algorithms utilize the stabilizing terms in the parameter update rule and present local convergence.

*Remark 20:* For the output feedback controller, an additional uncertainty due to estimated states is introduced during the learning period. Moreover, the computations are increased due the observer present at each subsystem. The state estimation error forces frequent events when compared to the state feedback controller where the state estimation error is absent. However, for practical applications, all the states are not measured and with output feedback, only the output vector is broadcast through the network when compared to the entire state vector. Typically,  $p_i \leq n_i$  in (1) and the packet size of the outputs are expected to be smaller than that of the states. Therefore, the output feedback controller requires much lower network resources when compared to state feedback controller.

*Remark 21:* The location of the observer is crucial and there are several locations which are feasible to place an observer operating with event-triggered feedback, as discussed in the literature [11], [15]. For the interconnected system, the extended observers discussed here are placed along with the controller for the following reasons b) a) only the output from each subsystem is broadcast through the network; b) using the outputs from all the subsystem, the overall state vector can be reconstructed at each subsystem, as required by the distributed controllers. These advantages are lost when the observers are placed along with the sensors at each subsystem. In order to eliminate an additional event-sampling mechanism at each subsystem, the observer states are held constant between the event sampling instants.

Simulation results are presented in the next section for two examples to substantiate the analytical design.

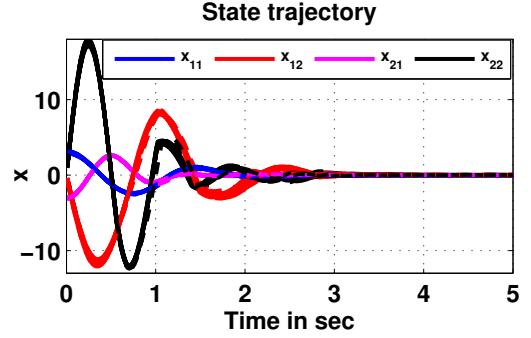


Fig. 2. State trajectories (Linear example)

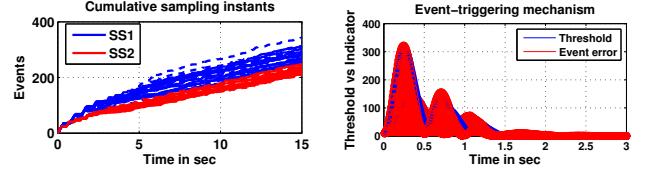


Fig. 3. Event-triggering mechanism

## V. SIMULATION RESULTS

In this section, two examples are considered to verify the analytical design presented in the paper. The first example includes a system of two inverted pendulums connected by spring. The applicability of the proposed control algorithm for linear system is verified by considering the linear dynamics first and then the nonlinear dynamics are considered. In the second example, a more practical nonlinear system with three interconnected subsystems is considered.

*Example 1:* The example used here has two inverted pendulum connected by spring [4], which can be represented of the form (2). A NN with one layer and 5 neurons together with polynomial basis set wherein the control variables  $\alpha_1 = 25, \beta = 0.01, L_i(x) = \frac{1}{2}x_i^T\beta x_i$  and  $\phi(x) = [x_{1,1}^2, x_{1,2}^2, x_{2,1}^2, x_{2,2}^2, x^T x]^T$ ; the initial conditions are defined in the interval  $[0,1]$  and the initial weights of the NN are chosen randomly from  $[-1,1]$ . The dynamics of the system are given by  $\dot{x}_{11} = x_{12}, \dot{x}_{12} = (\frac{m_1 gr}{J_1} - \frac{k r^2}{4 J_1}) \sin x_{11} + \frac{k r}{2 J_1} (l - b) + \frac{u_i}{J_1} + \frac{k r^2}{4 J_1} \sin x_{11}$ . For the linear dynamics, refer [9]. The parameters in the system dynamics are  $m_1=2, m_2=2.5, J_1=5, J_2=6.25, k=10, r=0.5, l=0.5$ , and  $g=9.8, b=0.5$ . The controller design parameters are chosen as  $R_1=.03, R_2=0.03, Q_i = 0.1X^T X$ .

The results in Fig.2 shows the distributed controller performance for the linear system for various initial conditions. Fig. 3 shows the cumulative events for the linear interconnected system, which demonstrates the advantage of event based feedback. Next, the results for the event-triggered controller are presented with the distributed control scheme for the nonlinear dynamical system. For the event-triggered controller, the initial states and the weights are chosen as in the previous

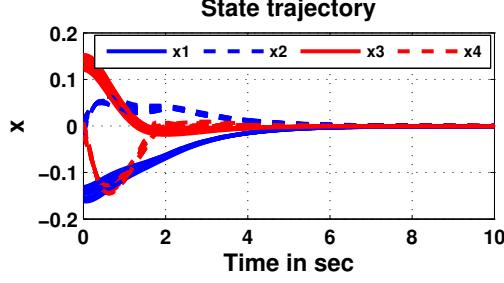


Fig. 4. State trajectories (Nonlinear example - 1)

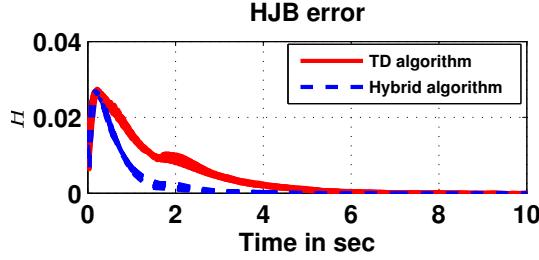


Fig. 5. HJB error (Nonlinear example - 1)

case. The design parameters are  $\Gamma = 0.95$ ,  $\alpha_1 = 20$ ,  $\beta = 0.01$ ,  $R_i = 0.03$ ,  $Q = 2X^T X$ . The system state trajectories with event-triggered controller are stable during the learning phase.

This can be verified from Fig. 4 for both the subsystems. The results in Fig. 4 include the state trajectories for various initial states. The HJB residual error for the TD ADP based controller and the proposed hybrid learning based controller are compared. It is evident from the results in Fig. 5 that the iterative weight updates between event-triggering instants seems to reduce the learning time.

The observer performance is presented in Fig. 6. The plots of estimated and actual outputs with the event-triggered feedback are compared when the hybrid learning algorithm is employed to generate the control policy online. The event triggered feedback and aperiodic update of the observer results in a piecewise continuous estimate of the actual states. The observer error convergence is essential for the stability of the controlled system.

Due to space consideration all the simulation figures are not included. Efficiency of the event-triggering condition designed for the two subsystems, SS1 for subsystem 1 and SS2 for subsystem 2; the convergence time for the observer estimation error and the HJB error for various initial conditions are recorded in Table 1.

**Example 2:** For the second example, a more practical system which is composed of three interconnected subsystems is considered. The three subsystems describe the dynamics of

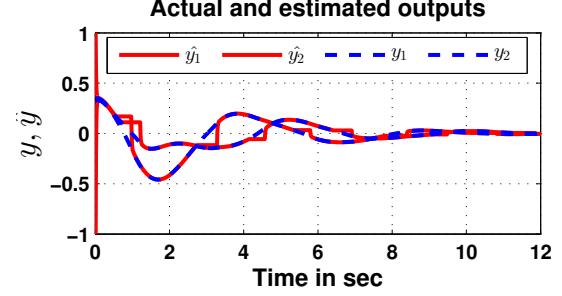


Fig. 6. Observer performance: Example 1

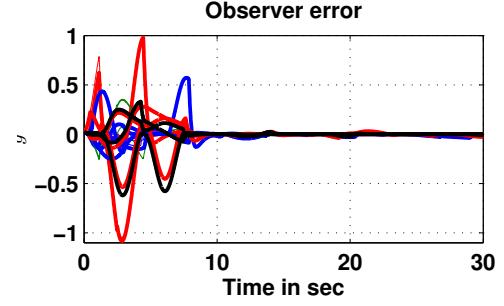


Fig. 7. Observer performance: Example 2

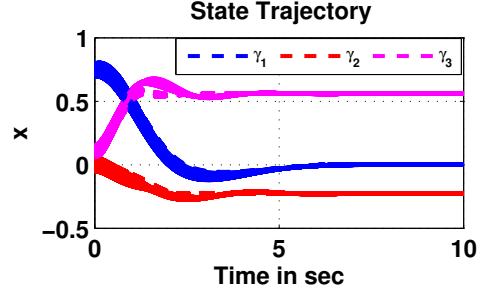


Fig. 8. State trajectory of walking robot

knee and thigh in a walking robot [8]. Let  $\gamma_1(t)$  be the relative angle between the two thighs,  $\gamma_2(t)$  and  $\gamma_3(t)$  be the right and left knee angles relative to the right and the left thigh. The dynamical equations of motion (in rad/sec) are

$$\begin{aligned} \ddot{\gamma}_1(t) &= 0.1[1 - 5.25\gamma_1^2(t)]\dot{\gamma}_1(t) - \gamma_1(t) + u_1(t) \\ \ddot{\gamma}_2(t) &= 0.01 \left[ 1 - p_2(\gamma_2(t) - \gamma_{2e})^2 \right] \dot{\gamma}_2(t) - 4(\gamma_2(t) - \gamma_{2e}) \\ &\quad + 0.057\gamma_1(t)\dot{\gamma}_1(t) + 0.1(\dot{\gamma}_2(t) - \dot{\gamma}_3(t)) + u_2(t) \\ \ddot{\gamma}_3(t) &= 0.01 \left[ 1 - p_3(\gamma_3(t) - \gamma_{3e})^2 \right] \dot{\gamma}_3(t) - 4(\gamma_3(t) - \gamma_{3e}) \\ &\quad + 0.057\gamma_1(t)\dot{\gamma}_1(t) + 0.1(\dot{\gamma}_3(t) - \dot{\gamma}_2(t)) + u_3(t). \end{aligned}$$

The parameter values used in the simulation are  $(\gamma_{2e}, \gamma_{3e}, p_2, p_3) = (-0.227, 0.559, 6070, 192)$ . The control objective is to design torque commands and bring the robot to a halt. The proposed control scheme with a NN to approximate  $V_i^*(X)$  at each subsystem is designed. The angles were initialized as  $40^\circ \pm 3^\circ$ ,  $3^\circ \pm 1^\circ$ ,  $-3^\circ \pm 1^\circ$  and the angular velocities were initialized at random to take values between 0 and 1. Two layer NNs with 12 neurons in the hidden layer are used at each subsystem. The NN weights

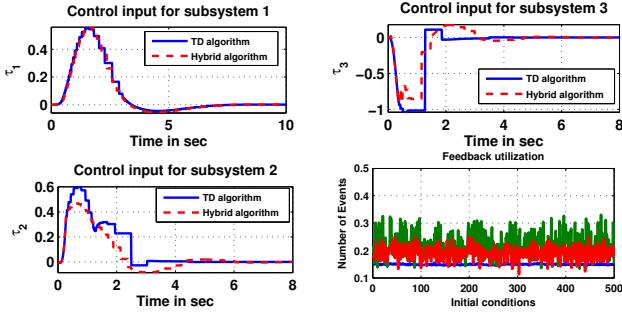


Fig. 9. Event triggered control

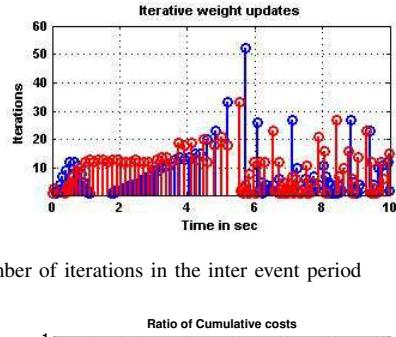


Fig. 10. Number of iterations in the inter event period

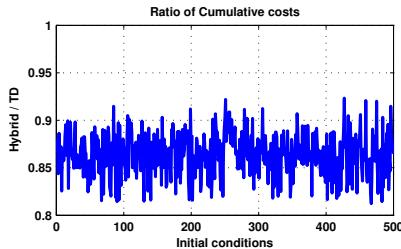


Fig. 11. Cost comparison (Example 2)

of the input layer were initialized at random to form random vector functional link network [29] and the second layer weights are initialized to take values between 0 and 1.

The states of each subsystem generated using the proposed learning approach for different initial conditions are recorded. It can be observed that the states reach their equilibrium point  $(0, -0.227, 0.559)$  every time, ensuring stable operation, for both state and output feedback control implementation (Fig. 8). The convergence of the observer estimation error can be verified from Fig. 7.

The hybrid algorithm converges faster and reaches steady state before the time driven ADP. The observer estimation error converged to a neighborhood of origin. In the analysis, different initial values for  $x_i(0)$  and  $\hat{X}_i(0)$  were chosen to test the algorithm and the results are tabulated. It is observed that whenever the observer error persists, performing iterative weight updates did not improve the learning rate. Therefore, the observer should be designed in such a way that the observer error converges faster and in this case the hybrid algorithm with output feedback controller outperformed the time driven ADP (Table 1).

The control torques generated using the hybrid learning algorithm with event triggered feedback and TD ADP are presented in Fig. 9. Also, the feedback utilization (ratio the event triggered feedback instants and the sensor samples) are presented for simulations carried out for 500 different initial conditions (Fig. 9).

The cumulative cost is calculated using the cost function defined in (4). The comparison of the cumulative cost calculated for the hybrid learning approach with that of the TD ADP reveals that the proposed hybrid scheme results in a lower cumulative cost. Fig. 11 shows the ratio of costs due to hybrid algorithm over TD algorithm for different initial conditions. For the output feedback case, due to the presence of the observer estimation error, the convergence of the HJB error takes more time when compared to state feedback. The improvement in the learning scheme is due to the learning process in the inter-sampling period. For analysis, the sensor sampling time was fixed at 10ms and the control scheme was simulated to record the number of times the weight update rule was executed in the inter-event period (Fig. 10). It can be seen that the inter-event time is not uniform and hence, the number of weight updates are varying.

Initially, the events are not spaced out and therefore, the iterative updates do not take place, but with time, the events become spaced out, but still with varying intervals. This results in a varying number of iterative weight updates. The comparison of HJB residual error for TD ADP and the learning scheme proposed in this paper reveals that the learning scheme introduced in this paper requires less time for convergence. Table 1 summarizes the comparison of the two learning algorithms. Feedback utilization is the ratio of events with respect to the sensor samples, when the sensor operates with a sampling period of 10ms.

## VI. CONCLUSIONS

This paper presents an approximation based distributed controller with event triggered state and output feedback that seeks optimality for a class of nonlinear interconnected system. The event-triggered control execution significantly reduces the communication and computational resource utilization by reducing the frequency of feedback instants. The proposed hybrid learning scheme seems to accelerate the learning of the NN weights with event-triggered feedback while reducing the communication costs.

The event triggering condition is independent of the estimated parameters and an additional estimator at the event-triggering mechanism is not required. The event-triggering mechanism is decentralized, asynchronous and ensures that the system is stable during the inter-event period. The requirement of initial stabilizing control policy is relaxed by utilizing the dynamics of the system.

## VII. APPENDIX

*Proof of Lemma 1:* Consider the Lyapunov candidate function  $L_i(\tilde{X}_i) = \frac{1}{2}\tilde{X}_i^T\gamma_i\tilde{X}_i + \frac{1}{4}(\tilde{X}_i^T\gamma_i\tilde{X}_i)^2$ , with the first derivative given by  $\dot{L}_i(\tilde{X}_i) = \tilde{X}_i^T\gamma_i\dot{\tilde{X}}_i + \tilde{X}_i^T\gamma_i\tilde{X}_i\tilde{X}_i^T\gamma_i\dot{\tilde{X}}_i$ .

Example	Algorithm	Cumulative cost (Normalized)		Convergence time in sec		Feedback utilization	
		State feedback (SF)	Output feedback (OF)	HJB error		Observer error	SF
				SF	OF		OF
1	TD	1	1	10.13	12.89	3.13	0.3716
	Hybrid	0.988	0.912	6.35	10.74	2.90	0.398
2	TD	1	1	4.8	37.20	30.654	0.2
	Hybrid	0.86	0.5916	4.1	31.62	27.13	0.3

TABLE I  
SIMULATION ANALYSIS

Using the estimation error dynamics (20), and Assumptions 3-4, we get

$$\begin{aligned} \dot{L}_i &\leq \tilde{X}_i^T \gamma_i (F(X_i) - F(\hat{X}_i) + (G(X_i) - G(\hat{X}_i))U_{i,e} \\ &\quad - \mu_i C(X_{i,e} - \hat{X}_i)) + \tilde{X}_i^T \gamma_i \tilde{X}_i \tilde{X}_i^T \gamma_i (F(X_i) - F(\hat{X}_i) \\ &\quad + (G(X_i) - G(\hat{X}_i))U_{i,e} - \mu_i C(X_{i,e} - \hat{X}_i)). \end{aligned} \quad (24)$$

We use the definition of the control policy and apply the norm operator in (24). Further, Young's inequality is utilized to obtain

$$\begin{aligned} \dot{L}_i &\leq -(\|\gamma_i\| \|\mu_i\| \|C\| - \|\gamma_i\| L_f - \frac{3}{2}) \|\tilde{X}_i\|^2 \\ &\quad - (\|\gamma_i\|^2 \|\mu_i\| \|C\| - \|\gamma_i\|^2 L_f - \frac{6}{2}) \|\tilde{X}_i\|^4 \\ &\quad + \frac{1}{8} G_M^4 \|\gamma_i\|^8 \left\| R_i^{-1} G^T(\hat{X}_e) \nabla_x^T \Phi(\hat{x}_e) \tilde{\Theta}_i \right\|^4 \\ &\quad + \frac{1}{2} G_M^2 \|\gamma_i\|^2 \left\| R_i^{-1} G^T(\hat{X}_e) \nabla_x^T \Phi(\hat{x}_e) \tilde{\Theta}_i \right\|^2 \\ &\quad + \frac{1}{8} \|\gamma_i\|^8 \|\mu_i\|^4 \|C\|^4 \|e\|^4 + \frac{1}{2} \|\gamma_i\|^2 \|\mu_i\|^2 \|C\|^2 \|e\|^2 \\ &\quad + \frac{1}{8} G_M^4 \|\gamma_i\|^8 \|U_i^*\|^4 + \frac{1}{2} G_M^2 \|\gamma_i\|^2 \|U_i^*\|^2 \end{aligned} \quad (25)$$

where  $L_f, e$  are Lipschitz constant, measurement error, respectively. Further simplification reveals,

$$\begin{aligned} \dot{L}_i &\leq -\eta_{i,o1} \|\tilde{X}_i\|^2 - \eta_{i,o2} \|\tilde{X}_i\|^4 + \xi_{i1,obs} + \frac{1}{8} G_M^2 \\ &\quad \|\gamma_i\|^2 (G_M^2 \|\gamma_i\|^4 \left\| R_i^{-1} G^T(\hat{X}_e) \nabla_x^T \Phi(\hat{x}_e) \tilde{\Theta}_i \right\|^4 \\ &\quad + 4 \left\| R_i^{-1} G^T(\hat{X}_e) \nabla_x^T \Phi(\hat{x}_e) \tilde{\Theta}_i \right\|^2) \end{aligned} \quad (26)$$

where  $\eta_{i,o1} = \|\gamma_i\| \|\mu_i\| \|C\| - \|\gamma_i\| L_f - 1.5$ ,  $\eta_{i,o2} = \|\gamma_i\|^2 \|\mu_i\| \|C\| - \|\gamma_i\|^2 L_f - 3$ ,  $\xi_{i1,obs} = \frac{1}{8} \|\gamma_i\|^8 \|\mu_i\|^4 \|C\|^4 \|e\|^4 + \frac{1}{2} \|\gamma_i\|^2 \|\mu_i\|^2 \|C\|^2 \|e\|^2 + \frac{1}{8} G_M^4 \|\gamma_i\|^8 \|U_i^*\|^4 + \frac{1}{2} G_M^2 \|\gamma_i\|^2 \|U_i^*\|^2$ . Since the control policy is assumed to be optimal,  $\bar{U}_i = 0$  and therefore,  $\tilde{\Theta} = 0$ . This concludes the proof.

*Proof of Theorem 1:* Consider the Lyapunov function

$$L_i(x_i, \tilde{\theta}_i, \tilde{X}_i) = L_{i1}(x_i) + L_{i2}(\tilde{\theta}_i) + L_{i3}(\tilde{X}_i).$$

Consider the term  $L_i(\tilde{\theta}_i)$ , taking the first derivative to get

$$\begin{aligned} \dot{L}_{i2} &= \frac{\alpha_{i1}}{\hat{\rho}_{i,e}^2} \tilde{\theta}_i^T \hat{\sigma}_{i,e} \{-\tilde{\theta}_i^T \hat{\sigma}_{i,e} + \frac{1}{4} \tilde{\theta}_i^T \nabla_x \phi(\hat{x}_e) \hat{D}_{i,\varepsilon} \nabla_x^T \phi(\hat{x}_e) \tilde{\theta}_i \\ &\quad + Q_i(\hat{X}_e) - Q_i(X) - \varepsilon_{i_{HJB}} + \theta_i^{*T} [\nabla_x \phi(\hat{x}_e) \bar{f}_i(\hat{x}_e) \\ &\quad - \nabla_x \phi(x) \bar{f}_i(x)] + \frac{1}{4} \theta_i^{*T} [\nabla_x \phi(x) D_i \nabla_x^T \phi(x) \\ &\quad - \nabla_x \phi(\hat{x}_e) \hat{D}_{i,\varepsilon} \nabla_x^T \phi(\hat{x}_e)] \theta_i^*\} + \tilde{\theta}_i^T \varpi_{i1} \end{aligned} \quad (27)$$

with  $\hat{\rho}_{i,e} = \hat{\sigma}_{i,e}^T \hat{\sigma}_{i,e} + 1$ ,  $\varpi_{i1}$  is the sum of stabilizing term and the sigma modification term in the weight adaptation rule. Using the Lipschitz constant  $L_Q$  (for  $Q_i(X)$ ) and on simplification, we get

$$\begin{aligned} \dot{L}_{i2} &\leq -\frac{\alpha_{i1}}{\hat{\rho}_{i,e}^2} (\tilde{\theta}_i^T \hat{\sigma}_{i,e})^2 + \frac{\alpha_{i1}}{4 \hat{\rho}_{i,e}^2} \tilde{\theta}_i^T \hat{\sigma}_{i,e} P_{i1} + \tilde{\theta}_i^T \varpi_{i1} \\ &\quad + \frac{\alpha_{i1}}{\hat{\rho}_{i,e}^2} \left\| \tilde{\theta}_i^T \hat{\sigma}_{i,e} \right\| L_Q \left\| \tilde{X}_{i,e} \right\| + \frac{\alpha_{i1}}{\hat{\rho}_{i,e}^2} \tilde{\theta}_i^T \hat{\sigma}_{i,e} \{-\varepsilon_{i_{HJB}} \\ &\quad + \theta_i^{*T} [\nabla_x \phi(\hat{x}_e) (\bar{f}_i(\hat{x}_e) - \bar{f}_i(x)) + (\nabla_x \phi(\hat{x}_e) - \nabla_x \phi(x)) \bar{f}_i(x)] \\ &\quad + \frac{1}{4} \theta_i^{*T} [\nabla_x \phi(x) D_i \nabla_x^T \phi(x) - \nabla_x \phi(\hat{x}_e) \hat{D}_{i,\varepsilon} \nabla_x^T \phi(\hat{x}_e)] \theta_i^*\} \\ &\quad \text{with } P_{i1} = \tilde{\theta}_i^T \nabla_x \phi(\hat{x}_e) \hat{D}_{i,\varepsilon} \nabla_x^T \phi(\hat{x}_e) \tilde{\theta}_i. \end{aligned}$$

Here,  $\tilde{X}_{i,e}$  is the event triggered state-estimation error. This is defined as  $\tilde{X}_{i,e} = X_{i,e} - \hat{X}_{i,e} = X_i - \hat{X}_i - e$ . Applying the norm operator, we get

$$\begin{aligned} \dot{L}_{i2} &\leq -\frac{\alpha_{i1}}{\hat{\rho}_{i,e}^2} (\tilde{\theta}_i^T \hat{\sigma}_{i,e})^2 + \frac{\alpha_{i1}}{4 \hat{\rho}_{i,e}^2} \tilde{\theta}_i^T \hat{\sigma}_{i,e} P_{i1} + \tilde{\theta}_i^T \varpi_{i1} + \\ &\quad \frac{\alpha_{i1}}{\hat{\rho}_{i,e}^2} \left\| \tilde{\theta}_i^T \hat{\sigma}_{i,e} \right\| L_Q \left\| \tilde{X}_{i,e} \right\| + \frac{\alpha_{i1}}{\hat{\rho}_{i,e}^2} \left\| \tilde{\theta}_i^T \hat{\sigma}_{i,e} \right\| [ \left\| \theta_i^{*T} \nabla_x \phi(\hat{x}_e) \right\| \\ &\quad L_{\bar{f}_i} \left\| \tilde{X}_{i,e} \right\| + \theta_{iM}^* L_{\phi_i} \left\| \tilde{X}_{i,e} \right\| \left\| \bar{f}_i(x) \right\| ] + \frac{\left\| \alpha_{i1} \tilde{\theta}_i^T \hat{\sigma}_{i,e} \right\|}{\hat{\rho}_{i,e}^2} \\ &\quad \left\| \nabla_x \varepsilon \dot{x}_i^* \right\| + \frac{\left\| \alpha_{i1} \tilde{\theta}_i^T \hat{\sigma}_{i,e} \right\|}{2 \hat{\rho}_{i,e}^2} B_{i1} \\ &\quad \text{where } B_{i1} = \left\| \nabla_x \varepsilon D_i \nabla_x^T \varepsilon \right\| + V_{xiM}^2 D_{iM}, \\ &\quad L_{\bar{f}_i} \text{ and } L_{\phi_i} \text{ are Lipschitz constants of } \bar{f}_i, \nabla_x \phi. \end{aligned} \quad (28)$$

The second term can be simplified as

$$\begin{aligned}
& \frac{\alpha_{i1}}{4\hat{\rho}_{i,e}^2}\tilde{\theta}_i^T\hat{\sigma}_{i,e}P_{i1} = \frac{\alpha_{i1}}{4\hat{\rho}_{i,e}^2}\tilde{\theta}_i^T\{\nabla_x\phi(\hat{x}_e)\hat{f}_{i,e} - \frac{1}{2}\nabla_x\phi(\hat{x}_e) \\
& \hat{D}_{i,\varepsilon}\nabla_x^T\phi(\hat{x}_e)\theta_i^* + \frac{1}{2}\nabla_x\phi(\hat{x}_e)\hat{D}_{i,\varepsilon}\nabla_x^T\phi(\hat{x}_e)\tilde{\theta}_i\}P_{i1} \\
& \leq \frac{\alpha_{i1}}{4\hat{\rho}_{i,e}^2}(\|\tilde{\theta}_i^T\nabla_x\phi(\hat{x}_e)\|L_{\bar{f}_i}\|\tilde{X}_{i,e}\| + \frac{\alpha_{i1}}{8\hat{\rho}_{i,e}^2}\|\tilde{\theta}_i^T\nabla_x\phi(\hat{x}_e)\| \\
& D_{i,M}L_{\phi_i}\|\tilde{X}_{i,e}\|\theta_{iM}^*)\|P_{i1}\| + \frac{\alpha_{i1}}{4\hat{\rho}_{i,e}^2}\|\nabla_x\phi(\hat{x}_e)\bar{f}_i - \\
& \frac{1}{2}\nabla_x\phi(\hat{x}_e)\hat{D}_{i,\varepsilon}\nabla_x^T\phi(x)\theta_i^* + \frac{1}{2}\nabla_x\phi(\hat{x}_e)\hat{D}_{i,\varepsilon}\nabla_x^T\phi(\hat{x}_e)\tilde{\theta}_i\}P_{i1}. \tag{29}
\end{aligned}$$

Substituting (29) in (28), we obtain

$$\begin{aligned}
\dot{L}_{i2} & \leq -\frac{\alpha_{i1}}{\hat{\rho}_{i,e}^2}(\tilde{\theta}_i^T\hat{\sigma}_{i,e})^2 + \tilde{\theta}_i^T\varpi_{i1} + \frac{1}{2}L_Q^2\|\tilde{X}_{i,e}\|^2 + \frac{1}{8}B_{i1}^2 \\
& + \frac{\alpha_{i1}^2}{2\hat{\rho}_{i,e}^4}\|\tilde{\theta}_i^T\hat{\sigma}_{i,e}\|^2 + \frac{\alpha_{i1}^2}{2\hat{\rho}_{i,e}^4}\|\tilde{\theta}_i^T\hat{\sigma}_{i,e}\|^2 + \frac{\alpha_{i1}}{8\hat{\rho}_{i,e}^2}\|P_{i1}\|^2 + \\
& \frac{\alpha_{i1}^2}{\hat{\rho}_{i,e}^4}\|\tilde{\theta}_i^T\hat{\sigma}_{i,e}\|^2 + \frac{1}{2}\|\nabla_x\varepsilon\dot{x}_i^*\|^2 + \frac{\alpha_{i1}^2}{2\hat{\rho}_{i,e}^4}\|\tilde{\theta}_i^T\hat{\sigma}_{i,e}\|^2V_{xiM}^2 + \\
& \frac{1}{2}L_{\bar{f}_i}^2\|\tilde{X}_{i,e}\|^2 + \frac{\alpha_{i1}^2}{8\hat{\rho}_{i,e}^2}\|\tilde{\theta}_i^T\nabla_x\phi(\hat{x}_e)\|^2L_{\bar{f}_i}^2\|\tilde{X}_{i,e}\|^2 + \frac{1}{8\hat{\rho}_{i,e}^2} \\
& \|P_{i1}\|^2 + \frac{\alpha_{i1}^2}{8\hat{\rho}_{i,e}^2}\|\tilde{\theta}_i^T\nabla_x\phi(\hat{x}_e)\|^2\|\bar{f}_i\|^2 + \frac{1}{8\hat{\rho}_{i,e}^2}\|P_{i1}\|^2 \\
& + \frac{\alpha_{i1}^2}{16\hat{\rho}_{i,e}^2}\|\tilde{\theta}_i^T\nabla_x\phi(\hat{x}_e)\|^2D_{iM}^2L_{\phi_i}^2\|\tilde{X}_{i,e}\|^2\theta_{iM}^{*2} \\
& + \frac{1}{16\hat{\rho}_{i,e}^2}\|P_{i1}\|^2 + \frac{\alpha_{i1}^2}{16\hat{\rho}_{i,e}^2}\|\tilde{\theta}_i^T\nabla_x\phi(\hat{x}_e)\|^2D_{iM}^2V_{xiM}^2 \\
& + \frac{1}{16\hat{\rho}_{i,e}^2}\|P_{i1}\|^2 + \frac{1}{2}\theta_{iM}^{*2}L_{\phi_i}^2\|\tilde{X}_{i,e}\|^2\|\bar{f}_i\|^2.
\end{aligned}$$

Using Youngbs inequality and on simplification, we get

$$\begin{aligned}
\dot{L}_{i2} & \leq -\frac{\alpha_{i1}}{\hat{\rho}_{i,e}^2}(\tilde{\theta}_i^T\hat{\sigma}_{i,e})^2 + \tilde{\theta}_i^T\varpi_{i1} + L_Q^2\|\tilde{X}_i\|^2 \\
& + \frac{2\alpha_{i1}}{\hat{\rho}_{i,e}^4}\|\tilde{\theta}_i^T\hat{\sigma}_{i,e}\|^2 + \frac{\alpha_{i1}}{8\hat{\rho}_{i,e}^2}\|P_{i1}\|^2 + \frac{1}{2}\|\nabla_x\varepsilon\dot{x}_i^*\|^2 \\
& + L_{\bar{f}_i}^2\|\tilde{X}_i\|^2 + \frac{3}{8\hat{\rho}_{i,e}^2}\|P_{i1}\|^2 + \frac{1}{2}\theta_{iM}^{*2}L_{\phi_i}^2\|\tilde{X}_{i,e}\|^2\|\bar{f}_i\|^2 \\
& + \frac{\alpha_{i1}^4}{8\hat{\rho}_{i,e}^2}L_{\bar{f}_i}^4\|\tilde{X}_i\|^4 + \frac{\alpha_{i1}^4}{16\hat{\rho}_{i,e}^2}\|\bar{f}_i\|^4 + \frac{\alpha_{i1}^4}{16\hat{\rho}_{i,e}^2}D_{iM}^4L_{\phi_i}^4\|\tilde{X}_i\|^4\theta_{iM}^{*4} \\
& + \frac{6}{32\hat{\rho}_{i,e}^2}\|\tilde{\theta}_i^T\nabla_x\phi(\hat{x}_e)\|^4 + \frac{\alpha_{i1}^2}{2\hat{\rho}_{i,e}^4}\|\tilde{\theta}_i^T\hat{\sigma}_{i,e}\|^2V_{xiM}^2 + B_{i2} \\
& B_{i2} = \frac{\alpha_{i1}^4}{32\hat{\rho}_{i,e}^2}D_{iM}^4V_{xiM}^4 + \frac{1}{8}B_{i1}^2 + L_{\bar{f}_i}^2\|e_i\|^2 + \frac{\alpha_{i1}^4}{8\hat{\rho}_{i,e}^2}L_{\bar{f}_i}^4\|e_i\|^4 \\
& + L_Q^2\|e_i\|^2 + \frac{\alpha_{i1}^4}{16\hat{\rho}_{i,e}^2}D_{iM}^4L_{\phi_i}^4\|e_i\|^4\theta_{iM}^{*4}.
\end{aligned}$$

Rearranging the equation, after simplifying, the first derivative becomes

$$\begin{aligned}
\dot{L}_{i2} & \leq (\frac{\alpha_{i1}}{8\hat{\rho}_{i,e}^2} + \frac{3}{8\hat{\rho}_{i,e}^2} + \frac{6}{32\hat{\rho}_{i,e}^2D_{iM}^2})\|P_{i1}\|^2 \\
& - (\kappa_i - \frac{2\alpha_{i1}^2\|\hat{\sigma}_{i,e}\|^2}{\hat{\rho}_{i,e}^4} - \frac{\alpha_{i1}^2V_{xiM}^2\|\hat{\sigma}_{i,e}\|^2}{2\hat{\rho}_{i,e}^4} + \frac{\alpha_{i1}\|\hat{\sigma}_{i,e}\|^2}{\hat{\rho}_{i,e}^2} - \frac{1}{2})\|\tilde{\theta}_i^T\|^2 \\
& + \frac{1}{2}\|\nabla_x\varepsilon\dot{x}_i^*\|^2 + L_Q^2\|\tilde{X}_i\|^2 + L_{\bar{f}_i}^2\|\tilde{X}_i\|^2 + \frac{\|\kappa_i\theta_i^*\|^2}{2} \\
& + \frac{1}{2}\theta_{iM}^{*2}L_{\phi_i}^2\|\tilde{X}_{i,e}\|^2\|\bar{f}_i\|^2 + \frac{\alpha_{i1}^4}{8\hat{\rho}_{i,e}^2}L_{\bar{f}_i}^4\|\tilde{X}_i\|^4 + \frac{\alpha_{i1}^4}{16\hat{\rho}_{i,e}^2}\|\bar{f}_i\|^4 \\
& + \frac{\alpha_{i1}^4}{16\hat{\rho}_{i,e}^2}D_{iM}^4L_{\phi_i}^4\|\tilde{X}_i\|^4\theta_{iM}^{*4} + B_{i2} - \frac{1}{2}\beta_i\tilde{\theta}_i\nabla_x(\hat{x}_e)\hat{D}_{i,\varepsilon}L_{ix}(\hat{x}_{i,e})
\end{aligned}$$

Taking the derivative of the first term in the Lyapunov candi-

date function  $L_{i1}$ , we have

$$\begin{aligned}
\dot{L}_{i1}(x_i) & = L_{ix}(x_i)\dot{x}_i = L_{ix}(x_i)[\bar{f}_i(x) + g_i(x_i)\hat{u}_{i,e}] \\
& = L_{ix}(x_i)\dot{x}_i^* + L_{ix}(x_i)B_{i3} + \frac{1}{2}L_{ix}(x_i)\hat{D}_{i,\varepsilon}\nabla_x^T\phi(\hat{x}_e)\tilde{\theta}_i \\
& + \|L_{ix}(x_i)\|D_{iM}L_{\phi_i}\|\tilde{X}_i\|\theta_{iM}^*, \\
B_{i3} & = \frac{1}{2}(D_i(\nabla_x^T\phi(x)\theta_i^* + \nabla_x^T\varepsilon) + D_{iM}\nabla_x^T\phi(x)\theta_i^* \\
& + D_{iM}L_{\phi_i}e\theta_i^*. \tag{30}
\end{aligned}$$

Using (30), with  $L_i(x_i) = \frac{1}{2}x_i^Tx_i + \frac{1}{4}(x_i^Tx_i)^2$ , we have,  $\dot{L}_i(x_i, \tilde{\theta}_i) = \dot{L}_{i1}(x_i) + \dot{L}_{i2}(\tilde{\theta}_i)$ , grouping similar terms to get

$$\begin{aligned}
\dot{L}_i(x_i, \tilde{\theta}_i) & \leq -\eta_{ix^2}\|x_i\|^2 - \eta_{ix^4}\|x_i\|^4 - (\zeta_1 - 1)\|x_i\|^6 \\
& - \eta_{i\theta^2}\|P_{i1}\|^2 - \eta_{i\theta^2}\|\tilde{\theta}_i^T\|^2 + 3\beta_i^2\|\tilde{x}_i\|^6 + \xi_{i1cl}(e) \\
& + (D_{iM}^2L_{\phi_i}^2\theta_{iM}^2 + L_Q^2 + \frac{1}{2}\beta_i^2L_{iL}^2 + L_{\bar{f}_i}^2)\|\tilde{X}_i\|^2 \\
& + (\theta_{iM}^4L_{\phi_i}^4 + \frac{\alpha_{i1}^4}{8\hat{\rho}_{i,e}^2}L_{\bar{f}_i}^4 + \frac{\alpha_{i1}^4}{16\hat{\rho}_{i,e}^2}D_{iM}^4L_{\phi_i}^4\theta_{iM}^4)\|\tilde{X}_i\|^4 \tag{31}
\end{aligned}$$

where the constants are defined as follows  $\eta_{i\theta^2} = \kappa_i - 2\alpha_{i1}^2 - \frac{\alpha_{i1}^2V_{xiM}^2}{2} + \alpha_{i1} - \frac{1}{2} - \frac{D_{iM}^2\nabla_x^T\phi_M}{2}; \eta_{i\theta^4} = -(\frac{\alpha_{i1}}{8\hat{\rho}_{i,e}^2} + \frac{3}{8\hat{\rho}_{i,e}^2} + \frac{6}{32\hat{\rho}_{i,e}^2D_{iM}^2}); \eta_{ix^2} = \zeta_1 - 2 - 2\beta_i^2L_{iL}^2 - \frac{1}{4}L_{iL}^2 - \frac{1}{2}\psi\|\nabla_x\varepsilon M N\|^2; \eta_{ix^4} = 2\zeta_1 - \frac{3}{4} - \frac{\alpha_{i1}^4}{16\hat{\rho}_{i,e}^2}L_{\bar{f}_i}^4N^4 - \frac{1}{4}L_{\bar{f}_i}^4N^4; \xi_{i1cl}(e) = B_{i4} + \frac{1}{2}\|B_{i3}\|^2 + \frac{1}{4}\|B_{i3}\|^4, L_{iL} \text{ is the Lipschitz constant for } L_{ix}. \text{ Recalling the results from Lemma 1 and using (31), we get}$

$$\begin{aligned}
\dot{L}_i & \leq -\eta_{i\theta^4}\|P_{i1}\|^2 - \eta_{ix^2}\|x_i^T\|^2 - \eta_{i,o1}\|\tilde{X}_i\|^2 - \bar{\zeta}_1\|x_i\|^6 \\
& - \eta_{ix^4}\|x_i^T\|^4 - \eta_{i\theta^2}\|\tilde{\theta}_i^T\|^2 - \eta_{i\tilde{x}^2}\|\tilde{X}_i\|^4 - \eta_{i\tilde{x}^2}\|\tilde{X}_i\|^2 + \xi_{icl} \tag{32}
\end{aligned}$$

where

$$\begin{aligned}
\eta_{ix^2} & = (\eta_{i,o2} - (D_{iM}^2L_{\phi_i}^2\theta_{iM}^2 + L_Q^2 + \frac{1}{2}\beta_i^2L_{iL}^2 + L_{\bar{f}_i}^2)) \\
\eta_{i\tilde{x}^2} & = (\eta_{i,o1} - (\theta_{iM}^4L_{\phi_i}^4 + \frac{\alpha_{i1}^4}{8\hat{\rho}_{i,e}^2}L_{\bar{f}_i}^4 + \frac{\alpha_{i1}^4}{16\hat{\rho}_{i,e}^2}D_{iM}^4L_{\phi_i}^4\theta_{iM}^4)) \\
\xi_{icl} & = \xi_{i1,obs} + \frac{1}{16}\|R_i^{-1}G^T\|^4 + \xi_{i1cl}(e) \\
\eta_{i\theta^4} & = \eta_{i\theta4} - \frac{NR_i^{-1}G_M}{64D_{iM}^2} - \frac{N}{16D_{iM}^2}, \bar{\zeta}_1 = \zeta_1 - 1.
\end{aligned}$$

The parameters  $\alpha_{i1}, \beta_i, \kappa_i, \mu_i$  can be chosen to ensure that the constants in (32) are positive. The sigma modification term in the weight tuning equation gives the negative term in  $\tilde{\theta}_i$ , independent of the states.

*Proof of Theorem 2:* First, recalling the results from the previous theorem, it can be observed that when the event-sampling error is set to zero, the bounds obtained in Theorem 1 will be further reduced. Now, consider the time-driven algorithm between any two event triggering instants.

*Case 1:* In the event based TD learning scheme, the weights of the NN are held constant and are not tuned between events. Hence, the derivative of the second term of the Lyapunov function will be zero. Using the event-sampling condition for the output feedback and using the definition of the observer estimation error, we get  $L_{1i}(\tilde{X}_i) = \frac{1}{2}(X_i^T\gamma_i X_i - 2X_i^T\gamma_i \hat{X}_i + \hat{X}_i^T\gamma_i \hat{X}_i)$ . Now, using the event-sampling condition and  $\dot{X}(t) = \hat{X}(t_k), t_k \leq t < t_{k+1}$ , we arrive at a bounding function,  $L_{1i}(\tilde{X}_i) \leq \sum_{i=1}^N L_i(x_i) + \hat{X}_i^T\gamma_i \hat{X}_i$ . Now using the Lyapunov function from Theorem 1, the first derivative

is obtained as  $\dot{L}_i(x_i, \tilde{\theta}_i, \tilde{X}_i) = \dot{L}_i(x_i) + \dot{L}_i(\tilde{\theta}_i) + \dot{L}_i(\tilde{X}_i)$ . Substituting the bounds obtained above reveals

$$\dot{L}_i(x_i(t), \tilde{\theta}_i(t), \tilde{X}_i(t)) \leq 2 \sum_{i=1}^N -\Gamma_i L_{i1}(x_i(t_k)).$$

*Case 2(Event-triggered hybrid learning algorithm):* In this case, the weights of the value function estimator are tuned using the past feedback information using (21). Select the Lyapunov function from Theorem 1, now the first derivative is given by

$$\begin{aligned} \dot{L}_i(x_i(t), \tilde{\theta}_i(j), \tilde{X}_i(t)) &\leq 2 \sum_{i=1}^N -\Gamma_i L_{i1}(x_i(t_k)) + H_{x^4} \|X\|^2 \\ &- \left( \frac{|\alpha_{i1}|}{8\hat{\rho}_{i,e}^2} - \frac{3}{8\hat{\rho}_{i,e}^2} - \frac{1}{4\hat{\rho}_{i,e}^2 D_{iM}^2} \right) \|P_{i1}\|^2 + H_{x^4} \|X\|^4 \\ &+ H_{\tilde{x}^2} \|\tilde{X}_i\|^2 - \left( \kappa_i - \frac{|\alpha_{i1}| \|\hat{\sigma}_i\|^2}{\hat{\rho}_i^2} - \frac{3\alpha_{i1}^2}{2\hat{\rho}_i^2} - \frac{1}{2} \right) \|\tilde{\theta}_i^T\|^2 \\ &+ H_{\tilde{x}^4} \|\tilde{X}_i\|^4 + B_{i2} + \frac{\|\kappa_i \theta_i^*\|^2}{2} \end{aligned}$$

with  $j$  being the iteration index for the weights  $\tilde{\theta}_i$  and  $H_{x^4} = 0.25 \|\theta_i^{*T} L_{\phi_i} L_{\bar{f}_i}\|^2$ ,  $H_{\tilde{x}^2} = 0.5 \|L_Q\|^2 + \|L_{\bar{f}_i}\|^2$ ,  $H_{x^2} = \frac{1}{2} \|\nabla_x \varepsilon \psi\|^2 + \frac{1}{16\hat{\rho}_i^2} \|\alpha_{i1} L_{\bar{f}_i}\|^2$ ,  $B_{i3} = B_{i2} + \frac{1}{2} \|\kappa_i \theta_i^*\|^2$ ,  $H_{\tilde{x}^4} = \frac{1}{32\hat{\rho}_i^2} \|\alpha_{i1} D_{iM} L_{\phi_i} \theta_i^*\|^4 + \frac{1}{16\hat{\rho}_i^2} \|\alpha_{i1} L_{\bar{f}_i}\|^4 + \frac{1}{4} \|\theta_i^{*T} L_{\phi_i} L_{\bar{f}_i}\|^2$ . We obtain the first derivative as

$$\begin{aligned} \dot{L}_i(x_i(t), \tilde{\theta}_i(j), \tilde{X}_i(t)) &\leq 4 \sum_{i=1}^N -\Gamma_i L_{i1}(x_i(t_k)) \\ &- \left( \frac{|\alpha_{i1}|}{8\hat{\rho}_{i,e}^2} - \frac{3}{8\hat{\rho}_{i,e}^2} - \frac{1}{4\hat{\rho}_{i,e}^2 D_{iM}^2} \right) \|P_{i1}\|^2 \\ &- \left( \kappa_i - \frac{|\alpha_{i1}| \|\hat{\sigma}_i\|^2}{\hat{\rho}_i^2} - \frac{3\alpha_{i1}^2}{2\hat{\rho}_i^2} - \frac{1}{2} \right) \|\tilde{\theta}_i^T\|^2 + H_M \|\tilde{X}_i\|^2 + B_{i3} \end{aligned} \quad (33)$$

where  $H_M$  is the maximum of  $\{H_{\tilde{x}^2}, H_{\tilde{x}^4}, H_{x^2}, H_{x^4}\}$ . With the proposed iterative weight tuning, the Lyapunov first derivative is decreasing when the states and the weight estimation errors are outside the ultimate bound obtained from (33).

*Proof of Corollary :* When  $e_i = 0$ , recalling the results obtained for the output feedback case in Theorem 1, it is evident that when the event-sampling error is bounded, ISS like results can be obtained with state estimation error set to zero. Further, setting the measurement error to zero, the bounds can be obtained for the state feedback controller operating in continuous time. Now, consider the inter event period with TD (case 1) and hybrid algorithm (case 2).

*Case 1:* The weights of the NN are not updated between events in time-driven ADP, the derivative of the second term will be zero. Therefore, the first derivative can be written as  $\dot{L}_{HJB} = \sum_{i=1}^N (\beta \dot{L}_{i1}(x) + 0)$ . From the event-sampling condition, the first derivative is given as

$$\dot{L}_i(x_i(t)) \leq -\alpha L_{i1}(x(t_k)), \quad t \in [t_k, t_{k+1}).$$

Hence, it can be concluded that the Lyapunov derivative is negative semi-definite and reveals that the Lyapunov function non-increasing between events.

*Case 2 (Event-triggered hybrid learning algorithm):* Select the Lyapunov function candidate as in case 1. The first derivative is  $L_{HJB} = \sum_{i=1}^N (L_{ix}(x_i) \dot{x}_i + \dot{L}_{i\theta})$ . The derivatives can be used from Theorem 2, case 2 with the observer estimation

error set to 0. This gives a stronger result when compared to the TD ADP.

## REFERENCES

- [1] S. Mehraeen and S. Jagannathan, "Decentralized optimal control of a class of interconnected nonlinear discrete-time systems by using online hamilton-jacobi-bellman formulation," *IEEE Transactions on Neural Networks*, vol. 22, no. 11, pp. 1757–1769, Nov 2011.
- [2] D. Wang, D. Liu, H. Li, and H. Ma, "Neural-network-based robust optimal control design for a class of uncertain nonlinear systems via adaptive dynamic programming," *Information Sciences*, vol. 282, pp. 167–179, 2014.
- [3] D. Liu, D. Wang, and H. Li, "Decentralized stabilization for a class of continuous-time nonlinear interconnected systems using online learning optimal control approach," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 25, no. 2, pp. 418–428, Feb 2014.
- [4] J. T. Spooner and K. M. Passino, "Decentralized adaptive control of nonlinear systems using radial basis neural networks," *IEEE Transactions on Automatic Control*, vol. 44, no. 11, pp. 2050–2057, Nov 1999.
- [5] S. Huang, K. K. Tan, and T. H. Lee, "Decentralized control design for large-scale systems with strong interconnections using neural networks," *IEEE Transactions on Automatic Control*, vol. 48, no. 5, pp. 805–810, May 2003.
- [6] K. S. Narendra and S. Mukhopadhyay, "To communicate or not to communicate: A decision-theoretic approach to decentralized adaptive control," in *Proceedings of the 2010 American Control Conference*, June 2010, pp. 6369–6376.
- [7] X. Wang and M. D. Lemmon, "Event-triggering in distributed networked control systems," *IEEE Transactions on Automatic Control*, vol. 56, no. 3, pp. 586–601, March 2011.
- [8] W. B. Dunbar, "Distributed receding horizon control of dynamically coupled nonlinear systems," *IEEE Transactions on Automatic Control*, vol. 52, no. 7, pp. 1249–1263, July 2007.
- [9] M. Guinaldo, D. V. Dimarogonas, K. H. Johansson, J. Sanchez, and S. Dormido, "Distributed event-based control for interconnected linear systems," in *2011 50th IEEE Conference on Decision and Control and European Control Conference*, Dec 2011, pp. 2553–2558.
- [10] X. Wang and M. Lemmon, "On event design in event-triggered feedback systems," *Automatica*, vol. 47, no. 10, pp. 2319 – 2322, 2011.
- [11] P. Tallapragada and N. Chopra, "On event triggered tracking for nonlinear systems," *IEEE Transactions on Automatic Control*, vol. 58, no. 9, pp. 2343–2348, Sept 2013.
- [12] M. Mazo and P. Tabuada, "Decentralized event-triggered control over wireless sensor/actuator networks," *IEEE Transactions on Automatic Control*, vol. 56, no. 10, pp. 2456–2461, Oct 2011.
- [13] E. Garcia and P. J. Antsaklis, "Model-based event-triggered control for systems with quantization and time-varying network delays," *IEEE Transactions on Automatic Control*, vol. 58, no. 2, pp. 422–434, Feb 2013.
- [14] A. Sahoo, H. Xu, and S. Jagannathan, "Neural network-based event-triggered state feedback control of nonlinear continuous-time systems," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 27, no. 3, pp. 497–509, March 2016.
- [15] X. Zhong and H. He, "An event-triggered adp control approach for continuous-time system with unknown internal states," *IEEE Transactions on Cybernetics*, vol. PP, no. 99, pp. 1–12, 2016.
- [16] D. P. Bertsekas, D. P. Bertsekas, D. P. Bertsekas, and D. P. Bertsekas, *Dynamic programming and optimal control*. Athena Scientific Belmont, MA, 1995, vol. 1, no. 2.
- [17] P. J. Werbos, "Optimization methods for brain-like intelligent control," in *Decision and Control, 1995., Proceedings of the 34th IEEE Conference on*, vol. 1, Dec 1995, pp. 579–584 vol.1.
- [18] R. S. Sutton and A. G. Barto, *Reinforcement learning: An introduction*. MIT press Cambridge, 1998, vol. 1, no. 1.
- [19] D. V. Prokhorov, R. A. Santiago, and D. C. Wunsch, "Adaptive critic designs: A case study for neurocontrol," *Neural Networks*, vol. 8, no. 9, pp. 1367–1372, 1995.
- [20] H. Xu and S. Jagannathan, "Stochastic optimal controller design for uncertain nonlinear networked control system via neuro dynamic programming," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 24, no. 3, pp. 471–484, March 2013.
- [21] F. L. Lewis, D. Vrabie, and K. G. Vamvoudakis, "Reinforcement learning and feedback control: Using natural decision methods to design optimal adaptive controllers," *IEEE Control Systems*, vol. 32, no. 6, pp. 76–105, Dec 2012.

- [22] A. G. Barto, W. B. Powell, J. Si, and D. C. Wunsch, "Handbook of learning and approximate dynamic programming," 2004.
- [23] T. Dierks and S. Jagannathan, "Optimal control of affine nonlinear continuous-time systems," in *Proceedings of the 2010 American Control Conference*, June 2010, pp. 1568–1573.
- [24] Z. Chen and S. Jagannathan, "Generalized hamilton-jacobi-bellman formulation -based neural network control of affine nonlinear discrete-time systems," *IEEE Transactions on Neural Networks*, vol. 19, no. 1, pp. 90–106, Jan 2008.
- [25] D. Wang, D. Liu, Q. Zhang, and D. Zhao, "Data-based adaptive critic designs for nonlinear robust optimal control with uncertain dynamics," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. PP, no. 99, pp. 1–12, 2015.
- [26] N. Vignesh and S. Jagannathan, "Distributed event-sampled approximate optimal control of interconnected affine nonlinear continuous-time systems," in *2016 American Control Conference (ACC)*, July 2016, pp. 3044–3049.
- [27] D. Liu, D. Wang, D. Zhao, Q. Wei, and N. Jin, "Neural-network-based optimal control for a class of unknown discrete-time nonlinear systems using globalized dual heuristic programming," *IEEE Transactions on Automation Science and Engineering*, vol. 9, no. 3, pp. 628–634, July 2012.
- [28] V. Narayanan and S. Jagannathan, "Approximate optimal distributed control of uncertain nonlinear interconnected systems with event-sampled feedback," in *2016 IEEE 55th Conference on Decision and Control (CDC)*, Dec 2016, pp. 5827–5832.
- [29] F. Lewis, S. Jagannathan, and A. Yesildirak, *Neural network control of robot manipulators and non-linear systems*. CRC Press, 1998.
- [30] H. K. Khalil and J. Grizzle, *Nonlinear systems*. Prentice hall New Jersey, 1996, vol. 3.
- [31] B. Xu, C. Yang and Z. Shi, "Reinforcement Learning Output Feedback NN Control Using Deterministic Learning Technique," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 25, no. 3, pp. 635–641, July 2014.