IEEE TRANSACTIONS ON
CYBERNETICS

Event-triggered Distributed Control of Nonlinear Interconnected Systems Using Online Reinforcement Learning with Exploration

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

***Abstract*—In this paper, a distributed control scheme for an interconnected system composed of uncertain input affine nonlinear subsystems with event triggered state feedback is presented by using a novel hybrid learning scheme-based approximate dynamic programming (ADP) with online exploration. First, an approximate solution to the Hamilton-Jacobi-Bellman (HJB) equation is generated with event sampled neural network approximation and subsequently, a near optimal control policy for each subsystem is derived. Artificial neural networks (NNs) are utilized as function approximators to develop a suite of identifiers and learn the dynamics of each subsystem. The NN weight tuning rules for the identifier and event-triggering condition are derived using Lyapunov stability theory. Taking into account, the effects of NN approximation of system dynamics and bootstrapping, a novel NN weight update is presented to approximate the optimal value function. Finally, a novel strategy to incorporate exploration in online control framework, using identifiers, is introduced to reduce the overall cost at the expense of additional computations during the initial online learning phase. System states and the NN weight estimation errors are regulated and local uniformly ultimately bounded (UUB) results are achieved. The analytical results are substantiated using simulation studies.**

I. Introduction

Advanced control schemes are necessary for efficient and cost effective operation of industrial systems with uncertain dynamics. The adaptive/approximate dynamic programming (ADP) schemes aim to address the problem of optimization over time through learning without needing apriori knowledge of the system dynamics [1]. In the ADP schemes, an approximate optimal value function as the solution to the Hamilton-Jacobi-Bellman (HJB) equation is generated from which an optimal control policy for a nonlinear system is constructed using nonstandard techniques [2], [3] that are inspired by the reinforcement learning (RL) [4], [5].

The RL theory was developed based on the naturalistic learning observed in biological species [4]. Naturally, RL-ADP approaches are expected to mimic human intelligence for control with four desired characteristics [1]. The ability to solve optimal control problem over time, through learning; involve a critic module to generate reinforcement signal; use the reinforcement signal to generate optimal action; and finally, an adaptive component to emulate the system and estimate the system internal states.

Significant effort in developing ADP schemes, with the four desired characteristics, has been put forth in the literature [6]–[12]. However, in applications involving real-time online control, iterative learning approach like policy/value iteration (PI/VI) to generate control actions is undesirable due to the large iterations required for convergence [9]–[11]. Reducing the computations considerably, the time-driven (TD) ADP control approach introduced in [11] was designed without using iterative learning. This approach was motivated by the one step temporal difference learning (TDL) of RL [4], [6] and traditional adaptive control theory [13].

Due to reduction in the number of iterations of the TD ADP method, the control sequence achieved optimality asymptotically and not at each time step. Later, distributed optimal control schemes for interconnected system was considered in the literature by using iterative learning approaches [14], [15]. However, with the advent of networked control systems, efficient utilization of communication resource and faster learning are also desired in an ADP scheme, especially for distributed control applications. The control schemes in [14], [15] and the references therein were presented using continuous feedback requiring substantial communication resources when the feedback is closed via a communication network.

To avoid frequent control updates without sacrificing system stability, event-triggered control design was introduced [16]–[20] and extended for distributed feedback control [20]; filtering [21]; robust control [19] and optimal control [22]. A review of various event-triggered control schemes in the literature is available in [23]. In the adaptive event triggered controller presented by Sahoo et al. [16] using NNs, the event-triggering mechanism was designed as a function of estimated neural network (NN) weights. The rationale behind such a design was to increase the events during the initial learning period, to facilitate learning. Later, an event-triggered control scheme using PI was proposed [17], [18].

However, the event triggered ADP schemes [16]–[18], are inefficient due to following reasons: a) the learning time is increased due to intermittent feedback, as the frequency of events decides the approximation accuracy [24]; b) the sampling instants are dynamic, hence, the inter-sampling intervals are time-varying, restricting the use of iterative learning schemes [17], [18]. Therefore, to accelerate the learning process with event-triggered feedback and ensure online, real time implementation, we proposed a flexible hybrid learning

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: (vnxv4@mst.edu, sarangap@mst.edu).

Digital Object Identifier: 10.1109/TCYB.2017.2741342

framework in our previous effort [25], when the system dynamics were known. Practically, most systems have certain portion of the dynamics uncertain.

Hence, we studied and reported preliminary results for the hybrid learning scheme [26], relaxing the requirement of accurate knowledge of the system dynamics. Despite improving the system performance, by reducing the cost during the initial online learning phase, the learning scheme presented in [26] is still inefficient because: a) the sensor samples measured during the inter-event period are not utilized in the learning process as the feedback instants are decided by the event triggering mechanism [25], [26]; b) online exploration, which is essential during the online learning phase is not utilized for better estimate of value function. The exploration vs exploitation dilemma is a classical problem in RL theory and it is also observed in the forward-in-time ADP schemes. In all the ADP based learning schemes, the exploratory signals are applied directly to the system which is not always desirable [1], [10], [11], [14], [16]–[18].

Therefore, in this paper, we overcome these deficiencies in the existing learning schemes from a RL perspective by introducing a novel distributed NN identifier which is different from the NN identifier in [26]. Further, a new weight update rule for learning and enhancing the approximate optimal value function estimate with an online exploration strategy by using the identifiers is introduced. Hence, the cumulative cost during the learning phase is reduced at the expense of additional computations, which can be considered as a trade-off. These changes resulted in a novel design of an approximate control scheme when compared with [25], [26] and event-triggered ADP schemes [17], [18], [24].

The analytical results for the enhanced learning scheme with online exploration proposed in this paper are provided. Firstly, in contrast to the adaptive event-triggering schemes [16]–[18], [25], [26], it is demonstrated that the closed-loop system is input-to-state stable (ISS) which is needed to ensure that the event-trigger mechanism does not induce Zeno behavior. Finally, local uniformly ultimately bounded (UUB) regulation of the system, identifier states to a neighborhood of origin and convergence of the developed policy to a neighborhood of the optimal policy are achieved with the proposed distributed controller with exploration. Simulation results are presented to demonstrate the advantages of the proposed control scheme.

This paper is organized as follows: Section II introduces the dynamics of the system being investigated and presents a brief background on distributed optimal control formulation for interconnected system. Section III presents discussions on existing event-triggered ADP schemes and our previous effort [25], [26] from an RL perspective. Section IV presents the proposed learning control scheme with online exploration. Stability analysis and simulation results are included in Sections V and VI, respectively. The conclusions drawn from this study are reported in Section VII.

II. BACKGROUND AND PROBLEM STATEMENT

A. Notations

The subscript $(\bullet)_i$ will be used to denote the variables of the i^{th} subsystem in the interconnected system and $(\hat{\bullet})$ is used to indicate that the variable is an estimated quantity; $(\tilde{\bullet})$ denotes that the quantity is an estimation or approximation error. The variables $\Re, \mathbb{N}$ denote the sets of real and natural numbers respectively; $\Re^n$ denotes the n dimensional Euclidean space; $\Re^{n\times m}$ denotes the product space generated by $\Re^n, \Re^m$. In the analysis, when $x \in \Re^n$, $\|x\|$ denotes the Euclidean norm; for $A \in \Re^{n\times m}$, $\|A\|$ denotes its Frobenius norm. The analysis of the event triggered controller will follow the sampled data approach.

B. System description

Consider a nonlinear input affine continuous-time system composed of N interconnected subsystems, described by the differential equation

$$\dot{x}_i = f_i(x_i) + g_i(x_i)u_i + \sum_{\substack{j=1\\ j\neq i}}^{N} \Delta_{ij}(x_i, x_j),\ x_i(0) = x_{i0} \tag{1}$$

where $x_i(t) \in B_i \subseteq \Re^{n_i\times 1}$ represents the state vector of the i^{th} subsystem and $\dot{x}_i(t)$ its time derivative, B_i is a compact set, $u_i(t) \in \Re^{m_i}$ is the control input, $f_i : B_i \to \Re^{n_i}$, $g_i : B_i \to \Re^{n_i\times m_i}$ are uncertain nonlinear maps and $\Delta_{ij} : \Re^{n_i\times n_j} \to \Re^{n_i}$ is the uncertain nonlinear interconnection between i^{th} and j^{th} subsystem. The augmented system dynamics are

$$\dot{x} = F(x) + G(x)u,\ x(0) = x_0 \tag{2}$$

where $F = [(f_1 + \sum_{j=2}^{N} \Delta_{1j})^T, ., (f_N + \sum_{j=1}^{N-1} \Delta_{Nj})^T]^T$, $x = [x_1^T, ., x_N^T]^T \in B \subseteq \Re^n$, $B = \bigcup_{i=1}^{N} B_i$, $u = [u_1^T, .., u_N^T]^T \in \Re^m$, $m = \sum_{i=1}^{N} m_i$, $n = \sum_{i=1}^{N} n_i$ and $G = diag([g_1(x_1).., g_N(x_N)])$. The following assumptions are needed for the control design.

Assumption 1: The dynamics (1) and (2) are stabilizable with equilibrium point at the origin. Full state measurements are available for control. The communication network which facilitates information sharing among subsystems is lossless.

Assumption 2: The nonlinear map $g_i(x_i)$ is bounded such that $0 < g_{im} < \|g_i(x_i)\| \leq g_{iM}$, in B_i for every subsystem.

Assumption 3: The functions $f_i(x_i)$, $\Delta_{ij}(x_i, x_j)$, $g_i(x_i)$ are locally Lipschitz continuous on compacts.

In the next subsection, the notion of event-triggered feedback and greedy policy design with aperiodic event based feedback is presented.

C. Event-triggered feedback and Optimal control

Consider a sequence of time instants, $\{t_k\}_{k=0}^{\infty}$, to denote the event-sampling instants. Let $x_i(t_k^i)$ be the state of the i^{th} subsystem at time instant t_k^i. Between successive event-sampling instants t_k^i, t_{k+1}^i, the state vector is denoted as $\breve{x}_i(t) = x_i(t_k^i),\ \forall k \in 0\cup\mathbb{N}$. Using the zero-order-hold (ZOH), the last updated states and control are held at actuators and controllers between events. To denote the difference between the actual system states and the state available at the controller, an event-sampling error is defined as

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

By rewriting $\breve{x}_i(t)$ using (3), the feedback between events can be defined as a continuous function $\breve{x}_i(t) = x_i(t) - e_i(t)$. Next, define the infinite horizon cost function of the augmented system (2), as

$$V(x(t)) = \int_t^\infty \big[Q(x) + u^T(\tau)Ru(\tau)\big]d\tau \tag{4}$$

where $Q(x) > 0,\ \forall x \in B\backslash\{0\},\ Q(0) = 0,\ R > 0$ are the penalty functions of appropriate dimensions. Let $V(.)$ and its time-derivative be continuous on a compact set B. Then, $\dot{V}(x(t)) = -\big[Q(x) + u^T(t)Ru(t)\big]$. Using the infinitesimal version of (4), define the Hamiltonian function $H(x,u) = \big[Q(x) + u^T(t)Ru(t)\big] + (\partial V^T/\partial x)\dot{x}$. The optimal control policy which minimizes (4) (assuming a unique minimum exists) is obtained by using the stationarity condition as $u^* = -\frac{1}{2}R^{-1}G^T(x)\partial V^*/\partial x$ and it is called greedy policy with respect to (4). The Hamiltonian function can be defined between two event-triggering instants, $[t_k, t_{k+1})$, as

$$H(x(t), u(t_k)) = \big[Q(x) + u^T Ru\big] + (\partial V^T/\partial x)\dot{x}. \tag{5}$$

The greedy policy with event-triggered state becomes

$$u^*(t) = -\tfrac{1}{2}R^{-1}G^T(\breve{x})(\partial V^*/\partial \breve{x}) \tag{6}$$

with $\breve{x}(t) = x(t_k),\ t \in [t_k, t_{k+1})$.

Remark 1: Substituting (6) in (5) reveals the continuous time equivalent of the Bellman equation which is called the HJB equation and its solution, the optimal value function $V^*(x(t))$, is required to obtain the greedy policy (6). Using a zero-order hold (ZOH), we can ensure that the control is piecewise continuous (6).

Now, for the interconnected system (2) under consideration, the i^{th} subsystem dynamics (1) are influenced by the states of the j^{th} subsystem satisfying $\Delta_{ij}(x_i, x_j) \neq 0$. To compensate for this interaction, $u_i(t)$ is desired to be a function of both $x_i(t), x_j(t)$ [27].

Proposition 1: [26] Consider the i^{th} subsystem in (1) and the cost function (4) for (2), then $\exists u_i^*(t) \in \Re^{m_i}$, given by

$$u_i{}^* = -0.5{R_i}^{-1}{g_i}^T(x_i)(\partial {V_i}^*(x)/\partial x_i),\ \forall i \in 1, 2, ..N. \tag{7}$$

as a function of $x_i(t), x_j(t)$, for all $j \in 1, 2, .., N,\ : \Delta_{ij} \neq 0$, where ${V_i}^*(x)$ represent the optimal value function of the i^{th} subsystem, R_i is a positive definite matrix, such that the cost function (4) is minimized.

Remark 2: The control policy (7) is obtained by rewriting the cost function of the overall system as the sum of cost functions of individual subsystems [26].

The greedy policy for the augmented system (2) can be obtained using (7) at each subsystem, given the system dynamics and the optimal value function. Since the system dynamics and the optimal value function are unknown function approximators are used to approximate the same.

D. Event sampled NN approximation

With the objective of finding the approximate optimal value function as an approximate solution to the HJB using aperiodic event-triggered feedback, the event-based NN approximation [24] is utilized. Define a smooth function, $\chi : B \to \Re$, in a compact set $B \subseteq \Re^n$. Given $\varepsilon_M > 0$, $\exists \theta^* \in \Re^{p\times 1} : \chi(x) = {\theta^*}^T\phi(x_e) + \varepsilon_e$. The event-triggered approximation error ε_e is defined as $\varepsilon_e = {\theta^*}^T(\phi(x_e + e) - \phi(x_e)) + \varepsilon(x)$, satisfying $\|\varepsilon_e\| < \varepsilon_M,\ \forall x_e \in B$, where x, x_e are continuous and event triggered variables, e is the measurement error due to event sampling, $\varepsilon(x)$ is the bounded NN reconstruction error and $\phi(x_e)$ is an appropriately chosen basis function.

Remark 3: An important relationship between the accuracy of NN approximation and frequency of events is revealed by the representation of the NN approximation with event-triggered aperiodic inputs [24], introducing a trade-off between the sampling frequency and approximation accuracy.

The following assumption is required for the ADP design.

Assumption 4: The solution for the HJB (5) is unique, real-valued, smooth and satisfies $V^*(x) = \sum_{i=1}^N V_i^*(x)$. Further, $\phi(x)$ is chosen such that $\phi(0) = 0$, the activation function and its derivative and the constant, target NN weights are assumed to be bounded [10], [11].

The parameterized representation of the optimal value function using NN weights θ^* and basis function $\phi(x_e)$ with event based inputs is given as

$$V^*(x) = {\theta^*}^T\phi(x_e) + \varepsilon(x_e) \tag{8}$$

where $\varepsilon(x_e)$ is the event driven reconstruction error. Define the target NN weights as θ_i^* at the i^{th} subsystem. Using a parameterized representation (8) for $V_i^*(x)$, HJB equation [11], [26] can be derived as

$$\begin{aligned}{\theta_i^*}^T\nabla_x\phi(x)\bar{f}_i - \frac{{\theta_i^*}^T\nabla_x\phi(x)D_i{\nabla^T}_x\phi(x)\theta_i^*}{4} \\ + \varepsilon_{i_{HJB}} + Q_i(x) = 0\end{aligned} \tag{9}$$

where $Q_i(x) > 0$, $D_i = g_i(x_i){R_i}^{-1}{g_i}^T(x_i)$, $\bar{f}_i = f_i(x_i) + \sum_{\substack{j=1\\ j\neq i}}^N \Delta_{ij}$ and $\varepsilon_{i_{HJB}} = \nabla_x\varepsilon_i^T(\bar{f}_i - 0.5D_i(\nabla^T{}_x\phi(x)\theta_i^* + \nabla_x\varepsilon_i) + 0.25D_i\nabla_x\varepsilon_i)$. The estimated value function is given by $\hat{V}_i(x) = \hat{\theta}_i^T\phi(x)$, where $\hat{\theta}_i$ is the NN estimated weights and its gradient along the states is given by $\partial\hat{V}_i/\partial x_i = \hat{\theta}_i^T\nabla_x\phi(x)$ and $\nabla_x\phi(x)$ is the gradient of the activation function $\phi(x)$ along x. The Hamiltonian function using $\hat{V}_i(x_e) = \hat{\theta}_i^T\phi(x_e)$ reveals

$$\begin{aligned}\hat{H} =& Q_i(x_e) + \hat{\theta}_i^T\nabla_x\phi(x_e)\bar{f}_i \\ &- 0.25\hat{\theta}_i^T\nabla_x\phi(x_e)D_{i,\varepsilon}{\nabla^T}_x\phi(x_e)\hat{\theta}_i\end{aligned} \tag{10}$$

where $D_{i,\varepsilon} = D_{i,\varepsilon}(x_{i,e}) = g_i(x_{i,e}){R_i}^{-1}{g_i}^T(x_{i,e})$. The estimated optimal control input is obtained from (10) as

$$u_{i,e} = -0.5{R_i}^{-1}g_i^T(x_{i,e})\hat{\theta}_i^T\nabla_x\phi(X_e),\ \forall i \in 1, 2, ..N. \tag{11}$$

Note that (10) is used as the forcing function to tune $\hat{\theta}_i$. The NN identifier design with event triggered feedback is introduced in the next subsection. The identifiers are utilized to generate the uncertain nonlinear functions and also for the purpose of exploration, which will be discussed in section IV.

III. EVENT DRIVEN ADAPTIVE DYNAMIC PROGRAMMING

In this section, first, NN identifiers are designed at each subsystem to approximate the uncertain nonlinear functions in

(1). Then, the event-triggered hybrid learning algorithm for constructing an approximately optimal control sequence using the identifier NN is introduced.

A. Identifier design for the interconnected system

For approximating the subsystem dynamics, consider a distributed identifier at each subsystem, which operates with event triggered feedback information

$$\dot{\hat{x}}_i = \hat{f}_i(\hat{x}_i) + \hat{g}_i(\hat{x}_i)u_{i,e} + \sum_{\substack{j=1 \\ j\neq i}}^{N} \hat{\Delta}_{ij}(\hat{x}_i, \hat{x}_j) - A_i\tilde{x}_{i,e} \quad (12)$$

where $\tilde{x}_{i,e} = x_{i,e} - \hat{x}_i$, is the event-driven state estimation error and $A_i > 0$ is a positive definite matrix which stabilizes the NN identifier. Using NN approximation, the parametric equations for the nonlinear functions in (1) are $g_i(x_i) = W_{ig}\sigma_{ig}(x_i) + \varepsilon_{ig}(x_i)$, $\bar{f}_i(x) = W_{if}\sigma_{if}(x) + \varepsilon_{if}(x)$; where $W_{i\bullet}$ denotes the target NN weights, $\sigma_{i\bullet}$ denotes the bounded NN activation functions and $\varepsilon_{i\bullet}$ denotes the bounded reconstruction errors. Using the estimate of the NN weights, $\hat{W}_{i\bullet}$, define $\hat{\bar{f}}_i(x) = \hat{W}_{if}\sigma_{if}(x)$ and $\hat{g}_i(\hat{x}_i) = \hat{W}_{ig}\sigma_{ig}(\hat{x}_i)$. Now, to analyze the stability of (12), define the state estimation error $\tilde{x}_i(t) = x_i(t) - \hat{x}_i(t)$. Using (12) and (1), the dynamic equation describing the evolution of $\tilde{x}_i(t)$ is revealed as

$$\begin{aligned}\dot{\tilde{x}}_i = {} & \tilde{W}_{if}\sigma_{if}(x) + W_{if}\tilde{\sigma}_{if} - \tilde{W}_{if}\tilde{\sigma}_{if} + [\tilde{W}_{ig}\sigma_{ig}(x_i) + \\ & W_{ig}\tilde{\sigma}_{ig} - \tilde{W}_{ig}\tilde{\sigma}_{ig}]u_{i,e} + \varepsilon_{ig}u_{i,e} + \varepsilon_{if} + A_i\tilde{x}_i + A_ie_i\end{aligned} \quad (13)$$

with $\tilde{\sigma}_{i\bullet} = \sigma_{i\bullet}(x) - \sigma_{i\bullet}(\hat{x})$, $\tilde{W}_{i\bullet} = W_{i\bullet} - \hat{W}_{i\bullet}$.

Remark 4: Note that the approximation of $\bar{f}(x)$ requires the states of the i^{th}, j^{th} subsystem satisfying $\Delta_{ij}(x_i, x_j) \neq 0$. Therefore, the inputs to the NN are $\hat{x}_i, x_{j,e}, \tilde{x}_i$. Due to the presence of $x_{j,e}$ as input, the identifier is considered to be distributed.

With the proposed NN identifiers at each subsystem, the control design equations (10) and (11) can be re-derived as $\hat{H} = Q_i(x_e) + \hat{\theta}_i^T\nabla_x\phi(x_e)\hat{\bar{f}}_i - 0.25\hat{\theta}_i^T\nabla_x\phi(x_e)\hat{D}_{i,\varepsilon}\nabla^T{}_x\phi(x_e)\hat{\theta}_i$ and $u_{i,e} = -0.5R_i{}^{-1}\hat{g}_i^T(x_{i,e})\hat{\theta}_i^T\nabla_x\phi(x_e)$, $\hat{D}_{i,\varepsilon} = \hat{g}_i(x_{i,e})R_i{}^{-1}\hat{g}_i^T(x_{i,e})$. To this end, all the design equations to learn the greedy policy $u_i^*(t)$ without requiring the nonlinear functions $\bar{f}_i, g_i$ and V_i^* are developed. Next, define the following positive definite, radially unbounded Lyapunov candidate function for the identifier

$$J_{iI}(\tilde{x}_i, \tilde{W}_{if}, \tilde{W}_{ig}) = J_{i\tilde{x}} + J_{i\tilde{f}} + J_{i\tilde{g}} \quad (14)$$

with $J_{i\tilde{x}} = 0.5\mu_{i1}\tilde{x}_i^TP_i\tilde{x}_i$, $J_{i\tilde{f}} = 0.5\mu_{i2}\tilde{W}_{if}^T\tilde{W}_{if} + 0.25\mu_{i4}(\tilde{W}_{if}^T\tilde{W}_{if})^2$, $J_{i\tilde{g}} = 0.5\mu_{i3}\tilde{W}_{ig}^T\tilde{W}_{ig} + 0.25\mu_{i5}(\tilde{W}_{ig}^T\tilde{W}_{ig})^2 + 0.125\mu_{i6}(\tilde{W}_{ig}^T\tilde{W}_{ig})^4$; where $\mu_{ij}, P_i > 0$, $j = 1, 2, ., 6$. Local UUB regulation of $\tilde{x}_i(t), \tilde{W}_{i\bullet}(t)$ is achieved when (13) is injected with a non-zero bounded input $e_i(t)$ and this result is summarized next.

Lemma 1: Consider the identifier dynamics (12). Using the estimation error, $\tilde{x}_i(t)$, as a forcing function, define NN weight tuning using the Levenberg-Marquardt scheme with sigma modification term to avoid parameter drift as

$$\begin{aligned}\dot{\hat{W}}_{if} &= \frac{\alpha_{if}\sigma_{if}\tilde{x}_{i,e}^T}{c_{if} + \left\|\tilde{x}_{i,e}^T\right\|^2} - \kappa_{if}\hat{W}_{if}, \\ \dot{\hat{W}}_{ig} &= \frac{\alpha_{ig}\sigma_{ig}u_{i,e}\tilde{x}_{i,e}^T}{c_{if} + \left\|\tilde{x}_{i,e}^T\right\|^2\left\|u_{i,e}^T\right\|^2} - \kappa_{ig}\hat{W}_{ig}\end{aligned} \quad (15)$$

where $\alpha_{if}, \alpha_{ig}, \kappa_{if}, \kappa_{ig}, c_{if}$ are positive design constants. The error dynamics using (15) are obtained as

$$\begin{aligned}\dot{\tilde{W}}_{if} &= \frac{-\alpha_{if}\sigma_{if}\tilde{x}_{i,e}^T}{c_{if} + \left\|\tilde{x}_{i,e}^T\right\|^2} + \kappa_{if}\hat{W}_{if}, \\ \dot{\tilde{W}}_{ig} &= \frac{-\alpha_{ig}\sigma_{ig}u_{i,e}\tilde{x}_{i,e}^T}{c_{if} + \left\|\tilde{x}_{i,e}^T\right\|^2\left\|u_{i,e}^T\right\|^2} + \kappa_{ig}\hat{W}_{ig}.\end{aligned} \quad (16)$$

Let all the assumptions introduced in the previous sections hold. If $u_{i,e}(t) = u_{i,e}^*$, then $\alpha_{if}, \alpha_{ig}, \kappa_{if}, \kappa_{ig}, A_i > 0$ can be chosen such that (13) and (16) are stable and $\tilde{x}_i(t), \tilde{W}_{i\bullet}(t)$ are locally uniformly ultimately bounded (UUB).

Proof: See appendix.

Remark 5: The assumption that the control input is optimal and the measurement error acting as an input $e_i(t)$ is bounded will be relaxed in the closed loop stability analysis (See Section V). The stability of the identifier in the presence of measurement errors is required to employ the identifiers for the purpose of exploration, wherein the measurement errors in (13) are replaced by bounded exploratory signals.

Now, an event-triggered implementation of the distributed controller design for (2) using hybrid learning algorithm is presented.

B. Event based hybrid learning scheme

A brief discussion on the hybrid learning scheme [26] with uncertain dynamics is presented here. An event triggering mechanism is required at each subsystem to determine the discrete time instants when: 1) the i^{th} subsystem controller receives $x_i(t)$; 2) $u_i(t)$ is updated with the latest states at the actuator and 3) $x_i(t)$ is broadcast to the neighboring subsystems. Define a positive definite, continuous function $J_i(x_i) = x_i^T\Gamma_ix_i$, with $\Gamma_i > 0$. For $0 < \alpha_i < 1$ and $k \in \mathbb{N}$, design the event-triggering mechanism to satisfy the condition

$$J_{ix}(x_i(t)) \leq (1 + t_k^i - t)\alpha_iJ_{ix}(x_i(t_k^i)), \quad t \in [t_k^i, t_{k+1}^i). \quad (17)$$

with $t_0^i = 0, \ \forall i \in 1, 2, .., N$.

Remark 6: Note that t_k^i and t_k^j, for $i \neq j$ are independent. The objective of this paper is to develop learning algorithms which accelerates the learning process when the feedback from the system is available only at aperiodic, event triggered time instants. Therefore, the learning algorithms presented in the paper are independent of the event triggering condition.

The optimality of the value function in the event based temporal difference (TD) algorithm in [24] is directly related with the frequency of event-triggering instants. To improve the estimate of the optimal value function, past data can be used in between events to further bring down the HJB residual error which reduces the NN weight estimation error

$\tilde{\theta}_i = \theta_i^* - \hat{\theta}_i$. Also note that the time between consecutive events is not constant. Therefore, the RL based iterative algorithms which perform iterative learning until a stopping criterion is satisfied require strong conditions on the inter-event period. This stopping criterion is pre-decided as a minimum threshold on the HJB errors.

In the hybrid learning scheme, the weights of the value function approximator NN are tuned during $t_k^i < t < t_{k+1}^i$, using the HJB residual error calculated at t_k^i. With the approximated dynamics using identifiers, the weight update rule for the proposed hybrid scheme is given by

$$\dot{\hat{\theta}}_i = \begin{cases} -(\alpha_{iv}\hat{\psi}_i\hat{H}_i)/(1+\hat{\psi}_i^T\hat{\psi}_i)^2 + 0.5\mu_{i1}\nabla_x\phi\hat{D}_i^T P_i\tilde{x}_i \\ \qquad -\kappa_3\hat{\theta}_i + 0.5\alpha_{iv}\nabla_x\phi\hat{D}_i x_i, \;\; t = t_k^i \\ -(\alpha_{iv}\hat{\psi}_i\hat{H}_i)/(1+\hat{\psi}_i^T\hat{\psi}_i)^2, \;\; t_k^i < t < t_{k+1}^i, \;\; \forall k \in \mathbb{N}. \end{cases} \tag{18}$$

where $\hat{\psi}_i = \partial\hat{H}_i/\partial\hat{\theta}_i$, $\hat{D}_i = \hat{g}_i(x_i){R_i}^{-1}\hat{g}_i^T(x_i)$ and $\mu_{i1}, P_i, \kappa_3, \alpha_{iv} > 0$ are design constants. As a consequence of the weight updates in the interval (t_k, t_{k+1}), the convergence time for the learning algorithm is reduced.

Remark 7: The estimated Hamiltonian in (18) utilizes the approximation $\hat{\bar{f}}_i, \hat{g}_i$ to calculate the HJB error. The term $0.5\mu_{i1}\nabla_x\phi\hat{D}_i^T P_i\tilde{x}_i$ in (18) can be viewed as a compensation for the identification errors (13) and $\kappa_3\hat{\theta}_i$ in (18) is the sigma modification term to relax the persistent excitation (PE) condition and avoid parameter drift.

Remark 8: If the $t_{k+1}^i - t_k^i$ is large, sufficient time is available to tune the NN weights such that the HJB error reduced to a value very close to zero. This provides a value function estimate very close to the optimal value function.

The proposed hybrid learning scheme is best suitable for online implementation. Nevertheless, the hybrid learning scheme seem to be inefficient due to the fact that it does not utilize the feedback information and the reward signal available during the inter-event period. The classical problem of exploration vs exploitation and a modified/enhanced learning algorithm which overcomes the drawbacks of the hybrid learning scheme are introduced next.

IV. LEARNING WITH EXPLORATION FOR ONLINE CONTROL

The basic idea behind the enhanced hybrid learning scheme is presented first and the role of the identifiers will be highlighted. The identifiers presented in the previous section are used to approximate the subsystem dynamics. In contrast, in this section the NN identifiers which approximate the overall system dynamics will be designed at each subsystem to aid in the implementation of the modified weight update rule which will be introduced in this section. Finally, the role of exploration and the challenges involved in online control will be discussed.

A. Enhanced hybrid learning

The state and control information along the state trajectory during the inter-event period is unused in the existing algorithms leading to inefficient learning. Instead, this information during the inter-event period can be stored and used to update the weights of the value function NN at the event sampling instant. It should be noted that the state information during the inter-event period is not available at the controller/learning mechanism though it is measured and utilized at the event triggering mechanism.

Therefore, the state and control information can be stored at the trigger mechanism and transmitted to the controller at the event sampling instants. This means that for the interconnected system, the states are to be transmitted from the sensor to the controller at each subsystem and broadcasted to other subsystems. As a consequence, the communication overhead is increased as the packet size will increase due to fewer events.

To mitigate this problem, the identifier located at each subsystem can be used to generate this data and can be used in the learning process. However, the use of online identifier and the controller together results in an unreliable set of data for the value function estimator as demonstrated later in the simulation section. By tuning the identifier weights first, the data generated by the identifier can be utilized for learning the optimal value function. Let the sensor sampling frequency be defined as τ_s. Consider the weight tuning rule

$$\dot{\hat{\theta}}_i = \begin{cases} -\dfrac{\alpha_{i1v}\hat{\psi}_i\hat{H}_i}{(1+\hat{\psi}_i^T\hat{\psi}_i)^2} - \dfrac{\alpha_{i2v}\hat{\Psi}_i\bar{H}_i}{(1+\hat{\Psi}_i^T\hat{\Psi}_i)^2} - \kappa_3\hat{\theta}_i + \\ 0.5\alpha_{iv}\nabla_x\phi\hat{D}_i x_i + 0.5\mu_{i1}\nabla_x\phi(x)\hat{D}_i^T P_i\tilde{x}_i, \;\; t = t_k^i \\ -(\alpha_{iv}\hat{\psi}_i\hat{H}_i)/(1+\hat{\psi}_i^T\hat{\psi}_i)^2, \;\; t_k^i < t < t_{k+1}^i, \;\; \forall k \in \mathbb{N}. \end{cases} \tag{19}$$

with the design variables similar to (18) and $\bar{H}_i, \hat{\Psi}_i$ are the estimated Hamiltonian and its derivative with respect to the NN weights calculated using the estimated states during the inter-event period. Since $\bar{H}_i$ is a function of the overall states, a NN identifier which approximates the overall system can provide the overall state estimate at each subsystem and the design of such an identifier is briefly presented next.

Remark 9: From the simulation analysis, it is observed that gains satisfying $\alpha_{i1v} > \alpha_{i2v}$ yields better results.

B. Identifiers for the enhanced hybrid learning scheme

Consider the NN identifier at each subsystem as

$$\dot{\hat{X}}_i = \hat{F}_i(\hat{X}_i) + \hat{G}_i(\hat{X}_i)U_{i,e} - A_i\tilde{X}_{i,e} \tag{20}$$

where the subscript i indicates variables available at the i^{th} subsystem; $\hat{F}_i, \hat{G}_i$ are the approximated functions of the overall dynamics F, G; $\hat{X}$ is the estimate of x in (2) and U is the augmented control u. In contrast to (12), the identifier described by (20) estimates the states of the interconnected system (2) to collect the state information and calculate the reinforcement signal for the inter event period. The actual and estimated weights for the functions F_i, G_i can be defined as in Section III. A and equations similar to (13)-(16) can be derived for the observer in (20).

Remark 10: The observer design procedure for (20) is similar to that in (12). Therefore, all the details are not included. However, there are a few major differences in the NN design. Since the observer in (20) approximates the nonlinear mapping of the overall system, first, the NN takes as input, the vector $[\hat{x}_i^T\ \hat{x}_j^T]^T, \forall j = 1, 2, ..N$ instead of $\hat{x}_i$; second, the

number of neurons in the hidden layer are to be increased as the domain of the nonlinear map being approximated are of higher dimensions.

The local UUB of the identifier presented in Section III is applicable to the identifier designed in this section. Therefore, to avoid redundancy, the results are not re-derived at this point. With this NN identifier, the weight update rule (19) can be realized.

Remark 11: The use of function approximators to learn the optimal value function and system dynamics adds to the uncertainty of bootstrapping [4] in finding the optimal control inputs. In addition, since the learning scheme is based on asynchronous generalized policy iteration (GPI) [4], the initial weights of the function approximators affect the state trajectory and cumulative cost (return).

Remark 12: The proposed enhanced hybrid learning scheme can be viewed from the RL perspective as follows: in the inter event period, the system generates reinforcement signal along the state trajectories which are not fed back to the controller. This bexperienceb is not utilized by the learning schemes presented in [18], [24], [26]. Therefore, the additional term, $\alpha_{i2v}\hat{\Psi}_i\bar{H}_i/(1+\hat{\Psi}_i^T\hat{\Psi}_i)^2$, in (19) uses the bexperienceb in the inter event period to provide a better optimal value function estimate.

C. Role of identifiers and exploration in online control

One of the classical problems in the RL literature [4], [28] is the dilemma of exploration vs exploitation. To understand this problem let us consider the RL decision making problem. The decision making process consists of constructing maps of states to expected future reward using reinforcement signals [4]. The future actions are influenced by this prediction of future reward, i.e. using the feedback signal, the HJB error is computed and the approximate optimal value function is updated based on the HJB error; the estimated value function is then used to obtain the future control action. If the control action is of the form (11), then it is a greedy policy and hence, exploitative. This is due to the fact that the control policy exploits the current knowledge of the optimal value function and minimizes the Hamiltonian (10). In contrast, if a control policy that is not greedy is applied to the system, then the control policy is said to be explorative. One has to ensure stability when such a policy is used in online control.

The PE condition is an important requirement for the ADP control methods in [10] for the convergence of the estimated parameters to its target values. This condition ensures that sufficient data is collected to learn the unknown function before the system states settle at an equilibrium point. Adaptive control theorists developed sigma and epsilon modification techniques [3], [13] to prevent parameter drift and relax PE condition requirement. However, from a learning perspective the sigma and epsilon modification techniques inhibit the learning algorithm from exploring.

To perturb the system and to satisfy the PE condition a control policy of the form $\varpi_e(t) = u(t)+\xi(t)$ was used in the learning algorithms presented in [10], [11] and the references therein, where $\xi(t)$ is seen as an exploratory signal and u is a stabilizing/greedy control policy. For example, random noise signal was used as $\xi(t)$ in the simulations; while [29] explicitly considered the control law with $\xi(t)$ to develop an actor-critic based ADP design. To relax the PE condition, sufficient data can be collected to satisfy the rank condition, as indicated in traditional adaptive control [13]. It should also be considered that exploration signal $\xi(t)$ is not easy to design. Although several exploration policies are investigated for finite Markov decision processes [4] and offline learning schemes [4], [28], an exploration policy which can provide guaranteed time for convergence to a near optimal policy for an online control problem is not available.

Also, in control, issues of stability and robustness are non-trivial. The system can become unstable in the process of exploration due to the application of $\xi(t)$ in the control action. Inspired by the work on efficient exploration in [28], a novel technique to incorporate exploration in the learning controller is developed next.

D. Exploration using identifiers

The TD learning [4], [11], [24] and the hybrid learning schemes [22], [26] reduce the HJB error but $\hat{H}_i(x,u_i)\neq 0$ every time the control action is updated; i.e., optimality is achieved only in the limit $(t\to\infty, \hat{V}\to V^*)$. Further, in asynchronous learning [4], the optimal value function is learnt only along the state trajectory and not the entire state space. Therefore, the initial weights of the value function approximator affect the cumulative cost of operating the system. To minimize the cost during the learning period an exploration strategy using identifiers is presented next.

First, consider the identifier described by (20). We will consider two sets of initial weights, one of which will be used by the controller to generate the control action $\varpi_{ie}^{(1)}(t) = u_i^{(1)}(t) + \xi_i^{(1)}(t)$, such that $\xi_i^{(1)}(t) = 0$; the other one will be exploratory policy $\varpi_{ie}^{(2)}(t) = u_i^{(2)}(t) + \xi_i^{(2)}(t)$ with $\xi_i^{(2)}(t)\neq 0$, used with the identifier. Fig. 1 is a simplified block diagram representation for implementing the proposed exploration strategy. It can be observed that in order to incorporate exploration without affecting the performance of the existing controller, an addition identifier and value function estimator are required.

Let $\hat{\Theta}_{1i}, \hat{\Theta}_{2i}$ be the weight vectors at the i^{th} subsystem. Calculate the Hamiltonian as $\hat{H}_i^{(p)}(\hat{x}_e) = Q_i + \hat{\Theta}_{pi}^T\nabla_x\phi(\hat{x}_e)\hat{\bar{f}}_i - \frac{1}{4}\hat{\Theta}_{pi}^T\nabla_x\phi(\hat{x}_e)\hat{D}_i\nabla^T{}_x\phi(\hat{x}_e)\hat{\Theta}_{pi}$ where $p=1,2$ for each initial weights. We can construct the cost function trajectory with the value function estimator using the NN weights $\hat{\Theta}_{1i}, \hat{\Theta}_{2i}$ for both the policies $\varpi_{ie}^{(1)}, \varpi_{ie}^{(2)}$. Similar to (7), the stationarity condition can provide the $u_{i,e}$ from $\hat{H}_i^{(p)}$. Using the Hamiltonian error, the NN weights are tuned using the weight update rule (19).

Thus, we can obtain two policies, one exploitative and the other using an exploration policy. For example a random exploration policy can be used. For each initial NN weights, a cost function, control policy, Hamiltonian error and state trajectory is generated. During the learning period, using the

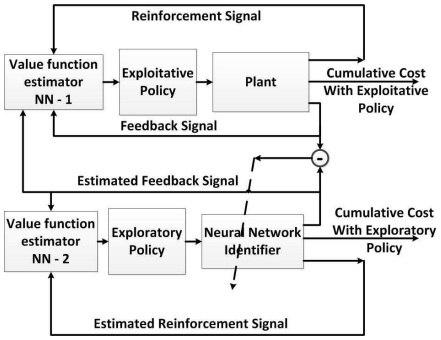

Fig. 1. Block diagram representation of exploration strategy.

performance index, the cumulative cost be calculated, for $p \in \{1, 2\}$, using the integral

$$V_i^{(p)}(t) = \int_t^{t_{switch}} \left[Q_i(x) + \varpi_{ie}^{(p)^T}(\tau) R_i \varpi_{ie}^{(p)}(\tau) \right] d\tau. \quad (21)$$

Note that the value function trajectories for the two policies start at the same initial cost and evolve based on the function $Q_i(x) + \varpi_{ie}^{(p)^T}(\tau) R_i \varpi_{ie}^{(p)}(\tau)$. Let the time instant $t = t_{switch}$ denote the time at which the difference between the cumulative rewards due to the two control policies start to increase steadily. Define $\hat{V}_i^*(\Theta) = \min\{V_i^1(t), V_i^2(t)\}$. Using the value function approximator NN that corresponds to the estimate $\hat{V}_i^*(\Theta)$ generate the greedy policy at the event based sampling instants $t_k^i \geq t_{switch},\ \forall k \in \mathbb{N}$. If both the policies result in the same cumulative cost $V_i^1(t), V_i^2(t)$, the reliability of the cost function estimate can be evaluated by using their HJB error. Choose the estimated value function $\hat{V}_i^*$ such that $\hat{\theta}_i$ satisfies the condition $\hat{\theta}_i = \min(\arg\min_{\Theta_1}(\hat{H}_i^1), \arg\min_{\Theta_2}(\hat{H}_i^2))$. Thus, $\hat{V}_i^*$ which is close to the optimal value function is used to generate the control action and potentially minimize the cost during the learning period. Note that the exploration policy need not necessarily yield a reduced cost function trajectory during the learning period. However, it is observed during the simulation analysis that the appropriate choice of exploration policy can significantly reduce the cost during the learning period.

Remark 13: In contrast to [28], the exploration strategy presented here evaluates the cumulative cost due to the two policies and by relying on the cumulative cost observed from the past experience, chooses the approximated value function learnt using the policy which resulted in lower cumulative cost.

Remark 14: The sigma/epsilon modification term $(\kappa_3 \hat{\theta}_i)$ added in the learning rule (19) ensures that the approximated value function reach a neighborhood of the optimal value function, without compromising the stability. Further, the control action ϖ_{ie} generated using the proposed learning algorithm without the exploration strategy ($\xi_i = 0$) is always exploitative as $\varpi_{ie} = u_{i,e}$, minimizing the cost function (4). Therefore, injecting exploratory signal ξ_i, to the identifier and searching for a better policy using the proposed exploration strategy is not going to affect the system performance or stability. In contrast, it can only improve the optimality of the control action. Therefore, it is a very efficient tool for online learning and control applications.

Remark 15: The learning schemes which collect data online using stabilizing controller and then use the data collected to update the value functions can also use this scheme during the initial learning period to collect sufficient data points. The advantage is that the control policy minimizes the cost function even during the learning period and sufficiently brichb data can be collected using the proposed exploration strategy [10].

Using Lyapunov based analysis, the stability results for the closed loop system is presented next.

V. STABILITY ANALYSIS

In this section, first, a more generic result which establishes the fact that the continuously updated closed-loop system admits a local input-to-state practically stable Lyapunov function in the presence of bounded external input (measurement error). This result is required to ensure that the event triggering mechanism does not exhibit zeno behavior. Further, it is shown using two cases that as the event sampling instants increase, the states, weight estimation errors and the identifier errors reach a neighborhood of origin. Using the fact [11] that the optimal controller renders the closed-loop dynamics bounded reveals

$$\|f(x) + g(x)u^*\| \leq \|\delta(x)\| = C_1 \|x\| \quad (22)$$

where $\delta(x) \in \Re^n$, $C_1 \in \Re$. This will be used in the derivations to demonstrate system stability.

Theorem 1: Consider the subsystem dynamics (1). Define the NN weight update rule (18) for the value function approximator and (15) for the identifiers. Let the assumptions 1 to 4 hold. Then, $\alpha_{iv}, \mu_i, \kappa_3 > 0$ can be designed such that $\tilde{\theta}_i, \tilde{W}_{if}, \tilde{W}_{ig}$ and $x, \tilde{x}_i$ are locally uniformly ultimately bounded when a bounded measurement error is introduced in the feedback.

Proof: See appendix.

Theorem 2: Consider the augmented nonlinear system (2) and its component subsystems (1). Define the NN weight update rule (18) for the value function approximator and (15), for the identifiers. Let events be generated based on the event-triggering condition defined by (17). Then, the control parameters can be designed such that $\tilde{\theta}_i, \tilde{W}_{if}, \tilde{W}_{ig}$ and $x, \tilde{x}_i$ are locally uniformly ultimately bounded.

The proof of Theorem 2 is a special case of Theorem 3 and therefore, all the details are not provided to avoid redundancy.

Remark 16: From the results of Theorem 1 the closed-loop system admits a Lyapunov function which satisfies the local input-to-state practical stablility (ISpS) when the measurement error is bounded. By analyzing the same Lyapunov function during the inter-event period, using the event-triggering condition, the boundedness of the measurement can be established.

Remark 17: Appropriate choice of design parameters will result in lower bounds on $x, \tilde{x}_i$ and $\tilde{\theta}_i, \tilde{W}_{if}, \tilde{W}_{ig}$. Redundant events can be avoided using a dead-zone operator [24].

Remark 18: Define the minimum time between two events as $\tau_{\min} = \min\{t_{k+1} - t_k\},\ \forall k \in \mathbb{N}$. Then $\tau_{\min} > 0$ as a result of Assumption 3, Theorems 1 and 2 [24].

Now the close-loop stability results with the modified learning algorithm and exploration is presented.

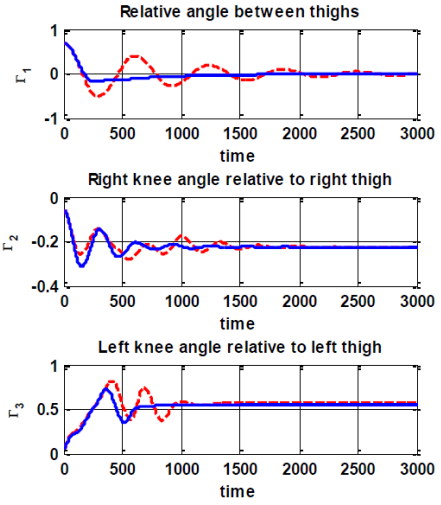

Fig. 2. State Trajectories (Dotted (red) Lines b Hybrid Algorithm vs Enhanced hybrid algorithm) (Time in 10^{-2}s).

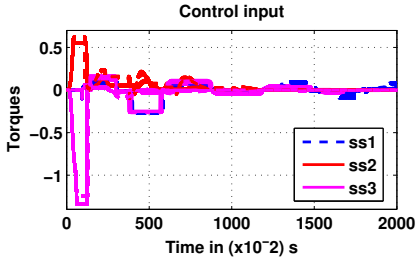

Fig. 3. Control torques.

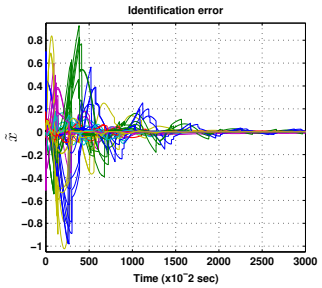

Fig. 4. Identifier approximation error.

Theorem 3: Consider the augmented nonlinear system (2) and its component subsystems (1). Define the NN weight update rule (15) for the identifiers (20). Define the event-triggering condition (17), further, let the NN weights of the value function estimator be tuned based on the rule (19). Under the assumptions prescribed in the previous sections, the control parameters can be designed such that the NN weight estimation errors $\tilde{\theta}_i, \tilde{W}_{if}, \tilde{W}_{ig}$, the interconnected system states and $\tilde{x}(t)$ are locally uniformly ultimately bounded.

Proof: See appendix.

A walking robotic system is used as the simulation example to verify the theoretical results.

VI. SIMULATION RESULTS

In this section, three coupled nonlinear subsystems are considered for application of the distributed ADP algorithms presented in this paper. The three subsystems are physically meaningful in that they capture the thigh and knee dynamics of a walking robot experiment [15]. In the following, $\gamma_1(t)$ is the relative angle between the two thighs, $\gamma_2(t)$ is the right knee angle (relative to the right thigh) and $\gamma_3(t)$ is the left knee angle (relative to left thigh). The controlled equations of motion in units of (rad/sec) are $\ddot{\gamma}_1(t) = 0.1[1 - 5.25\gamma_1^2(t)]\dot{\gamma}_1(t) - \gamma_1(t) + u_1(t)$,

$$\ddot{\gamma}_2(t) = 0.01\left[1 - p_2(\gamma_2(t) - \gamma_{2e})^2\right]\dot{\gamma}_2(t) - 4(\gamma_2(t) - \gamma_{2e}) + 0.057\gamma_1(t)\dot{\gamma}_1(t) + 0.1(\dot{\gamma}_2(t) - \dot{\gamma}_3(t)) + u_2(t),$$

$$\ddot{\gamma}_3(t) = 0.01\left[1 - p_3(\gamma_3(t) - \gamma_{3e})^2\right]\dot{\gamma}_3(t) - 4(\gamma_3(t) - \gamma_{3e}) + 0.057\gamma_1(t)\dot{\gamma}_1(t) + 0.1(\dot{\gamma}_3(t) - \dot{\gamma}_2(t)) + u_3(t)$$

where $\ddot{\gamma}_i$ correspond to the dynamics of the i^{th} subsystem (SSi). The control objective is to bring the robot to a stop in a stable manner. The parameter values $(\gamma_{2e}, \gamma_{3e}, p_2, p_3)(t)$ can be considered in the model taking on the values (-0.227,0.559,6070,192).

The control scheme proposed in this paper requires 3 NNs at every subsystem. All the NNs were designed to have two layers and formed random vector functional link networks [3]. The NN that approximated $f_i(x) + \Delta_{ij}(x)$, was designed with 25 neurons in the hidden layer. The other two NNs that approximated g_i, V_i^* were designed with 7,6 hidden layer neurons respectively. The following initial conditions were set for the simulation: $x_i(0) \in [-1, 1]$, $\hat{x}(0) = 0$, $\hat{\theta}_i(0), \hat{W}_{if}(0), \hat{W}_{ig}(0) \in [0, 1]$.

The controller parameters are: $\alpha_{i1v} = 40$, $\alpha_{i2v} = 0.03$, $\mu_i = 1.95$, $P_i = 2, \kappa_3 = 0.001$, $Q_i = 20$, $R_i = 1, A_i = 80$, $C_{if} = 0.5$, $\kappa_{if} = \kappa_{ig} = 0.0001$, $\alpha_{if} = \alpha_{ig} = 100$ and $\Gamma_i = 0.99$.

The robotic system is simulated with the torques generated using the control algorithm with hybrid and enhanced (modified) hybrid approach and exploration. It can be observed that the states reach their equilibrium point faster in the modified hybrid approach (Fig. 2). The magnitude of the control torque for the hybrid ADP based learning scheme and the enhanced hybrid approach are compared in Fig.3 using the event triggered feedback. The enhanced hybrid scheme converges faster due to the improved learning as a result of using the reinforcement signals during the inter event period for tuning the NN weights.

Convergence of the identification error ensures that the reinforcement signals used to learn optimal value function and policy are reliable. To test the analytical results for the identifier, 500 different initial conditions and exploration signals like random noise and trigonometric functions of different frequency but restricted in magnitude to 0.1 were used. The states estimation errors converged on each of these simulations as seen in Fig. 4.

The optimal value function is learnt using the consistency condition dictated by the HJB equation. A lower Hamilto-

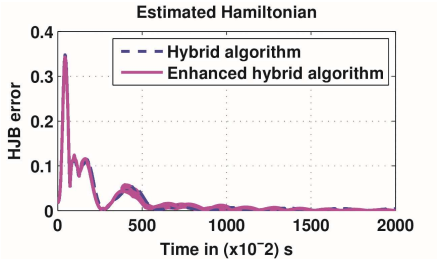

Fig. 5. Comparison of HJB error.

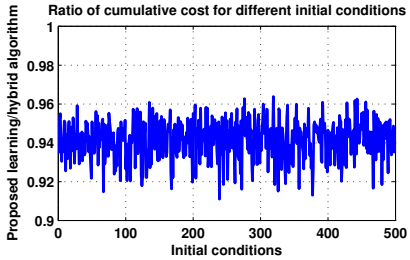

Fig. 6. Comparison of cost.

nian/HJB residual error implies that the value function weight estimate is close to the target weights. Evidently, from Fig. 5, the enhanced/modified weight tuning rule improves the optimality due to faster convergence of the HJB residual error.

To verify the proposed learning scheme, the cumulative cost calculated for the hybrid learning algorithm and the modified update rule taking into account the states and reinforcement evaluated in the inter event period are compared in Fig. 6. For 500 randomly chosen initial values of states of the system and identifier, the ratio of the cumulative cost at the end of 20s for hybrid and the proposed learning algorithm is recorded in Fig.6. Due to the dependence of the learning scheme on the identifier, the convergence of the identification errors should precede the convergence of the controller.

The improvement in the learning scheme is a result of the weights updated between events using the past data and the exploration strategy. Finally, four additional NN approximators were utilized, each initialized with the weights randomly selected in [0,2]. To demonstrate the efficiency of the proposed strategy in off-setting the effects of initial NN weights, each of the randomly picked weights were used to generate a control policy and the cost function over time using additional identifiers for each NN weights. These cost trajectories are compared with the cost function trajectory of the system with the exploration strategy presented in the paper. The variable $\Theta_1^*(t)$ is the estimated NN weights which are used to generate the control action sequence, selected by the exploration strategy, online. This seems to optimize the performance of the system better than the other policies as seen in Fig. 7.

Since multiple NNs are used to generate the cost function trajectories using the identifier states, computations are increased. However, the effect of the initial weights of the NN approximator on the cost function trajectory is reduced and the learning algorithm eventually uses the optimal approximated value function which yields the best sequence of control policy, in terms of the cost function. To test the event triggering mechanism, the sensors were sampled at 1 ms and the number of events generated are recorded. The ratio of total number of events from the 3 subsystems with the total number of sensor samples collected are computed as 0.5108 for the enhanced hybrid learning scheme and 0.4981 for the hybrid learning scheme. This demonstrates the benefits of the enhanced NN weight update rule when compared with the hybrid learning rule as almost 51% of the sensor information sampled at the event triggering mechanism is not used by the learning algorithm in the hybrid learning scheme.

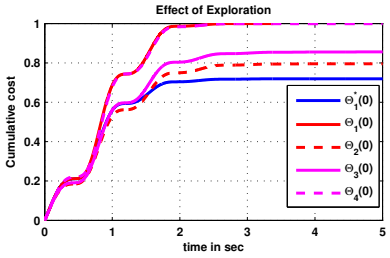

Fig. 7. Cost function trajectories.

VII. CONCLUSIONS

A novel enhanced hybrid learning scheme is introduced with exploration by using a model which in turn is utilized for the control of interconnected systems. Local UUB regulation of the system states, NN weights estimation errors and the identification errors are achieved with the proposed. The NN identifiers approximated the system nonlinearities and also aided in evaluating the exploration signals to gather useful information about the system dynamics which improved the optimality of the control actions. The proposed learning scheme seems to match and better the performance of continuous time TD ADP learning scheme with limited feedback information with some addition computations.

APPENDIX

Proof of Lemma 1: Consider the following Lyapunov candidate function $J_{iI}(\tilde{x}_i, \tilde{W}_{if}, \tilde{W}_{ig}) = J_{i\tilde{x}} + J_{i\tilde{f}} + J_{i\tilde{g}}$ with $J_{i\tilde{x}} = \frac{1}{2}\mu_{i1}\tilde{x}_i^T P_i \tilde{x}_i$, $J_{i\tilde{f}} = \frac{1}{2}\mu_{i2}\tilde{W}_{if}^T\tilde{W}_{if} + \frac{1}{4}\mu_{i4}(\tilde{W}_{if}^T\tilde{W}_{if})^2$ and $J_{i\tilde{g}} = \frac{1}{2}\mu_{i3}\tilde{W}_{ig}^T\tilde{W}_{ig} + \frac{1}{4}\mu_{i5}(\tilde{W}_{ig}^T\tilde{W}_{ig})^2 + \frac{1}{8}\mu_{i6}(\tilde{W}_{ig}^T\tilde{W}_{ig})^4$, where $\mu_{ij}, P_i, j = 1, 2, ., 6$, are positive constants of appropriate dimensions. Consider the first term in the Lyapunov function. Taking the derivative and substituting the estimation error dynamics yields

$$\dot{J}_{i\tilde{x}} = \mu_{i1}\tilde{x}_i^T P_i A_i \tilde{x}_i + \mu_{i1}\tilde{x}_i^T P_i(\tilde{W}_{if}\sigma_{if}(x_i) - \tilde{W}_{if}\tilde{\sigma}_{if} + [\tilde{W}_{ig}\sigma_{ig}(x_i) - \tilde{W}_{ig}\tilde{\sigma}_{ig}]u_{i,e} + \varepsilon_{ig}u_{i,e} + W_{ig}\tilde{\sigma}_{ig}u_{i,e} + \varepsilon_{if} + W_{if}\tilde{\sigma}_{if} + A_i e_i)$$

where e_i is the vector of event triggering errors. Applying the norm operator and choosing the design matrix A_i as Hurwitz results in

$$\dot{J}_{i\tilde{x}} \leq -\lambda_{\min}(\mu_{i1}\bar{q}_i)\|\tilde{x}_i\|^2 + \|\mu_{i1}\|\left\|\tilde{x}_i^T\right\|\|P_i\|\left\|\tilde{W}_{if}\right\| \|\sigma_{if}(\hat{x}_i)\| + \|\mu_{i1}\|\left\|\tilde{x}_i^T\right\|\|P_i\|[\left\|\tilde{W}_{ig}\right\|\|\sigma_{ig}(\hat{x}_i)\| + \|\varepsilon_{ig}\| + \|W_{ig}\|\|\sigma_{ig}(x_i) - \sigma_{ig}(\hat{x}_i)\|]\|u_{i,e}\| + \|\mu_{i1}\|\left\|\tilde{x}_i^T\right\|\|P_i\| [\|\varepsilon_{if}\| + \|W_{if}\|\|\sigma_{if}(x_i) - \sigma_{if}(\hat{x}_i)\| + \|A_i\|\|e_i\|]$$

where the solution to the Lyapunov equation for the pair (A_i, P_i), $2\bar{q}_i$ is used; λ_{min} indicates the minimum eigenvalue; $\|\sigma_{i\bullet}\| \leq N_{io\bullet}$, the number of hidden layer neurons, the subscript M with the weight variables denote the bounds on the target/ideal weights, $\|e_i\| \leq e_{iM}$ and $\varepsilon_{i\bullet M}$ is the bound on the reconstruction errors. Using the Youngs inequality $(\forall a, b, \epsilon > 0, ab \leq \frac{a^2}{2\epsilon} + \frac{\epsilon b^2}{2})$, the Lyapunov derivative becomes

$$\dot{J}_{i\tilde{x}} \leq -\lambda_{\min}(\mu_{i1}\bar{q}_i - \frac{7}{2})\|\tilde{x}_i\|^2 + \frac{1}{2}\|\mu_{i1}\|^2\|P_i\|^2\left\|\tilde{W}_{if}\right\|^2 N_{iof} + \frac{1}{2}\|\mu_{i1}\|^2\|P_i\|^2\left\|\tilde{W}_{ig}\right\|^2 N_{iog}\|u_{i,e}\|^2 + 0.5\|\mu_{i1}\|^2 \|P_i\|^2\varepsilon_{igM}^2\|u_{i,e}\|^2 + 2\|\mu_{i1}\|^2\|P_i\|^2\|W_{ig}\|^2 N_{iog}\|u_{i,e}\|^2 + 0.5\|\mu_{i1}\|^2\|P_i\|^2(\varepsilon_{ifM}^2 + 4W_{ifM}^2 N_{iof} + \|A_i\|^2 e_{iM}^2).$$

Utilizing the definition of the control policy, we have

$$\|u_{i,e}\|^2 \leq \left\|R_i^{-1}\right\|^2 W_{igM}^2 N_{iog}\left\|\nabla_x^T\Phi(x_{i,e})\theta_i^*\right\|^2 + \left\|R_i^{-1}\right\|^2 N_{iog}\left\|\tilde{W}_{ig}\right\|^2\left\|\nabla_x^T\Phi(x_{i,e})\theta_i^*\right\|^2 + \left\|R_i^{-1}\right\|^2 N_{iog}\left\|\tilde{W}_{ig}\right\|^2 \left\|\nabla_x^T\Phi(x_{i,e})\tilde{\theta}_i\right\|^2 + \left\|R_i^{-1}\right\|^2 W_{igM}^2 N_{iog}\left\|\nabla_x^T\Phi(x_{i,e})\tilde{\theta}_i\right\|^2.$$

Using this in the Lyapunov function derivative and expanding the terms and applying Youngbs inequality, the first derivative is simplified as

$$\begin{aligned}\dot{J}_{i\tilde{x}} \leq & -\lambda_{\min}(\mu_{i1}\bar{q}_i - 3.5)\|\tilde{x}_i\|^2 + 0.5\left\|\tilde{W}_{if}\right\|^4 \\ & + 1.5\left\|\tilde{W}_{ig}\right\|^4 + \eta_{io\tilde{x}B} + (\frac{1}{2}\left\|R_i^{-1}\right\|^4 + 4)\left\|\tilde{W}_{ig}\right\|^8 \\ & + (\frac{11}{2} + \frac{1}{8}\|\mu_{i1}\|^4\|P_i\|^4 N_{iog}^4)\left\|\nabla_x^T\Phi(x_{i,e})\tilde{\theta}_i\right\|^4\end{aligned} \tag{23}$$

where the bounds on the gradient of the optimal value function, V_{ixM}^*, is used to define the term $\eta_{io\tilde{x}B}$ as

$$\eta_{io\tilde{x}B} = 0.125\|\mu_{i1}\|^4\|P_i\|^4\left\|R_i^{-1}\right\|^4 W_{igM}^4 N_{iog}^2 V_{ixM}^{*^4}(\frac{\varepsilon_{igM}^4}{V_{ixM}^{*^4}} + 16N_{iog}^2 + 4N_{iog}^2 W_{igM}^2/V_{ixM}^{*^4} + N_{iog}^2 + \varepsilon_{igM}^4/W_{igM}^4 + N_{iog}^2/W_{igM}^4) + 0.125\|\mu_{i1}\|^4\|P_i\|^4 N_{iof}^2 + 0.125\|\mu_{i1}\|^8 \|P_i\|^8\left\|R_i^{-1}\right\|^8 W_{igM}^8 N_{iog}^4(N_{iog}^2 + 0.0625N_{iog}^2 + \frac{\varepsilon_{igM}^8}{16W_{igM}^8}) + 0.5\|\mu_{i1}\|^2\|P_i\|^2(\varepsilon_{igM}^2\left\|R_i^{-1}\right\|^2 W_{igM}^2 N_{iog} V_{ixM}^{*^2} + 4W_{igM}^4 N_{iog}^2\left\|R_i^{-1}\right\|^2 V_{ixM}^{*^2} + \varepsilon_{ifM}^2 + 4W_{ifM}^2 N_{iof} + \|A_i\|^2 e_{iM}^2).$$

Now consider the second term in the Lyapunov candidate function. Taking the derivative and using the weight estimation error dynamics reveals

$$\dot{J}_{i\tilde{f}} = -\frac{\mu_{i2}\tilde{W}_{if}^T\alpha_{if}\sigma_{if}\tilde{x}_{i,e}^T}{c_{if} + \tilde{x}_{i,e}^T\tilde{x}_{i,e}} + \mu_{i2}\tilde{W}_{if}^T\kappa_{if}\hat{W}_{if} - \frac{\mu_{i4}(\tilde{W}_{if}^T\tilde{W}_{if})\tilde{W}_{if}^T\alpha_{if}\sigma_{if}\tilde{x}_{i,e}^T}{c_{if} + \tilde{x}_{i,e}^T\tilde{x}_{i,e}} + \mu_{i4}(\tilde{W}_{if}^T\tilde{W}_{if})\tilde{W}_{if}^T\kappa_{if}\hat{W}_{if}$$

Using the fact that $a/(1 + a^T a) \leq 1,\ \forall a \in \Re$ and the Youngs inequality, we get

$$\begin{aligned}\dot{J}_{i\tilde{f}} \leq & -(\lambda_{\min}(\mu_{i2}\kappa_{if}) - 1)\left\|\tilde{W}_{if}\right\|^2 \\ & -(\lambda_{\min}(\mu_{i4}\kappa_{if}) - 2)\left\|\tilde{W}_{if}\right\|^4 + \eta_{iofB}\end{aligned} \tag{24}$$

where the bounded term is given by

$$\eta_{iofB} = 0.5\|\mu_{i2}\|^2\|\alpha_{if}\|^2(N_{iof}^2 + W_{ifM}^2\|\kappa_{if}\|^2/\|\alpha_{if}\|^2) + 0.125\|\mu_{i4}\|^4\|\alpha_{if}\|^4(N_{iof}^4 + W_{ifM}^4\|\kappa_{if}\|^4/\|\alpha_{if}\|^4).$$

Finally, consider the last term in the Lyapunov candidate function. Taking the derivative and substituting the weight estimation error dynamics yields

$$\begin{aligned}\dot{J}_{i\tilde{g}} = & \mu_{i3}\tilde{W}_{ig}^T(-(\alpha_{ig}\sigma_{ig}u_{i,e}\tilde{x}_{i,e}^T/\hat{\rho}) + \kappa_{ig}\hat{W}_{ig}) \\ & + \mu_{i5}(\tilde{W}_{ig}^T\tilde{W}_{ig})(\tilde{W}_{ig}^T(-(\alpha_{ig}\sigma_{ig}u_{i,e}\tilde{x}_{i,e}^T/\hat{\rho}) + \kappa_{ig}\hat{W}_{ig})) \\ & + \mu_{i6}(\tilde{W}_{ig}^T\tilde{W}_{ig})^3(\tilde{W}_{ig}^T(-(\alpha_{ig}\sigma_{ig}u_{i,e}\tilde{x}_{i,e}^T/\hat{\rho}) + \kappa_{ig}\hat{W}_{ig}))\end{aligned}$$

Similar to the simplification procedure above, we get

$$\begin{aligned}\dot{J}_{i\tilde{g}} \leq & -\lambda_{\min}(\mu_{i3}\kappa_{ig} - 1)\left\|\tilde{W}_{ig}\right\|^2 - \lambda_{\min}(\mu_{i5}\kappa_{ig} - 2)\left\|\tilde{W}_{ig}\right\|^4 \\ & - \lambda_{\min}(\mu_{i6}\kappa_{ig} - 3)\left\|\tilde{W}_{ig}\right\|^8 + \eta_{iogB}\end{aligned} \tag{25}$$

where $\hat{\rho} = c_{if} + \tilde{x}_{i,e}^T\tilde{x}_{i,e}u_{i,e}^T u_{i,e}$ and the bounded term

$$\eta_{iogB} = 0.0078\mu_{i6}^8\alpha_{ig}^8(N_{iog}^4 + W_{igM}^8\kappa_{ig}^8/\alpha_{ig}^8) + 0.5\alpha_{ig}^2\mu_{i3}^2 (N_{iog} + W_{igM}^2\kappa_{ig}^2/\alpha_{ig}^2) + 0.125\mu_{i5}^4\alpha_{ig}^4(N_{iog}^2 + W_{igM}^4\frac{\kappa_{ig}^4}{\alpha_{ig}^4}).$$

The first derivative of the Lyapunov function is obtained as

$$\begin{aligned}\dot{J}_{iI} \leq & -\lambda_{\min}(\mu_{i1}\bar{q}_i - \frac{7}{2})\|\tilde{x}_i\|^2 - \lambda_{\min}(\mu_{i3}\kappa_{ig} - 1)\left\|\tilde{W}_{ig}\right\|^2 \\ & + \eta_{ioB} - \lambda_{\min}(\mu_{i5}\kappa_{ig} - \frac{7}{2})\left\|\tilde{W}_{ig}\right\|^4 - (\lambda_{\min}(\mu_{i2}\kappa_{if}) - 1) \\ & \left\|\tilde{W}_{if}\right\|^2 - \lambda_{\min}(\mu_{i6}\kappa_{ig} - 3 - (0.5\left\|R_i^{-1}\right\|^4 + 4))\left\|\tilde{W}_{ig}\right\|^8 \\ & - (\lambda_{\min}(\mu_{i4}\kappa_{if}) - 2.5)\left\|\tilde{W}_{if}\right\|^4 \\ & + (5.5 + 0.125\|\mu_{i1}\|^4\|P_i\|^4 N_{iog}^4)\left\|\nabla_x^T\Phi(x_{i,e})\tilde{\theta}_i\right\|^4.\end{aligned}$$

Since the control policy is optimal, $\tilde{\theta}_i = 0$ in the control policy and the final Lyapunov derivative expression reveals

$$\begin{aligned}\dot{J}_{iI} \leq & -\lambda_{\min}(\mu_{i1}\bar{q}_i - 3.5)\|\tilde{x}_i\|^2 - \lambda_{\min}(\mu_{i3}\kappa_{ig} - 1)\left\|\tilde{W}_{ig}\right\|^2 \\ & - \lambda_{\min}(\mu_{i5}\kappa_{ig} - 3.5)\left\|\tilde{W}_{ig}\right\|^4 - (\lambda_{\min}(\mu_{i2}\kappa_{if}) - 1)\left\|\tilde{W}_{if}\right\|^2 \\ & - (\lambda_{\min}(\mu_{i4}\kappa_{if}) - 2.5)\left\|\tilde{W}_{if}\right\|^4 \\ & - \lambda_{\min}(\mu_{i6}\kappa_{ig} - 3 - (0.5\left\|R_i^{-1}\right\|^4 + 4))\left\|\tilde{W}_{ig}\right\|^8 + \eta_{ioB}\end{aligned}$$

where the bounds are defined as $\eta_{ioB} = \eta_{iogB}+\eta_{iofB}+\eta_{io\tilde{x}B}$. This reveals that the identification and weight estimation errors of the identifiers at each subsystem are locally UUB.

Proof of Theorem 1 (local ISS): Consider the Lyapunov function for the interconnected system $J = \sum_{i=1}^{N} J_i(x_i, \tilde{\theta}_i, \tilde{X}_i, \tilde{W}_{if}, \tilde{W}_{ig})$, with the individual terms defined as $J_i = J_{ix} + J_{i\tilde{\theta}} + J_{iI}(\tilde{X}_i, \tilde{W}_{if}, \tilde{W}_{ig})$, $J_{ix} = 0.5\alpha_{iv}x_i^T x_i$, $J_{i\tilde{\theta}} = 0.5\tilde{\theta}_i^T\gamma_i\tilde{\theta}_i$. The derivative of $J_{i\tilde{\theta}}$ can be obtain using the weight estimation error dynamics as

$$\dot{J}_{i\tilde{\theta}} = (\tilde{\theta}_i^T\gamma_i\alpha_{iv}\hat{\psi}_{i,e}\hat{H}_{i,e}/\bar{\rho}^2) + \tilde{\theta}_i^T\gamma_i\kappa_3\hat{\theta}_i - 0.5(\beta_{iv}\tilde{\theta}_i^T\gamma_i\nabla_x\phi\hat{D}_{i,\varepsilon}x_{i,e} + \mu_{i1}\tilde{\theta}_i^T\gamma_i\nabla_x\phi(x_e)\hat{D}_{i,\varepsilon}^T P_i\tilde{x}_{i,e}).$$

This derivation follows the derivation in [11]. In [11], the derivations do not include the identification error and the event triggering error, these additional terms due to event triggering and the identifiers are grouped as A_1, B_1 in the next step and are simplified. Substituting the expression for $\hat{H}_{i,e}$, $\hat{\psi}_{i,e}$ and simplification of terms reveals that the first term $(\tilde{\theta}_i^T\gamma_i\alpha_{iv}\hat{\psi}_{i,e}\hat{H}_{i,e})/\bar{\rho}^2$ is

$$\dot{J}_{i\tilde{\theta}} \leq \begin{array}{l} (\frac{1}{2}\gamma_i\left\|\tilde{\theta}_i^T\nabla_x\phi(x_e)\right\|^2 + \frac{\gamma_i}{\bar{\rho}^2}(4\alpha_{iv}^4 + \frac{1}{16} + 2\varepsilon_{iM}^4)\|\dot{x}_i^*\|^4) \\ \frac{\gamma_i}{32\bar{\rho}^2}(-\upsilon_{\tilde{\theta}}\left\|\tilde{\theta}_i^T\nabla_x\phi(x_e)\right\|^4 + (\alpha_{iv}^2+2)16\|B_1\|^2) \\ + 10\gamma_i\alpha_{iv}^2\left\|\left(\tilde{\theta}_i^T A_1(\tilde{f},\tilde{g},x)\right)\right\|^2/8\bar{\rho}^2 + B_{i\tilde{\theta}}/\bar{\rho}^2 \end{array}$$

where $\bar{\rho} = 1+\hat{\psi}_{i,e}^T\hat{\psi}_{i,e}$, $\upsilon_{\tilde{\theta}} = 4\alpha_{iv}\left\|\hat{D}_{i,\varepsilon,\min}\right\|^2 - 17\left\|\hat{D}_{i,\varepsilon}\right\|^2 - 0.625$, $B_{i\tilde{\theta}} = +\alpha_{iv}^2D_{iM}^2\varepsilon_M^2 + 2\alpha_{iv}^4D_{iM}^4\varepsilon_M^4 + 2\varepsilon_{iM}^8D_{i\max}^4 + (0.5\alpha_{iv}^2+2)2\varepsilon_{iM}^4D_{i\max}^2 + 32(0.5\alpha_{iv}^2+2)^2\varepsilon_{iM}^2 + 64\alpha_{iv}^4 + 40\gamma_i\alpha_{iv}^2L_{if}^4E_i^4 + 4.1\gamma_i\alpha_{iv}^2D_{iM}^4V_{ixM}^4 + 172\gamma_i\alpha_{iv}^2V_{ixM}^4D_{iM}^4 + 4713\gamma_i\alpha_{iv}^2V_{ixM}^8D_{iM}^4\left\|R_i^{-1}\right\|^4$.

Now, simplifying the terms B_1, A_1

$$\|B_1\|^2 \leq 4.5\left\|\tilde{\theta}_i^T\nabla_x\phi(x_e)\right\|^4 + 20\|\tilde{g}_i^T(x_e)\|^8 + 8.5\left\|\tilde{\bar{f}}_i(x_e)\right\|^4 + 25.6\|\tilde{g}_i^T(x_e)\|^8V_{ixM}^4\left\|R^{-1}\right\|^4 + 0.75\|\tilde{g}_i^T(x_e)\|^4 + B_{i\tilde{\theta}1} \quad (26)$$

$$1.25\left\|\left(\tilde{\theta}_i^T A_1\right)\right\|^2 \leq 0.47\left\|\tilde{\theta}_i^T\nabla_x\phi(x_e)\right\|^4 + 20\left\|\tilde{\bar{f}}_i(x_e)\right\|^4 + 12.5\|\tilde{g}_{i,e}^T\|^8 + 18V_{ixM}^4\left\|R_i^{-1}\right\|^4\|\tilde{g}_i(x_e)\|^8 + 40L_{if}^4E_i^4 + 177V_{ixM}^4D_{iM}^4 + 4713V_{ixM}^8D_{iM}^4\left\|R_i^{-1}\right\|^4. \quad (27)$$

Collecting the bounded terms from (27) and (26), we have $B_{i\tilde{\theta}1} = 256V_{ixM}^8\left\|R^{-1}\right\|^4 + 16D_{iM}^2V_{ixM}^4 + 512V_{ixM}^8D_{iM}^2\left\|R^{-1}\right\|^2 + 16L_{iq}^2E_i^2 + 16D_{iM}^4V_{ixM}^4 + 32L_{if}^4E_i^4 + 16V_{ixM}^2L_{if}^2 + 32V_{ixM}^4 + 256D_{iM}^4V_{ixM}^4 + 6553.6D_{iM}^4V_{ixM}^8\left\|R^{-1}\right\|^4 + 4D_{iM}^2V_{ixM}^4$. Using the relation $\tilde{\theta}_i^T\gamma_i\alpha_{iv}\hat{\psi}_{i,e}\hat{H}_{i,e}/\bar{\rho}^2$, (26) and (27), we get

$$\dot{J}_{i\tilde{\theta}} \leq \left(\begin{array}{l} -(\gamma_i\kappa_3 - \frac{1}{2} - \frac{\gamma_i}{2\bar{\rho}^2}\nabla_x\phi_{\min}^2)\left\|\tilde{\theta}_i\right\|^2 + \frac{\gamma_i}{\bar{\rho}^2}(\frac{8\alpha_{iv}^4}{2} + \\ \frac{2}{32} + 2\varepsilon_{iM}^4)\|\dot{x}_i^*\|^4 - (\frac{\upsilon_{\tilde{\theta}}}{32} - \frac{\alpha_{iv}^2}{2} - \frac{9}{4}(\alpha_{iv}^2+2)) \\ \left\|\tilde{\theta}_i^T\nabla_x\phi(x_e)\right\|^4\frac{\gamma_i}{\bar{\rho}^2} + \frac{\gamma_i}{\bar{\rho}^2}(10(\alpha_{iv}^2+2) + \frac{5}{4}\alpha_{iv}^2) \\ \|\tilde{g}_{i,e}^T\|^8 + \frac{3}{8\bar{\rho}^2}\gamma_i(\alpha_{iv}^2+2)\|\tilde{g}_{i,e}^T\|^4 + (\frac{17}{4}(\alpha_{iv}^2+2) \\ + 20\alpha_{iv}^2)\left\|\tilde{\bar{f}}_i(x_e)\right\|^4\gamma_i/\bar{\rho}^2 + (12.8\gamma_i(\alpha_{iv}^2+2) \\ + 18\gamma_i\alpha_{iv}^2)V_{ixM}^4\left\|R_i^{-1}\right\|^4\|\tilde{g}_i(x_e)\|^8/\bar{\rho}^2 + \frac{B_{i\tilde{\theta}}}{\bar{\rho}^2} \\ - \frac{\mu_{i1}\gamma_i}{2}\tilde{\theta}_i^T\nabla_x\phi(x_e)\hat{D}_{i,\varepsilon}^T P_i\tilde{x}_{i,e} \\ - \frac{\beta_{iv}\gamma_i}{2}\tilde{\theta}_i^T\nabla_x\phi\hat{D}_{i,\varepsilon}x_{i,e}. \end{array}\right)$$

Consider the first term of the Lyapunov function of the i^{th} subsystem, taking its derivative and substituting the subsystem dynamics, we get

$$\begin{array}{l} \dot{J}_{ix} = \alpha_{iv}x_i^T[\bar{f}_i(x_i) + g_i(x_i)u_{i,e}]. \\ \leq -(\bar{\bar{q}}\alpha_{iv})\|x_i\|^2 + 0.125(5\left\|x_i^T\right\|^2 + 4N_{iog}^4\left\|\tilde{W}_{ig}^T\right\|^8 \\ + 2\left\|\nabla_x^T\phi(x_e)\tilde{\theta}_i\right\|^4) + \frac{1}{8}(N_{iog}^2\left\|\tilde{W}_{ig}^T\right\|^4 + \alpha_{iv}^8D_{iM}^4\left\|R_i{}^{-1}\right\|^4) \\ + .5(\alpha_{iv}^2D_{iM}^2\varepsilon_M^2 + \alpha_{iv}^2D_{iM}^2V_{ixM}^2) \\ + 0.5(\alpha_{iv}^4D_{iM}^4 + \alpha_{iv}^2V_{ixM}^2D_{iM}^2 + \alpha_{iv}^4V_{ixM}^4D_{iM}^2\left\|R_i{}^{-1}\right\|^2). \end{array}$$

In order to combine the Lyapunov derivative of the online value function estimator and identifiers (Lemma 1). Define

$$\begin{array}{l} B_{i\tilde{\theta}} = \alpha_{iv}^2D_{iM}^2\varepsilon_M^2 + 2\alpha_{iv}^4D_{iM}^4\varepsilon_M^4 + 2\varepsilon_{iM}^8D_{i\max}^4 + \\ (0.5\alpha_{iv}^2+2)2\varepsilon_{iM}^4D_{i\max}^2 + 32(\frac{\alpha_{iv}^2}{2}+2)^2\varepsilon_{iM}^2 + 64\alpha_{iv}^4 + \\ 40\gamma_i\alpha_{iv}^2L_{if}^4E_i^4 + 4.1\gamma_i\alpha_{iv}^2D_{iM}^4V_{ixM}^4 + 172\gamma_i\alpha_{iv}^2V_{ixM}^4D_{iM}^4 \\ + 4713\gamma_i\alpha_{iv}^2V_{ixM}^8D_{iM}^4\left\|R_i^{-1}\right\|^4 + .5\gamma_i^2\kappa_3^2\theta_{iM}^2 + 0.5\gamma_i \\ (\alpha_{iv}^2+2)B_{i\tilde{\theta}1} + \bar{\rho}^2(\eta_{ioB} + \gamma_i^4 + 2\|P_ie_i\|^2 + 0.25\|e_i\|^2), \\ \upsilon_{\tilde{f}} = \lambda_{\min}(\mu_{i4}\kappa_{if}) - \frac{5}{2} - \frac{4\gamma_i}{\bar{\rho}^2}(6.07\alpha_{iv}^2 + 2.14)N_{iof}^2, \\ \upsilon_{\tilde{g}^2} = 3 - \frac{1}{2}\left\|R^{-1}\right\|^4 + 4 - \frac{18\gamma_i}{\bar{\rho}^2}(\frac{71.2}{100}(\alpha_{iv}^2+2) + \alpha_{iv}^2) \\ N_{iog}^4(1 + \left\|R_i^{-1}\right\|^4V_{ixM}^4), \\ \upsilon_x = (\bar{\bar{q}}\alpha_{iv} - 0.875 - \gamma_i(4\alpha_{iv}^4 + 0.0625 + 2\varepsilon_{iM}^4)C_i^4/\bar{\rho}^2), \\ \eta_{icl} = \frac{B_{i\tilde{\theta}}}{\bar{\rho}^2} + \frac{1}{8}\alpha_{iv}^8D_{iM}^4\left\|R_i{}^{-1}\right\|^4 + \frac{1}{2}\alpha_{iv}^2D_{iM}^2\varepsilon_M^2 \\ + \frac{1}{2}\alpha_{iv}^2D_{iM}^2V_{ixM}^2 + 0.5(\alpha_{iv}^4D_{iM}^4 + \alpha_{iv}^2V_{ixM}^2D_{iM}^2 \\ + \alpha_{iv}^4V_{ixM}^4D_{iM}^2\left\|R_i{}^{-1}\right\|^2), \upsilon_{\tilde{\theta}2} = \frac{\upsilon_{\tilde{\theta}}\gamma_i}{32} - \frac{\gamma_i\alpha_{iv}^2}{2} \\ - 2.25\gamma_i(\alpha_{iv}^2+2) - N\bar{\rho}^2(5.5 + 0.225 \\ \|\mu_{i1}\|^4\|P_i\|^4N_{iog}^4) - \frac{1}{128}\mu_{i1}^4\bar{\rho}^2\left\|\hat{D}_{i,\varepsilon}^T\right\|^4 - \frac{1}{8}\beta_{iv}^4\bar{\rho}^2\left\|\hat{D}_{i,\varepsilon}\right\|^4, \\ \upsilon_{\tilde{g}} = \lambda_{\min}(\mu_{i5}\kappa_{ig}) - 3.5 - 3\gamma_i(\alpha_{iv}^2+2)N_{iog}^2/8\bar{\rho}^2. \end{array}$$

Using the results of Lemma 1 and combining $\dot{J}_{ix}, \dot{J}_{i\tilde{\theta}}$ reveals

$$\dot{J}_i \leq \begin{pmatrix} -(\lambda_{\min}(\mu_{i1}\bar{q}_i) - \frac{7}{2} - \frac{2\|P_i\|^2}{N})\left\|\tilde{X}_i\right\|^2 - (\gamma_i\kappa_3 \\ -0.5 - \frac{\gamma_i\nabla_x\phi_{\min}^2}{2\bar{\rho}^2})\left\|\tilde{\theta}_i\right\|^2 - (\upsilon_{\tilde{g}} - \frac{N_{iog}^2}{8})\left\|\tilde{W}_{ig}\right\|^4 \\ -(\frac{\upsilon_{\tilde{\theta}2}}{\bar{\rho}^2} - \frac{1}{4})\nabla_x\phi_{\min}^4\left\|\tilde{\theta}_i^T\right\|^4 \\ -(\lambda_{\min}(\mu_{i2}\kappa_{if}) - 1)\left\|\tilde{W}_{if}\right\|^2 - \upsilon_{\tilde{f}}\left\|\tilde{W}_{if}\right\|^4 \\ -(\lambda_{\min}(\mu_{i6}\kappa_{ig}) - \upsilon_{\tilde{g}^2} - \frac{N_{iog}^4}{2})\left\|\tilde{W}_{ig}\right\|^8 \\ -(\lambda_{\min}(\mu_{i3}\kappa_{ig}) - 1)\left\|\tilde{W}_{ig}\right\|^2 - \upsilon_x\|x_i\|^2 + \eta_{icl} \end{pmatrix}$$

where the derivative $\dot{J}_i$ is negative definite as long as $\|x_i\| > \sqrt{\eta_{icl}/\upsilon_x} \equiv \eta_{iX1}$ or $\left\|\tilde{X}_i\right\| > \sqrt{\eta_{icl}/(\lambda_{\min}(\mu_{i1}\bar{q}_i) - 3.5 - 2\|P_i\|^2/N)}$ or $\left\|\tilde{W}_{if}\right\| > \sqrt[4]{\eta_{icl}/\upsilon_{\tilde{f}}}$ or $\left\|\tilde{\theta}_i^T\right\| > \sqrt[4]{\frac{\eta_{icl}}{(\upsilon_{\tilde{\theta}2}/\bar{\rho}^2 - 0.25)\nabla_x\phi_{\min}^4}} \equiv \eta_{i\tilde{\Theta}1}$, or $\left\|\tilde{W}_{ig}\right\| > \sqrt[8]{\frac{\eta_{icl}}{(\lambda_{\min}(\mu_{i6}\kappa_{ig}) - \upsilon_{\tilde{g}^2} - 0.5N_{iog}^4)}}$. The overall bounds are obtained as $\eta_{X1} = \bigcup_{i=1}^{N}\eta_{iX1}$ and $\eta_{\Theta 1} = \bigcup_{i=1}^{N}\eta_{i\tilde{\Theta}1}$. This concludes the proof.

Proof of Theorem 3: Consider the Lyapunov candidate function $J(x, \tilde{\Theta}, \tilde{X}, \tilde{W}) = \sum_{i=1}^{N} J_i(x_i, \tilde{\theta}_i, \tilde{X}_i, \tilde{W}_{if}, \tilde{W}_{ig})$, with $J_i(x_i, \tilde{\theta}_i, \tilde{X}_i, \tilde{W}_{if}, \tilde{W}_{ig}) = J_{ix} + J_{i\tilde{\theta}} + J_{iI}(\tilde{X}_i, \tilde{W}_{if}, \tilde{W}_{ig})$. We will consider two cases corresponding to measurement error being zero and non-zero.

Case 1: Consider the Lyapunov function term for the identifier $J_{iI}(\tilde{X}_i, \tilde{W}_{if}, \tilde{W}_{ig})$. From Lemma 1, we have

$$\dot{J}_{i\tilde{x}} \leq -\lambda_{\min}(\mu_{i1}\bar{q}_i - 3.5)\left\|\tilde{X}_i\right\|^2 + .5\left\|\tilde{W}_{if}\right\|^4 + 1.5\left\|\tilde{W}_{ig}\right\|^4 + (.5\left\|R^{-1}\right\|^4 + 4)\left\|\tilde{W}_{ig}\right\|^8 + (5.5 + .125\|\mu_{i1}\|^4\|P_i\|^4 N_{iog}^4)\left\|\nabla_x^T\Phi(X_i)\tilde{\Theta}_i\right\|^4 + \eta_{io\tilde{x}B}$$

$$\eta_{io\tilde{x}B} = 0.125\|\mu_{i1}\|^4\|P_i\|^4\left\|R^{-1}\right\|^4 W_{igM}^4 N_{iog}^2 V_{ixM}^{*4} (\varepsilon_{igM}^4/V_{ixM}^{*4} + 16N_{iog}^2 + 4N_{iog}^2 W_{igM}^2/V_{ixM}^{*4} + N_{iog}^2 + \varepsilon_{igM}^4/W_{igM}^4 + N_{iog}^2/W_{igM}^4) + 0.125\|\mu_{i1}\|^4\|P_i\|^4 N_{iof}^2 + 0.125\|\mu_{i1}\|^8\|P_i\|^8\left\|R^{-1}\right\|^8 W_{igM}^8 N_{iog}^4 (N_{iog}^2 + 0.0625N_{iog}^2 + 0.0625\varepsilon_{igM}^8/W_{igM}^8) + 0.5\|\mu_{i1}\|^2\|P_i\|^2(\varepsilon_{igM}^2\left\|R^{-1}\right\|^2 W_{igM}^2 N_{iog} V_{ixM}^{*2} + 4W_{igM}^4 N_{iog}^2\left\|R^{-1}\right\|^2 V_{ixM}^{*2} + \varepsilon_{ifM}^2 + 4W_{ifM}^2 N_{iof}).$$

Now consider the second term in the Lyapunov candidate function. Taking the derivative and using the weight estimation error dynamics reveals

$$\dot{J}_{i\tilde{f}} \leq -(\lambda_{\min}(\mu_{i2}\kappa_{if}) - 1)\left\|\tilde{W}_{if}\right\|^2 - (\lambda_{\min}(\mu_{i4}\kappa_{if}) - 2)\left\|\tilde{W}_{if}\right\|^4 + \eta_{iofB}$$

$$\eta_{iofB} = 0.5\|\mu_{i2}\|^2\|\alpha_{if}\|^2(N_{iof}^2 + W_{ifM}^2\|\kappa_{if}\|^2/\|\alpha_{if}\|^2) + 0.125\|\mu_{i4}\|^4\|\alpha_{if}\|^4(N_{iof}^4 + W_{ifM}^4\|\kappa_{if}\|^4/\|\alpha_{if}\|^4).$$

Finally, consider the last term in the Lyapunov candidate function. Taking the derivative and substituting the weight estimation error dynamics yields

$$\dot{J}_{i\tilde{g}} \leq -\lambda_{\min}(\mu_{i3}\kappa_{ig} - 1)\left\|\tilde{W}_{ig}\right\|^2 - \lambda_{\min}(\mu_{i5}\kappa_{ig} - 2)\left\|\tilde{W}_{ig}\right\|^4 - \lambda_{\min}(\mu_{i6}\kappa_{ig} - 3)\left\|\tilde{W}_{ig}\right\|^8 + \eta_{iogB}$$

$$\eta_{iogB} = \frac{1}{125}\mu_{i6}^8\alpha_{ig}^8(N_{iog}^4 + W_{igM}^8\kappa_{ig}^8/\alpha_{ig}^8) + \frac{1}{2}\alpha_{ig}^2\mu_{i3}^2(N_{iog} + W_{igM}^2\kappa_{ig}^2/\alpha_{ig}^2) + 0.125\mu_{i5}^4\alpha_{ig}^4(N_{iog}^2 + W_{igM}^4\kappa_{ig}^4/\alpha_{ig}^4).$$

The first derivative of $J_{iI}(\tilde{X}_i, \tilde{W}_{if}, \tilde{W}_{ig})$ is thus obtained as

$$\dot{J}_{iI} \leq -\lambda_{\min}(\mu_{i1}\bar{q}_i - \frac{7}{2})\left\|\tilde{X}_i\right\|^2 - \lambda_{\min}(\mu_{i3}\kappa_{ig} - 1)\left\|\tilde{W}_{ig}\right\|^2 - \lambda_{\min}(\mu_{i5}\kappa_{ig} - \frac{7}{2})\left\|\tilde{W}_{ig}\right\|^4 - (\lambda_{\min}(\mu_{i2}\kappa_{if}) - 1)\left\|\tilde{W}_{if}\right\|^2 - (\lambda_{\min}(\mu_{i4}\kappa_{if}) - 2.5)\left\|\tilde{W}_{if}\right\|^4 - \lambda_{\min}(\mu_{i6}\kappa_{ig} - 3 - (0.5\left\|R^{-1}\right\|^4 + 4))\left\|\tilde{W}_{ig}\right\|^8 + (5.5 + 0.125\|\mu_{i1}\|^4\|P_i\|^4 N_{iog}^4)\left\|\nabla_x^T\Phi(X_i)\tilde{\Theta}_i\right\|^4 + \eta_{ioB}.$$

Now, combining the Lyapunov derivative of the online value function estimator $\dot{J}_{i\tilde{\theta}}$ from the previous theorem and identifiers, we get

$$\dot{J}_{i\tilde{\theta}} + \dot{J}_{iI} \leq \begin{pmatrix} -(\lambda_{\min}(\mu_{i1}\bar{q}_i) - 3.5 - 2\|P_i\|^2/N)\left\|\tilde{X}_i\right\|^2 \\ -(\gamma_i\kappa_3 - 0.5 - \gamma_i\nabla_x\phi_{\min}^2/2\bar{\rho}^2)\left\|\tilde{\theta}_i\right\|^2 \\ +(\gamma_i(4\alpha_{iv}^4 + \frac{1}{16} + 2\varepsilon_{iM}^4)\frac{C_i^4}{\bar{\rho}^2} + \frac{1}{4})\|x_i\|^2 \\ -\upsilon_{\tilde{\theta}2}\nabla_x\phi_{\min}^4\left\|\tilde{\theta}_i^T\right\|^4/\bar{\rho}^2 - \upsilon_{\tilde{g}}\left\|\tilde{W}_{ig}\right\|^4 \\ -((\mu_{i2}\kappa_{if}) - 1)\left\|\tilde{W}_{if}\right\|^2 - \upsilon_{\tilde{f}}\left\|\tilde{W}_{if}\right\|^4 \\ -(\lambda_{\min}(\mu_{i6}\kappa_{ig}) - \upsilon_{\tilde{g}^2})\left\|\tilde{W}_{ig}\right\|^8 \\ -(\lambda_{\min}(\mu_{i3}\kappa_{ig}) - 1)\left\|\tilde{W}_{ig}\right\|^2 + B_{i\tilde{\theta}}/\bar{\rho}^2 \end{pmatrix}$$

where the terms $\upsilon_x = (\bar{\bar{q}}\alpha_{iv} - 0.875 - \gamma_i(4\alpha_{iv}^4 + 0.0625 + 2\varepsilon_{iM}^4)C_i^4/\bar{\rho}^2)$, $B_{i\tilde{\theta}} = \alpha_{iv}^2 D_{iM}^2\varepsilon_M^2 + 2\alpha_{iv}^4 D_{iM}^4\varepsilon_M^4 + 2\varepsilon_{iM}^8 D_{i\max}^4 + (0.5\alpha_{iv}^2 + 2)2\varepsilon_{iM}^4 D_{i\max}^2 + 32(0.5\alpha_{iv}^2 + 2)^2\varepsilon_{iM}^2 + 64\alpha_{iv}^4 + 4.1\gamma_i\alpha_{iv}^2 D_{iM}^4 V_{ixM}^4 + 172\gamma_i\alpha_{iv}^2 V_{ixM}^4 D_{iM}^4 + 4713\gamma_i\alpha_{iv}^2 V_{ixM}^8 D_{iM}^4\left\|R_i^{-1}\right\|^4 + 0.5\gamma_i(\alpha_{iv}^2 + 2)B_{i\tilde{\theta}1} + 0.5\gamma_i^2\kappa_3^2\theta_{iM}^2 + \bar{\rho}^2(\eta_{ioB} + \gamma_i^4)$, $\eta_{icl} = 0.125\alpha_{iv}^8 D_{iM}^4\left\|R_i{}^{-1}\right\|^4 + 0.5(\alpha_{iv}^2 D_{iM}^2\varepsilon_M^2 + \alpha_{iv}^2 D_{iM}^2 V_{ixM}^2)B_{i\tilde{\theta}}/\bar{\rho}^2 + 0.5\alpha_{iv}^4 D_{iM}^4 + 0.5\alpha_{iv}^2 V_{ixM}^2 D_{iM}^2 + 0.5\alpha_{iv}^4 V_{ixM}^4 D_{iM}^2\left\|R_i{}^{-1}\right\|^2$. It can be observed that with the conditions $\|x_i\| > \sqrt{\frac{\eta_{icl}}{\upsilon_x}} \equiv \eta_{iX}$ or $\left\|\tilde{X}_i\right\| > \sqrt{\frac{\eta_{icl}}{(\lambda_{\min}(\mu_{i1}\bar{q}_i) - \frac{7}{2} - \frac{2}{N}\|P_i\|^2)}}$ or $\left\|\tilde{\theta}_i^T\right\| > \sqrt[4]{\frac{\eta_{icl}}{((\upsilon_{\tilde{\theta}2}/\bar{\rho}^2) - 0.25)\nabla_x\phi_{\min}^4}} \equiv \eta_{i\tilde{\Theta}}$ or $\left\|\tilde{W}_{if}\right\| > \sqrt[4]{\eta_{icl}/\upsilon_{\tilde{f}}}$

or $\left\|\tilde{W}_{ig}\right\| > \sqrt[8]{\eta_{icl}/(\lambda_{\min}(\mu_{i6}\kappa_{ig}) - \upsilon_{\tilde{g}^2} - 0.5N_{iog}^4)} \equiv \eta_{\tilde{g}}$. The Lyapunov first derivative is less than zero and

$$\dot{J}_i \leq \begin{pmatrix} -(\lambda_{\min}(\mu_{i1}\bar{q}_i) - 3.5 - 2\|P_i\|^2/N)\left\|\tilde{X}_i\right\|^2 - \\ (\gamma_i\kappa_3 - 0.5 - \gamma_i\nabla_x\phi_{\min}^2/2\bar{\rho}^2)\left\|\tilde{\theta}_i\right\|^2 + \eta_{icl} \\ -(\upsilon_{\tilde{\theta}2}/\bar{\rho}^2 - 0.25)\nabla_x\phi_{\min}^4\left\|\tilde{\theta}_i^T\right\|^4 - \upsilon_{\tilde{f}}\left\|\tilde{W}_{if}\right\|^4 \\ -(\upsilon_{\tilde{g}} - \frac{1}{8}N_{iog}^2)\left\|\tilde{W}_{ig}\right\|^4 - (\lambda_{\min}(\mu_{i2}\kappa_{if}) - 1) \\ \left\|\tilde{W}_{if}\right\|^2 - (\lambda_{\min}(\mu_{i6}\kappa_{ig}) - \upsilon_{\tilde{g}^2} - \frac{1}{2}N_{iog}^4)\left\|\tilde{W}_{ig}\right\|^8 \\ -(\lambda_{\min}(\mu_{i3}\kappa_{ig}) - 1)\left\|\tilde{W}_{ig}\right\|^2 - \upsilon_x\|x_i\|^2 \end{pmatrix}$$

The overall bounds for the interconnected system states and the optimal value function estimation error is obtained as $\eta_X = \bigcup_{i=1}^{N} \eta_{iX}$ and $\eta_\Theta = \bigcup_{i=1}^{N} \eta_{i\tilde{\Theta}}$. The objective is to ensure that the event sampled implementation of the closed loop system is stable and the states and weight estimation errors, identification errors reach the bound described by case 1.

Case 2: Consider the event triggering condition $J_{ix}(t) \leq (1+t_k^i-t)\alpha_i J_{ix}(t_k^i)$, $t \in [t_k^i, t_{k+1}^i)$, $\forall k \in \{0, \mathbb{N}\}$. It is easy to see that the first derivative of the triggering condition results in $\dot{J}_{ix}(t) \leq -\alpha_i J_{ix}(t_k^i)$, $t \in [t_k^i, t_{k+1}^i)$, $\forall k \in \{0, \mathbb{N}\}$. Next, consider the identifier Lyapunov function and the Lyapunov function for the value function estimator. Using the definition of the event triggering error, we have

$$\dot{J}_{i\tilde{\theta}} + \dot{J}_{iI} \leq \begin{pmatrix} -(\lambda_{\min}(\mu_{i1}\bar{q}_i) - \frac{7}{2} - \frac{2}{N}\|P_i\|^2)\left\|\tilde{X}_i\right\|^2 \\ -(\gamma_i\kappa_3 - \frac{1}{2} - \frac{\gamma_i}{2\bar{\rho}^2}\nabla_x\phi_{\min}^2)\left\|\tilde{\theta}_i\right\|^2 \\ +(\frac{\gamma_i}{\bar{\rho}^2}(\frac{8\alpha_{iv}^4}{2} + \frac{2}{32} + 2\varepsilon_{iM}^4)C_i^4 + \frac{1}{4})\|x_i\|^2 \\ -\frac{1}{\bar{\rho}^2}\upsilon_{\tilde{\theta}2}\nabla_x\phi_{\min}^4\left\|\tilde{\theta}_i^T\right\|^4 - (\upsilon_{\tilde{g}})\left\|\tilde{W}_{ig}\right\|^4 \\ -(\lambda_{\min}(\mu_{i2}\kappa_{if}) - 1)\left\|\tilde{W}_{if}\right\|^2 - \upsilon_{\tilde{f}}\left\|\tilde{W}_{if}\right\|^4 \\ -(\lambda_{\min}(\mu_{i6}\kappa_{ig}) - \upsilon_{\tilde{g}^2})\left\|\tilde{W}_{ig}\right\|^8 \\ -(\lambda_{\min}(\mu_{i3}\kappa_{ig}) - 1)\left\|\tilde{W}_{ig}\right\|^2 + B_{i\tilde{\theta}}/\bar{\rho}^2 \\ +40\gamma_i\alpha_{iv}^2 L_{if}^4(\|x_i(t_k)\|^4 + \|x_i\|^4) \end{pmatrix}$$

where $B_{i\tilde{\theta}} = +\alpha_{iv}^2 D_{iM}^2\varepsilon_M^2 + 2\alpha_{iv}^4 D_{iM}^4\varepsilon_M^4 + 2\varepsilon_{iM}^8 D_{i\max}^4 + (\alpha_{iv}^2 0.5 + 2)2\varepsilon_{iM}^4 D_{i\max}^2 + 32(0.5\alpha_{iv}^2 + 2)^2\varepsilon_{iM}^2 + 64\alpha_{iv}^4 + 4.1\gamma_i\alpha_{iv}^2 D_{iM}^4 V_{ixM}^4 + 172\gamma_i\alpha_{iv}^2 V_{ixM}^4 D_{iM}^4 + 4713\gamma_i\alpha_{iv}^2 V_{ixM}^8 D_{iM}^4\left\|R_i^{-1}\right\|^4 + 0.5\gamma_i(\alpha_{iv}^2 + 2)B_{i\tilde{\theta}1} + 0.5\gamma_i^2\kappa_3^2\theta_{iM}^2 + \bar{\rho}^2(\eta_{ioB} + \gamma_i^4)$, $\eta_{icl2} = (\gamma_i(4\alpha_{iv}^4 + 0.0625 + 2\varepsilon_{iM}^4)C_i^4/\bar{\rho}^2 + 0.25 + \alpha_i 40\gamma_i\alpha_{iv}^2 L_{if}^4\|x_i(t_k)\|^4 + 40\gamma_i\alpha_{iv}^2 L_{if}^4\|x_i(t_k)\|^4 + B_{i\tilde{\theta}}/\bar{\rho}^2$. Combining all the Lyapunov function derivatives we get

$$\dot{J}_i \leq \begin{pmatrix} -\alpha_i J_{ix}(t_k) - (\lambda_{\min}(\mu_{i1}\bar{q}_i) - \frac{7}{2} - \frac{2}{N}\|P_i\|^2)\left\|\tilde{X}_i\right\|^2 \\ -(\gamma_i\kappa_3 - \frac{1}{2} - \frac{\gamma_i}{2\bar{\rho}^2}\nabla_x\phi_{\min}^2)\left\|\tilde{\theta}_i\right\|^2 + \eta_{icl2} \\ -\frac{1}{\bar{\rho}^2}\upsilon_{\tilde{\theta}2}\nabla_x\phi_{\min}^4\left\|\tilde{\theta}_i^T\right\|^4 - \upsilon_{\tilde{g}}\left\|\tilde{W}_{ig}\right\|^4 - \upsilon_{\tilde{f}}\left\|\tilde{W}_{if}\right\|^4 \\ -(\lambda_{\min}(\mu_{i2}\kappa_{if}) - 1)\left\|\tilde{W}_{if}\right\|^2 - (\mu_{i6}\kappa_{ig} - \upsilon_{\tilde{g}^2}) \\ \left\|\tilde{W}_{ig}\right\|^8 - (\lambda_{\min}(\mu_{i3}\kappa_{ig}) - 1)\left\|\tilde{W}_{ig}\right\|^2 \end{pmatrix}$$

It can be observed that the bounds obtained for the states, weight estimation errors and identifier errors are larger than the corresponding bounds obtained in case 1. In order to conclude the proof, it is sufficient to show that the bounds in the inter event period are decreasing as the events $\{t_k^i\} \to \infty$. It can be observed that the Lyapunov function based event triggering condition is a continuous function. Therefore, $x(t)$ is decreasing as long as $\|x_i\| > \eta_{iX}$ and if the gains satisfy the stability conditions obtained in Theorem 1 and 2.

Therefore, the bounds η_{icl2} converges to η_{icl} as $t \to \infty$. Thus, combining case 1 and case 2, the closed loop Lyapunov function derivative, $\dot{J} < 0$, as long as the conditions derived in the Theorems are satisfied. If the NN weights are tuned using (19), the terms corresponding to the data collected in the interevent period will result in an expression for $\dot{J}_{i\tilde{\theta}}$ similar to the expressions in Theorem 1 with slightly different bounds and coefficients, without affecting the final result on the stability. Due to space consideration, the details are not included here. Finally, $\left\|V^* - \hat{V}\right\| \leq \left\|\tilde{\Theta}\right\|\|\Phi(x)\| + \varepsilon_M \leq \eta_{\Theta 1}\Phi_M + \varepsilon_M \equiv \eta_{\tilde{V}}$ and $\|u^* - u\| \leq \lambda_{\max}(R^{-1})((V_{xM} + \eta_{\tilde{V}})\eta_{\tilde{g}}\sqrt{N_{iog}} + \varepsilon_{gM} + G_M\eta_{\tilde{V}})$. From Lemma 1, identifiers exhibit local-ISS like behavior and if the exploratory signal is bounded, the exploratory policy and the identifier states used to update the NN weights with the exploratory policy will be locally UUB. This concludes the proof.

REFERENCES

[1] P. J. Werbos, "Optimization methods for brain-like intelligent control," in *Proceedings of 1995 34th IEEE Conference on Decision and Control*, vol. 1, Dec 1995, pp. 579–584 vol.1.

[2] A. G. Barto, R. S. Sutton, and C. W. Anderson, "Neuronlike adaptive elements that can solve difficult learning control problems," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. SMC-13, no. 5, pp. 834–846, Sept 1983.

[3] F. Lewis, S. Jagannathan, and A. Yesildirak, *Neural network control of robot manipulators and non-linear systems*. CRC Press, 1998.

[4] R. S. Sutton and A. G. Barto, *Reinforcement learning: An introduction*. MIT press Cambridge, 1998, vol. 1, no. 1.

[5] J. Si and Y.-T. Wang, "Online learning control by association and reinforcement," *IEEE Transactions on Neural Networks*, vol. 12, no. 2, pp. 264–276, Mar 2001.

[6] P. Dayan, "The convergence of TD(λ) for general λ," *Machine Learning*, vol. 8, no. 3, pp. 341–362, May 1992. [Online]. Available: http://dx.doi.org/10.1023/A:1022632907294

[7] K. Doya, "Reinforcement learning in continuous time and space," *Neural Computation*, vol. 12, no. 1, pp. 219–245, 2000. [Online]. Available: http://dx.doi.org/10.1162/089976600300015961

[8] S. Ferrari and R. F. Stengel, "An adaptive critic global controller," in *Proceedings of the 2002 American Control Conference (IEEE Cat. No.CH37301)*, vol. 4, May 2002, pp. 2665–2670 vol.4.

[9] S. N. Balakrishnan, J. Ding, and F. L. Lewis, "Issues on stability of adp feedback controllers for dynamical systems," *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, vol. 38, no. 4, pp. 913–917, Aug 2008.

[10] 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.

[11] 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.

[12] H. Modares, F. L. Lewis, and Z. P. Jiang, "Optimal output-feedback control of unknown continuous-time linear systems using off-policy reinforcement learning," *IEEE Transactions on Cybernetics*, vol. 46, no. 11, pp. 2401–2410, Nov 2016.

[13] S. Sastry and M. Bodson, *Adaptive control: stability, convergence and robustness*. Courier Corporation, 2011.

[14] H. Ma, Z. Wang, D. Wang, D. Liu, P. Yan, and Q. Wei, "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: Systems*, vol. 46, no. 6, pp. 750–758, June 2016.

[15] 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.

[16] A. Sahoo, H. Xu, and S. Jagannathan, "Neural network-based adaptive event-triggered control of nonlinear continuous-time systems," in *2013 IEEE International Symposium on Intelligent Control (ISIC)*, Aug 2013, pp. 35–40.

[17] K. G. Vamvoudakis, "An online actor/critic algorithm for event-triggered optimal control of continuous-time nonlinear systems," in *2014 American Control Conference*, June 2014, pp. 1–6.

[18] L. Dong, X. Zhong, C. Sun, and H. He, "Event-triggered adaptive dynamic programming for continuous-time systems with control constraints," *IEEE Transactions on Neural Networks and Learning Systems*, vol. PP, no. 99, pp. 1–12, 2016.

[19] T. Liu and Z. P. Jiang, "A small-gain approach to robust event-triggered control of nonlinear systems," *IEEE Transactions on Automatic Control*, vol. 60, no. 8, pp. 2072–2085, Aug 2015.

[20] 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.

[21] X. Ge and Q.-L. Han, "Distributed event-triggered h filtering over sensor networks with communication delays," *Information Sciences*, vol. 291, pp. 128 – 142, 2015. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0020025514008524

[22] V. Narayanan and S. Jagannathan, "Distributed adaptive optimal regulation of uncertain large-scale interconnected systems using hybrid q-learning approach," *IET Control Theory & Applications*, vol. 10, no. 12, pp. 1448–1457, 2016.

[23] X. M. Zhang, Q. L. Han, and B. L. Zhang, "An overview and deep investigation on sampled-data-based event-triggered control and filtering for networked systems," *IEEE Transactions on Industrial Informatics*, vol. 13, no. 1, pp. 4–16, Feb 2017.

[24] A. Sahoo, H. Xu, and S. Jagannathan, "Approximate optimal control of affine nonlinear continuous-time systems using event-sampled neurodynamic programming," *IEEE Transactions on Neural Networks and Learning Systems*, vol. PP, no. 99, pp. 1–14, 2016.

[25] V. Narayanan 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.

[26] ——, "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.

[27] 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.

[28] B. C. Da Silva and A. G. Barto, "TD-$\triangle\pi$: A model-free algorithm for efficient exploration." Twenty-Sixth Conference on Artificial Intelligence (AAAI-2012), Toronto, Ontario, Canada, 2012.

[29] J. Y. Lee, J. B. Park, and Y. H. Choi, "Integral reinforcement learning for continuous-time input-affine nonlinear systems with simultaneous invariant explorations," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 26, no. 5, pp. 916–932, May 2015.