Summer 2017

Event-triggered near optimal adaptive control of interconnected systems

Vignesh Narayanan

Department: Electrical and Computer Engineering

EVENT-TRIGGERED NEAR OPTIMAL ADAPTIVE CONTROL OF INTERCONNECTED SYSTEMS

by

VIGNESH NARAYANAN

A DISSERTATION

Presented to the Graduate Faculty of the

MISSOURI UNIVERSITY OF SCIENCE AND TECHNOLOGY

In Partial Fulfillment of the Requirements for the Degree

DOCTOR OF PHILOSOPHY

in

ELECTRICAL ENGINEERING

2017

Approved by

Dr. Jagannathan Sarangapani, Advisor
Dr. Kelvin T. Erickson
Dr. Robert G. Landers
Dr. Douglas A. Bristow
Dr. Vy K. Le

PUBLICATION DISSERTATION OPTION

This dissertation has been prepared using the publication option and is composed of five papers.

Paper I : Distributed adaptive optimal regulation of uncertain large-scale interconnected systems using hybrid Q-learning approach, 15-45.

Paper II : Event-Triggered Distributed Approximate Optimal State and Output Control of Affine Nonlinear Interconnected Systems, 46-86.

Paper III : Event-triggered Distributed Control of Nonlinear Interconnected Systems Using Online Reinforcement Learning with Exploration, 87-126.

Paper IV : Adaptive Optimal Event-triggered Control of Linear Dynamic Systems, 127-156.

Paper V : Approximate Optimal Event-triggered Control of Nonlinear Systems, 157-190.

ABSTRACT

Increased interest in complex interconnected systems like smart-grid, cyber manufacturing have attracted researchers to develop optimal adaptive control schemes to elicit a desired performance when the complex system dynamics are uncertain. In this dissertation, motivated by the fact that aperiodic event sampling saves network resources while ensuring system stability, a suite of novel event-sampled distributed near-optimal adaptive control schemes are introduced for uncertain linear and affine nonlinear interconnected systems in a forward-in-time and online manner.

First, a novel stochastic hybrid Q-learning scheme is proposed to generate optimal adaptive control law and to accelerate the learning process in the presence of random delays and packet losses resulting from the communication network for an uncertain linear interconnected system. Subsequently, a novel online reinforcement learning (RL) approach is proposed to solve the Hamilton-Jacobi-Bellman (HJB) equation by using neural networks (NNs) for generating distributed optimal control of nonlinear interconnected systems using state and output feedback. To relax the state vector measurements, distributed observers are introduced. Next, using RL, an improved NN learning rule is derived to solve the HJB equation for uncertain nonlinear interconnected systems with event-triggered feedback. Distributed NN identifiers are introduced both for approximating the uncertain nonlinear dynamics and to serve as a model for online exploration.

Next, the control policy and the event-sampling errors are considered as non-cooperative players and a min-max optimization problem is formulated for linear and affine nonlinear systems by using zero-sum game approach for simultaneous optimization of both the control policy and the event based sampling instants. The net result is the development of optimal adaptive event-triggered control of uncertain dynamic systems.

ACKNOWLEDGMENTS

I would like to sincerely thank my advisor, Dr. Jagannathan Sarangapani, whose support made this work possible. Additionally, I would like to thank Dr. Kelvin T. Erickson, Dr. Robert G. Landers, Dr. Douglas A. Bristow and Dr. Vy K. Le for serving on my committee. I also thank the NSF and the Intelligent Systems Center for financially supporting the research. Furthermore, I would like to thank Dr. L. Acar, Dr. A. Sahoo, Dr. Hao Xu, Dr. Travis Dierks and Krishnan Raghavan, whose guidance and encouragement were instrumental to my success. Finally, I thank my parents, family and friends whose support has been immeasurable.

TABLE OF CONTENTS

Page

SECTION

PAPER

SECTION

LIST OF ILLUSTRATIONS

Figure Page

PAPER I

PAPER II

PAPER III

PAPER IV

PAPER V

LIST OF TABLES

Table Page

PAPER I

PAPER II

PAPER III

PAPER IV

PAPER V

SECTION

1. INTRODUCTION

In the past decade, significant advances have occurred in the areas of computing power, communication and control. These advances are enabling design of critical infrastructures related to manufacturing, energy and transportation which are ‘smart’in the sense that they monitor themselves, communicate and self-govern [3]. Due to the complexity of tasks performed by such critical systems, an efficient decision making component is required which can reduce the communication, computational cost and guarantee a certain degree of performance.

Nature is an ultimate source of motivation for researchers and most efficient solution for complex problems can be found among several biological systems. Researchers have been inspired by the biological systems and studied their functioning to improve the efficiency of engineering systems and the way they are controlled. Especially, the science of decision making in the face of uncertainties is a complex task which is carried out by every biological species all the time. The use of artificial neural networks (NNs) in computation and control is one of the many tools that emerged from the bio-inspired research. A computational model of the decision making process which can be observed in the biological systems is given in Fig. 1.1.

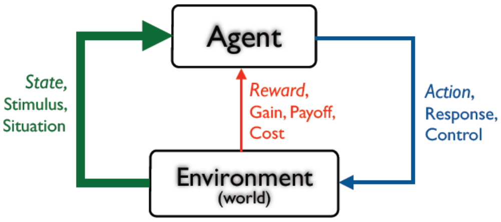


Fig. 1.1. Decision making process [25].

Any biological species interacts with the environment and learns to respond to a stimulus which results in maximum reward. This computational model is studied by several researches in the field of reinforcement learning (RL) and control theory [30] to develop optimal decision making schemes which continuously learn and adapt online. These learning schemes are popularly known as adaptive dynamics programming [4, 5, 31]. A basic block diagram of such adaptive control scheme is given in Fig. 1. 2. The control actions generated by the decision making body/actor are evaluated by a critic based on a scalar reward and using the information from the critic, the actor learns to generate control actions which results in maximum reward.

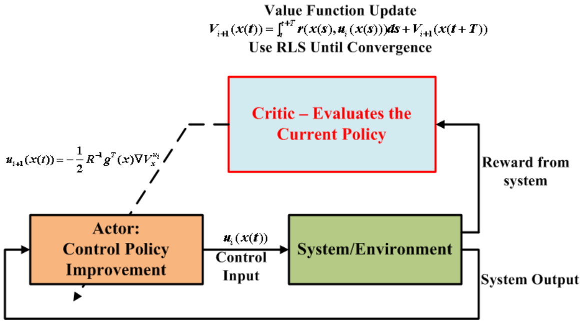

Value Function Update
$V_{i+1}(x(t)) = \int_t^{t+T} r(x(s), u_i(x(s)))ds + V_{i+1}(x(t+T))$
Use RLS Until Convergence
Critic – Evaluates the Current Policy
Reward from system
$u_{i+1}(x(t)) = -\frac{1}{2} R^{-1} g^T(x) \nabla V_x^{u_i}$
Actor: Control Policy Improvement
$u_i(x(t))$
Control Input
System/Environment
System Output

Fig. 1.2. Simplistic block diagram of reinforcement learning.

Currently, decentralized and distributed control of interconnected systems is finding itself in wide range of applications like robotic systems, power grids, traffic control systems, urban drainage system to name a few (Fig. 1.3). For systems which are spatially distributed, having a centralized decision making body is very costly due to the communication and computational resources required to ensure that the centralized controller has access to all the subsystems-to gain the feedback information and send the control commands. Moreover, with such control architecture, adding or removing subsystems is a tedious task and failure/maintenance of the controller requires shutting down the entire system. With the advent of networked control system, using decentralized/distributed controllers which communicate with each other offers more flexibility and scalability.

Moreover, for the large-scale critical infrastructures comprising of interconnected systems, the centralized RL computational model in Fig.1.1 and the control scheme in Fig.1.2 which considers the overall system is not very efficient. Going back to the biological systems, some of the studies conducted on social creatures like honeybees [6], birds [7], ants and fish [8] reveal that these groups of organisms, when performing a complex task, work together to achieve a common goal efficiently by incorporating the principle of 'division of labor'. Recognizing these benefits, researchers have developed control and learning schemes for interconnected large-scale systems composed of individual subsystems. Just as the biological organisms described above distribute the overall objective among them to complete their respective tasks more effectively, decentralized and distributed controllers are designed at each subsystem to achieve a local goal in such a way that the overall objective of the interconnected system is realized.

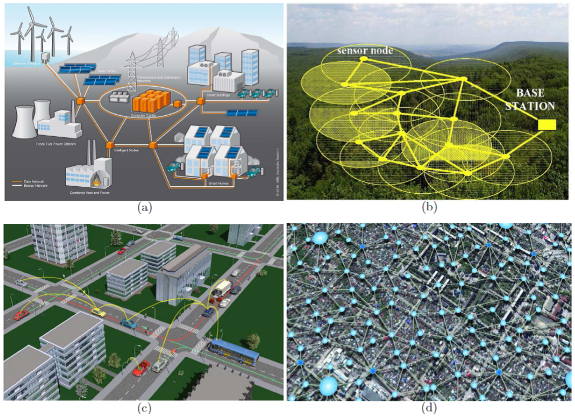


Fig. 1.3. Examples of large-scale interconnected systems a) Smart grid; b) Wireless sensor networks; c) traffic network; d) Water distribution network [2].

However, to do this task efficiently without requiring the system dynamics, optimization theory inspired by reinforcement learning is considered. In addition to the use of optimization, event-driven computation [9], inspired by the biological processes of human brain, is applied in designing processors to reduce the computations and power usage. For example, using a bio-inspired design with event driven computing, the power density of IBM neuromorphic chip is reported to be 1/10000th of the existing microprocessors. Moreover, with the advent of networked control systems (NCS), many interconnected systems share a communication network to transmit information between nodes. These nodes may be another subsystem, sensor modules, actuator or a controller. In addition to the benefits in terms of computations and power usage, event driven computing saves network resources in NCS. A simplified block diagram for implementing an event triggered control scheme is given in Fig. 1.4. The states of the plant/system (x_k) are continuously monitored by a trigger mechanism which decides when the controller requires the feedback information to update the control action (u_k). A zero-order-hold (ZOH) is used to hold the last updated state and control at the controller and actuator, respectively. Considering the benefits of optimizing a control task combined with reducing the computations and communication cost have stimulated the interests in many researchers and the focus has shifted from controlling a single system to controlling large-scale interconnected systems with event triggered feedback.

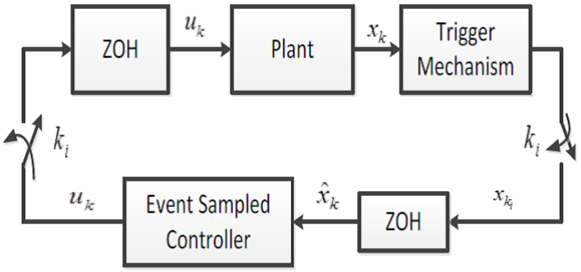


Fig. 1.4. Basic block diagram for event-triggered controller.

In summary, based on the current demands driven by scarcity of resources, economic considerations and huge technological advancements in computing power and communication systems, event-driven control of large-scale interconnected systems is as an active area of current research. While stabilization is the primary objective, cost-efficient performance and optimization in the face of uncertain system dynamics is required to fulfill these current demands. Next, an overview of existing control methodologies addressing the above issues is presented.

1.1. OVERVIEW OF CURRENT CONTROL METHODOLOGIES

Decentralized control scheme was first introduced by Siljak in 1978 [10]. Since then, several control schemes for decentralized and distributed control of large-scale interconnected systems have been proposed. A detailed survey of decentralized control scheme is carried out in [11], while a detailed account on distributed controllers is presented in [12]. Decentralized controllers, which ignore the effects of interconnections, are found to be inefficient when compared to the controllers that facilitate communication among subsystems and enable controllers at each subsystem to utilize feedback information from other subsystems [2, 28]. However, decentralized robust control approach using local state information has been studied in the literature due to their simplicity and scalability [1]. On the other hand, distributed controllers require information from neighboring subsystems to generate a control action. One of the issues encountered in the distributed control scheme is how often the subsystems should communicate their state and output information among other subsystems.

Over the years, the controllers for large-scale systems have evolved to stabilize the subsystems in the presence of uncertain interconnection matrix with limited communication [1]. Adaptive controllers were proposed to learn the interconnection terms, with which suitable compensation was provided, but they were limited to handle weak interconnections [10]. Later, reference models were utilized to provide information about the other subsystems [28]. However, it was demonstrated that ignoring the effects of interconnections and using the information from the reference models can result in unacceptable transient performance [28].

Further, developing optimal controllers for an interconnected system is a challenging design problem. One of the methods to design optimal controllers is by model-predictive control. By utilizing the communication network connecting subsystems, several distributed control algorithms to solve optimization problem for large-scale system using model-predictive control (MPC) have been proposed. Although MPC based control algorithms are popular due to their inherent ability to handle input and state constraints efficiently, distributed MPC algorithms are not as efficient as their centralized counterpart due to the effect of coupling between the subsystems in large-scale systems. Also, MPC based algorithms in general requires system model to predict the future output over a limited time horizon with which a desired cost-function is minimized iteratively.

The other approach to solve the optimal control problem for interconnected system is by using game-theoretic formulation. Multi-player game formulation is one of the control design approaches presented in the literature to solve the optimal control problem for interconnected systems [13]. The subsystems are considered to be cooperative and each subsystem is controlled to work in tandem to achieve a common goal while minimizing a cost function of the overall system. However, due to the nature of performance objective, implementing the control generated using the multiplayer game approach requires a centralized scheme and scalability of such controllers due to curse-of-dimensionality is an issue.

On the other hand, the distributed optimal controllers can be designed by breaking down the objective of the overall system into several components corresponding to each subsystem and designing controllers to satisfy the component objectives. However, such a control scheme requires a well-defined method to decompose the performance objective of the overall system for distributed control generation such that both the local cost function and the aggregated overall cost function is minimized [13].

Moreover, the control schemes presented in [12] and the references therein, assume that a dedicated communication network is available to continuously share the feedback information among the subsystems. This significantly increases the communication cost. Lately, aperiodic state dependent sampling is studied under various names, such as, multi rate sampling, Lebesgue sampling [14], and interrupt driven triggering [15]. Recently, this scheme is studied under a formal name of “event-triggered”[16] sampling and various theoretical and experimental results emphasizing its inherent advantages, in computation and communication saving, are available in the literature [17, 18, 19, 20, 21, 22].

In general, an emulation-based approach is used for the event-triggered system design where the controller is designed considering the periodic sampling and an event-trigger condition is developed maintaining the stability intact. As a general rule the system is assumed to be input to state stable (ISS) with respect to the measurement error, and an event-trigger condition is designed considering the difference between the current state and last sampled state as event-trigger error along with a state dependent threshold. Further, a non-zero positive lower bound on the inter event time is also guaranteed to avoid Zeno behavior. In addition, the event-triggered control approach is also extended to accommodate other design considerations, such as, output feedback design, decentralized designs, and trajectory tracking control [23]. All the above design approaches hold the system state or output between any two trigger instants for controller implementation and usually a zero order hold (ZOH) is used for this purpose. Alternatively, a model-based approach [21, 24] is developed where the system state vector is reconstructed and, subsequently, used for designing the control input. As the control input is based on the model, no feedback transmission is required unless there is a significant change in the system performance due to external disturbance or internal parameter variation. In this scheme, the authors in [24] presented an input generator as a model to predict the system states which are used to compute the control. Further, the authors in [21, 22] consider the nominal dynamics of the system with uncertainty, usually of smaller magnitude and bounded, to form a model. The asymptotic stability is guaranteed by designing the event-trigger condition. It is observed that the model based approach reduces the event-trigger instants or transmission more effectively when compared to the ZOH based approach, but, with a higher computational load due to induction of the model. Nevertheless, MPC based optimization and model-based event triggering is an approach that can work well given in tandem an accurate system model which is not always available.

Looking at the optimal control aspect of the event-based control, a few results are available [18, 19, 20]. The problem is formulated as an optimal stopping problem and an analytical solution is provided. Further, in [19], the authors characterized the certainty equivalence controller to be optimal in a linear quadratic Gaussian (LQG) frame work and derived a separation principle to design the control and optimal event-trigger instants separately.

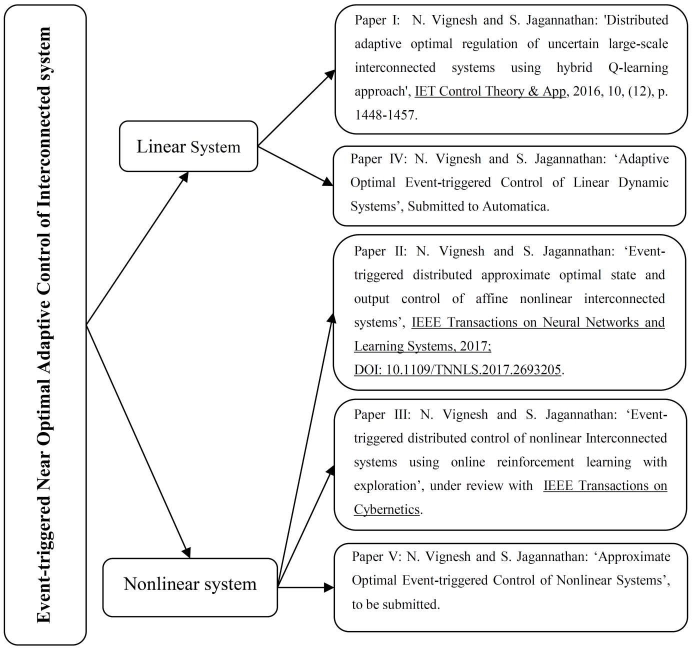
Despite these results from the literature on event-triggered control, these schemes consider either the complete knowledge of the system dynamics or system with a smaller uncertainty with known nominal dynamics. Moreover, the optimal solution of the event-based control requires the system dynamics to be solved in backward in time manner, requiring accurate knowledge of system dynamics to pre-calculate the sequence of control actions. Thus, a forward in time and online solution to the optimal control problem in an event-triggered context is required. To accommodate the uncertain dynamics and generate optimal control policy with event based feedback, a control scheme using NNs is proposed in [22]. The event driven function approximation property of the NN is studied. A strong relationship between the frequency of events and approximation accuracy, convergence of the learning algorithm is observed. Therefore, the time-driven ADP scheme is which generates online approximate optimal control takes longer time to converge. Despite several efforts in developing a forward in time solution to optimal control problems using reinforcement learning [25, 26], all the learning schemes use fixed iterative learning steps, rendering them inefficient for event triggered implementation. For example, policy iteration requires iterative updates until convergence of the Bellman error during both - the policy evaluation and policy improvement steps; value iteration requires iterative updates until convergence of the Bellman error for the policy evaluation step and a single iteration in the policy improvement step; the time-driven learning algorithms requires single step policy evaluation and improvement.

Motivated by the above facts, in this dissertation, a suite of novel event-triggered adaptive optimal control designs for linear and nonlinear interconnected system are presented. Adaptive and neural network based learning methods are used to learn the unknown parameters/dynamics and a forward in time solution is presented. In addition, Lyapunov stability analysis is carried out to guarantee the stability of the closed-loop event triggered system. First, the design of both control policy and the event-triggering mechanism is formulated as a two-player zero-sum game. Here, a novel cost function is introduced as a function of state vector, control policy and the measurement/event-triggering error. The control policy and the measurement error due to event-triggered feedback will be considered as two non-cooperative players. The saddle point solution to this min-max problem results in the minimization of the control policy while maximizing the measurement error.

The resulting measurement error from this min max optimization problem is utilized as the dynamic threshold in an event-trigger condition to determine the sampling instants. Since the control policy explicitly accounts for the worst-case event-triggering error, the stability and the performance of the system is preserved. Moreover, since the inter-event time is directly proportional to the event-triggering error and utilizing the maximum event trigger error as a dynamic threshold results in optimizing the inter-event time. This net result is an optimal event-triggered controller which explicitly takes into account both generation of event-triggered sampling instants and control policy. In addition, Lyapunov stability analysis is carried out to guarantee the stability of the closed-loop event triggered system. Next, the organization of the thesis is presented.

1.2. ORGANIZATION

In this dissertation, the control of large-scale interconnected systems is undertaken while incorporating learning and adaptation with limited feedback information. This dissertation is presented in five papers and their relation to one another is illustrated in Fig. 1.5. The underlying common theme of each paper is the control of interconnected system with event triggered execution of control tasks incorporating optimization and learning component.


Event-triggered Near Optimal Adaptive Control of Interconnected system
Linear System
Nonlinear system

Paper I: N. Vignesh and S. Jagannathan: 'Distributed adaptive optimal regulation of uncertain large-scale interconnected systems using hybrid Q-learning approach', IET Control Theory & App, 2016, 10, (12), p. 1448-1457.

Paper IV: N. Vignesh and S. Jagannathan: ‘Adaptive Optimal Event-triggered Control of Linear Dynamic Systems’, Submitted to Automatica.

Paper II: N. Vignesh and S. Jagannathan: ‘Event-triggered distributed approximate optimal state and output control of affine nonlinear interconnected systems’, IEEE Transactions on Neural Networks and Learning Systems, 2017; DOI: 10.1109/TNNLS.2017.2693205.

Paper III: N. Vignesh and S. Jagannathan: ‘Event-triggered distributed control of nonlinear Interconnected systems using online reinforcement learning with exploration’, under review with IEEE Transactions on Cybernetics.

Paper V: N. Vignesh and S. Jagannathan: ‘Approximate Optimal Event-triggered Control of Nonlinear Systems’, to be submitted.

Fig. 1.5. Outline of the dissertation.

The first paper deals with large-scale interconnected linear systems with uncertain dynamics. A control scheme is proposed for distributed implementation of centralized adaptive optimal control when the subsystems share their state information through a lossy communication network. The random delays and packet drop-outs are modelled and the system dynamics are re-formulated and stochastic model free hybrid Q-learning algorithm for facilitating convergence of the learning algorithm with event sampled feedback is presented. In the second paper, the hybrid learning algorithm is extended for nonlinear interconnected systems. State and output feedback based approximate optimal controllers are proposed which reduce the number of times the control task is executed while guaranteeing desired levels of performance. Instead of learning the centralized optimal value functions, distributed value functions are approximated using NNs and the Hamiltonian-Jacobi-Bellman (HJB) equation in an online, forward-in-time manner.

Due to the reduction in the feedback instants, the traditional time-driven learning schemes suffered from an increased time for convergence and the hybrid learning schemes improved the learning process by decoupling the dependence of convergence time, approximation accuracy with number of event triggering instants. To match the time taken by the learning algorithm with continuous feedback for convergence and to potentially improve the optimality of the control actions, in the third paper, reinforcement learning theory is studied and a novel learning scheme using generalized policy iteration is proposed and an online exploration strategy using identifiers is presented. On the other hand, in the fourth paper, a robust adaptive optimal learning control scheme is proposed to generate decentralized control actions at each subsystem. In contrast to the papers 1-3, the decentralized scheme is advantageous as the design of Lyapunov function for each subsystem is easier than to determine a Lyapunov function for the large-scale system. Further, existence of convex value function for a large-scale system is not always guaranteed [2]; even if there exists a convex value function for the overall system, due to curse of dimensionality, adding a subsystem to the interconnected system exponentially increases the complexity of the solution for the optimal control problem which renders the learning problem intractable. However, designing decentralized controllers without imposing any restrictions on the interconnection strengths is a challenge and small-gain theorem for large scale interconnected system is employed to provide robustness against the interconnections and quantify the performance bounds that can be achieved with the proposed control policy. Finally, online implementation of integral reinforcement learning in the event triggered feedback framework is proposed.

In the first three papers, the event-triggered sampling instants are designed using the Lyapunov function which explicitly accounts for the system stability but provides no information or control over the system performance when the control policy is not updated between the event based sampling instants. Therefore, in the fourth and the fifth papers, adaptive near-optimal control schemes are proposed to simultaneously optimize the control actions and ensure satisfactory system performance even when the control actions are not frequently updated. Here, the control policy and the event-triggering errors are modelled as non-cooperative players and a novel cost function is proposed as a function of states, control policy and the event-triggering error. A zero-sum game based approach is followed to develop a saddle point solution to the min-max optimization problem wherein the control policy is applied to the system and the maximizing error obtained as the solution to the optimization problem is utilized as a dynamic threshold for the event-triggering error to determine the sampling instants. In contrast to the Papers I-III, the zero-sum game based control schemes proposed in the Paper IV, V are advantageous as the trade-off between the frequency of feedback and the system performance is optimized. While Paper IV focuses on linear systems, Paper 5 presents a simultaneous optimization approach for nonlinear system, wherein the event generating mechanism and the controllers are designed to balance the system performance and frequency of events. In the Paper IV, a model-free approach using hybrid Q-learning scheme is presented and the design approach is extended for distributed control of interconnected linear systems. In Paper V, an approximation based hybrid learning scheme is proposed to learn the NN weights which approximates the solution to the Hamilton-Jacobi-Isaacs equation. A decentralized event-triggering solution is presented for distributed control of nonlinear interconnected system. In all the Paper, comprehensive simulation studies and Lyapunov based stability analysis is carried out and the results are presented.

1.3. CONTRIBUTIONS

This dissertation provides contributions to the field of control of interconnected systems. The control laws developed in this dissertation in the context of event-triggered feedback use adaptive optimal control and reinforcement learning theory. The major contributions of Papers 1-3 include: (a) development of a novel hybrid Q-learning scheme using event-sampled states, input vector and their history; (b) the derivation of a time-driven and hybrid Q-learning scheme for an uncertain large-scale interconnected system enclosed by a communication network without any assumptions on coupling terms; (c) a decentralized event-sampling condition based on Lyapunov function without needing a mirror estimator at the sensor; d) development of an approximately optimal controller for nonlinear interconnected system using state and output feedback with event-triggered ADP approach in the presence of communication; e) design of a novel hybrid learning scheme, with full state measurements and for the case when only the outputs are available, to reduce the convergence time of the learning scheme ; f) design of an adaptive event-triggering mechanism using locally available information; g) design of extended nonlinear observers that utilizes the event-triggered output vector at each subsystem to relax the need for the entire state-vector to be measured and broadcasted; h) development of a RL based novel learning control scheme suitable for event-triggered control implementation; i) a suite of NN identifier designs to reconstruct unknown nonlinear functions in the system dynamics; j) a novel NN weight adaptation rule to reconstruct and learn the approximated optimal value function; k) an online exploration strategy using identifiers and e) stability analysis using Lyapunov theory.

Further, the contributions of the Papers IV and V include: a) a novel optimal event-triggering and controller co-design using zero-sum game formulation for linear and a class of nonlinear systems; b) development of an optimal adaptive online Q-learning scheme for generating the optimal control policy while maximizing the event-triggering intervals in a forward-in-time manner when the system dynamics are uncertain; c) development of an online NN learning scheme for generating optimal control and event-triggering policies when the system dynamics are uncertain; d) extension of the approximate optimal event-triggered design to the distributed control of interconnected systems; e) derivation of inter-event time or event triggered sampling instants for the cases of known and uncertain system dynamics; f) Lyapunov stability analysis and verification of the proposed design using numerical examples via simulation.

PAPER

I. DISTRIBUTED ADAPTIVE OPTIMAL REGULATION OF UNCERTAIN LARGE-SCALE INTERCONNECTED SYSTEMS USING HYBRID Q-LEARNING APPROACH

ABSTRACT

In this paper, a novel hybrid Q-learning algorithm is introduced for the design of a linear adaptive optimal regulator for a large-scale interconnected system with event-sampled inputs and state vector. Here, the time-driven Q-learning along with proposed iterative parameter learning updates are utilized within the event-sampled instants to both improve efficiency of the optimal regulator and obtain a more generalized online Q-learning framework. The network-induced losses due to the presence of a communication network among the subsystems are considered along with the uncertain system dynamics. Stochastic model-free Q-learning and dynamic programming are utilized in the hybrid learning mode for the optimal regulator design. The asymptotic convergence of the system state vector and boundedness of the parameter vector is demonstrated using Lyapunov analysis. Further, when the regression vector of the Q-function estimator satisfies the persistency of excitation (PE) condition, the Q-function parameters converge to the expected target values. The analytical design is evaluated using numerical examples via simulation. The net result is the design of a data-driven event-sampled adaptive optimal regulator for an uncertain large-scale interconnected system.

1. INTRODUCTION

Optimal control [1] using adaptive dynamic programming (ADP) [2, 3, 4, 5, 6, 7] has drawn more attention because of the forward-in-time solution to the optimal control problems for uncertain systems. The ADP based control schemes use reinforcement learning to solve the Bellman or Hamilton-Jacobi-Bellman (HJB) equation[4] through online parameterization and obtain optimal control policy. Among the ADP-based Q-learning schemes, [2] proposed a policy iteration approach using the Bellman equation. Later, the Q-learning scheme was extended in [3] to zero-sum-game formulation by using model-free policy iteration.

Policy/value iteration based techniques use significant number of iterative parameter updates within a sampling interval to maintain system stability, and its online implementation is not practically viable [5]. Therefore, online implementation for such iterative techniques was presented in [4], where the parameters are updated after collecting sufficient data-points. In contrast, the effort from [5] followed by [6, 7] introduced a time-based model-free ADP scheme where the past data of the cost-to-go errors are used for constructing the optimal value function.

On the other hand, control of large-scale interconnected systems [8] has been an active area of research. Large-scale systems are complex systems composed of geographically distributed subsystems connected through a communication network. The traditional centralized controller design for such systems is often impractical for computational reasons and lack of control integrity [9]. Therefore, various decentralized/distributed control schemes have been developed in the literature such that each subsystem has an independent controller [9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22]. The complexity in the control design arises due to the structural constraint in the form of interconnection/coupling matrix [8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20], which determines how the states/control of one subsystem influence the dynamics of the other subsystems.

Over the years, the controllers for large-scale systems have evolved to stabilize the subsystems in the presence of uncertain interconnection matrix with limited communication [8, 9, 10]. Adaptive controllers were proposed to learn the interconnection terms, with which suitable compensation was provided [9, 10, 11, 12], but they were limited to handle weak interconnections. Later, reference models were utilized to provide information about the other subsystems. However, it is reported in [13] that if the subsystems do not communicate their state information with each other and use a reference model to obtain this information, unsatisfactory transient performance will occur. Further, utilizing the communication network connecting the subsystems, several distributed control algorithms to solve optimization problem for large-scale system using model-predictive control (MPC) have been proposed in [14, 15, 16, 17, 18] and the references therein.

Although MPC based control algorithms are popular due to their inherent ability to handle input and state constraints efficiently, distributed MPC algorithms are not as efficient as their centralized counterpart due to the effect of coupling between the subsystems in the large-scale systems [14, 16, 18]. Also, MPC based algorithms in general requires system model to predict the future output over a limited time horizon with which a desired cost-function is minimized iteratively [14, 15, 16, 17, 18]. In contrast, the Q-function based control algorithm developed in this work neither requires an accurate model for the system nor utilizes significant iterations to solve the optimization problem. It should be noted that the above mentioned works [14, 16, 17, 18, 19] use periodic feedback and utilize the system dynamics to generate control. However, it is not feasible to communicate the state information periodically due to the communication cost involved.

Recently, it was demonstrated that event-based sampling is advantageous over periodic sampling in terms of computational cost [6, 21, 22, 23]. The aperiodic event-based sampling instants are determined by using a trigger condition while maintaining stability of the system. Such an event-sampled approach for control design was extended to large-scale interconnected systems in [20, 21, 22] by assuming either weak interconnections [20, 21] or control gain satisfying a strong matching condition, in order to decouple the subsystems [22].

The presence of a communication network among the subsystems and in the feedback-loop introduces random time delays and data-dropouts [24, 25, 26], which degrades control performance. It was shown in [7] that a linear time-invariant system with a communication network within its feedback loop can be represented as a stochastic time-varying linear system with uncertain dynamics. To the best knowledge of the authors, a time-based Q-learning scheme with intermittent feedback is not reported for such uncertain large-scale interconnected systems.

Therefore, in this paper, a novel hybrid model-free Q-learning scheme using event-sampled state and input vector is introduced for a large-scale interconnected system that is enclosed by a communication network. This algorithm enables a finite number of proposed Q-function parameter updates iteratively within the event-sampled instants to attain optimality faster without explicitly increasing the events when compared with the algorithm in [6].

In the proposed algorithm, the temporal-difference based ADP schemes [6, 7] and the policy/value iterations based ADP schemes [4, 27] become special cases. It also relaxes the assumptions on the estimated control input utilized in [6, 7]. This makes the learning algorithm more flexible than the existing model-free Q-learning based ADP schemes for online control. Since the Q-function parameters at each subsystem are estimated online with event-sampled input, state information along with past history and the data obtained from other subsystems through the communication network, an overall system model is not required. This makes the control scheme data-driven [28]. It is important to note that the infinite horizon cost function associated can be evaluated only for an admissible control policy [4]. This requires the control policy obtained using the learning process to be admissible at every step.

The contributions of the paper include: (a) development of a novel hybrid Q-learning scheme using event-sampled states, input vector and their history; (b) the derivation of a time-driven and hybrid Q-learning scheme for an uncertain large-scale interconnected system enclosed by a communication network without any assumptions on coupling terms; (c) a decentralized event-sampling condition based on Lyapunov function without needing a mirror estimator at the sensor; and (d) demonstration of closed-loop stability for such system using Lyapunov analysis.

This paper uses, $\Re$ to denote the set of all real numbers, Euclidean norm for vectors and Frobenius norm for matrices. The next section introduces the system description followed by the derivation of time-driven Q-learning scheme for large-scale interconnected systems with periodic feedback.

2. BACKGROUND

2.1 System Description

Consider a linear time-invariant continuous-time system having N interconnected subsystems shown in Fig. 2.1 with subsystem dynamics described by

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

where $x_i, \dot{x}_i \in \Re^{n_i \times 1}$ represent the state vector and state derivatives respectively, $u_i \in \Re^{m_i}$, $A_i \in \Re^{n_i \times n_i}$ and $B_i \in \Re^{n_i \times m_i}$ denote control input, internal dynamics and control gain matrices of the i^{th} subsystem, $A_{ij} \in \Re^{n_i \times n_j}$ represents the interconnection matrix between the i^{th} and j^{th} subsystem, $i \in 1, 2, ..N$. The overall system description can be expressed in a compact form as

$$\dot{X}(t) = AX(t) + BU(t), X(0) = X_0, \tag{2}$$

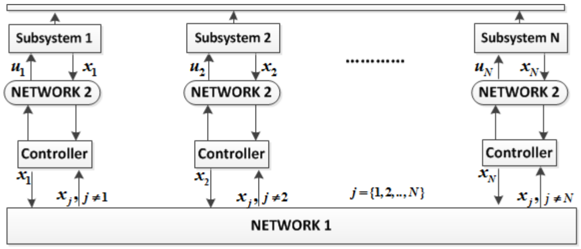


Fig. 2.1. Large-scale interconnected system.

where $X \in \Re^n, U \in \Re^m, B \in \Re^{n\times m}, A \in \Re^{n\times n}, \dot{X} = [\dot{x}_1^T, ., \dot{x}_N^T]^T$, $A = \begin{pmatrix} A_1 \ldots A_{1N} \\ : \; \ldots \; : \\ A_{N1} \ldots \, A_N \end{pmatrix}, B =$ $diag[B_1, ., B_N], U = [u_1^T, ., u_N^T]^T$. The system dynamics A_i, B_i and the interconnection matrix A_{ij} are considered uncertain. In the large-scale interconnected system, the subsystems communicate with each other via Network 1, while each subsystem is also enclosed by Network 2. Effects of the network-induced losses can be modeled along with the system dynamics by utilizing the standard and mild assumptions as listed in [7, 24].

Assumption 1 *The system (2) is considered controllable and the states are measurable. Further, the order of subsystems is considered known.*

With the network-induced delays and data-dropout, the original plant can be represented as

$$\dot{X}(t) = AX(t) + \gamma_{ca}(t)BU(t - \tau(t)), \qquad X(0) = X_0, \tag{3}$$

where $\gamma_{ca}(t)$ is the data-dropout indicator, which becomes $I^{n\times n}$ when the control input is received at the actuator and $0^{n\times n}$ when the control policy is lost at time t. This only includes the data loss in Network 2 and $\tau(t)$ is the total delay. Now, integrating the system dynamics with network parameters over the sampling interval [7, 24], we get

$$X_{k+1} = A_d X_k + \gamma_{ca,k} B_0^k U_k + \gamma_{ca,k-1} B_1^k U_{k-1} + \ldots + \gamma_{ca,k-\bar{d}} B_{\bar{d}}^k U_{k-\bar{d}}, \qquad X(0) = X_0, \tag{4}$$

where $X_k = X(kT_s), A_d = e^{AT_s}$, $\bar{d}$ is the delay bound, U_k is the control input. B_0^k, B_i^k, $\forall i = \{1, 2, \ldots \bar{d}\}$ are all defined as in [7]. From the discretized system representation, we can define an augmented state vector consisting of state and past control inputs as $\bar{X}(k) = [X_k^T \; U_{k-1}^T \cdots U_{k-\bar{d}}^T]^T \in \Re^{n+\bar{d}m}$.

The new augmented system representation is given by

$$\bar{X}_{k+1} = A_{\bar{x}k}\bar{X}_k + B_{\bar{x}k}U_k, \quad \bar{X}(0) = \bar{X}_0, \tag{5}$$

with the matrices $A_{\bar{x}k} = \begin{bmatrix} A_d\gamma_{ca,k-1}B_1^k \cdots \gamma_{ca,k-\bar{d}}B_{\bar{d}}^k & & & \\ 0 & \cdots & 0 & 0 \\ \vdots & I_m & \vdots & \vdots \\ 0 & \cdots & I_m & 0 \end{bmatrix}$ and $B_{\bar{x}k} = \begin{bmatrix} \gamma_{ca,k}B_0^k \\ I_m \\ \vdots \\ 0 \end{bmatrix}$.

Remark 1 *Note that the system dynamics are now stochastic due to the network-induced delays and data-dropouts. The assumptions regarding the controllability, observability and the existence of unique solution for the stochastic Riccati equation (SRE) are now dependent on the Grammian functions [1].*

Hence, the following assumption is needed to proceed further.

Assumption 2 *The system is both uniformly completely observable and controllable [1].*

The time-driven Q-learning and adaptive optimal regulation of such stochastic linear time-varying interconnected system is presented next.

2.2 Periodically Sampled Time-Driven Q-Learning

For the system dynamics (5), the infinite horizon cost function is defined as

$$J_k = \underset{\tau,\gamma}{E}\left[\frac{1}{2}\sum\nolimits_{t=k}^{\infty} \bar{X}_k^T P_{\bar{x}}\bar{X}_k + U_k^T R_{\bar{x}} U_k\right] \tag{6}$$

where $P_{\bar{x}} = diag(P, \frac{R}{\bar{d}}, .., \frac{R}{\bar{d}}), R_{\bar{x}} = \frac{R}{\bar{d}}$. The penalty matrices P, R are positive semidefinite and positive definite respectively. $\underset{\tau,\gamma}{E}(.)$ denotes the expected value of the stochastic process (.).

The cost function (6) can also be represented as $J_k = \underset{\tau,\gamma}{E}[\bar{X}_k^T S_k \bar{X}_k]$ with S_k being the symmetric positive semi-definite solution of the SRE [1]. The next step is to define the action-dependent Q-function for the stochastic system (5) with the cost-to-go function (6) as

$$Q(\bar{X}_k, U_k) = \underset{\tau,\gamma}{E}\left[r(\bar{X}_k, U_k) + J_{k+1}\left|\bar{X}_k\right.\right] = \underset{\tau,\gamma}{E}\{[\bar{X}_k^T U_k^T]G_k[\bar{X}_k^T U_k^T]^T\} \tag{7}$$

where $r(\bar{X}_k, U_k) = \bar{X}_k^T P_{\bar{x}} \bar{X}_k + U_k^T R_{\bar{x}} U_k$ and G_k is a time-varying matrix. From the Bellman equation, we get

$$\begin{bmatrix} \bar{X}_k \\ U_k \end{bmatrix}^T \underset{\tau,\gamma}{E}(G_k) \begin{bmatrix} \bar{X}_k \\ U_k \end{bmatrix} = \begin{bmatrix} \bar{X}_k \\ U_k \end{bmatrix}^T \begin{bmatrix} P_{\bar{x}} + \underset{\tau,\gamma}{E}(A_{\bar{x}k}^T S_{k+1} A_{\bar{x}k}) & \underset{\tau,\gamma}{E}(A_{\bar{x}k}^T S_{k+1} B_{\bar{x}k}) \\ \underset{\tau,\gamma}{E}(B_{\bar{x}k}^T S_{k+1} A_{\bar{x}k}) & R_{\bar{x}} + \underset{\tau,\gamma}{E}(B_{\bar{x}k}^T S_{k+1} B_{\bar{x}k}) \end{bmatrix} \begin{bmatrix} \bar{X}_k \\ U_k \end{bmatrix} \tag{8}$$

where $\underset{\tau,\gamma}{E}(G_k) = \begin{bmatrix} \underset{\tau,\gamma}{E}(G_k^{\bar{x}\bar{x}}) \, \underset{\tau,\gamma}{E}(G_k^{\bar{x}U}) \\ \underset{\tau,\gamma}{E}(G_k^{U\bar{x}}) \underset{\tau,\gamma}{E}(G_k^{UU}) \end{bmatrix}$

From the matrix equation (8), the time-varying control gain can be expressed as

$$K_k = \underset{\tau,\gamma}{E}\{[R_{\bar{x}} + B_{\bar{x}k}^T S_{k+1} B_{\bar{x}k}]^{-1} B_{\bar{x}k}^T S_{k+1} A_{\bar{x}k}\} = \underset{\tau,\gamma}{E}\{(G_k^{UU})^{-1} G_k^{U\bar{x}}\} \tag{9}$$

The Q-function (7) in parametric form is given by

$$Q(\bar{X}_k, U_k) = \underset{\tau,\gamma}{E}(z_k^T G_k z_k) = \underset{\tau,\gamma}{E}(\Theta_k^T \xi_k) \tag{10}$$

where $z_k = [(\gamma_{sc,k}\bar{X}_k)^T U_k^T]^T \in \Re^{\bar{l}}$ with $\bar{l} = m + n + m\bar{d}$, $\xi_k = z_k^T \otimes z_k$ is the regression vector. $\otimes$ denotes Kronecker product, and $\Theta_k \in \Omega_\Theta \subset \Re^{l_g}$ is formed by vectorization of the parameter matrix G_k. $\gamma_{sc,k}$ is a packet loss indicator, defined similar to $\gamma_{ca,k}$. The estimate of the optimal Q-function is expressed as

$$\hat{Q}(\bar{X}_k, U_k) = \underset{\tau,\gamma}{E}(z_k^T \hat{G}_k z_k) = \underset{\tau,\gamma}{E}(\hat{\Theta}_k^T \xi_k) \tag{11}$$

where $\hat{\Theta}_k \in \Re^{l_g}$ is the estimate of expected target parameter Θ_k. By Bellman's principle of optimality, the optimal value function satisfies

$$0 = \underset{\tau,\gamma}{E}(J_{k+1}^* \left|\bar{X}_k\right.) - \underset{\tau,\gamma}{E}(J_k^*) + \underset{\tau,\gamma}{E}(r(\bar{X}_k, U_k)) = \underset{\tau,\gamma}{E}(r(\bar{X}_k, U_k)) + \underset{\tau,\gamma}{E}(\Theta_k^T \Delta\xi_k), \tag{12}$$

where $\Delta\xi_k = \xi_{k+1} - \xi_k$, and $\underset{\tau,\gamma}{E}(J_{k+1}^* \left|\bar{X}_k\right.)$ is the expected cost-to-go at $k+1^{st}$ instant, given the state information of the k^{th} instant. Since the estimated Q-function does not satisfy (12), the temporal difference (TD) error will be observed as

$$e_B(k) = \underset{\tau,\gamma}{E}(r(\bar{X}_k, U_k) + \hat{\Theta}_k^T \Delta\xi_k) \tag{13}$$

Remark 2 *In the iterative learning schemes [2, 4] the parameters of the Q-function estimator (QFE) is updated by minimizing the error in (13) until the error converges to a small value for every time step k. On the contrary, time-driven ADP schemes [6, 7] calculate the Bellman error at each step and update once at the sampling instant and the stability of the closed-loop system is established under certain mild assumptions on the estimated control policy.*

The overall cost function (6) for the large-scale system (5), can be represented as the sum of the individual cost of all the subsystems as, $J_k = \sum_{i=1}^{N} J_{i,k}$, where $J_{i,k} = \underset{\tau,\gamma}{E}\{\frac{1}{2}\sum_{s=k}^{\infty} \bar{x}_{i,k}^T P_{\bar{x},i}\bar{x}_{i,k}$ $+u_{i,k}^T R_{\bar{x},i}u_{i,k}\}$ is the quadratic cost function for i^{th} subsystem with $\bar{x}_i$ representing the augmented states of the i^{th} subsystem, $P_{\bar{x}} = diag\{P_{\bar{x},1}\cdots P_{\bar{x},N}\}$ and $R_{\bar{x}} = diag\{R_{\bar{x},1}\cdots R_{\bar{x},N}\}$. The optimal control sequence to minimize the quadratic cost function (6) in a decentralized framework is not straightforward because of the interconnection dynamics. The optimal control policy for each subsystem, which minimizes the cost function (6), is obtained by using the SRE of the overall system given the system dynamics $A_{\bar{x}k}$ and $B_{\bar{x}k}$, as

$$u_{i,k}^* = \underset{\tau,\gamma}{E}\{-K_{i,k}^*\bar{x}_{i,k} - \sum_{j=1,j\neq i}^{N} K_{ij,k}^*\bar{x}_{j,k}\} \tag{14}$$

where $K_{i,k}^*$ are the diagonal elements and $K_{ij,k}^*$ are the off-diagonal elements, of K_k^* in (9). In the following lemma, it is shown that, with the control law (14) designed at each subsystem, the overall system is asymptotically stabilized in the mean square.

Lemma 1 *Consider the i^{th} subsystem of the large-scale interconnected system (5). Assuming that the system matrices $A_{\bar{x}k}, B_{\bar{x}k}$ are known along with Assumption-2. The optimal control policy obtained from (14) renders the individual subsystems asymptotically stable in the mean square.*

Proof: Note that the optimal control input is stabilizing [1]. Therefore, the closed-loop system matrix $(A_{\bar{x}k} - B_{\bar{x}k}K_k^*)$ is Schur. The Lyapunov equation $(A_{\bar{x}k} - B_{\bar{x}k}K_k^*)^T\bar{P}(A_{\bar{x}k} - B_{\bar{x}k}K_k^*) - \bar{P} = -\bar{F}$, has a positive definite solution $\bar{F}$. Consider the Lyapunov function candidate $L_k = \underset{\tau,\gamma}{E}(\bar{X}_k^T\bar{P}\bar{X}_k)$, with $\bar{P}$ being positive definite. The first difference, using the overall system dynamics with optimal control input is $\Delta L_k = -\underset{\tau,\gamma}{E}(\bar{X}_k^T \bar{F} \bar{X}_k)$. Since, $\bar{F}$ can be chosen as a diagonal matrix, the first difference in terms of the subsystems can be expressed as

$$\Delta L_k = -\sum\nolimits_{i=1}^{N} \underset{\tau,\gamma}{E}(\bar{x}_{i,k}^T \bar{F}_i \bar{x}_{i,k}) \leq -\sum\nolimits_{i=1}^{N} \bar{q}_{\min} \underset{\tau,\gamma}{E} \left\| \bar{x}_{i,k} \right\|^2 \tag{15}$$

where $\bar{q}_{\min}$ is the minimum singular value of $\bar{F}$. This implies the subsystems are asymptotically stable in the mean square. The results of this lemma will be used in the stability analysis of the interconnected system where the need for the accurate knowledge of $A_{\bar{x}k}, B_{\bar{x}k}$ will be relaxed. The controller design using a novel hybrid Q-learning based ADP approach for such large-scale interconnected system in the presence of network-induced losses and with intermittent feedback will be discussed next.

3. DISTRIBUTED EVENT-BASED HYBRID Q-LEARNING SCHEME

In this section, a novel hybrid learning scheme, which utilizes time-driven Q-learning based ADP approach, for the control of large-scale interconnected system to improve the convergence time with event-sampled state and input vector will be introduced. In the proposed algorithm, the idle-time between two events is utilized to perform limited parameter updates iteratively in order to minimize the Bellman error. With the finite number of iterations between any two events varying, the control policy need not necessarily converge to an admissible policy and the stability of the closed-loop system cannot be established either using the traditional iterative ADP schemes [4] or the time-driven Q-learning schemes [6, 7].

An additional challenge is to estimate the Q-function parameters in (11) for the system defined in (5) with intermittent feedback and in the presence of network-induced losses. Since subsystems broadcast their states via the communication network, each local subsystem can estimate the Q-function of the overall system so that a predefined reference model is not needed. Subsequently, the optimal control gains and the decoupling gains for each subsystem can be computed without using the complete knowledge of the system dynamics and interconnection matrix.

Although, the estimation of the Q-function at each subsystem increases the computation, this additional computation can be considered as trade-off for relaxing the assumption on the strength of interconnection terms and estimating optimal control policy. With the following assumption, the Q-function estimator design will be presented for intermittent feedback.

Assumption 3 *The target parameters are assumed to be slowly-varying [29].*

3.1 Time-Driven Q-Learning With Intermittent Feedback

In the case of an event-sampled system, the system state vector $\bar{X}_k$ is sent to the controller at event-sampled instants. To denote the event-sampling instants, we define a sub-sequence $\{k_l\}_{l\in\mathbb{N}}, \forall k \in \{0, \mathbb{N}\}$ with $k_0 = 0$ being the initial sampling instant and $\mathbb{N}$ is the set of natural numbers.

The system state vector $\bar{X}_{k_l}$ sent to the controller is held by ZOH until the next sampling instant, and it is expressed as $\bar{X}_k^e = \bar{X}_{k_l}, \quad k_l \le k < k_{l+1}$. The corresponding error referred to as an event-sampling error can be expressed as

$$e_{ET}(k) = \bar{X}_k - \bar{X}_k^e, \quad k_l \le k < k_{l+1}, \quad l = 1, 2, \cdots \tag{16}$$

Since the estimation of G_k must use $\bar{X}_k^e$, the Q-function estimate can be expressed as

$$\hat{Q}(\bar{X}_k^e, U_k) = \underset{\tau,\gamma}{E}\,(z_k^{e,T}\hat{G}_k z_k^e) = \underset{\tau,\gamma}{E}\,(\hat{\Theta}_k^T \xi_k^e), \qquad k_l \le k < k_{l+1} \tag{17}$$

where $z_k^e = \left[(\gamma_{sc,k}\bar{X}_k^e)^T U_k^T\right]^T \in \Re^{\bar{l}}$ and $\xi_k^e = z_k^{e,T} \otimes z_k^e$ being the event-sampled regression vector and $\hat{\Theta}$ is the result of vectorization of the matrix $\hat{G}_k$. The Bellman error with event-sampled state is

$$e_B(k) = \underset{\tau,\gamma}{E}\left[r(\bar{X}_k^e, U_k) + \hat{\Theta}_k^T \Delta\xi_k^e\right], \quad k_l \le k < k_{l+1} \tag{18}$$

where $r(\bar{X}_k^e, U_k) = \bar{X}_k^{e,T} P_{\bar{x}} \bar{X}_k^e + U_k^T R_{\bar{x}} U_k$, $\Delta\xi_k^e = \xi_{k+1}^e - \xi_k^e$. The Bellman error (28) can be rewritten as

$$e_B(k) = \underset{\tau,\gamma}{E}\left\{r(\bar{X}_k, U_k) + \hat{\Theta}_k^T \Delta\xi_k \ + \Xi_s\left(\bar{X}_k, e_{ET}(k), \hat{\Theta}_k\right)\right\} \tag{19}$$

where $\Xi_s\left(\bar{X}_k, e_{ET}(k), \hat{\Theta}_k\right) = r(\bar{X}_k - e_{ET}(k), U_k) - r(\bar{X}_k, U_k) + \hat{\Theta}_k^T(\Delta\xi_k^e - \Delta\xi_k)$.

Remark 3 *By comparing (19) with (13), the Bellman error in (19) has an additional error term which is $\Xi_s\left(\bar{X}_k, e_{ET}(k), \hat{\Theta}_k\right)$. This additional error consists of errors in cost-to-go, and the regression vector, which are driven by $e_{ET}(k)$. Hence, the estimation of QFE parameters depends upon the frequency of the event-sampling instants.*

The QFE estimated parameter vector, $\hat{\Theta}_k^i$, is tuned only at the event-sampling instants. The superscript i denotes the overall system parameters at the i^{th} subsystem and the estimated control policy can be computed as

$$U_k^i = -\hat{K}_k^i \bar{X}_k^{i,e} = -(\hat{G}_k^{i,uu})^{-1}(\hat{G}_k^{i,ux})\bar{X}_k^{i,e} \tag{20}$$

By using (20), the event-based estimated control input for the i^{th} subsystem is given by

$$u_{i,k} = -\hat{K}_{i,k}\bar{x}_{i,k}^e - \sum\nolimits_{j=1,j\neq i}^{N} \hat{K}_{ij,k}\bar{x}_{j,k}^e, \qquad k_l \le k < k_{l+1} \quad \forall i \in \{1, 2, ..N\} \tag{21}$$

Remark 4 *It should be noted that the optimal controllers designed at each subsystem takes into account the structural constraint which are present in the form of the interconnection matrix. However, the consideration of input, state and time constraints [1] as a part of the optimal control problem is reserved for future work.*

With the following assumption, the parameter update rule for the Q-function estimator will be presented.

Assumption 4 *The target parameter vector Θ_k is assumed to be bounded by positive constant, such that $\|\Theta_k\| \le \Theta_M$. The regression function $Z^i(\bar{X}_k)$ is locally Lipschitz for all $\bar{X}_k \in \Omega_x$.*

3.2 Parameter Update At Event-Sampling Instants

The QFE parameter vector $\hat{\Theta}^i_k$, is tuned by using the past data of the Bellman error (19) that is available at the event-sampling instants. Therefore, the auxiliary Bellman error at the event-sampling instants is expressed as $\Xi_B{}^{i,e}(k) = \Pi^{i,e}_k + \hat{\Theta}^{i,T}_k Z^{i,e}_k$, for $k = k_l$, where $\Pi^{i,e}_k = [r(\bar{X}^i_{k^i_l}, U^i_{k^i_l}) \;\; r(\bar{X}^i_{k^i_{l-1}}, U^i_{k^i_{l-1}} \;\; \cdots \;\; r(\bar{X}^i_{k^i_{l-v-1}}, U^i_{k^i_{l-v-1}})] \in \Re^{1\times v}$ and $Z^{i,e}_k = [\Delta\xi^i_{k^i_l}, \;\; \Delta\xi^i_{k^i_{l-1}}, \;\; \cdots \;\; \Delta\xi^i_{k^i_{l-v-1}}] \in \Re^{l_g \times v}$.

Remark 5 *A larger time history may lead to faster convergence, but it results in higher computation. The number of history values v is not fixed and a value $v < l$ is found suitable during simulation studies.*

Next, select the update law [29] for the QFE parameter vector $\hat{\Theta}^i_k$ tuned only at the event-sampling instants, as

$$\hat{\Theta}^i_k = \hat{\Theta}^i_{k-1} + \frac{W^i_{k-2} Z^{i,e}_{k-1} \Xi_B^{i,e^T}(k-1)}{1 + Z^{i,e^T}_{k-1} W^i_{k-2} Z^{i,e}_{k-1}}, \quad k = k_l \tag{22}$$

where

$$W^i_k = W^i_{k-1} - \frac{W^i_{k-1} Z^{i,e}_{k-1} Z^{i,e^T}_{k-1} W^i_{k-1}}{1 + Z^{i,e^T}_{k-1} W^i_{k-1} Z^{i,e}_{k-1}}, \quad k = k_l \tag{23}$$

with $W^i_0 = \beta I$, $\beta > 0$, a large positive value. The aperiodic execution of (22), saves computation, when compared to the traditional adaptive Q-learning techniques. The superscript i indicating the overall system parameters at the i^{th} subsystem, will be dropped from hereon. In the time-driven Q-learning scheme [6], the parameters of the QFE are not updated during the inter-event period. On the contrary, in the hybrid learning algorithm, the parameters are updated during the inter-event period and the update rules are presented next.

3.3 Iterative Parameter Update

The recursive least square (RLS) algorithm was used in [2, 4] to perform iterative updates within any two periodic sampling instants, using policy iteration. The update equation iteratively searches for a control policy that minimizes the Bellman error. Analytical results are provided in [2, 4] to show that each iterative update resulted in a control policy that is better than or as good as the existing control policy, in minimizing the Bellman error.

Since significant numbers of iterative updates are not viable for online control, the time-driven Q-learning [6, 7] was proposed which uses gradient descent based update equations at the sampling instants to minimize the Bellman error. It was shown in [6, 7] that as the sampling instants increases, the parameter estimation error converges to zero. In order to improve the estimation error convergence rate, the RLS update (22),(23) is used at the sampling instants in this work and convergence result similar to the time-driven Q-learning holds for the proposed algorithm without the iterative parameter updates presented next.

To utilize the time between two event-sampling instants, parameters are updated iteratively to minimize the error that was calculated during the previous event, which is expressed as $\Xi_B{}^{j,e}(k) = \Pi_k^{j,e} + \hat{\Theta}_k^{j,T} Z_k^{j,e}, \;\; k = k_l$, where j is the iteration index. The Q-function parameters are updated using the equations

$$\hat{\Theta}(k_l^j) = \hat{\Theta}(k_l^{j-1}) + \frac{W(k_{l-2}^{j-2})Z(k_{l-1}^{j-1})\Xi_B^T(k_{l-1}^{j-1})}{1 + Z^T(k_{l-1}^{j-1})W(k_{l-2}^{j-1})Z(k_{l-1}^{j-1})} \tag{24}$$

$$W(k_l^j) = W(k_{l-1}^{j-1}) - \frac{W(k_{l-1}^{j-1})Z(k_{l-1}^{j-1})Z^T(k_{l-1}^{j-1})W(k_{l-1}^{j-1})}{1 + Z^T(k_{l-1}^{j-1})W(k_{l-1}^{j-1})Z(k_{l-1}^{j-1})} \tag{25}$$

Whenever there is an event, the Q-function parameter vector which is updated iteratively using (24),(25) is passed on to the QFE to calculate the new Bellman error. The estimated control gain matrix can be obtained from the estimated parameter vector $\hat{\Theta}_k$ in (22) at each event-sampled instants. In terms of the estimated parameters, the control gains are given by (20), where

$$\hat{K}_k = (\hat{G}_k^{uu})^{-1}\hat{G}_k^{ux} = \begin{bmatrix} \hat{K}_1 \cdots \hat{K}_{1N} \\ \vdots \;\; \ddots \;\; \vdots \\ \hat{K}_{N1} \cdots \hat{K}_N \end{bmatrix} \tag{26}$$

is the estimated control gain. It is important to note that this control gain is obtained directly from the Q-function parameters which are constructed with the past data and the current feedback information, without using the system dynamics.

In the proposed algorithm, the update equations (24),(25) together with (28) search for an improved control policy during every inter-event period . Utilizing the Bellman error equation (28) to evaluate the existing control policy, the Q-function is iteratively updated between two event-sampling instants. However, in contrast to the algorithms in [2, 4], the iteration index j in (24),(25) depends on the event-sampling mechanism, resulting in finite, varying number of iterative updates between any two events.

Remark 6 *The control policy for the individual subsystem is given by (21). Since it is possible that $\hat{G}^{uu}_{k}$ might be rank-deficient during the learning phase, the following conditions are checked before the control law is updated. If $\hat{G}^{uu}_{k-1}$ is singular or if $\hat{G}^{uu}_{k-1} - R_{\bar{x}}$ is not positive definite, then, $\hat{G}^{uu}_{k-1}$ is replaced by $R_{\bar{x}}$ in the control policy. The conditions can be checked easily by calculating the eigenvalues of $\hat{G}^{uu}_{k-1}$.*

Remark 7 *The QFE parameter tuning law (22), (23) requires the state vectors X_{k_l} to $X_{k_{l-v-1}}$ for the computation of regression vector at $k = k_l$. Therefore, the past values are required to be stored at the value function estimator.*

With the update rules presented in this section and the control gains selected from (36), the assumption in [6, 7] that the inverse of $\hat{G}^{uu}_{k}$ exists when the updates utilize the time history of the regression function and Bellman error is also relaxed. The analytical results for the proposed learning algorithm is presented next.

3.4 Stability Analysis

Defining the QFE parameter estimation error $\underset{\tau,\gamma}{E}(\tilde{\Theta}_k) = \underset{\tau,\gamma}{E}(\Theta_k - \hat{\Theta}_k)$, the error dynamics using (22), (24) can be represented as

$$\underset{\tau,\gamma}{E}(\tilde{\Theta}^{0}_{k_{l+1}}) = \underset{\tau,\gamma}{E}(\tilde{\Theta}^{j}_{k} + \frac{W^{j}_{k} Z^{j,e}_{k} \Xi^{j,e^{T}}_{B}(k)}{1 + Z^{ej,T}_{k} W^{j}_{k} Z^{j,e}_{k}}), \quad k = k^{0}_{l} \tag{27}$$

$$\underset{\tau,\gamma}{E}(\tilde{\Theta}^{j+1}_{k_l}) = \underset{\tau,\gamma}{E}(\tilde{\Theta}^{j}_{k} + \frac{W^{j}_{k} Z^{j,e}_{k} \Xi^{j,e^{T}}_{B}(k)}{1 + Z^{ej,T}_{k} W^{j}_{k} Z^{j,e}_{k}}), \quad k^{0}_{l} < k < k^{0}_{l+1} \tag{28}$$

Remark 8 *When there is no data-loss, the Q-function estimator is updated and the control policy is updated as soon as it is computed. This requires the broadcast scheme to generate an acknowledgment signal whenever the packets are successfully received at the subsystems [22]. A suitable scheduling protocol has to ensure that the data lost in the network is kept minimal.*

Next an event-sampling condition has to be selected for the proposed scheme to work. Consider a quadratic function $f^i(k) = \bar{x}_i(k)^T \Gamma_i \bar{x}_i(k)$, with $\Gamma_i > 0$, for the i^{th} subsystem. The event-sampling condition should satisfy

$$f^i(k) \leq \lambda f^i(k_l + 1), \forall k \in [k_l + 1, k_{l+1}), \tag{29}$$

for stability, when $\lambda < 1$, as shown in the next section.

Remark 9 *The event-sampling condition presented here depends only on the local subsystem state information. The Lyapunov function based event-sampling condition is also presented in [23] for a single system. The hybrid learning algorithm presented in this paper is independent of the event-sampling condition.*

The following result will be used to prove the stability of the closed-loop system during the learning period.

Lemma 2 *Consider the system in (5) and the QFE (27). Define $\tilde{U}(k_{l-1}) = U(k_{l-1}) - \hat{U}(k_{l-1})$ and $\tilde{G}^{ux}_{k_{l-1}} = G^{ux}_{k_{l-1}} - \hat{G}^{ux}_{k_{l-1}}$. If the control policy is updated such that, whenever $\hat{G}^{uu}_{k_{l-1}} - R_{\bar{x}}$ is not positive definite or $\hat{G}^{uu}_{k_{l-1}}$ is singular, $\hat{G}^{uu}_{k_{l-1}}$ is replaced by $R_{\bar{x}}$ in the control policy, then*

$$\underset{\tau,\gamma}{E}\left(\tilde{U}(k_{l-1})\right) \leq \underset{\tau,\gamma}{E}\{2\left\|R_{\bar{x}}^{-1}\right\|\left\|G^{ux}_{k_{l-1}}\right\|\left\|\bar{X}_{k_{l-1}}\right\| + \left\|R_{\bar{x}}^{-1}\right\|\left\|\tilde{G}^{ux}_{k_{l-1}}\right\|\left\|\bar{X}_{k_{l-1}}\right\|\} \tag{30}$$

Proof: See Appendix.

Definition 1 *[29] A regression vector $\varphi(x_k)$ is said to be persistently exciting if there exists positive constants δ,$\underline{\alpha}$,$\bar{\alpha}$ and $k_d \geq 1$ such that $\underline{\alpha} I \leq \sum_{k=k_d}^{k+\delta} \varphi(x_k)\varphi^T(x_k) \leq \bar{\alpha} I$, where I is the identity matrix of appropriate dimension.*

Lemma 3 *Consider both the QFE in (27) with an initial admissible control policy* $U_0 \in \Re^m$. *Let the Assumption 1-4 hold, and the QFE parameter vector* $\hat{\Theta}(0)$ *be initialized in a compact set* Ω_Θ. *When the QFE is updated at the event-sampling instants using (22),(23) and during the inter-sampling period using (24),(25), the QFE parameter estimation error* $\underset{\tau,\gamma}{E}(\tilde{\Theta}^j_{k_l})$ *is bounded. Under the assumption that the regression vector* $\xi^j_{k_l}$ *satisfies the PE condition, the QFE parameter estimation error* $\tilde{\Theta}^j_{k_l}$ *for all* $\hat{\Theta}(0) \in \Omega_\Theta$ *converges to zero asymptotically in the mean square, with event-sampled instants* $k_l \longrightarrow \infty$.

Proof: See Appendix.

Remark 10 *Covariance resetting technique [29] is used to reset W whenever* $W \leq W_{min}$. *This condition will also be used in the Lyapunov analysis to ensure stability of the closed-loop system. With the covariance resetting, the parameter convergence proof in Lemma 3 will still be valid [29].*

Next, the Lyapunov analysis is used to derive the conditions for the stability of the closed-loop system, with the controller designed in this section.

Theorem 1 *Consider the closed-loop system (5), parameter estimation error dynamics (37) along with the control input (20). Let the Assumptions 1-4 hold, and let* $U(0) \in \Omega_u$ *be an initial admissible control policy. Suppose the last held state vector,* $\bar{X}^{e,j}_{k_l}$, *and the QFE parameter vector,* $\hat{\Theta}^j_{k_l}$ *are updated by using, (22),(23) at the event-sampled instants, and (24),(25) during the inter-sampling period. Then, there exists a constant* $\gamma_{\min} > 0$ *such that the closed-loop system state vector* $\bar{X}^j_{k_l}$ *for all* $\bar{X}(0) \in \Omega_x$ *converges to zero asymptotically in the mean square and the QFE parameter estimation error* $\tilde{\Theta}^j_{k_l}$ *for all* $\hat{\Theta}(0) \in \Omega_\Theta$ *remains bounded. Further, under the assumption that the regression vector* $\xi^j_{k_l}$ *satisfies the PE condition, the QFE parameter estimation error* $\tilde{\Theta}^j_{k_l}$ *for all* $\hat{\Theta}(0) \in \Omega_\Theta$ *converges to zero asymptotically in the mean square, with event-sampled instants* $k_l \longrightarrow \infty$, *provided the inequality* $\gamma_{\min} > \mu + \rho_1$ *is satisfied.* $\gamma_{min}, \mu, \rho_1$ *are positive constants, defined in the proof.*

Proof: See Appendix.

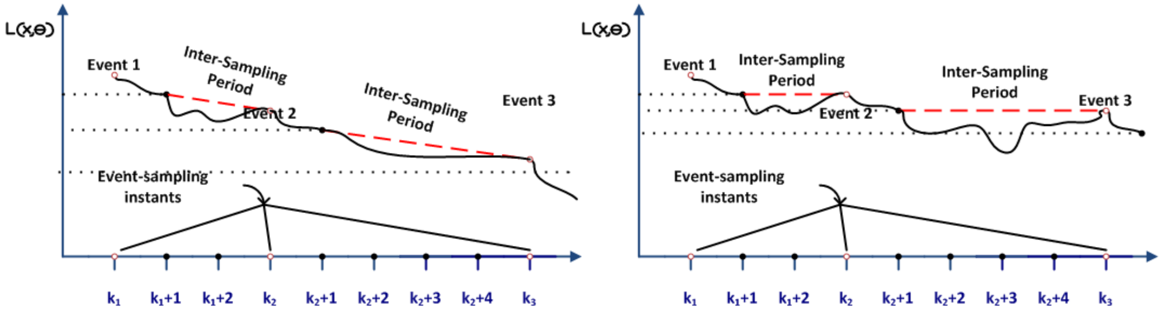
The evolution of the Lyapunov function is depicted in Fig. 2.2a. During the event-sampled instant, due to the updated control policy (21), the Lyapunov function decreases. Due to the event-sampling condition (39) and the iterative learning within the event-sampling instants, the Lyapunov function decreases during the inter-sampling period.

Since the iterative learning does not take place in the time-driven Q-learning [6], the first difference of the parameter estimation error is zero for the inter-event period. This makes the Lyapunov function negative semi-definite during this period. The evolution of the Lyapunov function is depicted in Fig. 2.2b for the time-driven Q-learning.

Remark 11 *The design constants $R_{\bar{x}}, W_{min}, W_0$ are selected based on the inequalities that are analytically derived in Theorem 1 using the bounds on $A_{\bar{x}k}, B_{\bar{x}k}, S_k$. Then, the constants Γ and $\bar{\Pi}$ can be found to ensure closed-loop system stability.*

Remark 12 *The requirement of PE condition is necessary so that the regression vector is non-zero until the parameter error goes to zero. By satisfying the PE condition in the regression vector, the expected value of the parameter estimation error $\tilde{\Theta}_k$ will converge to zero. This PE signal is viewed as the exploration signal in the reinforcement learning literature [4].*

Remark 13 *An initial identification process can be used to obtain the nominal values of $A_{\bar{x}k}, B_{\bar{x}k}$ which can be used to initialize the Q-function parameters.*



(a) (b)

Fig. 2.2. Evolution of the Lyapunov function (a) Hybrid learning. (b) TD learning.

Remark 14 *The algorithm proposed in this section can be used as a time-driven Q-learning scheme by not performing the iterative learning between the event-sampling instants, in stochastic framework. Also, if the iteration index, $j \to \infty$, for each k_l, the algorithm becomes the traditional policy iteration based ADP scheme.*

The event-sampling and broadcast algorithm for the subsystems followed by the proposed hybrid learning algorithm is summarized next.

3.5 Proposed Algorithm

For estimating the overall Q-function locally, we will use the following request-based event-sampling algorithm. Consider an event occurring at the i^{th} subsystem at the sampling instant k_l. This subsystem generates a request signal and broadcasts it with its state information to the other subsystems. Upon receiving the broadcast request, the other subsystems broadcast their respective state information to all the subsystems. This can be considered as a forced event at the other subsystems.

Remark 15 *The events at all the subsystem occur asynchronously based on the local event-sampling condition, whereas the Q-function estimator and control policy remain synchronized at each subsystem due to the forced event. The request signal is considered to be broadcasted without any delay in Network 1 in Fig. 2.1.*

The algorithm for the hybrid learning scheme is summarized as Algorithm 1. The proposed control scheme is tested via simulation and the results are presented next.

4. SIMULATION RESULTS AND DISCUSSION

A system of N interconnected inverted pendulums, coupled by a spring is considered for the verification of the analytical design. The dynamics are $\dot{x}(t) = \begin{bmatrix} 0 & 1 \\ \frac{g}{l} - \frac{a_i k}{ml^2} & 0 \end{bmatrix} x_i(t) + \begin{bmatrix} 0 \\ \frac{1}{ml^2} \end{bmatrix} u_i(t) + \sum_{j \in N_i} \begin{bmatrix} 0 & 0 \\ \frac{h_{ij} k}{ml^2} & 0 \end{bmatrix} x_j(t)$ where $l = 2$, $g = 10$, $m = 1$, $k = 5$ and $h_{ij} = 1$ for $\forall j \in \{1, 2, .., N\}$. The system is open loop unstable.

Algorithm 1 Hybrid Q-Learning for Intermittent feedback

```
 1: Initialize $\hat{\Theta}_0^j, W_0^j, U_0$
 2: for Event-sampling instants: $l = 0 \rightarrow \infty$ do
 3:     if Event = Yes then
 4:         Calculate Bellman Error $e_B(k_l^j)$
 5:         Update $\hat{\Theta}_{k_l}^j, W_{k_l}^j$
 6:         Update the control input at the actuator $U_{k_l}$
 7:         Pass the parameters $\hat{\Theta}_{k_l}^0, W_{k_l}^0, e_B(k_l^0)$ for iterations
 8:     else
 9:         for Iterative Index: $j = 0 \rightarrow \infty$ do
10:             Update $\hat{\Theta}_{k_l}^j, W_{k_l}^j$ with $e_B(k_l^j)$
11:             Calculate $e_B(k_l^{j+1})$
12:             if $e_B(k_l^{j+1})$-$e_B(k_l^j) < \epsilon$ or Event = Yes then
13:                 Pass the Parameters $\hat{\Theta}_{k_l}^j, W_{k_l}^j$ to QFE
14:                 Goto 4:
15:             end if
16:             $j = j + 1$
17:         end for
18:     end if
19:     if $e_B(k_{l+1}^0)$-$e_B(k_l^0) < \epsilon$ then
20:         Stop PE Condition
21:     end if
22:     $l = l + 1$
23: end for
```

Ideal network: The system is discretized with a sampling time of 0.1 sec. With $P_i = I_{2\times 2}$ and $R_i = 1$, $\forall i = 1, 2, 3$, the initial states for the system was selected as $x_1 = [2 \quad -3]^T$, $x_2 = [-1 \quad 2]^T$ and $x_3 = [-1 \quad 1]^T$ and $W(0) = 500, \lambda = 0.6, W_{min} = 250$. For the PE condition, Gaussian white noise with zero mean and 0.2 standard deviation was added to the control inputs. The initial parameters of the QFE is obtained by solving the SRE of the nominal model of the system. Under the ideal case, without network-induced losses, the comparison between time-driven Q-learning versus the proposed hybrid learning scheme shown in Fig. 2.3a, verifies that the convergence rate is faster in the hybrid learning scheme with event-sampled feedback. This is due to the iterative parameter update within the inter-event period.

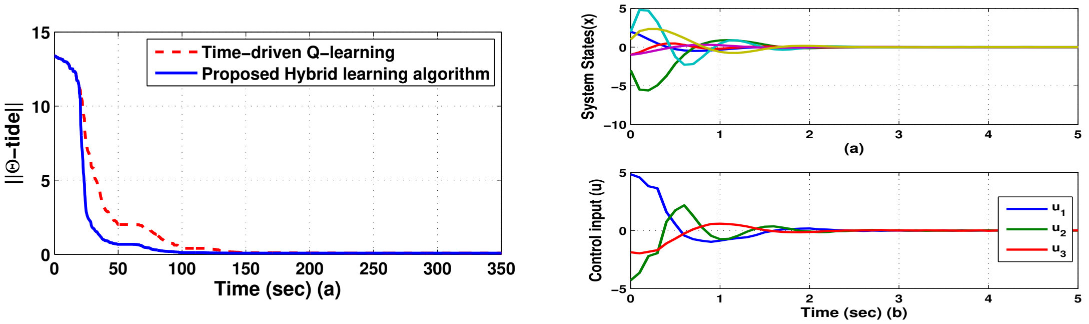


Fig. 2.3. (a) Estimation error comparison - ideal network (b) State and control trajectories with delays.

Monte-Carlo analysis: The simulation is carried out with random delays ($\bar{d} = 2$) introduced by the network. The delay is characterized by normal distribution with 80 ms mean and 10 ms standard deviation and a Monte-Carlo analysis is carried out for 500 iterations. In the case where the random delays are considered, the state and control trajectories are stable during the learning period as seen in Fig. 2.3b. The comparison between the time-driven Q-learning and the proposed hybrid learning schemes as seen in Fig. 2.4b shows that parameter error convergence in the hybrid scheme is much faster, which shows that the hybrid learning algorithm is more robust than the time-driven Q-learning in the presence of delays. This is partly due to augmented state vector and iterative parameter learning within the event-sampled instants.

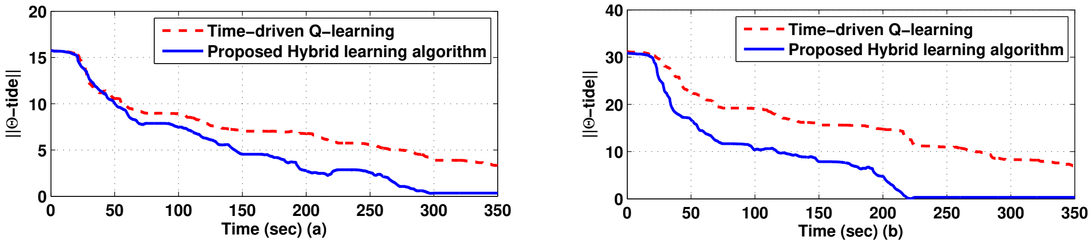


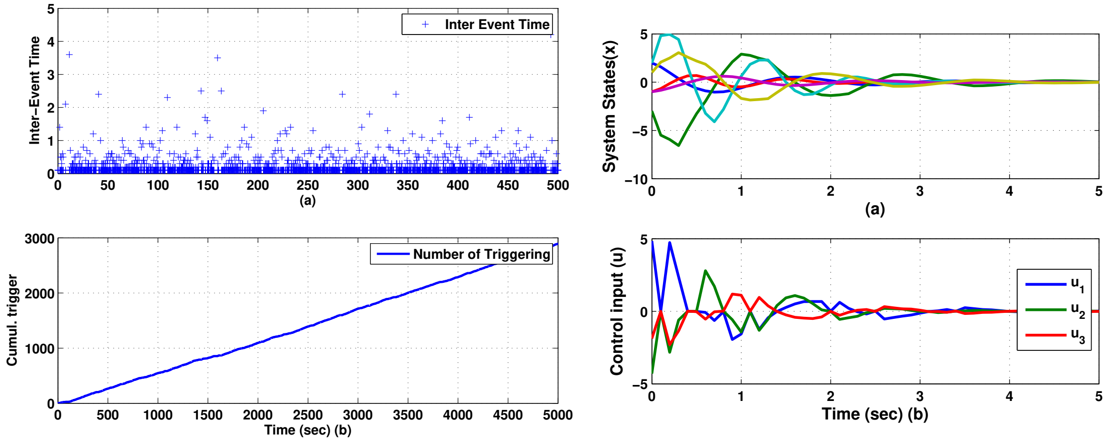
Fig. 2.4. Estimation error comparison (a) with 10% packet loss. (b) without packet loss.

Table 2.1. Comparison of parameter error convergence time.

Mean-delay (in ms)	% Data-drop out	Convergence time (in sec)	
		Time-driven Q-learning	Hybrid Learning algorithm
0	0	13.6	10.7
30	10	61.0	36.9
	25	246	190.0
80	10	632.0	317.0
	25	486.5	269.0
100	10	239.0	198.0
	25	637.3	239.8

Random packet-losses characterized with Bernoulli distribution is introduced keeping the probability of data lost as 10%. All design parameters are kept the same. Table 2.1, lists the convergence time for the parameter estimation error for the existing time-driven Q-learning algorithm and the proposed hybrid learning algorithm. The error threshold was defined as $10^{-}2$ and the design parameters were unchanged. In the ideal case, when there are no network losses, the difference in the convergence time for the two algorithms is small. As the network losses are increased, the parameter error converges to the threshold much faster with the proposed hybrid learning algorithm. It is clear that with the hybrid learning scheme the estimation error converges much quicker than the time-driven Q-learning scheme per the information given in the Table 2.1.

The total number of events, the state and control policy during the learning period is shown in Fig. 2.5a and 2.5b respectively. With the hybrid learning algorithm, the stability of the system is not affected during the learning period. As the events are spaced out, more number of iterative parameter updates takes place within the inter-event period. Simulation figures for all the cases are not included due to space consideration.



(a) (b)

Fig. 2.5. (a) Inter-event time and cumulative events. (b) State and control trajectories with packet-loss.

5. CONCLUSIONS

The proposed hybrid Q-learning based scheme for a large-scale interconnected system appears to guarantee a desired performance. The stability conditions for the closed-loop system during the learning period is derived using the Lyapunov stability analysis. Q-function parameters for the entire system are estimated at each subsystem with the event-sampled inputs, states and past state vectors. This control scheme does not impose any assumptions on the interconnection strengths. The mirror estimator is not used in the event-sampling mechanism and reference models for each subsystems are not needed. With the help of the simulation study, the proposed analytical design is verified. From the simulation results the proposed algorithm appears to provide advantages over the existing model-free Q-learning scheme for online control.

The proposed hybrid approach utilizes past input and state information for each subsystems and state information from other systems via communication network and therefore the net result is the design of a data-driven optimal regulator for a class of large-scale interconnected systems.

ACKNOWLEDGMENT

This research supported in part by NSF ECCS #1128281 and #1406533 and Intelligent Systems Center, at the Missouri University of Science and Technology, Rolla.

REFERENCES

[1] Lewis, F.L., and Syrmos, V.L.: 'Optimal control' (John Wiley & Sons, 1995).

[2] Bradtke, S.J., Ydstie, B.E., and Barto, A.G.: 'Adaptive linear quadratic control using policy iteration'. Proc. American Control Conference, July 1994, pp. 3475–3479.

[3] Al-Tamimi, A., Lewis, F.L., and Abu-Khalaf, M.:'Model-free Q-learning designs for linear discrete-time zero-sum games with application to H-infinity control', Automatica, 2007, **43**, (3), pp 473–481.

[4] Lewis, F.L., Vrabie, D., and Vamvoudakis, K.G.:'Reinforcement learning and feedback control: Using natural decision methods to design optimal adaptive controllers', IEEE Control Systems, 2012, **32**, (6), pp 76–105.

[5] Dierks, T., and Jagannathan, S.:'Online optimal control of affine nonlinear discrete-time systems with unknown internal dynamics by using time-based policy update', IEEE Transactions on Neural Networks and Learning Systems, 2012, **23**, (7), pp 1118–1129.

[6] Sahoo, A., and Jagannathan, S.: 'Event-triggered optimal regulation of uncertain linear discrete-time systems by using Q-learning scheme'. Proc. IEEE 53rd Annual Conference on Decision and Control, Los Angeles, CA, Dec. 2014, pp. 1233–1238.

[7] Xu, H., Jagannathan, S., and Lewis, F.L.:'Stochastic optimal control of unknown linear networked control system in the presence of random delays and packet losses', Automatica, 2012, **48**, (6), pp 1017–1030.

[8] Jamshidi, M.: ’Large-scale systems: modeling, control, and fuzzy logic’ (Prentice-Hall, Inc., 1996).

[9] Ioannou, P.:’Decentralized adaptive control of interconnected systems’, IEEE Transactions on Automatic Control, 1986, **31**, (4), pp 291–298.

[10] Šiljak, DD and Zečević, AI.:’Control of large-scale systems: Beyond decentralized feedback’, Annual Reviews in Control, 2005, **29**, (2), pp 169–179.

[11] Mehraeen, S., and Jagannathan, S.:’Decentralized optimal control of a class of interconnected nonlinear discrete-time systems by using online Hamilton-Jacobi-Bellman formulation’, IEEE Transactions on Neural Networks, 2011, **22**, (11), pp 1757–1769.

[12] Liu, D., Wang, D., and Li, H.:’Decentralized stabilization for a class of continuous-time nonlinear interconnected systems using online learning optimal control approach’, IEEE Transactions on Neural Networks and Learning Systems, 2014, **25**, (2), pp 418–428.

[13] Narendra, K.S., and Mukhopadhyay, S.: ’To communicate or not to communicate: A decision-theoretic approach to decentralized adaptive control’. Proc. American Control Conference, Baltimore, MD, July 2010, pp. 6369–6376.

[14] Venkat, A.N., Hiskens, I.A., Rawlings, J.B., *et al*:’Distributed MPC Strategies With Application to Power System Automatic Generation Control’, IEEE Transactions on Control Systems Technology, 2008, **16**, (6), pp 1192–1206.

[15] Camponogara, E., De, L., Marcelo L.:’Distributed optimization for MPC of linear networks with uncertain dynamics’, IEEE Transactions on Automatic Control, 2012, **57**, (3), pp 804–809.

[16] Song, Y., and Fang, X.:'Distributed model predictive control for polytopic uncertain systems with randomly occurring actuator saturation and packet loss', IET Control Theory & Applications, 2014, **8**, (5), pp 297–310.

[17] Zhou, X., Li, C., Huang, T., *et al*:'Fast gradient-based distributed optimisation approach for model predictive control and application in four-tank benchmark', IET Control Theory & Applications, 2015, **9**, (10), pp 1579–1586.

[18] Zheng, Y., Li, S., and Qiu, H.:'Networked coordination-based distributed model predictive control for large-scale system', IEEE Transactions on Control Systems Technology, 2013, **21**, (3), pp 991–998.

[19] Wang, X., Hong, Y., Huang, J., *et al*:'A distributed control approach to a robust output regulation problem for multi-agent linear systems', IEEE Transactions on Automatic Control, 2010, **55**, (12), pp 2891–2895.

[20] Chen, M.Z.Q., Liangyin Z., Su, H., *et al*:'A distributed control approach to a robust output regulation problem for multi-agent linear systems', IET Control Theory Applications, 2015, **9**, (5), pp 755–765.

[21] Wang, X., and Lemmon, M.D.:'Event-triggering in distributed networked control systems', IEEE Transactions on Automatic Control, 2011, **56**, (3), pp 586–601.

[22] Guinaldo, M., Lehmann, D., Sanchez, J., *et al*: 'Distributed event-triggered control with network delays and packet losses'. Proc. IEEE 51st Annual Conference on Decision and Control, Maui, HI, July 2012, pp. 1–6.

[23] Meng, X., and Chen, T.: 'Event-driven communication for sampled-data control systems'. Proc. American Control Conference, Washington, DC, June 2013, pp. 3002–3007.

[24] Halevi, Y., and Ray, A.:’Integrated communication and control systems: Part I - Analysis’, Journal of Dynamic Systems, Measurement, and Control, 1988, **110**, (4), pp 367–373.

[25] Zhang, W., Branicky, M.S., and Phillips, S.M.:’Stability of networked control systems’, Control Systems, IEEE, 2001, **21**, (1), pp 84–99.

[26] Shousong, H., and Qixin, Z.:’Stochastic optimal control and analysis of stability of networked control systems with long delay’, Automatica, 2003, **39**, (11), pp 1877–1884.

[27] Xiangnan Z., Zhen Ni, and Haibo He., *et al*: ’Event-triggered reinforcement learning approach for unknown nonlinear continuous-time system’. Proc. International Joint Conference on Neural Networks, Beijing, July 2014, pp. 3677-3684.

[28] Hou, Z., and Wang, Z.:’From model-based control to data-driven control: Survey, classification and perspective’, Information Sciences, 2013, **235**, (1), pp 3–35.

[29] Goodwin, G.C., and Sin, K.S.: ’Adaptive filtering prediction and control’ (Courier Corporation, 2014).

[30] Horn, R.A., and Johnson, C.R.: ’Matrix analysis’ (Cambridge university press, 2012).

APPENDIX

Proof of Lemma 2: The control input U_{k_l} always satisfies $\left\|U_{k_l}\right\| < \left\|R_{\bar{x}}{}^{-1}\right\| \left\|\hat{G}^{ux}_{k_l}\right\| \left\|X_{k_l}\right\|$. From Remark 6, two different possible control laws can emerge: One possibility is $\hat{G}^{uu}_{k_{l-1}}$ is non-singular and $\hat{G}^{uu}_{k_{l-1}} > R_{\bar{x}}$. Therefore, $\left\|U_{k_{l-1}}\right\| = \left\|(\hat{G}^{uu}_{k_{l-1}})^{-1}\hat{G}^{ux}_{k_{l-1}}\bar{X}_{k_{l-1}}\right\| \leq \left\|(\hat{G}^{uu}_{k_{l-1}})^{-1}\right\| \left\|\hat{G}^{ux}_{k_{l-1}}\right\| \left\|\bar{X}_{k_{l-1}}\right\|$. Since, Frobenius norm is used [30], we get

$$\left\|U_{k_{l-1}}\right\| \leq \left\|R_{\bar{x}}{}^{-1}\right\| \left\|\hat{G}^{ux}_{k_{l-1}}\right\| \left\|\bar{X}_{k_{l-1}}\right\| .$$

For the other possible condition

$$\left\|U_{k_{l-1}}\right\| = \left\|R_{\bar{x}}{}^{-1}\hat{G}^{ux}_{k_{l-1}}X_{k_{l-1}}\right\| \leq \left\|R_{\bar{x}}{}^{-1}\right\| \left\|\hat{G}^{ux}_{k_{l-1}}\right\| \left\|\bar{X}_{k_{l-1}}\right\| \tag{31}$$

Therefore, the error in the control law can be written as $\underset{\tau,\gamma}{E}(\tilde{U}_{k_{l-1}}) = \underset{\tau,\gamma}{E}(U^*{}_{k_{l-1}} - \hat{U}_{k_{l-1}})$. Taking the norm operator, using the definitions from (20) to get

$$\left\|\underset{\tau,\gamma}{E}(\tilde{U}_{k_{l-1}})\right\| \leq \underset{\tau,\gamma}{E}\{(\left\|(G^{uu}_{k_{l-1}})^{-1}G^{ux}_{k_{l-1}}\right\| + \left\|(\hat{G}^{uu}_{k_{l-1}})^{-1}\hat{G}^{ux}_{k_{l-1}}\right\|)\left\|\bar{X}_{k_{l-1}}\right\|\} \tag{32}$$

Since $G^{uu}_{k_{l-1}} > R_{\bar{x}}$, we get $\left\|\underset{\tau,\gamma}{E}(\tilde{U}_{k_{l-1}})\right\| \leq \underset{\tau,\gamma}{E}\left\|R_{\bar{x}}{}^{-1}\right\|(\left\|G^{ux}_{k_{l-1}}\right\| + \left\|\hat{G}^{ux}_{k_{l-1}}\right\|)\left\|\bar{X}_{k_{l-1}}\right\|$. By using the triangle inequality after representing the estimates in terms of the parameter error, we get

$$\left\|\underset{\tau,\gamma}{E}(\tilde{U}_{k_{l-1}})\right\| \leq \underset{\tau,\gamma}{E}\{2\left\|R_{\bar{x}}{}^{-1}\right\|\left\|G^{ux}_{k_{l-1}}\right\|\left\|\bar{X}_{k_{l-1}}\right\| + \left\|R_{\bar{x}}{}^{-1}\right\|\left\|\tilde{G}^{ux}_{k_{l-1}}\right\|\left\|\bar{X}_{k_{l-1}}\right\|\} \tag{33}$$

Thus the required inequality is obtained.

Proof of Lemma 3:

- During the event-sampling instants: Let the Lyapunov candidate function be

$$L_{i,\tilde{\Theta}}(k_l^j) = \underset{\tau,\gamma}{E}\,\tilde{\Theta}^T(k_l^j)W^{-1}(k_{l-1}^j)\tilde{\Theta}(k_l^j) \tag{34}$$

where j is the iteration index. From (23),using matrix inversion lemma, we have

$$W(k_l^j) = \frac{W(k_{l-1}^j)}{1 + Z(k_l^j)W(k_{l-1}^j)Z^T(k_l^j)} \tag{35}$$

Substituting in the error dynamics (37), we get $\tilde{\Theta}(k_l^j) = W(k_{l-1}^j)W^{-1}(k_{l-2}^j)\tilde{\Theta}(k_{l-1}^j)$. Using the definition of $\tilde{\Theta}$ and using (35), for all the value function estimators, the first difference becomes

$$\sum\nolimits_{i=1}^{N}\Delta L_{i,\tilde{\Theta}}(k_l^j) \leq -N\underset{\tau,\gamma}{E}\frac{\tilde{\Theta}^T(k_{l-1}^j)Z(k_{l-1}^j)Z^T(k_{l-1}^j)\tilde{\Theta}(k_{l-1}^j)}{1 + Z^T(k_{l-1}^j)W(k_{l-1}^j)Z(k_{l-1}^j)} \tag{36}$$

- During the inter-sampling instants: The parameters are updated iteratively using (24),(25). Let the Lyapunov candidate function be (34). Using similar arguments as in the previous case, the first difference is

$$\Delta L_{i,\tilde{\Theta}} = -\underset{\tau,\gamma}{E}\frac{\tilde{\Theta}^T(k_{l-1}^{j-1})Z(k_{l-1}^{j-1})Z^T(k_{l-1}^{j-1})\tilde{\Theta}(k_{l-1}^{j-1})}{1 + Z^T(k_{l-1}^{j-1})W(k_{l-2}^{j-2})Z(k_{l-1}^{j-1})} \tag{37}$$

Since the regression vector can become zero, we can only conclude that the Lyapunov function (34) is negative semi-definite. However, if the regression vector satisfies PE condition (Definition 1), $0 < \frac{Z(k_{l-1}^{j-1})Z^T(k_{l-1}^{j-1})}{1+Z^T(k_{l-1}^{j-1})W(k_{l-2}^{j-2})Z(k_{l-1}^{j-1})} \leq 1$ in (36),(37), this results in

$$\sum\nolimits_{i=1}^{N} \Delta L_{i,\tilde{\Theta}}(k_l^j) \leq -N\kappa_{min} \underset{\tau,\gamma}{E} \left\| \tilde{\Theta}^T(k_{l-1}^j) \right\|^2 \tag{38}$$

with $0 < \kappa_{min} \leq 1$. Thus, with the regression vector satisfying PE condition, the parameter estimation error is strictly decreasing both during the event-sampling instants and the inter-event period. This implies that as $k_l^j \to \infty$, the QFE parameter estimation error converges to zero asymptotically in the mean-square. This completes the proof. *Proof of Theorem 1:*

Case 1: The periodic feedback case will be analysed first. Let the Lyapunov function be

$$L(\bar{X}, \tilde{\Theta}) = \underset{\tau,\gamma}{E} \bar{X}_{k-1}^T \Gamma \bar{X}_{k-1} + \underset{\tau,\gamma}{E} \bar{\Pi} \sum\nolimits_{i=1}^{N} L_{i,\tilde{\Theta}} \tag{39}$$

$\bar{\Pi} = \eta \frac{\|W_0\|\rho_2}{N}$ with $\eta > 1$. Consider the first term, the first difference is written as $\Delta L_{\bar{x}} = \underset{\tau,\gamma}{E} \{\bar{X}_k^T \Gamma \bar{X}_k - \bar{X}_{k-1}^T \Gamma \bar{X}_{k-1}\}$. Substituting the system dynamics with the estimated control input, with the definition $K_k{}^* = \hat{K}_k - \tilde{K}_k$, we get

$$\Delta L_{\bar{x}} = \underset{\tau,\gamma}{E} \{\bar{X}_{k-1}^T (A_{\bar{x}k,c}^T - (B_{\bar{x}k}\tilde{K}_{k-1})^T)\Gamma(A_{\bar{x}k,c} - B_{\bar{x}k}\tilde{K}_{k-1})\bar{X}_{k-1} - \bar{X}_{k-1}^T \Gamma \bar{X}_{k-1}\} \tag{40}$$

where $A_{\bar{x}k,c} = A_{\bar{x}k} - B_{\bar{x}k}K_k^*$ is Schur with the optimal control policy U_k^* and there exists a positive definite solution $\bar{\Gamma}$ for the Lyapunov equation. The first difference is given by

$$\Delta L_{\bar{x}} \leq -\gamma_{\min} \underset{\tau,\gamma}{E} \left\|\bar{X}_{k-1}\right\|^2 + 2 \underset{\tau,\gamma}{E} \left\| \bar{X}_{k-1}^T (B_{\bar{x}k}\tilde{K}_{k-1})^T \Gamma \right\| \left\| A_{\bar{x}k,c}\bar{X}_{k-1} \right\| + \underset{\tau,\gamma}{E} \left\| \Gamma B_{\bar{x}k}\tilde{K}_{k-1}\bar{X}_{k-1} \right\|^2 \tag{41}$$

Applying Young's inequality, we get

$$\Delta L_{\bar{x}} \leq -\gamma_{\min} \underset{\tau,\gamma}{E} \left\|\bar{X}_{k-1}\right\|^2 + \underset{\tau,\gamma}{E} \left\|\epsilon_2 A_c \bar{X}_{k-1}\right\|^2 + \underset{\tau,\gamma}{E} \left\| (\Gamma + \frac{\Gamma^2}{\epsilon_2}) B_{\max} \tilde{U}_{k-1} \right\|^2,$$

where $\gamma_{\min}$ is the minimum singular value of $\bar{\Gamma}$ and ϵ_2 is a positive constant. Recalling Lemma 2

$$\underset{\tau,\gamma}{E} \left\|\tilde{U}(k-1)\right\|^2 \leq \underset{\tau,\gamma}{E} \{2 \left\|R_{\bar{x}}^{-1}\right\| \left\|G_{k-1}^{ux}\right\| \left\|\bar{X}_{k-1}\right\| + \left\|R_{\bar{x}}^{-1}\right\| \left\|\tilde{G}_{k-1}^{ux}\right\| \left\|\bar{X}_{k-1}\right\|\}^2 \tag{42}$$

Using Assumption 4, and with G_M as the bound on G_k^{ux}, we get

$$\begin{aligned} &\leq \underset{\tau,\gamma}{E} \{4G_M \left\| R_{\bar{x}}^{-1} \right\|^2 \left\| \bar{X}_{k-1} \right\|^2 + \left\| R_{\bar{x}}^{-1} \right\|^2 \left\| \tilde{G}_{k-1}^{ux} \right\|^2 \left\| \bar{X}_{k-1} \right\|^2 \\ &+2\varepsilon \left\| \bar{X}_{k-1} \right\|^2 + \frac{2G_M^2 \left\| R_{\bar{x}}^{-1} \right\|^4}{\varepsilon} \left\| \tilde{G}_{k-1}^{ux} \right\|^2 \left\| \bar{X}_{k-1} \right\|^2 \} \end{aligned} \tag{43}$$

where ϵ is a positive constant. On simplification, it yields that

$$\underset{\tau,\gamma}{E} \left\| \tilde{U}(k-1) \right\|^2 \leq \underset{\tau,\gamma}{E} \{(4G_M \left\| R_{\bar{x}}^{-1} \right\|^2 + 2\varepsilon) \left\| \bar{X}_{k-1} \right\|^2 + (\left\| R_{\bar{x}}^{-1} \right\|^2 + \frac{2G_M^2 \left\| R_{\bar{x}}^{-1} \right\|^4}{\varepsilon}) \left\| \tilde{G}_{k-1}^{ux} \right\|^2 \left\| \bar{X}_{k-1} \right\|^2 \} \tag{44}$$

Using the fact that $\underset{\tau,\gamma}{E} \left\| \tilde{G}_{k-1}^{ux} \right\| < \underset{\tau,\gamma}{E} \left\| \tilde{G}_{k-1} \right\|$, we obtain

$$\underset{\tau,\gamma}{E} \left\| \tilde{U}(k-1) \right\|^2 \leq \underset{\tau,\gamma}{E} \{(4G_M \left\| R_{\bar{x}}^{-1} \right\|^2 + 2\varepsilon) \left\| \bar{X}_{k-1} \right\|^2 + (\left\| R_{\bar{x}}^{-1} \right\|^2 + \frac{2G_M^2 \left\| R_{\bar{x}}^{-1} \right\|^4}{\varepsilon}) \left\| \tilde{\Theta}_{k-1} \right\|^2 \left\| \bar{X}_{k-1} \right\|^2 \} \tag{45}$$

Using (45), the first difference of the Lyapunov function becomes

$$\Delta L_x \leq -(\gamma_{\min} - \mu - \rho_1) \underset{\tau,\gamma}{E} \left\| \bar{X}_{k-1} \right\|^2 + \rho_2 \left\| \Gamma \right\| \underset{\tau,\gamma}{E} \left\| \tilde{\Theta}_{k-1} \right\|^2 \left\| \bar{X}_{k-1} \right\|^2, \tag{46}$$

where $\rho_1 = \left\| (\Gamma + \frac{\Gamma^2}{\epsilon_2}) B_{\max} \right\|^2 (4G_M \left\| R_{\bar{x}}^{-1} \right\|^2 + 2\varepsilon)$, $\mu = \left\| \epsilon_2 A_c \right\|^2$,

$$\rho_2 = \left\| \Gamma \right\| \left\| (1 + \frac{\Gamma}{\epsilon_2}) B_{\max} \right\|^2 (\left\| R_{\bar{x}}^{-1} \right\|^2 + \frac{2G_M^2 \left\| R_{\bar{x}}^{-1} \right\|^4}{\varepsilon}).$$

Recalling Lemma 3, when $0 < \left\| \Gamma \right\| \leq W_{min}$ (from *Remark* 10), substitute (35) in place of $\left\| \Gamma \right\|$ in (46). Since the history values are used, $\left\| Z_{k-1} \right\|^2 \geq \left\| \bar{X}_{k-1} \right\|^2$, then the first difference becomes

$$\Delta L \leq -(\gamma_{\min} - \mu - \rho_1) \underset{\tau,\gamma}{E} \left\| \bar{X}(k-1) \right\|^2 - (\bar{\Pi} N - \left\| W_0 \right\| \rho_2) \kappa_{min} \underset{\tau,\gamma}{E} \left\| \tilde{\Theta}(k-1) \right\|^2, \tag{47}$$

with $0 \leq \alpha \leq 1$. Substituting the value of $\bar{\Pi}$, the second term is always negative. Therefore, $L(k+1) < L(k), \forall k \in \mathbb{N}$.

Case 2: To extend the stability results for the event-based control scheme, it is required to prove that between any two aperiodic sampling instants, the Lyapunov function is non-increasing. Let the Lyapunov function be given by (60), taking the first difference

$$\Delta L_k = \underset{\tau,\gamma}{E} \{ \bar{X}_k^T \Gamma \bar{X}_k - \bar{X}_{k-1}^T \Gamma \bar{X}_{k-1} + \bar{\Pi} \sum\nolimits_{i=1}^{N} \Delta L_{i,\tilde{\Theta}} \} \quad k_l \leq k < k_{l+1}, \forall l \in \mathbb{N} \tag{48}$$

When the events occurring at k_l and $k_{l+1} = k_l + 1$, the Lyapunov function is decreasing due to (47). When the event-sampling does not occur consecutively at $k_l, k_l + 1$, the interval $[k_l, k_{l+1}) = [k_l, k_l + 1) \cup [k_l + 1, k_{l+1})$. During $[k_l, k_l + 1)$, the Lyapunov function is decreasing because of the control policy updated at k_l. In the interval $[k_l + 1, k_{l+1})$ due to the event-sampling algorithm the inequality in (39) is satisfied. Therefore, $\Delta L(\bar{X}, \tilde{\Theta}) = \underset{\tau,\gamma}{E}\{\bar{X}_k^T \Gamma \bar{X}_k - \lambda \bar{X}_{k_l+1}^T \Gamma \bar{X}_{k_l+1}\} + \Delta L_{i,\tilde{\Theta}}$. Using the results from Lemma 3 and for $\bar{\lambda} < 1$, we get

$$\Delta L_k = -(1-\bar{\lambda})\underset{\tau,\gamma}{E}\{\bar{X}_{k_l}^T \Gamma \bar{X}_{k_l}\} - N\kappa_{min}\underset{\tau,\gamma}{E}\left\|\tilde{\Theta}^T(k_{l-1}^j)\right\|^2 \tag{49}$$

Therefore, $\Delta L(\bar{x}, \tilde{\Theta}) < 0$ during the inter-sampling period. From Lemma 1,

$$\sum\nolimits_{i=1}^{N} \Delta L(\bar{x}_i, \tilde{\Theta}^i) < 0.$$

Combining Case 1 and Case 2, the Lyapunov equation satisfies the following inequality,

$$L(k_{l+1}) < L(k_l + 1) < L(k_l), \forall\{k_l\}_{l\in\mathbb{N}}. \tag{50}$$

This completes the proof.

II. EVENT-TRIGGERED DISTRIBUTED APPROXIMATE OPTIMAL STATE AND OUTPUT CONTROL OF AFFINE NONLINEAR INTERCONNECTED SYSTEMS

ABSTRACT

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

1. INTRODUCTION

Control of complex interconnected systems is one of the actively pursued areas of research in the control community [1, 2, 3, 4, 5, 6]. The composition of interacting subsystems presents a unique challenge in designing control algorithms for such interconnected systems. Various control schemes for such interconnected system have been developed which can be broadly categorized into strictly decentralized design, adaptive controllers, robust controllers and distributed controllers. The interactions between subsystems are assumed to be weak and decentralized controllers are designed [1] while in robust control approach, in addition to the decentralized control policy, a compensation term for the interconnections is added [2, 3].

In the design approach which uses adaptive controllers, the additional compensation term is adaptive and it is designed to learn the interconnection terms to cancel their effects [5, 6]. In summary, the controllers in [1, 2, 3, 4, 5, 6] are designed at each subsystem as function of local states. By communicating the states to other subsystems and using the states of the neighboring subsystems, it was demonstrated that the transient performance of each subsystem could be improved [6]; further, various distributed control schemes are given in [7, 8, 9] and the references therein.

One of the impediments for implementing distributed control algorithms is the communication cost involved due to sharing of states among subsystems. To mitigate these costs, event-triggered controllers were proposed [7, 9, 10, 11, 12, 13, 14, 15]. Initially, the focus of the event triggered control research was to design an event triggering mechanism to reduce the frequency of control implementation using latest sensor measurements without compromising the system stability. However, in addition to stability, optimality is desired.

When the system dynamics are linear and known, optimal control problem can be solved to obtain a backward-in-time solution using Riccati equation [16]. When the system dynamics are nonlinear, solution to the Hamilton-Jacobi-Bellman (HJB) equation is required for optimal policy. Since the HJB equation does not have a closed-form solution

[17], inspired by the reinforcement learning (RL) techniques [18], a suite of learning algorithms based on dynamic programming were proposed. These learning algorithms generate an approximation of the optimal value function and an approximate optimal control policy [1, 2, 3, 14, 15, 17, 19, 20, 21, 22, 23, 24, 25, 26, 27] and are broadly classified as approximate dynamic programming (ADP) schemes.

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

These results of ADP based learning controllers were used to develop decentralized control schemes for interconnected systems using continuous/periodic feedback which guarantee stability and optimality [1, 3]. The RL based online ADP methods [1, 3] applied to interconnected systems typically requires extensive computations and exchange of feedback information among subsystems through a communication network. Comparing with the traditional ADP design, the event based method samples the state and updates the controller only when it is necessary. Therefore, the computation and transmission costs are reduced.

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

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

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

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

Next, to relax the requirement of measuring the entire state vector, nonlinear observers are designed at each subsystem to estimate overall system state vector using outputs that are communicated only at event based sampling instants from the other subsystems. The hybrid learning scheme with the observers is analyzed using Lyapunov technique. It is shown that closed-loop system is stable with both event triggered state and output feedback. Finally, two simulation examples are used to evaluate the effectiveness of the analytical design presented in the paper.

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

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

2. BACKGROUND

2.1 System Dynamics

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

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

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

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

where $X \in S \subseteq \Re^{n \times 1}$, $U \in \Re^{m}$, $G(X) : \Re^{n} \to \Re^{n \times m}$, $F(X) : \Re^{n} \to \Re^{n}$, $U = [u_1{}^T, .., u_N{}^T]^T$, $m = \sum_{i=1}^{N} m_i$, $n = \sum_{i=1}^{N} n_i$, $\dot{X} = [\dot{x}_1^T, .., \dot{x}_N^T]^T$, $G(X) = diag(g_1(x_1), .., g_N(x_N))$, $F(X) = [(f_1(x_1) + \sum_{j=2}^{N} \Delta_{1j})^T, .., (f_N(x_N) + \sum_{j=1}^{N-1} \Delta_{Nj})^T]^T$ and S is a compact set obteined as a result of finite union of S_i. The following standard assumptions on the system dynamics are needed in order to proceed further.

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

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

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

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

Note that the dynamics of the system (2) and the component subsystems (1) do not explicitly describe how the subsystems are interconnected. Interconnected systems with state interactions will be the focus of this paper and the systems with output/control interactions are not dealt in this paper. The control scheme for the augmented system (2) is represented in the form of a block diagram in Fig. 3.1.

As seen in Fig. 3.1, at each subsystem an event-sampling mechanism monitors the subsystem states/outputs to determine the feedback/broadcast instants. For the case of output feedback, only the outputs from each subsystem are broadcast and the output vector is used to reconstruct the states of all the subsystems to be used in the controller. Due to the flexibility offered by the networked control architecture, the interconnected system represented in Fig. 1 consisting of two networks is preferred. The Network 1 enables information exchange between subsystems, while Network 2 is a local communication network which closes the feedback control loop of each subsystem. The communication resources involved in the control design of such systems motivated the use of event based feedback.

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

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

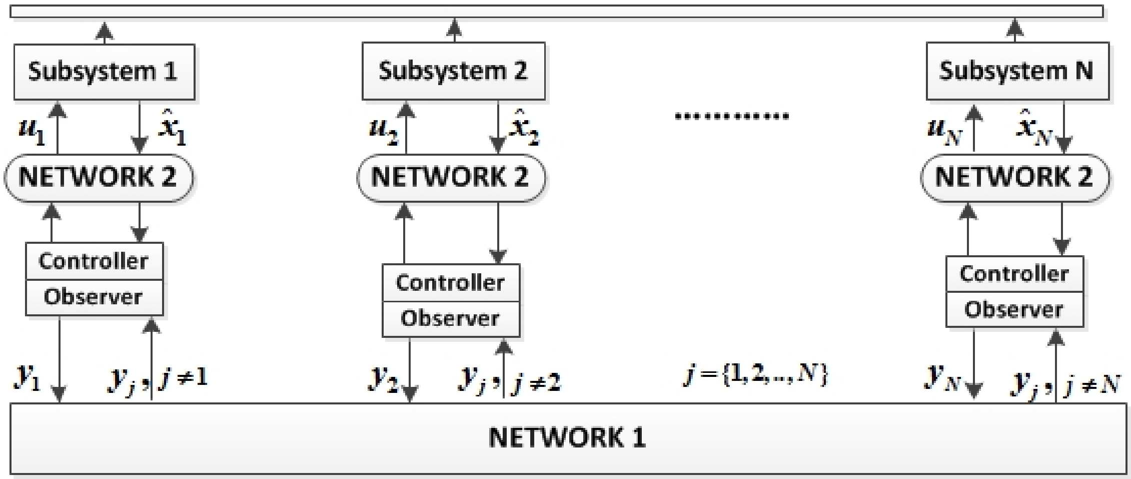


Fig. 3.1. Interconnected system.

This error is reset to zero at the sampling instants due to the feedback update. A brief background on the design of optimal controller using aperiodic, event-triggered feedback is presented in the next subsection.

2.2 Event Based Optimal Control Policy

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

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

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

$$\dot{V}(X(t)) = -\left[Q(X) + U^T(t)RU(t)\right]. \tag{5}$$

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

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

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

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

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

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

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

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

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

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

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

2.3 Neural Network Approximation Using Event Based Feedback

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

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

Given, $\chi : A \subseteq \Re^n \rightarrow \Re$, a smooth function in a compact set A and $\varepsilon_M > 0$, $\exists\theta^* \in \Re^{P\times 1} : \chi(x) = \theta^{*^T}\phi(x_e) + \varepsilon_e,\ t_k \leq t < t_{k+1}$, with $\|\varepsilon_e\| < \varepsilon_M$, $\forall x_e \in A$, where, $\phi(x_e)$ is a basis function driven by the inputs $x(t), e(t)$ and $\varepsilon_e = \theta^{*^T}(\phi(x_e + e) - \phi(x_e)) + \varepsilon$, the reconstruction error driven by $e(t)$. The error $e(t)$ is due to the difference between $x(t_k), x(t)$.

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

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

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

3. DISTRIBUTED CONTROLLER DESIGN

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

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

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

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

Proof: First, consider the infinite horizon value function defined by (4) for the augmented system in (2). Define $R = diag(R_1, R_2, .., R_N)$, $Q(X) = \sum_{i=1}^N Q_i(X)$, $U(X(t)) = [u_1^T, ., u_N^T]^T$,

$$\frac{\partial V^T}{\partial X} = \left[\frac{\partial V_1^T}{\partial x_1}, \frac{\partial V_2^T}{\partial x_2} .... \frac{\partial V_N^T}{\partial x_N}\right],$$

$V(X(t)) = \int_t^\infty \left(\sum_{i=1}^N \left[Q_i(X) + u_i{}^T R_i u_i\right]\right) d\tau = \int_t^\infty \sum_{i=1}^N V_i(X(\tau)) d\tau$. The Hamiltonian (8) becomes $H(X) = \left[\frac{\partial V_1^T}{\partial x_1}, ., \frac{\partial V_N^T}{\partial x_N}\right] [\dot{x}_1^T, ., \dot{x}_N^T]^T + \left(\sum_{i=1}^N \left[Q_i(X) + u_i{}^T R_i u_i\right]\right) = \sum_{i=1}^N H_i(X, u_i) : H_i(X, u_i) = \left(\frac{\partial V_i^T}{\partial x_i} \dot{x}_i + Q_i(X) + u_i{}^T R_i u_i\right)$. For optimality, each subsystem should generate a control policy from $H_i(x, u_i)$ as

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

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

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

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

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

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

The design using state feedback is presented next.

3.1 State Feedback Controller Design

We use the artificial NNs [29] to represent the optimal value function in a parametric form using NN weights and a set of basis function with a bounded approximation error. Using the parameterized representation, the value function is represented as $V(X) = \theta^T\phi(X) + \varepsilon(X)$, where $\phi(X)$ is a basis function and $\varepsilon(X)$ is the bounded approximation/reconstruction error.

Let the target NN weights be θ_i^* and the estimated NN weights be $\hat{\theta}_i$ at the i^{th} subsystem. The parameterized HJB equation with approximate optimal value function can be obtained as

$$\begin{aligned} &Q_i(X) + \theta_i^{*T}\nabla_x\phi(x)\bar{f}_i(x) - \frac{1}{4}\theta_i^{*T}\nabla_x\phi(x)D_i\nabla^T{}_x\phi(x)\theta_i^* \\ &+ \varepsilon_{i_{HJB}} = 0 \end{aligned} \tag{12}$$

where $\varepsilon_{i_{HJB}} = \nabla_x{\varepsilon_i}^T(\bar{f}_i(x) - \frac{D_i}{2}(\nabla^T{}_x\phi(x)\theta_i^* + \nabla_x\varepsilon_i)) + \frac{1}{4}\nabla_x{\varepsilon_i}^T D_i\nabla_x\varepsilon_i$, $\bar{f}_i(x) = f_i(x_i) + \sum_{\substack{j=1 \\ j \neq i}}^{N} \Delta_{ij}(x_i, x_j)$, the partial derivative of the optimal cost function V_i^{*T} with respect to x_i is $\nabla^T{}_x\phi(x)\theta_i^*$ and $D_i = D_i(x_i) = g_i(x_i){R_i}^{-1}{g_i}^T(x_i)$. Let $\|\nabla^T{}_x\phi(x)\theta_i^*\| \leq V_{xiM}, \|D_i\| \leq D_{iM}$. Now, using the estimated weights $\hat{\theta}_i$, the control input (10), can be written as $\hat{u}_i = -0.5R_i^{-1}g_i^T\hat{\theta}_i^T\nabla_x\phi(x)$ and the parameterized Hamiltonian equation is derived as

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

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

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

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

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

The event triggered state feedback controller design is introduced next.

3.2 Event Triggered State Feedback Controller

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

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

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

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

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

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

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

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

with $0 < \Gamma_i < 1$.

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

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

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

3.3 Event Triggered Output Feedback Controller

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

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

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

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

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

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

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

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

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

Lemma 4 *For the augmented system given in (2) composed of interconnected subsystems given in (1), consider the proposed observer (19) at each subsystem with the error dynamics (20) and let the measurement error (3) be bounded. The observer estimation error is locally UUB, provided the control policy is admissible and the observer gains are chosen such that* $\eta_{i,o1}, \eta_{i,o2} > 0$*, where the design variables* $\eta_{i,o1}, \eta_{i,o2}$ *are defined in the proof.*

Proof: See appendix.

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

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

3.4 Hybrid Learning Algorithm

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

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

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

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

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

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

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

For the stability analysis, first, using the fact that the optimal control policy results in a stable closed-loop system, a time-varying bound on the closed-loop dynamics are defined [23] as $\|F(X) + G(X)U^*\| \leq \psi \|X\|$, with $\psi > 0$. It was also shown in [23] that there exists positive constant ζ_1 such that, $\|L_x(X)\| \, \|f(X) + g(X)U^*\| \leq -\zeta_1 \|L_x(X)\|^2$, with the Lyapunov function $L(X)$, its derivative $L_x(X)$, with respect to the state vector. Choosing $L(X) = 0.5(X^T X)$, we get $\left\|X^T\right\| \|f(X) + g(X)U^*\| \leq -\zeta_1 \|X\|^2$, which will be used to analyze the proposed controller. With these results, the stability analysis of the proposed state-feedback controller, output feedback controller with event-triggered feedback will be presented in the next section.

4. STABILITY ANALYSIS

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

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

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

Similarly, for the case of output feedback, we have

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

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

Theorem 1 *Consider the nonlinear dynamics of the augmented system (2) with the equilibrium point at origin. Let the initial states* $x_{i0}, \hat{X}_{i0} \in S$ *and let* $\hat{\theta}_i(0)$ *be defined in a compact set* $\Omega_{i\theta}$*. Use the update rule defined in (14), with the estimated states, to tune the NN weights. With the estimated states evolving according to the observer dynamics given by (20), there exists* $\eta_{i's} > 0$ *such that* $\tilde{\theta}_i$*,* $X(t)$ *and the observer error dynamics are locally uniformly ultimately bounded by* ξ_{icl} *in the presence of a bounded external input. The constants,* $\eta_{i's}$ *and the bound,* ξ_{icl}*, are defined in the proof.*

Proof: Refer to the Appendix.

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

Theorem 2 *Consider the nonlinear interconnected system described by (2) wherein the initial states $x_{i0}, \hat{X}_{i0} \in S$, and let the NN weights be initialized in a compact set $\Omega_{i\theta}$. Consider the weight tuning rule defined in (21) using the estimated states and the event-triggering mechanism satisfying (18), with the measured outputs at each subsystem. With the estimated states evolving according to the observer dynamics given by (20), there exists $\eta_{i's} > 0$ such that $\tilde{\theta}_i$, $X(t)$ and the observer error dynamics are locally uniformly ultimately bounded by ξ_{icl} wherein the bound is obtained independent of the measurement error. The constants, $\eta_{i's}$ and the bound, ξ_{icl}, are defined in the proof.*

Proof: Refer to the Appendix.

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

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

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

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

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

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

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

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

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

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

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

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

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

5. SIMULATION RESULTS

In this section, two examples are considered to verify the analytical design presented in the paper. The first example includes a system of two inverted pendulums connected by spring. The applicability of the proposed control algorithm for linear system is verified by considering the linear dynamics first and then the nonlinear dynamics are considered. In the second example, a more practical nonlinear system with three interconnected subsystems is considered. *Example 1:* The example used here has two inverted pendulum connected by spring [4], which can be represented of the form (2). A NN with one layer and 5 neurons together with polynomial basis set wherein the control variables $\alpha_1 = 25, \beta = 0.01$, $L_i(x) = \frac{1}{2} x_i^T \beta x_i$ and $\phi(x) = \left[x_{1,1}{}^2, x_{1,2}{}^2, x_{2,1}{}^2, x_{2,2}{}^2, x^T x \right]^T$; the initial conditions are defined in the interval [0,1] and the initial weights of the NN are chosen randomly from [-1,1].

The dynamics of the system are given by $\dot{x}_{i1} = x_{i2}$, $\dot{x}_{i2} = (\frac{m_i g r}{J_i} - \frac{k r^2}{4 J_i}) \sin x_{i1} + \frac{k r}{2 J_i}(l - b) + \frac{u_i}{J_i} + \frac{k r^2}{4 J_i} \sin x_{j1}$. For the linear dynamics, refer [9]. The parameters in the system dynamics are m_1= 2,m_2= 2.5,J_1= 5,J_2= 6.25, $k = 10, r = 0.5, l = 0.5$, and $g = 9.8, b = 0.5$. The controller design parameters are chosen as R_1=.03,R_2= 0.03, $Q_i = 0.1 X^T X$. The

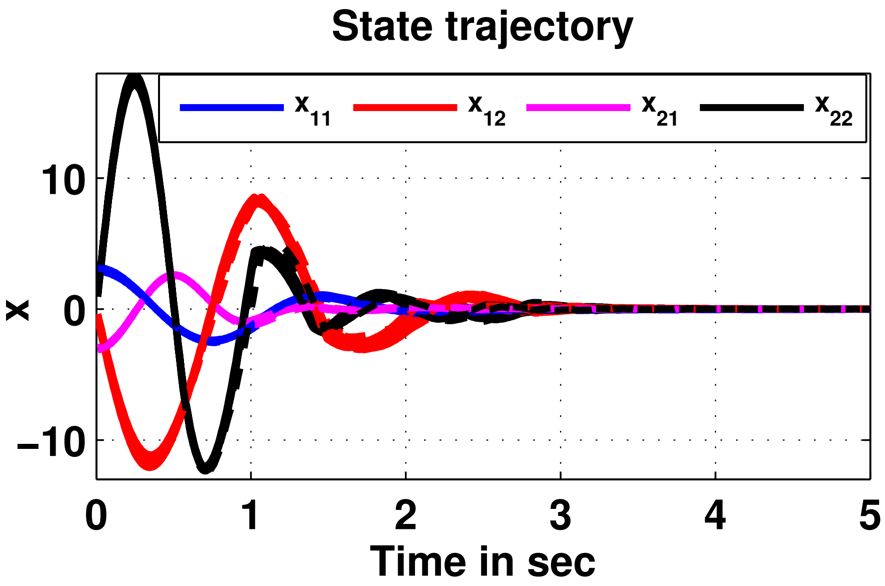


Fig. 3.2. State trajectories (Linear example).

results in Fig. 3.2 shows the distributed controller performance for the linear system for various initial conditions. Fig. 3.3 shows the cumulative events for the linear interconnected system, which demonstrates the advantage of event based feedback.

Next, the results for the event-triggered controller are presented with the distributed control scheme for the nonlinear dynamical system. For the event-triggered controller, the initial states and the weights are chosen as in the previous case. The design parameters are $\Gamma = 0.95, \alpha_1 = 20$, $\beta = 0.01$, $R_i = 0.03$, $Q = 2X^TX$. The system state trajectories with event-triggered controller are stable during the learning phase.

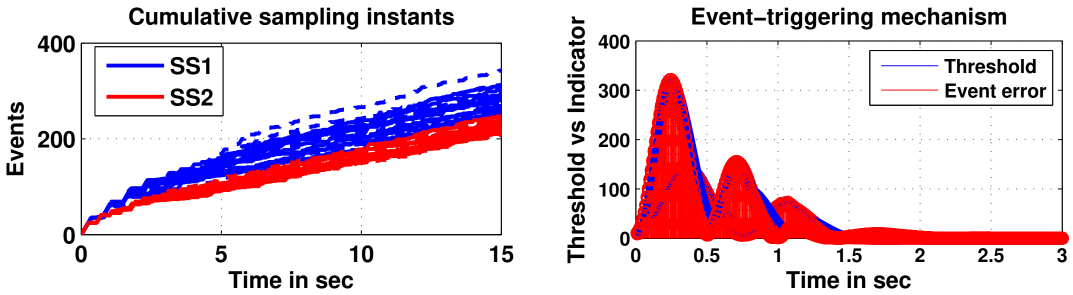


Fig. 3.3. Event-triggering mechanism.

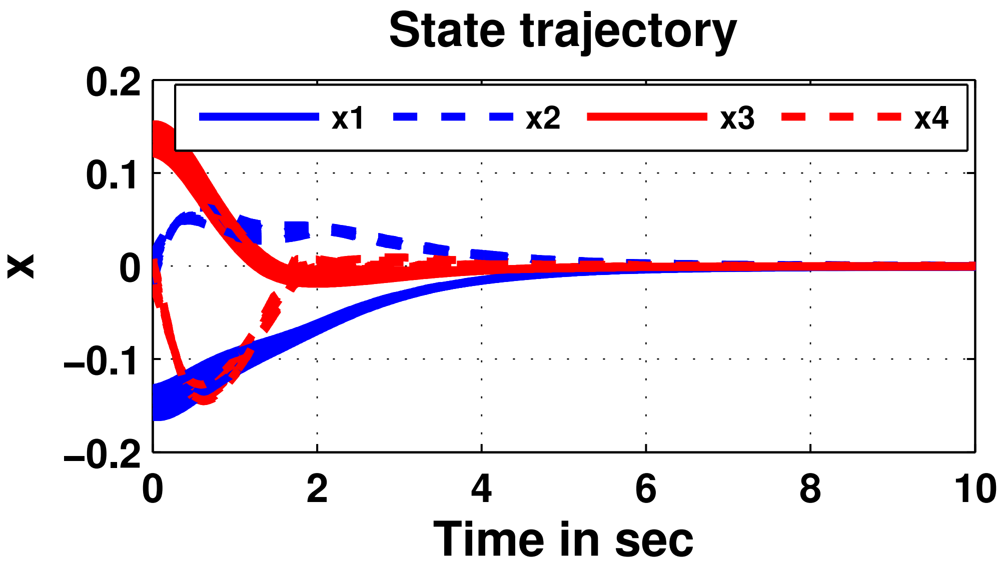


Fig. 3.4. State trajectories (Nonlinear example - 1).

This can be verified from Fig. 3.3 for both the subsystems. The results in Fig. 3.4 include the state trajectories for various initial states. The HJB residual error for the TD ADP based controller and the proposed hybrid learning based controller are compared. It is evident from the results in Fig. 3.5 that the iterative weight updates between event-triggering instants seems to reduce the learning time.

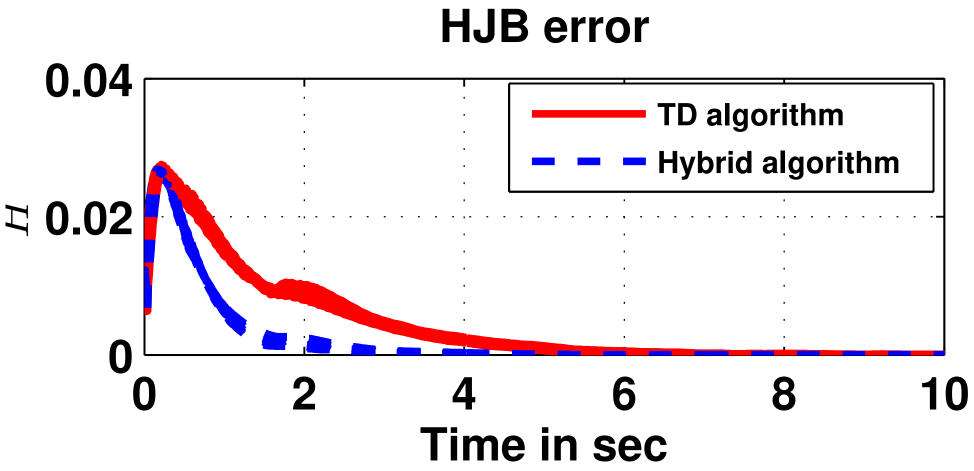


Fig. 3.5. HJB error (Nonlinear example - 1).

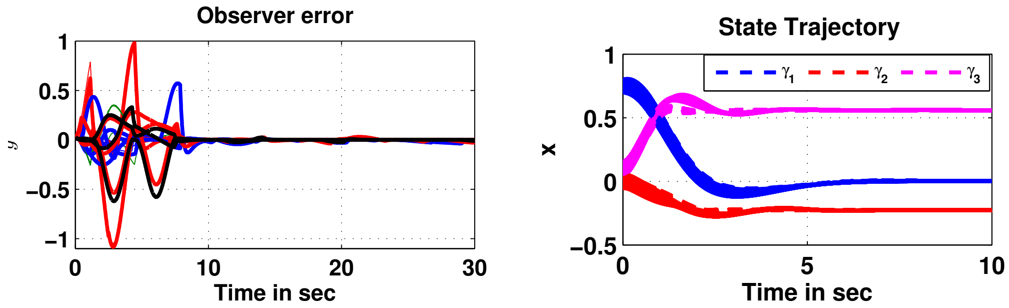


Fig. 3.6. Observer performance (Output estimation error): Example 2 and State trajectory of walking robot.

The observer performance is presented in Fig. 3.4. The plots of estimated and actual outputs with the event-triggered feedback are compared when the hybrid learning algorithm is employed to generate the control policy online. The event triggered feedback and aperiodic update of the observer results in a piecewise continuous estimate of the actual states. The observer error convergence is essential for the stability of the controlled system.

Due to space consideration all the simulation figures are not included. Efficiency of the event-triggering condition designed for the two subsystems, SS1 for subsystem 1 and SS2 for subsystem 2; the convergence time for the observer estimation error and the HJB error for various initial conditions are recorded in Table 3.1. The results of a second example considered for the simulation analysis is presented next.

Example 2: For the second example, a more practical system which is composed of three interconnected subsystems is considered. The three subsystems describe the dynamics of knee and thigh in a walking robot [8]. Let $\gamma_1(t)$ be the relative angle between the two thighs, $\gamma_2(t)$ and $\gamma_3(t)$ be the right and left knee angles relative to the right and the left thigh. The dynamical equations of motion (in rad/sec) are

$$\ddot{\gamma}_1(t) = 0.1[1 - 5.25\gamma_1^2(t)]\dot{\gamma}_1(t) - \gamma_1(t) + u_1(t)$$

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

The parameter values used in the simulation are $(\gamma_{2e}, \gamma_{3e}, p_2, p_3) = (-0.227, 0.559, 6070, 192)$. The control objective is to design torque commands and bring the robot to a halt. The proposed control scheme with a NN to approximate $V_i^*(X)$ at each subsystem is designed. The angles were initialized as $40^\circ \pm 3^\circ, 3^\circ \pm 1^\circ, -3^\circ \pm 1^\circ$ and the angular velocities were initialized at random to take values between 0 and 1. Two layer NNs with 12 neurons in the hidden layer are used at each subsystem. The NN weights of the input layer were initialized at random to form random vector functional link network [29] and the second layer weights are initialized to take values between 0 and 1.

The states of each subsystem generated using the proposed learning approach for different initial conditions are recorded. It can be observed that the states reach their equilibrium point (0,-0.227,0.559) every time, ensuring stable operation, for both state and output feedback control implementation (Fig. 3.6). The convergence of the observer estimation error can be verified from Fig. 3.6. This demonstrates that the distributed identifier at each subsystem is able to reconstruct the system internal states using the subsystem outputs which are available at discrete aperiodic time instants.

The hybrid algorithm converges faster and reaches steady state before the time driven ADP. The observer estimation error converged to a neighborhood of origin. In the analysis, different initial values for $x_i(0)$ and $\hat{X}_i(0)$ were chosen to test the algorithm and the results are tabulated. It is observed that whenever the observer error persists, performing

Table 3.1. Simulation analysis.

Example	Algorithm	Cumulative cost (Normalized)		Convergence time in sec			Feedback utilization	
		State feedback (SF)	Output feedback (OF)	HJB error		Observer error	SF	OF
				SF	OF			
1	TD	1	1	10.13	12.89	3.13	0.3716	0.8
	Hybrid	0.988	0.912	6.35	10.74	2.90	0.398	0.7824
2	TD	1	1	4.8	37.20	30.654	0.2	0.4825
	Hybrid	0.86	0.5916	4.1	31.62	27.13	0.3	0.55

iterative weight updates did not improve the learning rate. Therefore, the observer should be designed in such a way that the observer error converges faster and in this case the hybrid algorithm with output feedback controller outperformed the time driven ADP (Table 3.1).

The control torques generated using the hybrid learning algorithm with event triggered feedback and TD ADP are presented in Fig.3.7. Also, the feedback utilization (ratio the event triggered feedback instants and the sensor samples) are presented for simulations carried out for 500 different initial conditions (Fig. 3.7).

The cumulative cost is calculated using the cost function defined in (4). The comparison of the cumulative cost calculated for the hybrid learning approach with that of the TD ADP reveals that the proposed hybrid scheme results in a lower cumulative cost. Fig. 3.9 shows the ratio of costs due to hybrid algorithm over TD algorithm for different initial conditions. For the output feedback case, due to the presence of the observer estimation error, the convergence of the HJB error takes more time when compared to state feedback. The improvement in the learning scheme is due to the learning process in the inter-sampling period. For analysis, the sensor sampling time was fixed at 10ms and the control scheme was simulated to record the number of times the weight update rule was executed in the inter-event period (Fig. 3.8). It can be seen that the inter-event time is not uniform and hence, the number of weight updates are varying.

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

6. CONCLUSIONS

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

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

REFERENCES

[1] S. Mehraeen and S. Jagannathan, “Decentralized optimal control of a class of interconnected nonlinear discrete-time systems by using online hamilton-jacobi-bellman formulation,” *IEEE Transactions on Neural Networks*, vol. 22, no. 11, pp. 1757–1769, Nov 2011.

[2] D. Wang, D. Liu, H. Li, and H. Ma, “Neural-network-based robust optimal control design for a class of uncertain nonlinear systems via adaptive dynamic programming,” *Information Sciences*, vol. 282, pp. 167–179, 2014.

[3] D. Liu, D. Wang, and H. Li, “Decentralized stabilization for a class of continuous-time nonlinear interconnected systems using online learning optimal control approach,” *IEEE Transactions on Neural Networks and Learning Systems*, vol. 25, no. 2, pp. 418–428, Feb 2014.

[4] J. T. Spooner and K. M. Passino, “Decentralized adaptive control of nonlinear systems using radial basis neural networks,” *IEEE Transactions on Automatic Control*, vol. 44, no. 11, pp. 2050–2057, Nov 1999.

[5] S. Huang, K. K. Tan, and T. H. Lee, “Decentralized control design for large-scale systems with strong interconnections using neural networks,” *IEEE Transactions on Automatic Control*, vol. 48, no. 5, pp. 805–810, May 2003.

[6] K. S. Narendra and S. Mukhopadhyay, “To communicate or not to communicate: A decision-theoretic approach to decentralized adaptive control,” in *Proceedings of the 2010 American Control Conference*, June 2010, pp. 6369–6376.

[7] X. Wang and M. D. Lemmon, “Event-triggering in distributed networked control systems,” *IEEE Transactions on Automatic Control*, vol. 56, no. 3, pp. 586–601, March 2011.

[8] W. B. Dunbar, “Distributed receding horizon control of dynamically coupled nonlinear systems,” *IEEE Transactions on Automatic Control*, vol. 52, no. 7, pp. 1249–1263, July 2007.

[9] M. Guinaldo, D. V. Dimarogonas, K. H. Johansson, J. SÃąnchez, and S. Dormido, “Distributed event-based control for interconnected linear systems,” in *2011 50th IEEE Conference on Decision and Control and European Control Conference*, Dec 2011, pp. 2553–2558.

[10] X. Wang and M. Lemmon, “On event design in event-triggered feedback systems,” *Automatica*, vol. 47, no. 10, pp. 2319 – 2322, 2011.

[11] P. Tallapragada and N. Chopra, “On event triggered tracking for nonlinear systems,” *IEEE Transactions on Automatic Control*, vol. 58, no. 9, pp. 2343–2348, Sept 2013.

[12] M. Mazo and P. Tabuada, “Decentralized event-triggered control over wireless sensor/actuator networks,” *IEEE Transactions on Automatic Control*, vol. 56, no. 10, pp. 2456–2461, Oct 2011.

[13] E. Garcia and P. J. Antsaklis, “Model-based event-triggered control for systems with quantization and time-varying network delays,” *IEEE Transactions on Automatic Control*, vol. 58, no. 2, pp. 422–434, Feb 2013.

[14] A. Sahoo, H. Xu, and S. Jagannathan, “Neural network-based event-triggered state feedback control of nonlinear continuous-time systems,” *IEEE Transactions on Neural Networks and Learning Systems*, vol. 27, no. 3, pp. 497–509, March 2016.

[15] X. Zhong and H. He, “An event-triggered adp control approach for continuous-time system with unknown internal states,” *IEEE Transactions on Cybernetics*, vol. PP, no. 99, pp. 1–12, 2016.

[16] D. P. Bertsekas, D. P. Bertsekas, D. P. Bertsekas, and D. P. Bertsekas, *Dynamic programming and optimal control*. Athena Scientific Belmont, MA, 1995, vol. 1, no. 2.

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

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

[19] D. V. Prokhorov, R. A. Santiago, and D. C. Wunsch, “Adaptive critic designs: A case study for neurocontrol,” *Neural Networks*, vol. 8, no. 9, pp. 1367–1372, 1995.

[20] H. Xu and S. Jagannathan, “Stochastic optimal controller design for uncertain nonlinear networked control system via neuro dynamic programming,” *IEEE Transactions on Neural Networks and Learning Systems*, vol. 24, no. 3, pp. 471–484, March 2013.

[21] F. L. Lewis, D. Vrabie, and K. G. Vamvoudakis, “Reinforcement learning and feedback control: Using natural decision methods to design optimal adaptive controllers,” *IEEE Control Systems*, vol. 32, no. 6, pp. 76–105, Dec 2012.

[22] A. G. Barto, W. B. Powell, J. Si, and D. C. Wunsch, “Handbook of learning and approximate dynamic programming,” 2004.

[23] T. Dierks and S. Jagannathan, “Optimal control of affine nonlinear continuous-time systems,” in *Proceedings of the 2010 American Control Conference*, June 2010, pp. 1568–1573.

[24] Z. Chen and S. Jagannathan, “Generalized hamilton-jacobi-bellman formulation - based neural network control of affine nonlinear discrete-time systems,” *IEEE Transactions on Neural Networks*, vol. 19, no. 1, pp. 90–106, Jan 2008.

[25] D. Wang, D. Liu, Q. Zhang, and D. Zhao, “Data-based adaptive critic designs for nonlinear robust optimal control with uncertain dynamics,” *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. PP, no. 99, pp. 1–12, 2015.

[26] N. Vignesh and S. Jagannathan, “Distributed event-sampled approximate optimal control of interconnected affine nonlinear continuous-time systems,” in *2016 American Control Conference (ACC)*, July 2016, pp. 3044–3049.

[27] D. Liu, D. Wang, D. Zhao, Q. Wei, and N. Jin, “Neural-network-based optimal control for a class of unknown discrete-time nonlinear systems using globalized dual heuristic programming,” *IEEE Transactions on Automation Science and Engineering*, vol. 9, no. 3, pp. 628–634, July 2012.

[28] V. Narayanan and S. Jagannathan, “Approximate optimal distributed control of uncertain nonlinear interconnected systems with event-sampled feedback,” in *2016 IEEE 55th Conference on Decision and Control (CDC)*, Dec 2016, pp. 5827–5832.

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

[30] H. K. Khalil and J. Grizzle, *Nonlinear systems*. Prentice hall New Jersey, 1996, vol. 3.

APPENDIX

Proof of Lemma 1:

Consider the Lyapunov candidate function $L_i(\tilde{X}_i) = \frac{1}{2}\tilde{X}_i^T\gamma_i\tilde{X}_i + \frac{1}{4}(\tilde{X}_i^T\gamma_i\tilde{X}_i)^2$, with the first derivative given by $\dot{L}_i(\tilde{X}_i) = \tilde{X}_i^T\gamma_i\dot{\tilde{X}}_i + \tilde{X}_i^T\gamma_i\tilde{X}_i\tilde{X}_i^T\gamma_i\dot{\tilde{X}}_i$. Using the estimation error dynamics (20), and Assumptions 3-4, we get

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

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

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

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

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

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

Proof of Theorem 1:

Consider the Lyapunov function

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

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

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

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

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

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

$$\begin{aligned}\dot{L}_{i2} \leq {} & -\frac{\alpha_{i1}}{\hat{\rho}_{i,e}^2}(\tilde{\theta}_i^T\hat{\sigma}_{i,e})^2 + \frac{\alpha_{i1}}{4\hat{\rho}_{i,e}^2}\tilde{\theta}_i^T\hat{\sigma}_{i,e}P_{i1} + \tilde{\theta}_i^T\varpi_{i1} + \\ & \left\| \begin{aligned} & \frac{\alpha_{i1}}{\hat{\rho}_{i,e}^2}\tilde{\theta}_i^T\hat{\sigma}_{i,e}\{\nabla_x\varepsilon\dot{x}_i^* + \theta_i^{*T}[\nabla_x\phi(\hat{x}_e)L_{\bar{f}_i}\tilde{X}_{i,e} \\ & + L_{\phi_i}\tilde{X}_{i,e}\bar{f}_i(x)] + B_{i1} + L_Q\tilde{X}_{i,e}\} \end{aligned} \right\|, \\ & B_{i1} = \frac{1}{4}\nabla_x\varepsilon D_i\nabla_x^T\varepsilon + \frac{1}{4}\theta_i^{*T}[\nabla_x\phi(x)D_i\nabla_x^T\phi(x)]\theta_i^*.\end{aligned} \tag{28}$$

The second term can be simplified as

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

Substituting (30) in (29), we obtain

$$
\begin{aligned}
&\dot{L}_{i2}(\tilde{\theta}) \leq +\frac{\alpha_{i1}}{4\hat{\rho}_{i,e}^2}\{\tilde{\theta}_i^T \nabla_x \phi(\hat{x}_e) L_{\bar{f}_i} \tilde{X}_{i,e} + \tilde{\theta}_i^T \nabla_x \phi(\hat{x}_e) \bar{f}_i + 0.5 \\
&\tilde{\theta}_i^T \nabla_x \phi(\hat{x}_e) \hat{D}_{i,\varepsilon} L_{\phi_i} \tilde{X}_{i,e} \theta_i^* - 0.5 \tilde{\theta}_i^T \nabla_x \phi(\hat{x}_e) \hat{D}_{i,\varepsilon} \nabla_x^T \phi(x) \theta_i^* \\
&+ \frac{1}{2} \tilde{\theta}_i^T \nabla_x \phi(\hat{x}_e) \hat{D}_{i,\varepsilon} \nabla_x^T \phi(\hat{x}_e) \tilde{\theta}_i\} P_{i1} + \tilde{\theta}_i^T \varpi_{i1} - \\
&\frac{\alpha_{i1}}{\hat{\rho}_{i,e}^2} (\tilde{\theta}_i^T \hat{\sigma}_{i,e})^2 + \left\| \frac{\alpha_{i1}}{\hat{\rho}_{i,e}^2} \tilde{\theta}_i^T \hat{\sigma}_{i,e} \theta_i^{*T} \nabla_x \phi(\hat{x}_e) L_{\bar{f}_i} \tilde{X}_{i,e} \right\| \\
&+ \left\| \frac{\alpha_{i1}}{\hat{\rho}_{i,e}^2} \tilde{\theta}_i^T \hat{\sigma}_{i,e} \nabla_x \varepsilon \dot{x}_i^* \right\| + \left\| \frac{\alpha_{i1}}{\hat{\rho}_{i,e}^2} \tilde{\theta}_i^T \hat{\sigma}_{i,e} \theta_i^{*T} L_{\phi_i} \tilde{X}_{i,e} \bar{f}_i(x) \right\| \\
&+ \left\| \frac{\alpha_{i1}}{\hat{\rho}_{i,e}^2} \tilde{\theta}_i^T \hat{\sigma}_{i,e} + B_{i1} + L_Q \tilde{X}_{i,e} \right\|.
\end{aligned}
$$

Using Young's inequality and on simplification, we get

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

Rearranging the equation, after simplifying, the first derivative becomes

$$
\begin{aligned}
\dot{L}_{i2} &\leq (\frac{\alpha_{i1}}{8\hat{\rho}_{i,e}^2} + \frac{3}{8\hat{\rho}_{i,e}^2} + \frac{1}{4\hat{\rho}_{i,e}^2 D_{iM}^2}) \|P_{i1}\|^2 \\
&- (\kappa_i - \frac{|\alpha_{i1}| \, \|\hat{\sigma}_{i,e}\|^2}{\hat{\rho}_{i,e}^2} - \frac{3\alpha_{i1}^2}{2\hat{\rho}_{i,e}^2} - \frac{1}{2}) \|\tilde{\theta}_i^T\|^2 \\
&+ \frac{1}{2} \|\nabla_x \varepsilon \dot{x}_i^*\|^2 + \frac{1}{16\hat{\rho}_{i,e}^2} \|\alpha_{i1} \bar{f}_i\|^2 + B_{i2} + \frac{\|\kappa_i \theta_i^*\|^2}{2}
\end{aligned}
$$

$$
\begin{aligned}
&+ \frac{1}{2}\left\|\theta_i^{*T} L_{\phi_i} \tilde{X}_{i,e} \bar{f}_i(x)\right\|^2 + \frac{1}{2}\left\|L_Q \tilde{X}_i\right\|^2 \\
&+ \frac{1}{32\hat{\rho}_{i,e}^2}\left\|\alpha_{i1} D_{iM} L_{\phi_i} \tilde{X}_{i,e} \theta_i^*\right\|^4 + \frac{1}{2}\left\|L_{\bar{f}_i} \tilde{X}_{i,e}\right\|^2 \\
&+ \frac{1}{16\hat{\rho}_{i,e}^2}\left\|\alpha_{i1} L_{\bar{f}_i} \tilde{X}_{i,e}\right\|^4 - \frac{1}{2}\tilde{\theta}_i^T \nabla_x(\hat{x}_e) \hat{D}_{i,\varepsilon} L_{ix}(\hat{x}_{i,e}).
\end{aligned}
$$

Taking the derivative of the first term in the Lyapunov candidate function, we have

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

Using (31), $\dot{L}_i(x_i, \tilde{\theta}_i) = \dot{L}_{i1}(x_i) + \dot{L}_{i2}(\tilde{\theta}_i)$, grouping similar terms to get

$$
\begin{aligned}
\dot{L}_i(x_i, \tilde{\theta}_i) &\le -\eta_{i\theta 4}\|P_{i1}\|^2 - \eta_{ix^2}\left\|x_i^T\right\|^2 - \eta_{ix^4}\left\|x_i^T\right\|^4 \\
&- \eta_{i\theta^2}\left\|\tilde{\theta}_i^T\right\|^2 + (\frac{1}{4}\left\|\theta_i^{*T} L_{\phi_i}\right\|^4 + \frac{1}{32\hat{\rho}_{i,e}^2}\left\|\alpha_{i1} D_{iM} L_{\phi_i} \theta_i^*\right\|^4 \\
&+ \frac{1}{16\hat{\rho}_{i,e}^2}\left\|\alpha_{i1} L_{\bar{f}_i}\right\|^4 + \frac{1}{4}\|L_{iL}L_{iD}\|^2 + \frac{1}{4}\left\|\theta_i^*\right\|^4 + \frac{1}{8})\left\|\tilde{X}_i\right\|^4 \\
&+ (\frac{1}{4}\|L_{iL}eL_{iD}\|^2 + \frac{1}{4}\|L_{iL}L_{iD}e\|^2 + \frac{1}{2}\left\|L_Q\right\|^2 + \frac{1}{2}\left\|L_{\bar{f}_i}\right\|^2 \\
&+ \frac{1}{4}\|L_{iL}D_{iM}\|^2 + \frac{1}{2}\left\|L_{\phi_i}\theta_i^*\right\|^2)\left\|\tilde{X}_i\right\|^2 + \xi_{i1cl}(e)
\end{aligned}
\tag{31}
$$

where the constants are defined as follows $\eta_{i\theta^2} = (\kappa_i - |\alpha_{i1}| - 3\alpha_{i1}^2 - 0.5)$, $\xi_{i1cl} = B_{i4} + 0.5\|\beta_i B_{i3}\|^2 + 0.25\|\beta_i B_{i3}\|^4, \eta_{ix^2} = \beta_i\zeta_i - \frac{1}{16\hat{\rho}_{i,e}^2}\left\|\alpha_{i1} L_{\bar{f}_i} N\right\|^2 - \frac{1}{4}\|L_{iD}\|^2 - \frac{1}{2}\|\nabla_x\varepsilon\zeta_1\|^2 + \frac{1}{2}$, $\eta_{ix^4} = (\beta_i\zeta_i - \frac{1}{4}\left\|L_{\bar{f}_i}N\right\|^4 - \frac{5}{8}\|D_{iM}\|^4 - \frac{3}{2} - \frac{1}{4}\left\|D_{iM}L_{\phi_i}\right\|^4)$. From Lemma 1 and (32), we get

$$
\begin{aligned}
\dot{L}_i(x_i, \tilde{\theta}_i, \tilde{X}_i) &\le -\eta_{i\tilde{\theta}^4}\|P_{i1}\|^2 - \eta_{ix^2}\left\|x_i^T\right\|^2 - \eta_{i,o1}\left\|\tilde{X}_i\right\|^2 \\
&- \eta_{ix^4}\left\|x_i^T\right\|^4 - \eta_{i\theta^2}\left\|\tilde{\theta}_i^T\right\|^2 - \eta_{i\tilde{x}^4}\left\|\tilde{X}_i\right\|^4 - \eta_{i\tilde{x}^2}\left\|\tilde{X}_i\right\|^2 + \xi_{icl}
\end{aligned}
\tag{32}
$$

where $\eta_{i\tilde{x}^2} = (\eta_{i,o1} - \frac{1}{4}\|L_{iL}eL_{iD}\|^2 - \frac{1}{4}\|L_{iL}L_{iD}e\|^2 - \frac{1}{2}\left\|L_Q\right\|^2 - \frac{1}{2}\left\|L_{\bar{f}_i}\right\|^2 - \frac{1}{4}\|L_{iL}D_{iM}\|^2 - \frac{1}{2}\left\|L_{\phi_i}\theta_i^*\right\|^2)$,

$\xi_{icl} = \xi_{i1,obs} + \frac{1}{16}\left\|R_i^{-1}G^T\right\|^4 + \xi_{i1cl}(e)$,

$$\eta_{i\tilde{\theta}^4} = \eta_{i\theta 4} - (0.25NR_i^{-1}G_M + N)\backslash 16D_{iM}^2, \quad \eta_{i\tilde{x}^4} = (\eta_{i,o2} - \frac{1}{4}\left\|\theta_i^{*T}L_{\phi_i}\right\|^4 - \frac{1}{32\hat{\rho}_{i,e}^2}\left\|\alpha_{i1}D_{iM}L_{\phi_i}\theta_i^*\right\|^4$$

$$- \frac{1}{16\hat{\rho}_{i,e}^2}\left\|\alpha_{i1}L_{\bar{f}_i}\right\|^4 - \frac{1}{4}\|L_{iL}L_{iD}\|^2 - \frac{1}{4}\left\|\theta_i^*\right\|^4 - \frac{1}{8}).$$

The parameters $\alpha_{i1,}\beta_i$, κ_i, μ_i can be chosen to ensure that the constants in (33) are positive. The sigma modification term in the weight tuning equation gives the negative term in $\tilde{\theta}_i$, independent of the states.

Proof of Theorem 2:

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

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

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

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

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

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

$$H_{\tilde{x}^4} = \frac{1}{32\hat{\rho}_i^2}\left\|\alpha_{i1} D_{iM} L_{\phi_i}\theta_i^*\right\|^4 + \frac{1}{16\hat{\rho}_i^2}\left\|\alpha_{i1} L_{\bar{f}_i}\right\|^4 + \frac{1}{4}\left\|\theta_i^{*T} L_{\phi_i} L_{\bar{f}_i}\right\|^2.$$

We obtain the first derivative as

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

where H_M is the maximum of $\{H_{\tilde{x}^2}, H_{\tilde{x}^4}, H_{x^2}, H_{x^4}\}$. With the proposed iterative weight tuning, the Lyapunov first derivative is decreasing when the states and the weight estimation errors are outside the ultimate bound obtained from (34).

Proof of Corollary :

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

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

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

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

Case 2 (Event-triggered hybrid learning algorithm): Select the Lyapunov function candidate as in case 1. The first derivative is $\dot{L}_{HJB} = \sum_{i=1}^{N} (L_{ix}(x_i)\dot{x}_i + \dot{L}_{i\tilde{\theta}})$. The derivatives can be used from Theorem 2, case 2 with the observer estimation error set to 0. This gives a stronger result when compared to the TD ADP.

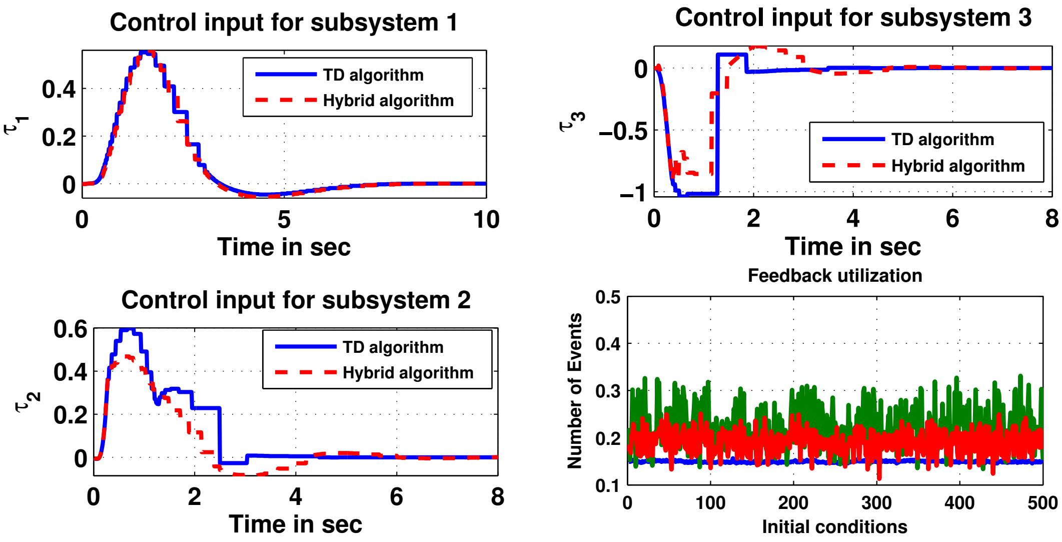


Fig. 3.7. Event triggered control.

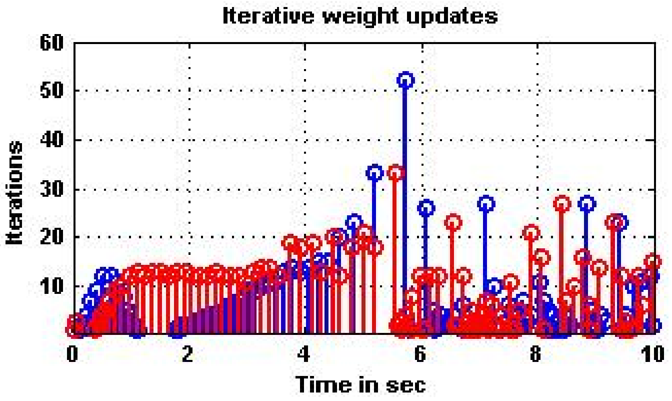


Fig. 3.8. Number of iterations in the inter event period.

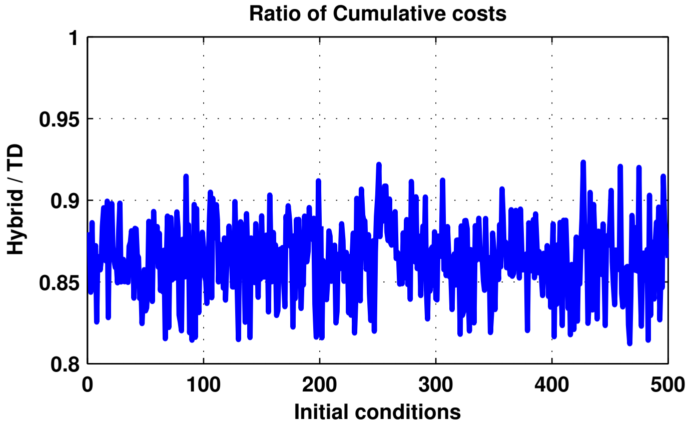


Fig. 3.9. Cost comparison (Example 2).

III. EVENT-TRIGGERED DISTRIBUTED CONTROL OF NONLINEAR INTERCONNECTED SYSTEMS USING ONLINE REINFORCEMENT LEARNING WITH EXPLORATION

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 on approximate dynamic programming with 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 (NN) 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 the Lyapunov stability theory. Taking into account, the effects of NN approximation of system dynamics and boot-strapping, 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 proposed to reduce the overall cost during the initial learning period. 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 exercise.

1. INTRODUCTION

Advanced control [1] schemes are necessary for efficient and cost effective operation of engineering systems in a variety of applications. The adaptive dynamic programming (ADP) design [1] aims to address the problem of optimization over time through learning without needing apriori knowledge of the system dynamics. Solution to the Hamilton-Jacobi-Bellman (HJB) equation [2] provides the optimal value function and optimal control policy for a nonlinear system. Due to the difficulty in solving HJB equation directly, nonstandard techniques [3] which are inspired by the reinforcement learning (RL) [4], are employed to construct an approximate solution.

In the applications involving real-time online control, iterative learning approach to generate control actions is undesirable [5, 6] due to the large iterations required for convergence. Reducing the computations considerably, the time-driven ADP approach introduced in [6] was designed to adjust its adaptive parameters once at each sampling instant. This approach was motivated by the one step temporal difference learning (TDL) of RL. Due to the reduction in the iterations, the optimality of the control sequence in the intermediate time steps was affected and learning algorithm converged asymptotically under some conditions [6]. Nevertheless, the stability results of the TDL scheme, despite the reduction in iterative learning steps and suboptimal control, attracted further investigation in online control applications [7, 8].

The RL-ADP approach is expected to mimic human intelligence as highlighted in [1] for control with four desired characteristics. The ability to solve the optimization over time, through learning; involving a critic to generate reinforcement signal; use of reinforcement signal to generate the control action; and finally, use an adaptive component to emulate the system and estimate the internal states. With the advent of networked control systems, computational efficiency, reduced communication resource utilization and reduced learning time are also desired for an advanced control scheme.

However, all the learning schemes in [5] and the references therein, are designed to work in periodic or continuous feedback framework requiring substantial computations. To mitigate computational aspects without sacrificing performance, event- triggered control [7] design was introduced. The event triggered controllers are guaranteed to retain the stability of the system despite using limited, aperiodic sensor measurements.

In the event triggered controller design presented in [7] using NNs with the time-driven (TD) ADP, the event-triggering mechanism was derived as a function of estimated weights of the NN. The rationale behind such a design was to increase the events during the initial learning period. As a consequence, the event-triggering mechanism generated frequent events when the NN weight estimation error was significant. This ensured that the learning process was unaffected by the aperiodic event based feedback. Developing an event-triggered learning control scheme which preserves stability and the learning efficiency while retaining fully the benefits of event triggered feedback is still a challenging design problem.

Later, an event-triggered control scheme with adaptive dynamics programming using policy iteration (PI) algorithm was proposed in [9]. In order to ensure faster convergence and real time implementation, a more flexible online learning framework was proposed in [8] for a large scale interconnected system. However, the scope of the work presented in [8] was limited to linear systems and the learning algorithm utilized the state-action value function or the Q-function [4] to determine the optimal control sequence.

ADP based distributed optimal control for interconnected system was considered in the literature [10, 11]. For such systems, distributed control [12] is preferred with a learning component. Though the learning schemes in [10, 11] are proven to be efficient, [11] presents a multi-player non-zero sum game formulation for the distributed control while [10] presents a robust design approach and both follows an iterative learning approach to obtain the optimal/Nash equilibrium solution. Moreover, all the control schemes [10, 11, 12, 13] are presented when continuous feedback is available.

In contrast, the main focus of this paper is to develop a distributed learning control scheme for a nonlinear interconnected system with the subsystems coupled by states such that both the stability and the learning efficiency is preserved when the feedback is aperiodic and event based. The uncertain dynamics and unknown nonlinear interconnections are reconstructed using identifier NNs designed at each subsystem. The proposed learning scheme enables TD learning, with the current feedback information, at the event-triggering instants and iterative learning, using past data, between events to enhance approximation accuracy of the NNs employed by the learning scheme.

However, the event based learning schemes in [7, 9] focuses on implementation of the controller in the event based feedback setting to reduce network resource utilization and reduce computations. This however reduces the efficiency of the learning mechanism in the following ways: a) the learning time is increased due to intermittent feedback; b) the sampling instants are dynamic, therefore, the sampling interval is time-varying, restricting the use of iterative learning algorithms with fixed iterations; c) All the sensor samples are not utilized during the learning process as the feedback instants are decided by the event triggering mechanism. Therefore, an improved learning algorithm utilizing the identifiers is introduced and a new learning rule is developed, which seems to considerably improve the learning efficiency of the event triggered ADP algorithms at the cost of additional computations.

Further, one of the classical problems of RL is the dilemma of exploration vs exploitation, a problem also observed in forward-in-time ADP based optimal controllers. This problem is highlighted and a few challenges involved in designing exploration strategies for control are discussed. A novel exploration strategy, inspired by [14], using identifiers is proposed. Normally, identifiers are designed to estimate the local states at each subsystem to implement the hybrid learning control scheme whereas for exploration, an identifier which can estimate the state vector of the overall interconnected system is required. Exploration enhances optimality at the expense of computations which can be considered as a trade-off.

Finally, 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 when the distributed controller using proposed learning scheme with exploration is utilized and this is demonstrated using Lyapunov analysis. Simulation results are presented to show the advantages of using the proposed learning control scheme with exploration and emphasize the challenges involved.

The major contributions of this paper include: a) A RL based novel learning control scheme suitable for event-triggered control implementation; b) a suite of NN identifier designs to reconstruct unknown nonlinear functions in the system dynamics; c) a novel NN weight adaptation rule to reconstruct and learn the approximated optimal value function; d) an online exploration strategy using identifiers and e) stability analysis using Lyapunov theory.

This paper is organized as follows: Section II introduces the dynamics of the system being investigated and presents a brief background on optimal control, the distributed optimal control formulation of interconnected system. Section III briefly talks about the existing event based ADP algorithms and introduces the hybrid learning scheme. Section IV presents the modified/improved hybrid learning algorithm and discussions on exploration; System stability analysis and simulation results are included in Section V and VI respectively. The conclusions drawn from this study are reported in Section VII.

2. BACKGROUND AND PROBLEM STATEMENT

2.1 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.

2.2 System Description

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

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

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

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

where $F = \left[(f_1 + \sum_{j=2}^{N} \Delta_{1j})^T, .., (f_N + \sum_{j=1}^{N-1} \Delta_{Nj})^T\right]^T$, $x = \left[x_1^T, .., x_N^T\right]^T \in B \subseteq \Re^{n}$, $B = \bigcup_{i=1}^{N} B_i$, $u = \left[u_1^T, .., u_N^T\right]^T \in \Re^{m}$, $m = \sum_{i=1}^{N} m_i$, $n = \sum_{i=1}^{N} n_i$ 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.

2.3 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

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

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, $\dot{V}(x(t)) = -\left[Q(x) + u^T(t)Ru(t)\right]$. Using the infinitesimal version of (4), define the Hamiltonian function $H(x, u) = \left[Q(x) + u^T(t)Ru(t)\right] + (\partial V^T/\partial x)\dot{x}$. 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

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

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 $x_i(t), x_j(t)$.

Proposition 1: [15] 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

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

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 [15].*

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.

2.4 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 [7] is utilized. Define a smooth function, $\chi : B \longrightarrow \Re$, 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 [7], 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 [5, 6].*

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 [6, 15] 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 $\varepsilon_{i_{HJB}} = \nabla_x \varepsilon_i^T (\bar{f}_i - 0.5 D_i ({\nabla^T}_x \phi(x) \theta_i^* + \nabla_x \varepsilon_i) + 0.25 D_i \nabla_x \varepsilon_i)$. 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

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

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.

3. 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.

3.1 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

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

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 equation describing the evolution of $\tilde{x}_i(t)$ is revealed as

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

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

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

with $J_{i\tilde{x}} = 0.5\mu_{i1}\tilde{x}_i^T P_i\tilde{x}_i$, $J_{i\tilde{f}} = 0.5\mu_{i2}\tilde{W}_{if}^T\tilde{W}_{if} + 0.25\mu_{i4}(\tilde{W}_{if}^T\tilde{W}_{if})^2$, $J_{i\tilde{g}} = 0.5\mu_{i3}\tilde{W}_{ig}^T\tilde{W}_{ig} + 0.25\mu_{i5}(\tilde{W}_{ig}^T\tilde{W}_{ig})^2 + 0.125\mu_{i6}(\tilde{W}_{ig}^T\tilde{W}_{ig})^4$; 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 5 *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*

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

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

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

If $u_{i,e}(t)$ is stabilizing, then $\exists\alpha_{if}, \alpha_{ig}, \kappa_{if}, \kappa_{ig}, A_i > 0$ 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 stabilizing 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.

3.2 Event Based Hybrid Learning Scheme

A brief discussion on the hybrid learning scheme [15] 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 $J_i(x_i) = x_i^T\Gamma_i x_i$, with $\Gamma_i > 0$. For $0 < \alpha_i < 1$ and $k \in \mathbb{N}$, design the event-triggering mechanism to satisfy the condition

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

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 [7] 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

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

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.

4. 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.

4.1 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.*

4.2 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 âĂŸexperienceâĂŹ is not utilized by the learning schemes presented in [7, 9, 15]. 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 âĂŸexperienceâĂŹ in the inter event period to provide a better optimal value function estimate.*

4.3 Role Of Identifiers And Exploration In Online Control

One of the classical problems in the RL literature [4, 14] 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 [5] 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, 16] 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 [5, 6] 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 [17] 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 [16]. 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, 14], 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 [14], a novel technique to incorporate exploration in the learning controller is developed next.

4.4 Exploration Using Identifiers

The TD learning [4, 6, 7] and the hybrid learning schemes [8, 15] 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 ($t \longrightarrow \infty, \hat{V} \longrightarrow V^*$). 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. 4.1 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 $\hat{H}_i^{(p)}(\hat{x}_e) = Q_i + \hat{\Theta}_{pi}^T \nabla_x \phi(\hat{x}_e) \hat{\hat{f}}_i - \frac{1}{4} \hat{\Theta}_{pi}^T \nabla_x \phi(\hat{x}_e) \hat{D}_i \nabla^T{}_x \phi(\hat{x}_e) \hat{\Theta}_{pi}$ where $p = 1, 2$ for each initial weights. We can construct the cost function trajectory with the value function estimator

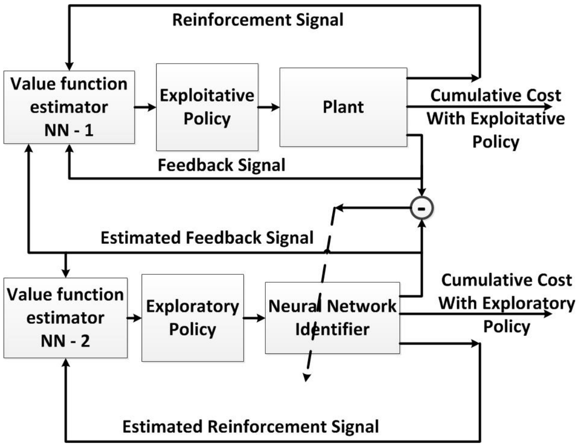

Reinforcement Signal
Value function estimator NN - 1
Exploitative Policy
Plant
Cumulative Cost With Exploitative Policy
Feedback Signal
-
Estimated Feedback Signal
Value function estimator NN - 2
Exploratory Policy
Neural Network Identifier
Cumulative Cost With Exploratory Policy
Estimated Reinforcement Signal

Fig. 4.1. Block diagram representation of exploration strategy.

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

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

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 [14], 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 'rich' data can be collected using the proposed exploration strategy [5].*

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

5. 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 [6] that the optimal controller renders the closed-loop dynamics bounded reveals

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

where $\delta(x) \in \Re^n$, $C_1 \in \Re$.

Theorem 1 *Consider the subsystem dynamics (1). Define the NN weight update rule (18) for the value function approximator and (15), for the identifiers. Then,* $\exists \alpha_{iv}, \mu_i, \kappa_3 > 0$ *and computable positive constants which define the bounds for* $\tilde{\theta}_i, \tilde{W}_{if}, \tilde{W}_{ig}$ *and* $x, \tilde{x}_i$ *and all the closed loop signals 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 when (17) is violated. Then, computable positive constants that define the bounds for* $\tilde{\theta}_i, \tilde{W}_{if}, \tilde{W}_{ig}$ *and* $x, \tilde{x}_i$ *exist and all the closed loop signals 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 [7].*

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 [7].*

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

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). Then positive constants can be computed that define bounds on the NN weight estimation error* $\tilde{\theta}_i, \tilde{W}_{if}, \tilde{W}_{ig}$, *the interconnected system states and* $\tilde{x}(t)$ *are locally bounded. Further, when the NN weights are tuned based on the rule (19), the state vector and weight estimation error for the value function estimator monotonically decreases for all* t. *Under the assumptions prescribed in the previous sections, the value function estimator error and the identifier states corresponding to each value function estimator NN remain locally uniformly ultimately bounded.*

Proof: See appendix.

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

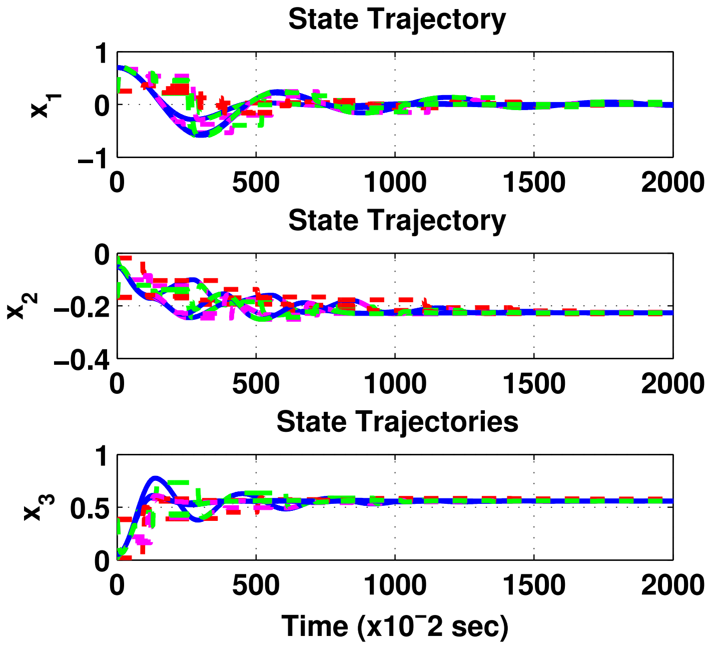


Fig. 4.2. State Trajectories (Dotted Lines - Hybrid vs Enhanced hybrid algorithm).

6. 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 [13]. 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).

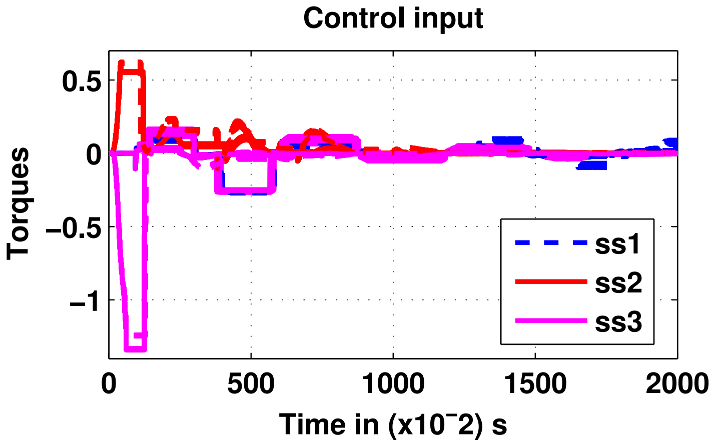


Fig. 4.3. Control torques (Dotted lines - Hybrid vs Enhanced hybrid scheme).

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: $\alpha_{i1_v} = 40$, $\alpha_{i2_v} = 0.03$, $\mu_i = 1.95$, $P_i = 2$, $\kappa_3 = 0.001$, $Q_i = 20$, $R_i = 1$, $A_i = 80$, $C_{if} = 0.5$, $\kappa_{if} = \kappa_{ig} = 0.0001$, $\alpha_{if} = \alpha_{ig} = 100$ 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. 4.2).

The magnitude of the control torque for the hybrid ADP based learning scheme and the enhanced hybrid approach are compared in Fig. 4.3 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.

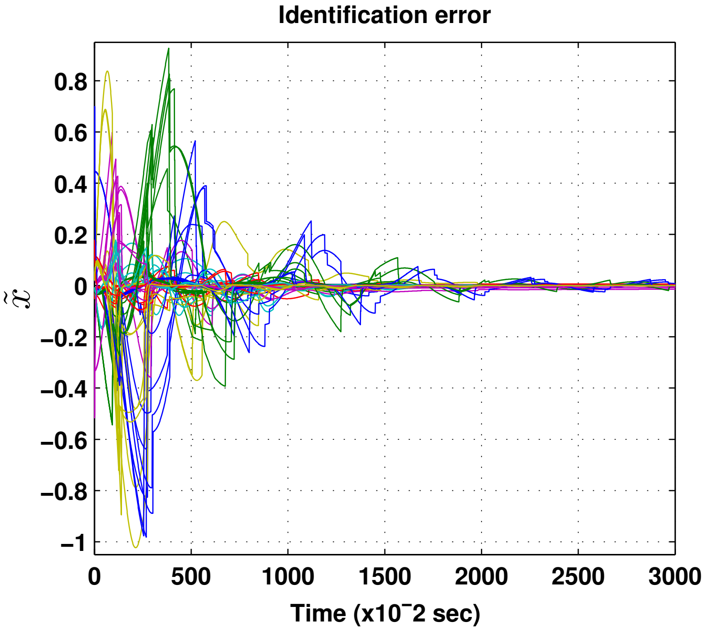


Fig. 4.4. Identifier approximation error.

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. 4.4.

The optimal value function is learnt using the consistency condition dictated by the HJB equation. A lower Hamiltonian/HJB residual error implies that the value function weight estimate is close to the target weights. Evidently, from Fig. 4.5, the enhanced/modified weight tuning rule improves the optimality due to faster convergence of the HJB residual error. This can be attributed to the fact that with the enhanced weight tuning, more information about the states and the corresponding value is utilized to tune the weights. This information is extracted from the reward signal obtained during the inter-event time with the estimated identifier states.

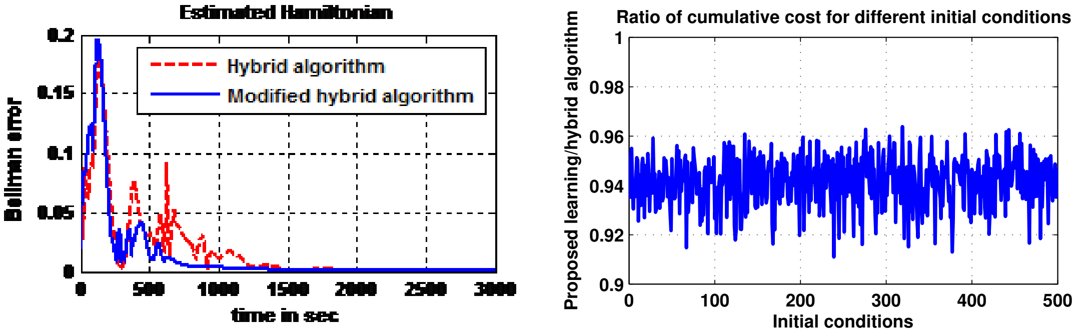


Fig. 4.5. Comparison of HJB error and Comparison of cost.

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. 4.5. 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. 4.5. 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. 4.6. 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 approxi-

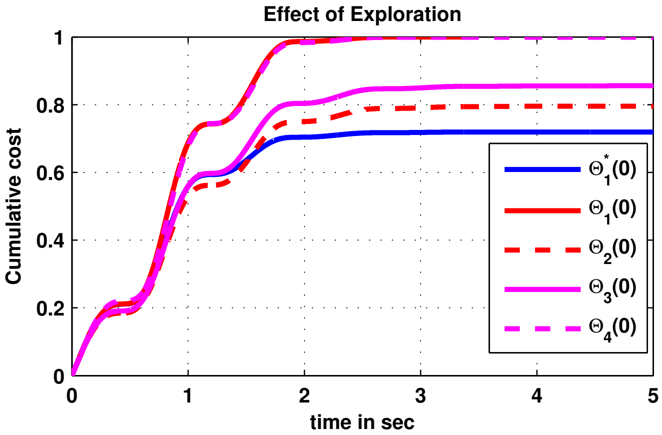


Fig. 4.6. Cost function trajectories.

mator 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. This again can be considered as choosing the cost function estimate which has yielded better reinforcement signals in the past. In doing so, the choice initial weights play far lesser role in the resulting cumulative cost during the learning period and hence, the transient performance.

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.

7. 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.

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] K. Doya, "Reinforcement learning in continuous time and space," *Neural computation*, vol. 12, no. 1, pp. 219–245, 2000.

[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] F. L. Lewis and D. Vrabie, "Reinforcement learning and adaptive dynamic programming for feedback control," *IEEE Circuits and Systems Magazine*, vol. 9, no. 3, pp. 32–50, Third 2009.

[6] 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.

[7] 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.

[8] 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.

[9] 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.

[10] 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.

[11] D. Liu, H. Li, and D. Wang, “Online synchronous approximate optimal learning algorithm for multi-player non-zero-sum games with unknown dynamics,” *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 44, no. 8, pp. 1015–1027, Aug 2014.

[12] 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.

[13] 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.

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

[15] V. Narayanan and S. Jagannathan, "Approximate optimal distributed control of uncertain nonlinear interconnected systems with event-sampled feedback," in *2016 IEEE 55th Conference on Decision and Control (CDC)*, Dec 2016, pp. 5827–5832.

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

[17] 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.

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

$$
\begin{aligned}
\dot{J}_{i\tilde{x}} = \mu_{i1}\tilde{X}_i^T P_i A_i \tilde{X}_i + \mu_{i1}\tilde{X}_i^T P_i(\tilde{W}_{if}\sigma_{if}(X_i) - \tilde{W}_{if}\tilde{\sigma}_{if} \\
+ [\tilde{W}_{ig}\sigma_{ig}(X_i) - \tilde{W}_{ig}\tilde{\sigma}_{ig}]U_{i,e} + \varepsilon_{ig}U_{i,e} + W_{ig}\tilde{\sigma}_{ig}U_{i,e} \\
+ \varepsilon_{if} + W_{if}\tilde{\sigma}_{if} + A_i E_i)
\end{aligned}
$$

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

$$
\begin{aligned}
\dot{J}_{i\tilde{x}} \le -\lambda_{\min}(\mu_{i1}\bar{q}_i)\left\|\tilde{X}_i\right\|^2 + \|\mu_{i1}\| \left\|\tilde{X}_i^T\right\| \|P_i\| \left\|\tilde{W}_{if}\right\| \\
\left\|\sigma_{if}(\hat{X}_i)\right\| + \|\mu_{i1}\| \left\|\tilde{X}_i^T\right\| \|P_i\| [\left\|\tilde{W}_{ig}\right\| \left\|\sigma_{ig}(\hat{X}_i)\right\| + \left\|\varepsilon_{ig}\right\| \\
+ \left\|W_{ig}\right\| \left\|\sigma_{ig}(X_i) - \sigma_{ig}(\hat{X}_i)\right\|] \left\|U_{i,e}\right\| + \|\mu_{i1}\| \left\|\tilde{X}_i^T\right\| \|P_i\| \\
[\left\|\varepsilon_{if}\right\| + \left\|W_{if}\right\| \left\|\sigma_{if}(X_i) - \sigma_{if}(\hat{X}_i)\right\| + \|A_i\| \|E_i\|]
\end{aligned}
$$

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; $\|\sigma_{i\bullet}\| \le N_{io\bullet}$, the number of hidden layer neurons, the subscript M with the weight variables denote the bounds on the target/ideal weights, $\|E_i\| \le E_{iM}$ and $\varepsilon_{i\bullet M}$ is the bound on the reconstruction errors.

Using the Youngs inequality ($\forall a, b, \epsilon > 0, ab \le \frac{a^2}{2\epsilon} + \frac{\epsilon b^2}{2}$), the Lyapunov derivative becomes

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

Utilizing the definition of the control policy, we have

$$
\begin{aligned}
\left\|U_{i,e}\right\|^2 \le& \left\|R^{-1}\right\|^2 W_{igM}^2 N_{iog}\left\|\nabla_x^T\Phi(x_{i,e})\Theta_i^*\right\|^2 + \left\|R^{-1}\right\|^2 N_{iog}\left\|\tilde{W}_{ig}\right\|^2\left\|\nabla_x^T\Phi(x_{i,e})\Theta_i^*\right\|^2 \\
&+ \left\|R^{-1}\right\|^2 N_{iog}\left\|\tilde{W}_{ig}\right\|^2\left\|\nabla_x^T\Phi(x_{i,e})\tilde{\Theta}_i\right\|^2 + \left\|R^{-1}\right\|^2 W_{igM}^2 N_{iog}\left\|\nabla_x^T\Phi(x_{i,e})\tilde{\Theta}_i\right\|^2.
\end{aligned}
$$

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

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

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

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

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

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

where the bounded term is given by

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

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 $\hat{\rho} = c_{if} + \tilde{X}_{i,e}^T\tilde{X}_{i,e}U_{i,e}^TU_{i,e}$ and the bounded term

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

The first derivative of the Lyapunov function is obtained as

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

Since the control policy is bounded and the final Lyapunov derivative expression reveals

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

where the bounds are defined as

$$\eta_{ioB} = \eta_{iogB} + \eta_{iofB} + \eta_{io\tilde{x}B} + (5.5 + 0.125\|\mu_{i1}\|^4\|P_i\|^4N_{iog}^4)\left\|\nabla_x^T\Phi(x_{i,e})\tilde{\Theta}_i\right\|^4.$$

This reveals that the identification and weight estimation errors of the identifiers at each subsystem are locally UUB if the control policy is bounded.

Proof of Theorem 1 (local ISS):

Consider the Lyapunov function for the interconnected system

$$J = \sum\nolimits_{i=1}^{N} J_i(x_{i,}\tilde{\theta}_i, \tilde{X}_i, \tilde{W}_{if}, \tilde{W}_{ig}),$$

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

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

This derivation follows the derivation in [6]. In [6], 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

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

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

Now, simplifying the terms B_1, A_1

$$\begin{aligned}&\|B_1\|^2 \leq 4.5\left\|\tilde{\theta}_i^T \nabla_x \phi(x_e)\right\|^4 + 20\left\|\tilde{g}_i^T(x_e)\right\|^8 + 8.5\left\|\tilde{\tilde{f}}_i(x_e)\right\|^4 \\ &+ 25.6\left\|\tilde{g}_i^T(x_e)\right\|^8 V_{ixM}^4 \left\|R^{-1}\right\|^4 + 0.75\left\|\tilde{g}_i^T(x_e)\right\|^4 + B_{i\tilde{\theta}1}. \\ &1.25\left\|\left(\tilde{\theta}_i^T A_1\right)\right\|^2 \leq 0.47\left\|\tilde{\theta}_i^T \nabla_x \phi(x_e)\right\|^4 + 20\left\|\tilde{\tilde{f}}_i(x_e)\right\|^4 + 12.5\left\|\tilde{g}_{i,e}^T\right\|^8 \\ &+ 18V_{ixM}^4 \left\|R_i^{-1}\right\|^4 \|\tilde{g}_i(x_e)\|^8 + 40L_{if}^4 E_i^4 + 177V_{ixM}^4 D_{iM}^4 + 4713V_{ixM}^8 D_{iM}^4 \left\|R_i^{-1}\right\|^4.\end{aligned} \tag{26}$$

Collecting the bounded terms from (24) and (23), we have $B_{i\tilde{\theta}1} = 256V_{ixM}^{8}\left\|R^{-1}\right\|^{4} + 16D_{iM}^{2}V_{ixM}^{4} + 512V_{ixM}^{8}D_{iM}^{2}\left\|R^{-1}\right\|^{2} + 16L_{iq}^{2}E_{i}^{2} + 16D_{iM}^{4}V_{ixM}^{4} + 32L_{if}^{4}E_{i}^{4} + 16V_{ixM}^{2}L_{if}^{2} + 32V_{ixM}^{4} + 256D_{iM}^{4}V_{ixM}^{4} + 6553.6D_{iM}^{4}V_{ixM}^{8}\left\|R^{-1}\right\|^{4} + 4D_{iM}^{2}V_{ixM}^{4}$. Using the expression $\tilde{\theta}_i^T\gamma_i\alpha_{iv}\hat{\psi}_{i,e}\hat{H}_{i,e}/\bar{\rho}^2$ and substituting the inequalities obtained in equations (23) and (24), we get

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

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

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

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

$$B_{i\tilde{\theta}} = \alpha_{iv}^2 D_{iM}^2 \varepsilon_M^2 + 2\alpha_{iv}^4 D_{iM}^4 \varepsilon_M^4 + 2\varepsilon_{iM}^8 D_{i\max}^4 + (0.5\alpha_{iv}^2 + 2)2\varepsilon_{iM}^4 D_{i\max}^2 + 32(\frac{\alpha_{iv}^2}{2} + 2)^2 \varepsilon_{iM}^2$$
$$+ 64\alpha_{iv}^4 + \gamma_i \alpha_{iv}^2 (40 L_{if}^4 E_i^4 + \frac{41}{10} D_{iM}^4 V_{ixM}^4 + 172 V_{ixM}^4 D_{iM}^4 + 4713 V_{ixM}^8 D_{iM}^4 \left\|R_i^{-1}\right\|^4)$$
$$+ .5\gamma_i^2 \kappa_3^2 \theta_{iM}^2 + 0.5\gamma_i(\alpha_{iv}^2 + 2)B_{i\tilde{\theta}1} + \bar{\rho}^2(\eta_{ioB} + \gamma_i^4 + 2\|P_i e_i\|^2 + 0.25\|e_i\|^2),$$
$$\upsilon_{\tilde{f}} = \lambda_{\min}(\mu_{i4}\kappa_{if}) - \frac{5}{2} - \frac{4\gamma_i}{\bar{\rho}^2}(6.07\alpha_{iv}^2 + 2.14)N_{iof}^2,$$
$$\upsilon_{\tilde{g}^2} = 3 - \frac{1}{2}\left\|R^{-1}\right\|^4 + 4 - \frac{18\gamma_i}{\bar{\rho}^2}(\frac{71.2}{100}(\alpha_{iv}^2 + 2) + \alpha_{iv}^2)N_{iog}^4(1 + \left\|R_i^{-1}\right\|^4 V_{ixM}^4),$$
$$\upsilon_x = (\bar{\bar{q}}\alpha_{iv} - 0.875 - \gamma_i(4\alpha_{iv}^4 + 0.0625 + 2\varepsilon_{iM}^4)C_i^4/\bar{\rho}^2),$$
$$\eta_{icl} = \frac{B_{i\tilde{\theta}}}{\bar{\rho}^2} + \frac{\alpha_{iv}^8}{8} D_{iM}^4 \left\|R_i{}^{-1}\right\|^4 + \frac{\alpha_{iv}^2}{2} D_{iM}^2(\varepsilon_M^2 + V_{ixM}^2) + \frac{1}{2}(\alpha_{iv}^4 D_{iM}^4 + \alpha_{iv}^2 V_{ixM}^2 D_{iM}^2$$
$$+ \alpha_{iv}^4 V_{ixM}^4 D_{iM}^2 \left\|R_i{}^{-1}\right\|^2),\ \upsilon_{\tilde{\theta}2} = \frac{\upsilon_{\tilde{\theta}}\gamma_i}{32} - \frac{\gamma_i \alpha_{iv}^2}{2} - 2.25\gamma_i(\alpha_{iv}^2 + 2) - N\bar{\rho}^2(5.5 + 0.225$$
$$\|\mu_{i1}\|^4 \|P_i\|^4 N_{iog}^4) - \frac{1}{128}\mu_{i1}^4 \bar{\rho}^2 \left\|\hat{D}_{i,\varepsilon}^T\right\|^4 - \frac{1}{8}\beta_{iv}^4 \bar{\rho}^2 \left\|\hat{D}_{i,\varepsilon}\right\|^4,$$
$$\upsilon_{\tilde{g}} = \lambda_{\min}(\mu_{i5}\kappa_{ig}) - 3.5 - 3\gamma_i(\alpha_{iv}^2 + 2)N_{iog}^2/8\bar{\rho}^2.$$

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 $\eta_{\Theta1} = \bigcup_{i=1}^{N} \eta_{i\tilde{\Theta}1}$. 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

$$\dot{J}_{i\tilde{f}} \leq -(\lambda_{\min}(\mu_{i2}\kappa_{if}) - 1)\left\|\tilde{W}_{if}\right\|^2$$
$$-(\lambda_{\min}(\mu_{i4}\kappa_{if}) - 2)\left\|\tilde{W}_{if}\right\|^4 + \eta_{iofB}$$
$$\eta_{iofB} = 0.5\|\mu_{i2}\|^2\left\|\alpha_{if}\right\|^2(N_{iof}^2 + W_{ifM}^2\left\|\kappa_{if}\right\|^2/\left\|\alpha_{if}\right\|^2)$$
$$+ 0.125\|\mu_{i4}\|^4\left\|\alpha_{if}\right\|^4(N_{iof}^4 + W_{ifM}^4\left\|\kappa_{if}\right\|^4/\left\|\alpha_{if}\right\|^4).$$

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

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

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

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

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 the time derivative of the combined Lyapunov function of the closed loop system as

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

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 $\{t_k^i\} \longrightarrow \infty$. 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 $t \longrightarrow \infty$. 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.

IV. ADAPTIVE OPTIMAL EVENT-TRIGGERED CONTROL OF LINEAR DYNAMIC SYSTEMS

ABSTRACT

Event-triggered control implementation is considered as an alternative to the traditional periodic implementation of control tasks. The advantage of such event-triggered control implementation is that the cost of communication and computations are considerably scaled down without affecting the fidelity of the controller. To determine the event-triggering instants, a state dependent threshold function is designed that bounds the event-triggering error which is defined as the deviation between the actual system state and the state information available at the controller. In this paper, a novel approach for optimizing the event-triggering instants and optimal state feedback controller co-design is proposed for linear dynamical systems using zero-sum game theory. The design task is formulated as a problem of finding the maximizing threshold for generating events and minimizing control policy to ensure satisfactory system performance. First, a solution to this min-max, optimization problem is proposed using the zero-sum game theory, when the system dynamics are known and then, a novel adaptive optimal solution using Q-learning is proposed for the case when the system dynamics are uncertain. Finally, an adaptive optimal decentralized event-triggering mechanism and distributed control co-design for a class of linear-interconnected system is presented. Theoretical results are substantiated by numerical examples via simulation.

1. INTRODUCTION

The study of event-triggered feedback and control implementation dates back to the early sixties [1]. A state based adaptive sampling method for a sampled data servo mechanism is first proposed in [5] wherein the adaptive sampling rate is controlled by the absolute value of the rate of change of the error signal with time. Further, the advantages of event-based sampling over the traditional feedback approach are presented for a first-order system by K.J. Astrom et al. [1]. Later, the theoretical framework for event-triggered control is formalized and an emulation based approach for event-triggered control implementation is presented for a networked control system and various results emphasizing its inherent advantages in computation and communication cost saving are studied [20].

In the emulation based design the continuous controller is presumed to be stabilizing and an event-triggering condition is developed to implement the controller such that the system stability is preserved. In the earlier works [9, 12, 13, 20, 24] the controlled system is assumed to be input-to-state stable (ISS) with respect to the measurement error and event-triggering conditions are designed to reduce the frequency of feedback instants while guaranteeing asymptotic stability. A non-zero positive lower bound on the inter-event times is also guaranteed to avoid accumulation point and zeno behavior. Further, various event-triggered control schemes are presented to accommodate other design considerations, such as, stochastic feedback control design [9], state-feedback design [12], state-estimation problem [18], decentralized event-triggering for wireless sensor networks [13] and distributed control design [24], trajectory tracking control [21] and robust controller design [4, 11].

In all the above design approaches, the sensor measurements and the control input are held between two consecutive events by a zero order hold (ZOH) circuit at the controller and actuator, respectively. In contrast, a model of the system is used to reconstruct the system state vector and, subsequently, used for designing the control input when the actual feedback is unavailable [6, 8]. As the control input is based on the model states, no feedback transmission is required unless there is a significant change in the system performance due to external disturbance or internal parameter variation. The asymptotic stability of system states is guaranteed by designing the event-triggering condition with this model-based approach. It is observed that the model-based approach reduces the number of events more effectively when compared to the ZOH based approach, but, requires an accurate model of the system.

It should be noted that in all the approaches, [1, 5, 9, 12, 20], [4, 6, 8, 11, 21, 24], the event-triggering instants are designed to ensure system stability. On the other hand, few results are available in the literature which presents an optimization based approach for generating events. Notably, the authors in [14] characterized a certainty equivalence controller to be optimal in a linear quadratic Gaussian (LQG) frame work. The optimal control input and the optimal event-triggering instants are designed using the separation principle. Furthermore, a suite of optimal event-triggering design is studied in [15].

In [15, 22], the event-triggering mechanism is optimized by formulating a cost function that penalizes the number of events and successive event-triggering instants are identified by minimizing the cost function assuming the knowledge of system dynamics. Further, an event-triggering mechanism is proposed for a discrete time model-predictive controller scheme wherein events are generated whenever the states evolve and leave a polytope obtained using quadratic programming. Further, authors in [13] propose a heuristic algorithm for generating the event-triggering instants. More recently, a dynamic event-triggering mechanism is proposed wherein an additional adaptive parameter is introduced in the event-triggering condition for a system with known dynamics and the inter-event time is obtained as the function of this adaptive parameter [7]. In summary, the optimal event-triggering and controller design are, in general, considered as independent problems and various approaches are proposed to tune the parameters of each of them to elicit a desired performance from the controlled system. However, to the best knowledge of the authors, the optimal co-design of controllers and the event-triggering mechanism is not considered in the literature.

Motivated by the above facts, in this paper, a novel optimal event-triggered control design scheme for linear systems is presented. Firstly, the design of control policy and the event-triggering mechanism is formulated as a two player zero sum game (min-max) problem. Therefore, a novel cost function is proposed as a function of states, control policy and the measurement/event-triggering error. The saddle point solution to this min-max problem results in a minimizing control policy and a maximizing measurement error policy that can be injected into the system. In the proposed design, the maximizing policy is utilized as the dynamic threshold to the event-triggering error to generate events while the minimizing control policy is applied to the system. Since the control policy explicitly accounts for the worst-case event-triggering error, the performance of the system is preserved. Moreover, since the inter-event time is directly proportional to the event-triggering error, utilizing the maximizing policy as a dynamic threshold for the event-triggering error, results in an increased inter-event time. This results in an optimal event-triggered controller which explicitly takes into account event-triggering and control policy design.

Further, an optimal adaptive event and control co-design scheme is proposed which relaxes the requirement of accurate knowledge of system dynamics. A Q-learning scheme is presented to determine the optimal control and event-triggering instants forward-in-time. Lyapunov stability analysis is used to guarantee stability of the closed-loop system. Finally, a decentralized event-triggered control implementation for distributed control of interconnected system is presented. An optimal distributed control law and an optimal decentralized event-triggering rule at each subsystem is generated. The problem is addressed for linear time invariant system in continuous time with uncertain dynamics.

The contributions of the paper include: 1) A novel optimal event-triggering and controller co-design using zero sum game; 2) Development of an adaptive optimal online Q-learning scheme for learning the optimal control and event-triggering policy; 3) Design of decentralized adaptive optimal event-triggering for distributed control of interconnected systems; 4) Lyapunov based stability analysis and verification of the proposed design using numerical examples via simulation.

The paper is organized as follows. In the section II, the system dynamics is introduced and the problem statement is presented. In section III, the main results are presented for the case when the system dynamics are known. In section IV, a model free Q-learning approach is proposed to solve the optimization problem forward-in-time, online, when the system dynamics are uncertain. Section V presents an optimal adaptive event-triggered distributed controller for interconnected system using the proposed method. Finally, simulation results are provided to show the effectiveness of the controller designed in section VI. Conclusions follow in section VII.

In this paper, $\Re$ denotes the set of all real numbers; **N** denotes the set of all natural numbers. Euclidean norm is used for vectors and Frobenius norm is used for matrices.The next section presents a brief background on the system dynamics and recall some of the results on ISS and optimal control.

2. BACKGROUND AND PROBLEM STATEMENT

2.1 System Description

Consider the dynamical system represented by

$$\dot{x}(t) = Ax(t) + Bu(t),, \;\; x(0) = x_0 \tag{1}$$

where $x \in \Re^n$ is the state vector of the system; $u : \Re^n \to \Re^m$ is the control input, $A : \Re^n \to \Re^n, B : \Re^n \to \Re^{n\times m}$ are linear maps representing internal dynamics and input gain matrix. The control signal for the system (1) is of the form

$$u(t) = Kx(t) \tag{2}$$

where K is a linear map.

In the event-triggered control framework, the feedback information is utilized only at certain discrete event-based sampling instants to update the control action. To represent these event-triggering instants, define a sequence of time instants $\{t_k\}_{k\in\{0,\mathbb{N}\}} \subseteq t$, such that $0 = t_0 < t_1 < ...$ Using a ZOH, the control input will be held at the actuator such that $u(x(t)) = u(\breve{x}(t))$ where in $\breve{x}(t) = x(t_k)$, $\forall t \in [t_k, t_{k+1})$. Hence, the control signals are piecewise continuous.

To determine the event-triggering rule, define the (measurement error) event-triggering error as the difference between the actual state measured and the states available at the controller

$$e(t) = \breve{x}(t) - x(t), \quad \forall t \in [t_k, t_{k+1}). \tag{3}$$

It can be observed that at the sampling instants $e_i(t) = 0$.

Assumption 1: The time required to read the state from the sensors and compute the control signal and update the actuators is considered negligible. Next, the optimal design of the sequence $\{t_k\}$, and the optimal control (2) co-design problem is defined.

2.2 Problem Statement

Consider the controllable linear time-invariant continuous-time system represented by (1). Let the control policy (2) be implemented with event-triggered feedback. Then, the system dynamics (1) can be re-written as

$$\dot{x}(t) = Ax + Bu(\breve{x}). \tag{4}$$

Define the performance measure for (1) to be

$$\|\zeta(t)\|^2 = x^T(t)Qx(t) + u^T(t)Ru(t) \tag{5}$$

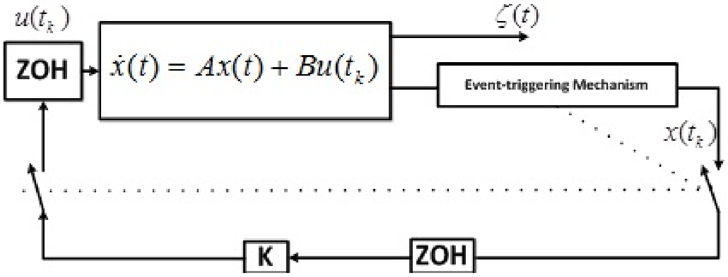


Fig. 5.1. Networked control system and event-triggered feedback.

where Q is a positive semidefinite and R being a positive definite matrices respectively.

A block diagram of the control architecture for event-triggered implementation of feedback control is given in the Fig. 5.1. The event-triggering mechanism determines the time-instants to close the feedback loop so that the latest state vector is used to generate the control signal. The gain and the control policy in (2) are generated by solving a minimization problem associated with the performance measure (5) whereas the event-execution rule determines the sampling instant sequence by designing an upper bound for the measurement error (3) based on the stability of the controller system.

In this paper, the objective is to develop an optimal state âĂŞfeedback control policy which minimizes (5) while simultaneously minimizing the number of events and meeting the performance of the system (5).

Remark 1: To determine the event-triggering instants, the state vector is sampled as a function of the event-triggering error by using the stability criterion [20]. Alternatively, in the inter-event period, Lyapunov function is utilized explicitly to determine the event-triggering instants [24]. Nevertheless, in general, stability of the system is considered to determine the event-triggering instants. The block diagram in Fig. 5.1 is a synchronous triggering scheme wherein the occurrence of an event closes the switch on either sides of the controller. Asynchronous event-triggering schemes require two event-triggering mechanisms one for each switch.

In the next section, a zero-sum game based control scheme is presented which satisfies the objectives defined in this section. The resulting synchronous event-triggering mechanism increases the time between successive events while the optimal control policy ensures system performance.

3. PROPOSED SOLUTION

In this section, the control input and the measurement error due to event-triggered feedback will be considered as two non-cooperative players applied to the system. A cost function is defined as a function of system state vector, control input vector and the measurement error. It will be demonstrated that the objectives listed in the previous section will be achieved by determining a saddle point solution to the optimization problem associated with the cost function. The maximizing measurement error will act as the threshold to generate events while the optimal control policy will be applied to the system with the feedback generated at these events. Existence of such saddle point solution to the min-max optimization problem depends on some fundamental properties of the system which are presented in Lemma 1.

Utilizing the system dynamics (4) and the definition of the measurement error (3), we can rewrite the dynamics as

$$\dot{x}(t) = Ax + Bu + D\eta \tag{6}$$

where $\eta = Ke$, $D = B$. Now, define the infinite horizon cost function using the performance measure (5) as

$$J(x, \eta, u) = \int_t^\infty [\|\zeta(t)\|^2 - \sigma^2 \eta^T \eta] d\tau. \tag{7}$$

where $\sigma > 0$ represents the attenuation constant. The objective is to find an optimal saddle-point solution (u^*, η^*) so that the optimal value function satisfies

$$V^*(x(t)) = \min_u \max_\eta J(u, \eta) = \max_\eta \min_u J(u, \eta). \tag{8}$$

Using the infinitesimal version of the cost function (7) and the system dynamics (6), the Hamiltonian function can be defined as

$$H = x^T Q x + u^T R u - \sigma^2 \eta^T \eta + V_x^T [Ax + Bu(t) + D\eta(t)] \tag{9}$$

where $V_x = \partial V / \partial x$, $V(x)$ is the value function defined by using the integral in (7). The optimal policies are obtained as

$$u(x, V_x^*) = -\frac{1}{2} R^{-1} B^T V_x^* \tag{10}$$

$$\eta(x, V_x^*) = \frac{1}{2\sigma^2} D^T V_x^* = \eta^*(x, V_x^*) \tag{11}$$

where V_x^* is the gradient of the optimal value function with respect to the states. Substituting the optimal policies in the Hamiltonian results in the continuous-time game algebraic Riccati equation (GARE).

Lemma 1:([3, 10]) Consider the infinite horizon cost function (7) and the linear system dynamics (6). Let the pair A, B be controllable and the pair A, C be observable, with $Q = C^T C$. Then, there exists a positive definite solution P^* for the GARE when $\sigma > \sigma^*$, where σ^* is the H_∞ gain of the system. Moreover, the optimal cost function is quadratic and satisfies $V^*(x_0) = x_0^T P^* x_0$.

Remark 2: Note that if there is a positive definite solution to the GARE, then the optimal cost function is finite and the control policy asymptotically stabilizes the system. The proof for Lemma 1 can be found in [2].

Lemma 2: Consider the infinite horizon cost function (7) and the linear time-invariant system dynamics (6). Let $P^* > 0$ be the positive definite solution for the GARE, then the optimal policy given by

$$u(x, P^*) = K^* x(t) = -\frac{1}{2} R^{-1} B^T P^* x(t) \tag{12}$$

generates an ISS Lyapunov function for (6) with respect to the measurement error $e(t)$. *Proof:* See Appendix.

Remark 3: The smooth function $L(x)$ satisfy, $x^T \lambda_{\min}(P^*)x \leq x^T P^* x \leq x^T \lambda_{\max}(P^*)x$, where $\lambda_{\min}(.)$and$\lambda_{\max}(.)$ represent the minimum and maximum singular values of the matrix (.). Further, from the proof of Lemma 2, we have $\|x\| \geq \gamma \|e\|$ implies $\dot{L}_x < 0$, where $\gamma = \|P^* DK\| / \|\delta_{x,m}\|$, $\delta_{x,m}$ is a constant defined in the proof of Lemma 2 using $Q$and$R$.

Next, the main results of this section are presented.

Theorem 1: Consider the infinite horizon cost function (7) and the linear time-invariant system dynamics (6). Let $P^* > 0$ be the positive definite solution for the GARE, then the optimal policy given by (12) be applied to the system with the following event-triggering condition given by

$$\|\eta(t)\| \leq \|\eta^*(t)\|, \quad t \in [t_k, t_{k+1}), \forall k \in \mathbb{N}. \tag{13}$$

Then, the closed-loop system is asymptotically stable when Q, R, σ are selected such that

$$\delta_{x,m} > \frac{1}{2\sigma^2} P^* D^2. \tag{14}$$

In addition, a positive minimum inter-event time, τ, exists, such that

$$\tau \geq \frac{1}{\|DK^*\|} \log(\frac{\|DK^*\|}{X_m} \|e^*\| + 1). \tag{15}$$

where $e^*(t) = K^+ \eta^*(t)$, K^+ is the generalized inverse of K^* in (12) and $X_m > 0$ is a positive scalar value.

Proof: See Appendix.

Remark 4: The matrix K^* may not have an inverse as it may not be a square matrix. The generalized inverse of K^* is used in (15) to quantify the minimum inter-event time and it is not used to derive the control policy or event-triggering condition.

Remark 5: The proposed event-trigger condition (13) allows the measurement error to grow until the system performance defined by the performance measure (5) is not deteriorated. This increases the inter-event time based on the relation between inter-event time and event-triggering threshold derived in the proof of Theorem 1. In comparison to the literature, the proposed design optimizes the system performance while reducing the frequency of feedback instants and controller implementation.

Remark 6: If the system dynamics are re-written as $\dot{x}(t) = Ax + Bu + D\eta$, with $D = BK$ and $\eta \in \Re^n$. The cost function (7) can be used to formulate a maximization problem such that the optimal cost $V^*(x(t)) = \max_e J(e)$ and the event-triggering condition can be defined as $\|e(t)\| \leq \|\eta^*(t)\|$. In this formulation, at each sampling instant, the control gain, K is fixed and a maximum threshold, η^*, for the event-triggering error, is determined. In contrast, the proposed min-max based optimization scheme determines the optimal control policy and the event-triggering condition, simultaneously.

Remark 7: The expression for the inter-event time derived in the proof of Theorem 1 can be used to find the successive event-triggering instants and can be used to develop an optimal self-triggering control scheme. Such a scheme does not require checking the event-triggering condition (13) continuously as the time instants $\{t_k\}$ are pre-computed.

In the next section, the system matrices A, B will be considered uncertain and an optimal adaptive event-triggered control design using hybrid Q-learning approach is presented.

4. CONTROLLER DESIGN Q-LEARNING

The saddle point solution for the proposed min-max problem formulated for the event-triggered control design can be learned online, forward-in-time using model free Q-learning approach. This also relaxes the requirement of accurate knowledge of matrices A, B in the system dynamics. A block diagram of the proposed learning scheme is given in Fig. 5.2. It can be observed that in order to learn the optimal control policy and the optimal event-triggering threshold, two Q-function estimators are required one at the controller and the other at the event-triggering mechanism.

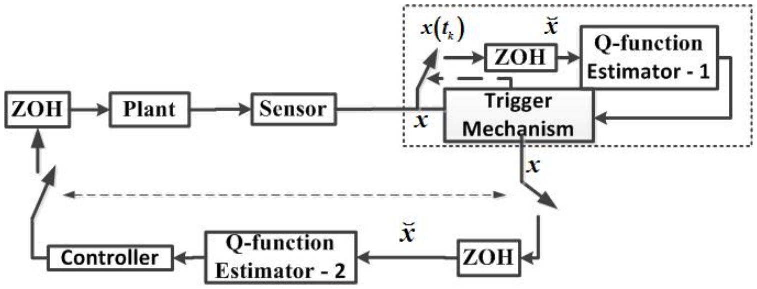


Fig. 5.2. Adaptive optimal event sampled control system.

First, the Q-function is formulated and to formulate the Q-function, consider the optimal cost function in quadratic form as

$$V(x(t)) = x^T(t)Px(t). \tag{16}$$

Taking the time-derivative of (16) and using the system dynamics (6), we have

$$\begin{aligned} \dot{V} &= x^T PAx + x^T PBu + x^T PD\eta + x^T A^T Px \\ &+u^T B^T Px + \eta^T D^T Px. \end{aligned} \tag{17}$$

Using the infinitesimal version of the cost function (7), we get

$$\dot{V} = -x^T Qx - u^T Ru + \sigma^2 \eta^T \eta. \tag{18}$$

Using (18) in (17) and adding (16) on both sides yields

$$\begin{aligned} V &= x^T PAx + x^T PBu + x^T PD\eta + x^T A^T Px+ \\ u^T B^T Px &+ e^T D^T Px + x^T Qx + u^T Ru - \sigma^2 \eta^T \eta + x^T Px. \end{aligned} \tag{19}$$

Using (19), the action-dependent value function or the Q-function can be defined as

$$\begin{aligned} &Q(x, u, \eta) = \\ &\begin{bmatrix} x(t) \\ u(t) \\ \eta(t) \end{bmatrix}^T \begin{bmatrix} A^T P + PA + Q + P & PB & PD \\ B^T P & R & 0 \\ D^T P & 0 & -\sigma^2 \end{bmatrix} \begin{bmatrix} x(t) \\ u(t) \\ \eta(t) \end{bmatrix} \end{aligned} \tag{20}$$

$$\begin{aligned} Q &= Z^T(t)GZ(t) = Z^T(t) \begin{bmatrix} G^{vv} G^{v\mu} G^{ve} \\ G^{\mu v} G^{\mu\mu} G^{\mu e} \\ G^{ev} G^{e\mu} G^{ee} \end{bmatrix} Z(t) \\ &= \theta^T \phi(t) \end{aligned} \tag{21}$$

where $\theta \in \Re^{\frac{(2n+m)(m+2n+1)}{2}\times 1}$ is the G matrix parameters represented in vector form, $\phi(t) \in \Re^{\frac{(2n+m)(m+2n+1)}{2}\times 1}$ is the kronecker product given by $[x^T(t)u^T(t)\eta^T(t)] \otimes [x^T(t)u^T(t)\eta^T(t)]^T$. The unknown parameter, θ can be estimated adaptively when the matrix G in (21) is uncertain. To derive an update equation, consider the time derivative of the optimal value function

$$\dot{V}^* = V_x^{*T}[Ax(t) + Bu^*(t) + D\eta^*(t)]. \tag{22}$$

Using the definition of the optimal policies (10) an (11)

$$\dot{V}^* = V_x^{*T}Ax(t) - 2u^{*T}(t)Ru^*(t) + 2\sigma^2\eta^{*T}(t)\eta^*(t). \tag{23}$$

From GARE, we have

$$-Q(x) + \frac{1}{4}V_x^{*T}BR^{-1}B^TV_x^* - \frac{1}{4\sigma^2}V_x^{*T}D^TDV_x^* = V_x^{*T}Ax. \tag{24}$$

Using (24) in (23) to get

$$\begin{array}{l} \dot{V}^* = -2u^{*T}(t)Ru^*(t) + 2\sigma^2\eta^{*T}(t)\eta^*(t) - x^TQx \\ +\frac{1}{4}x^TP^*BR^{-1}B^TP^*x - \frac{1}{4\sigma^2}x^TP^*D^TDP^*x \end{array}. \tag{25}$$

With the definitions (10) and (11), (25) is simplified as

$$\dot{V}^* = -x^TQx - u^{*T}(t)Ru^*(t) + \sigma^2\eta^{*T}(t)\eta^*(t). \tag{26}$$

Integrating both sides of (26), in the interval $[t_k, t_{k+1})$, reveals

$$\begin{array}{l} V^*(t_{k+1}) - V^*(t_k) = \\ \int_{t_k}^{t_{k+1}} (-x^TQx - u^{*T}Ru^* + \sigma^2\eta^{*T}\eta^*)d\tau. \end{array} \tag{27}$$

The Bellman equation is a fixed point equation with the optimal value function being the fixed point solution. Therefore, if the optimal value function and control policies in (27) are replaced by the estimated quantities, there would be an error, referred to as the Bellman or temporal difference error.

Define the estimate of the optimal value function as $\hat{V}$. Now, replacing the optimal value function in (27) with the estimated optimal value function reveals [10]

$$\begin{aligned} E_{k+1}(t) = &\int_{t_k}^{t_{k+1}} (x^T Qx + u^T Ru - \sigma^2 \eta^T \eta) d\tau \\ &+ \hat{V}(t_{k+1}) - \hat{V}(t_k) \end{aligned} \tag{28}$$

where E_{k+1} is the Bellman residual error/ temporal difference error calculated at the occurrence of $k+1$ event. Define the estimated parameter vector $\hat{\theta}$ and the estimated Q-function

$$\hat{Q}(x, u.e) = Z^T(t)\hat{G}(t)Z(t) = \hat{\theta}^T \phi(t). \tag{29}$$

The forcing function for the parameter update equation is obtained as

$$E_{k+1}(t) = \int_0^T (x^T Qx + u^T Ru - \sigma^2 \eta^T \eta) d\tau + \hat{\theta}^T \Delta\phi(\Delta T) \tag{30}$$

where $\Delta\phi(\Delta T) = \phi(t_{k+1}) - \phi(t_k)$ and $E_{k+1}(t)$ is the residual error calculated at the event-sampling instant t_{k+1}. Define parameter estimation error as $\tilde{\theta} = \theta - \hat{\theta}$. Then one step cost calculated using (27) with the target parameters θ and the one step cost calculated using (30) with the estimated parameter $\hat{\theta}$, reveals

$$\begin{aligned} &-E_{k+1}(t) = \theta^T(t)\Delta\phi(\Delta T) - \hat{\theta}^T(t)\Delta\phi(\Delta T) \\ &= \tilde{\theta}^T(t)\Delta\phi(\Delta T) \end{aligned} . \tag{31}$$

Theorem 2: Consider the infinite horizon cost function (7) and the linear time-invariant system dynamics (6). Let θ be the time-invariant, bounded, target parameter vector and $\hat{\theta}(0)$ be an initial estimated parameter vector defined in a compact set. Select the control policy

$$u(t) = -\frac{1}{2} R^{-1} \hat{G}^{\mu\nu}(t) x(t), \tag{32}$$

to be applied on the system with $\hat{G}^{\mu\nu}$ being the estimated submatrix in (21). Let the event-triggering condition be satisfying

$$\|\eta(t)\| \leq \|\hat{\eta}(t)\|, \quad t \in [t_k, t_{k+1}), \forall k \in \mathbb{N}. \tag{33}$$

where $\hat{\eta} = \frac{1}{2\sigma^2}\hat{G}^{ev}(t)x(t)$, $\hat{G}^{ev}$ is the estimate of the matrix G^{ev} in (21). Consider the Q-function parameter adaptation rule given by

$$\dot{\hat{\theta}} = -\alpha \frac{[\Delta\phi(\Delta T)]}{(1 + [\Delta\phi(\Delta T)]^T[\Delta\phi(\Delta T)])^2} E_{k+1}^T(t), \ \forall k = 0, 1, .. \tag{34}$$

Then, the state vector and the parameter estimation error converges asymptotically to zero with the event-triggering instants $k \longrightarrow \infty$, provided the design parameters α, Q, R, σ are chosen such that the following inequalities hold: $\bar{\delta}_x > 2K_M^2$, $\frac{\alpha}{\rho} > \frac{1}{2}\|R\|^2$ where $\rho = (1 + [\Delta\phi(\Delta T)]^T[\Delta\phi(\Delta T)])^2$, $\|K^*\| \le K_M$, $\bar{\delta}_x = [Q + \frac{1}{4}P^*BR^{-1}B^TP^* - \frac{1}{4\sigma^2}P^*DD^TP^*]$, $\alpha > 0$ is the learning step, $K_M > 0$ is a constant.

Proof: See Appendix.

Remark 8: The Bellman error, E_k is calculated at every event-triggering instant, t_k and the parameters are updated continuously using (34) both at the event sampling and inter-event intervals. The update rule utilizes the new information obtained at the event-triggering instant to calculate the Bellman error, $E_k(t)$ and the updates in the inter-event period, $[t_k, t_{k+1})$ tries to reduce the Bellman error calculated at the last event-triggering instant, t_k. In contrast to the traditional policy-iteration [10, 19] scheme, the parameter tuning proposed in (34) is a hybrid learning scheme [16] which can be implemented online.

In the next section, the hybrid Q-learning zeroâĂŞsum game theoretic formulation is extended to the distributed control of interconnected systems and the decentralized event-triggering conditions are presented.

5. EVENT-TRIGGERED DISTRIBUTED CONTROL

Consider an inter-connected system operating in continuous time with N interconnected subsystems, each of the form

$$\dot{x}_i(t) = A_i x_i(t) + B_i u_i(t) + \sum_{\substack{j=1 \\ j \neq i}}^{N} A_{ij} x_j(t), \quad x_i(0) = x_{i0} \tag{35}$$

where $x_i, \dot{x}_i \in \Re^{n_i \times 1}$ are the states and state derivatives of the i^{th} subsystem; $u_i \in \Re^{m_i}, A_i \in \Re^{n_i \times n_i}, B_i \in \Re^{n_i \times m_i}$, are the control inputs, internal dynamics and control gain matrices of the i^{th} subsystem, and $A_{ij} \in \Re^{n_i \times n_j}$ represents the interconnection between the i^{th} and j^{th} subsystems. The overall system description is given by

$$\dot{x}(t) = Ax(t) + Bu(t)x(0) = x_0 \tag{36}$$

where $x \in \Re^n, u \in \Re^m, B \in \Re^{n \times m}, A \in \Re^{n \times n}, u = [u_1^T, ., u_N^T]^T, \ A = (A_{ij})_{i,j=1,.,N}, A_{ii} = A_i, B = diag[B_1, ., B_N]$.

Assumption 2: The overall system described by (36) is controllable and all the states are measurable. The subsystems share their state information through a lossless network. Further, the order of the subsystems is known.

Lemma 3: Consider the subsystem (35) of the interconnected system (36). The control policy which stabilizes (36) renders the individual subsystems asymptotically stable.

Proof: See Appendix.

A distributed control law can be obtained if a centralized optimal controller is designed for the overall system, which is not feasible and hence, the Q-function is estimated at each subsystem to obtain a distributed control law at each subsystem, thereby minimizing the Hamiltonian and cost function of the overall subsystem. A hybrid Q-learning scheme which incorporates random delays and packet losses for distributed control is presented in [16]. The event-triggering mechanism, however, is designed using the Lyapunov function. In contrast, the control scheme proposed in this paper can be used to obtain an optimal adaptive even-triggering and distributed control co-design. To avoid redundancy only the main result is introduced next.

Corollary: Consider the infinite horizon cost function (7) and the linear time-invariant system dynamics (35). Let θ be the time-invariant, bounded, target parameter vector for the Q-function estimator. Let the control policy

$$u_i(t) = -\frac{1}{2}R_i^{-1}\hat{G}^{\mu\nu}(t)x(t), \tag{37}$$

be applied to the subsystems and let the decentralized event-triggering condition satisfies

$$\|e_i(t)\| \leq \left\|e_i^*(t)\right\|, \quad t \in [t_k, t_{k+1}), \forall k \in \mathbb{N} \tag{38}$$

where $R = diag(R_i)$, $e^* = [e_1^{*T} \ e_2^{*T} \ .. \ e_N^{*T}]^T = K^+\hat{\eta}(t)$. Let the Q-function parameters at each subsystem be updated using

$$\dot{\hat{\theta}} = -\alpha \frac{[\Delta\phi(\Delta T)]}{(1 + [\Delta\phi(\Delta T)]^T [\Delta\phi(\Delta T)])^2} E_{k+1}^T(t), \ \forall k = 0, 1, .. \tag{39}$$

Then, the state vector and the parameter estimation error converges asymptotically to zero with the event-triggering instants $k \to \infty$, provided the following inequalities hold: $\bar{\delta}_x > 2K_M^2$, $\frac{\alpha}{\rho} > \frac{1}{2}\|R\|^2$ where $\rho = (1 + [\Delta\phi(\Delta T)]^T[\Delta\phi(\Delta T)])^2$, $\|K^*\| \leq K_M$, $\bar{\delta}_x = [Q + \frac{1}{4}P^*BR^{-1}B^TP^* - \frac{1}{4\sigma^2}P^*DD^TP^*]$.

Proof: In order to complete this proof, Theorem 2 and the results of Lemma 4 are utilized to show that the subsystems states and the parameter estimation error at each subsystem converges to zero asymptotically.

In the next section, simulation results are provided to verify the theoretical claims.

6. SIMULATION RESULTS

Example 1: For the simulation results, first, consider the unstable linear batch reactor dynamics in continuous-time as

$$\dot{x} = \begin{bmatrix} 1.38 & -0.21 & 6.71 & -5.67 \\ -0.581 & -4.29 & 0 & 0.67 \\ 1.067 & 4.27 & -6.65 & 5.89 \\ 0.048 & 4.27 & 1.34 & -2.10 \end{bmatrix} x + \begin{bmatrix} 0 & 0 \\ 5.67 & 0 \\ 1.13 & -3.1 \\ 1.13 & 0 \end{bmatrix} u$$

with $x = [x_1 x_2 x_3 x_4]^T$ and $u = [u_1^T u_2^T]^T$. To verify the advantages of the proposed method, we compare the results of our approach with that of an LQR controller with the event-triggering condition of the form [20]. For the case of uncertain system dynamics, the proposed hybrid Q-learning based approach is compared with the traditional event-triggered Q-learning scheme [17, 23].

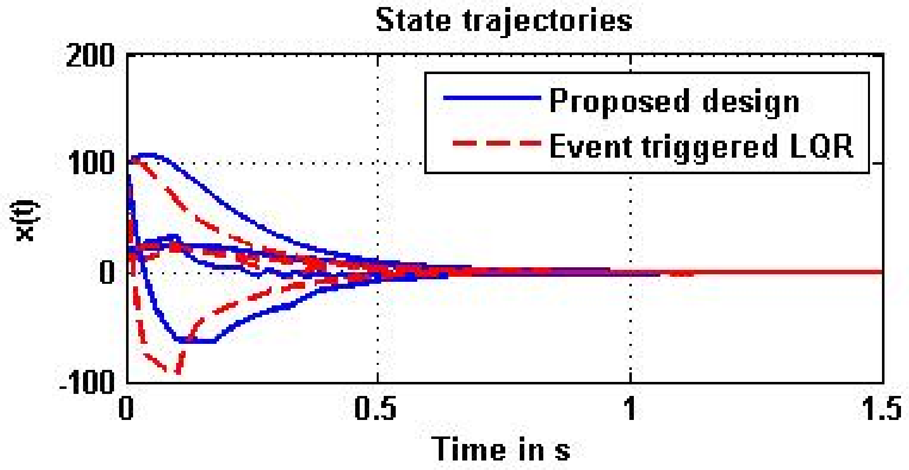


Fig. 5.3. Comparison of state trajectories.

Case 1: (Known system dynamics) The simulation analyses carried out with different deign parameters and the values of the penalizing matrices Q, R are taken as $10I, 0.2I$. The value of σ is taken as 0.7. These values are chosen as they resulted in comparable control effort and state trajectories.

Figs. 5.3 and 5.4 depict the convergence of the closed-loop system state vector, and control input. The proposed method is contrasted with the event-triggered implementation of LQR. The continuous time LQR is considered as a benchmark for the state and control trajectories as well as the cumulative cost.

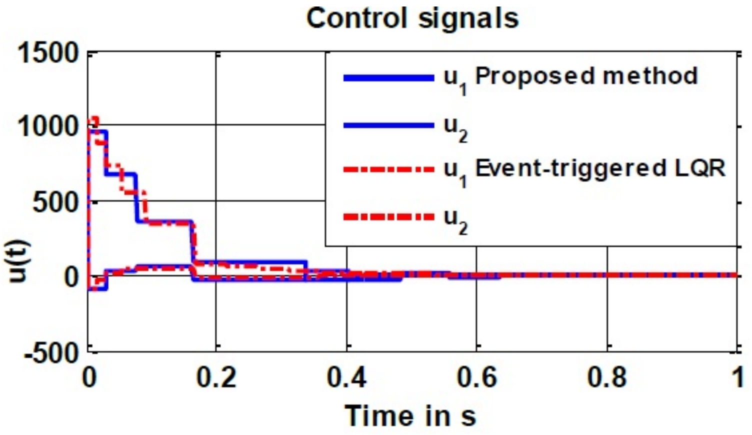


Fig. 5.4. Comparison of control signals.

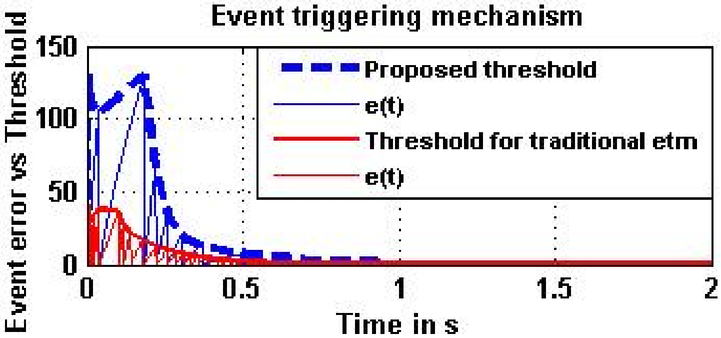


Fig. 5.5. Performance comparison of event-triggering mechanism (ETM).

The parameter σ is varied to demonstrate the advantages of the proposed adaptive optimal design with the traditional event-triggering approach and the results are summarized in Table 5.1 for the case when the system dynamics are uncertain. Due the space consideration, all the simulation figures are not included and the important results are summarized in Table 5.1, where ET is expanded as event-triggered, P.M is expanded as proposed method, IET is expanded as inter-event time.

Furthermore, the lower bound on the inter-event times is observed to be 0.001 s. It is clear from Fig. 5.5 that the event-triggering threshold with the proposed approach is considerably higher than the traditional approach. This in turn elongated the inter-event time thus reducing the resource utilization which is one of the primary objectives of the design. Fig. 5.6 shows the inter-event times. Due to the optimal choice of the control policy and the threshold for the event triggering error, the inter-event time is optimized and it is observed that this time can be designed to be larger with the proposed design for similar design parameter value.

Table 5.1. Analysis with event-triggering design parameter σ.

σ	Avg. IET in s		Cumulative cost		Number of events	
	ET LQR	P.M	ET LQR	P.M	ET LQR	P.M
0.55	0.0585	0.0857	9.96E+03	1.38E+04	127	87
0.7	0.0602	0.0906	1.12E+04	7.38E+03	123	82
0.75	0.0615	0.0934	1.06E+04	6.70E+03	121	79
0.85	0.0619	0.1	1.33E+04	6.09E+03	119	74
0.95	0.0589	0.1015	1.72E+04	5.99E+03	127	73

Example 2: (Load frequency control of three area power system) The states of a three-area power system model at each subsystem under consideration are - frequency change, incremental change in output power of the generator, change in governor valve position, incremental change in integral control, tie-line power deviation. For the detailed dynamics considered refer to Alrifai et al. 2011.

The controller design parameters are chosen as R$_i$= 0.1,$Q_i = 0.4$, $\sigma_i = 0.65$ and the parameter tuning rule with $\beta = 45$. All the system states are regulated using the proposed controller. The convergence of the state trajectories can be seen in Fig. 5.7.

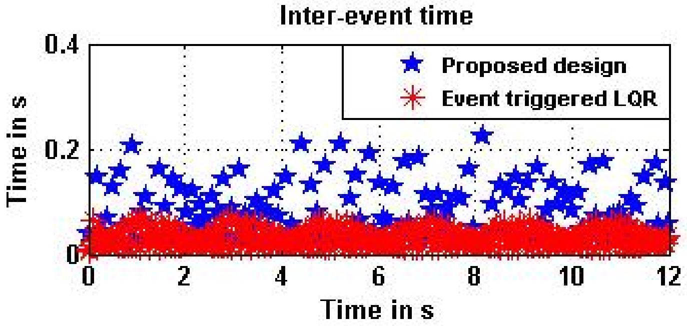


Fig. 5.6. Comparison of inter-event times.

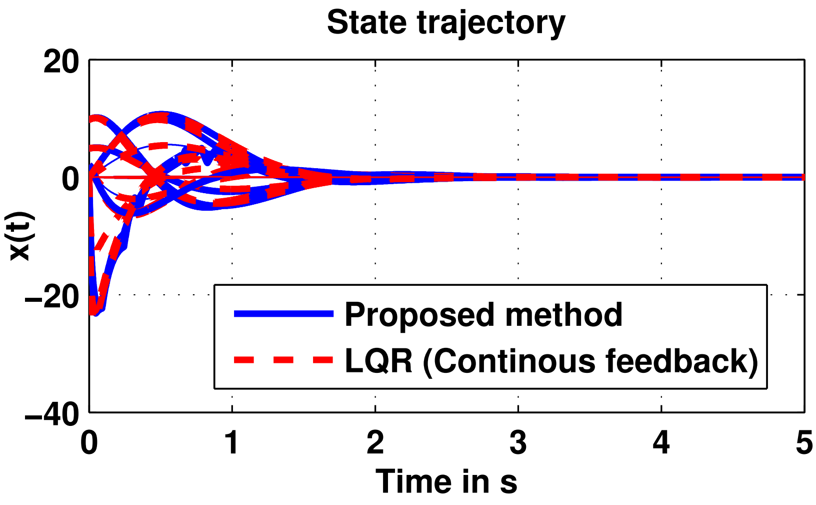


Fig. 5.7. Comparison of state trajectories.

Distributed controllers are designed at each subsystem and decentralized event-triggering mechanism is developed for every subsystem based on the proposed adaptive optimal design. For comparison, the traditional continuous feedback based LQR design is utilized. The parameter σ is varied to demonstrate the advantages of the proposed design with the traditional event-triggering approach and the results are summarized in Table 5.3.

It can be observed that the event-triggered implementation of the control policy and the continuous feedback LQR results in an asymptotically stable system. The piecewise continuous control input obtained with the proposed design and the continuous feedback control policy recorded in Fig. 5.8. The control effort for both the controllers is quite similar for all the three subsystems.

Table 5.2. Continuous time LQR controller.

Linear quadratic regulator - Continuous time	
Cumulative cost	Number of events
101.5169	5000

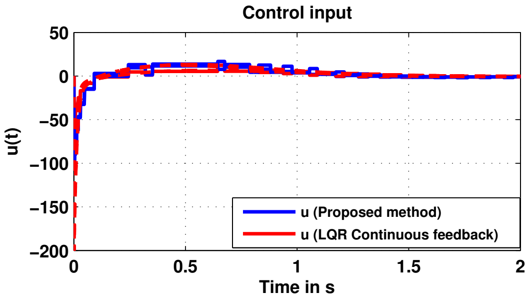


Fig. 5.8. Comparison of control inputs.

The parameter σ is varied and the resulting performance of the system in terms of cumulative cost, number of events, and the inter-event time are recorded in the Table 5.3, which clearly depicts the advantage of the proposed design. Also, comparing the cumulative cost with continuous control implementation in Table 5.2, it can be observed that for a comparative cost, the number events generated by the proposed method is 110 which is fractional and the control effort from Fig.5.8 demonstrates the advantage of the proposed method. The additional term in the cost function, which maximizes the event-triggering threshold provides an explicit relationship between event-triggering and system performance and hence, optimizes both of them.

Table 5.3. Analysis of decentralized optimal distributed control scheme.

σ	Avg. IET in s		Cumulative cost		Number of events	
	ET LQR	P.M	ET LQR	P.M	ET LQR	P.M
0.35	0.0576	0.0981	217.5966	151.4405	130	76
0.4	0.0512	0.0729	230.5441	144.3559	146	96
0.46	0.0567	0.0776	275.7723	126.436	128	95
0.52	0.0516	0.0591	234.7836	115.1525	143	121
0.6	0.0604	0.0699	394.1401	109.8525	124	104
0.65	0.0669	0.0647	861.1497	106.1883	110	115

7. CONCLUSIONS

This paper proposes a novel approach for simultaneously optimizing both the event-triggering sampling instants and state feedback controller using zero-sum game formulation. The proposed design scheme provides a tractable trade-off between the frequency of events and the system performance cost by utilizing the min-max optimization of the cost function. The inter-event time interval increases with the proposed event-triggering condition which considerably reduces the communication cost when compared to the traditional event-triggering schemes. The model-free Q-learning scheme generates the optimal control policy and the event-triggering condition even when the system dynamics are uncertain. Finally, the decentralized event-triggering condition enables distributed control of interconnected systems confirming the generic nature of the proposed design.

REFERENCES

[1] K. J. Astrom and B. M. Bernhardsson. Comparison of Riemann and Lebesgue sampling for first order stochastic systems. In *Proceedings of the 41st IEEE Conference on Decision and Control, 2002.*, volume 2, pp. 2011–2016, Dec 2002.

[2] Tamer Basar and Pierre Bernhard. H-infinity optimal control and related minimax design problems. *Birkhaüser, Boston, Massachusetts*, 1991.

[3] S. J. Bradtke, B. E. Ydstie, and A. G. Barto. Adaptive linear quadratic control using policy iteration. In *American Control Conference, 1994*, volume 3, pp. 3475–3479, June 1994.

[4] Claudio De Persis, Rudolf Sailer, and Fabian Wirth. Parsimonious event-triggered distributed control: A zeno free approach. *Automatica*, volume 49(7), pp. 2116–2124, 2013.

[5] R. Dorf, M. Farren, and C. Phillips. Adaptive sampling frequency for sampled-data control systems. *IRE Transactions on Automatic Control*, volume 7(1), pp. 38–47, Jan 1962.

[6] E. Garcia and P. J. Antsaklis. Model-based event-triggered control for systems with quantization and time-varying network delays. *IEEE Transactions on Automatic Control*, volume 58(2), pp.422–434, Feb 2013.

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

[8] WPMH Heemels and MCF Donkers. Model-based periodic event-triggered control for linear systems. *Automatica*, volume 49(3), pp. 698–711, 2013.

[9] Toivo Henningsson, Erik Johannesson, and Anton Cervin. Sporadic event-based control of first-order linear stochastic systems. *Automatica*, volume 44(11), pp. 2890–2895, 2008.

[10] Frank L Lewis and Vassilis L Syrmos. *Optimal control*. John Wiley & Sons, 1995.

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

[12] Jan Lunze and Daniel Lehmann. A state-feedback approach to event-based control. *Automatica*, volume 46(1), pp. 211–215, 2010.

[13] M. Mazo and P. Tabuada. Decentralized event-triggered control over wireless sensor/actuator networks. *IEEE Transactions on Automatic Control*, volume 56(10), pp. 2456–2461, Oct 2011.

[14] A. Molin and S. Hirche. On the optimality of certainty equivalence for event-triggered control systems. *IEEE Transactions on Automatic Control*, volume 58(2), pp. 470–474, Feb 2013.

[15] Adam Molin. Optimal event-triggered control with communication constraints. *Technische Universität München, München*, 2014.

[16] V. Narayanan and S. Jagannathan. Distributed adaptive optimal regulation of uncertain large-scale interconnected systems using hybrid q-learning approach. *IET Control Theory Applications*, volume 10(12), pp. 1448–1457, 2016.

[17] A. Sahoo and S. Jagannathan. Event-triggered optimal regulation of uncertain linear discrete-time systems by using q-learning scheme. In *53rd IEEE Conference on Decision and Control*, pp. 1233–1238, Dec 2014.

[18] Dawei Shi, Tongwen Chen, and Ling Shi. Event-triggered maximum likelihood state estimation. *Automatica*, volume 50(1), pp. 247–254, 2014.

[19] Richard S Sutton and Andrew G Barto. *Reinforcement learning: An introduction*, volume 1. MIT press Cambridge, 1998.

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

[21] P. Tallapragada and N. Chopra. On event triggered tracking for nonlinear systems. *IEEE Transactions on Automatic Control*, volume 58(9), pp. 2343–2348, Sept 2013.

[22] D. ToliÄĞ, R. Fierro, and S. Ferrari. Optimal self-triggering for nonlinear systems via approximate dynamic programming. In *2012 IEEE International Conference on Control Applications*, pp. 879–884, Oct 2012.

[23] K. G. Vamvoudakis and H. Ferraz. Event-triggered h-infinity control for unknown continuous-time linear systems using q-learning. In *2016 IEEE 55th Conference on Decision and Control (CDC)*, pp. 1376–1381, Dec 2016.

[24] X. Wang and M. D. Lemmon. Event-triggering in distributed networked control systems. *IEEE Transactions on Automatic Control*, volume 56(3), pp. 586–601, March 2011.

APPENDIX

Proof of Lemma 2

Choose the positive definite Lyapunov function $L(x) = \frac{1}{2}V^*(x)$. The time derivative along the system dynamics can be obtained as

$$\dot{L}(x) = \dot{L}_x(t) = V_x^{*^T}\dot{x} = [V_x^{*^T}Ax + V_x^{*^T}Bu + V_x^{*^T}D\eta] \tag{40}$$

Add and subtract $V_x^{*^T}Bu^* + V_x^{*^T}D\eta^*$ to get

$$\begin{aligned}&\dot{L}_x(t) = [V_x^{*^T}Ax + V_x^{*^T}Bu^* + V_x^{*^T}D\eta^*]\\&+V_x^{*^T}B(u - u^*) + V_x^{*^T}D(\eta - \eta^*)\end{aligned} \tag{41}$$

From (9), with (u^*, η^*, V^*), we have $H(x, u^*, \eta^*) = 0$ and

$$-x^TQx - u^{*T}Ru^* + \sigma^2\eta^{*T}\eta^* = V_x^{*T}[Ax + Bu^* + D\eta^*]. \tag{42}$$

Using (42) in (41) results in

$$\begin{aligned}&\dot{L}_x(t) = -x^TQx - u^{*T}Ru^* + \sigma^2\eta^{*T}\eta^*\\&+V_x^{*^T}B(u - u^*) + V_x^{*^T}D(\eta - \eta^*)\end{aligned}\;. \tag{43}$$

Using the definition, $\eta^* = \frac{1}{2\sigma^2}D^TP^*x$, and substituting (12) in (43), we get $\dot{L}_x(t) = -Q(x) - u^{*T}Ru^* - \sigma^2\eta^{*T}\eta^* + 2\sigma^2\eta^{*T}\eta$. By using the definition of optimal policy (10), and (11), we have

$$\dot{L}_x(t) = -x^T\delta_xx + x^TP^*D\eta \tag{44}$$

where $\delta_x = [Q + \frac{1}{4}P^* B R^{-1} B^T P^* + \frac{1}{4\sigma^2} P^* D D^T P^*]$. Applying norm operator to (44) reveals

$$\dot{L}_x(t) \leq -\delta_{x,m} \|x\|^2 + \|P^* D K\| \, \|x\| \, \|e\| \tag{45}$$

Proof of Theorem 1:

With the Lyapunov function candidate chosen similar to that in Lemma 2, from (44), we have

$$\dot{L}_x(t) = -x^T \delta_x x + x^T P^* D \eta. \tag{46}$$

Applying norm operator and using the event-triggering condition (13)

$$\dot{L}_x(t) \leq -\delta_{x,m} \|x\|^2 + \|P^* D\| \, \|x\| \, \|\eta^*\| \, . \tag{47}$$

Using the definition of η^* from (11), we get

$$\dot{L}_x(t) \leq -(\delta_{x,m} - \frac{1}{2\sigma^2} \|P^* D\|^2) \|x\|^2. \tag{48}$$

Now to derive the positive inter-event time, use (3) and taking the time-derivate reveals

$$\dot{e}(t) = \dot{\hat{x}}(t) - \dot{x}(t), \quad t \in [t_k, t_{k+1}). \tag{49}$$

Taking the norm operator and substituting the system dynamics reveals

$$\|\dot{e}(t)\| = \|\dot{x}(t)\| = \|Ax + Bu^* + D\eta\| \, . \tag{50}$$

From (48), since the states are asymptotically converging to zero, there exists a positive constant $X_m > 0$ such that $\|A + BK^*\| \, \|x\| \leq X_m$. Using this relation in (50) yields

$$\|\dot{e}(t)\| \leq \|DK^*\| \, \|e\| + X_m. \tag{51}$$

Integrating using the comparison lemma [21], reveals

$$\|e(t)\| \leq \frac{X_m}{\|DK^*\|} (e^{\|DK^*\|(t-t_k)} - 1), \quad t \geq t_k. \tag{52}$$

Substitute $t = t_{k+1}$ to obtain the minimum positive inter-event time

$$(t_{k+1} - t_k)_{\min} \geq \frac{1}{\|DK^*\|} \log(\frac{\|DK^*\|}{X_m} \|e^*\| + 1). \tag{53}$$

Proof of Theorem 2:

Consider the Lyapunov candidate

$$L(x, \theta) = L_x(t) + L_\theta(t) \tag{54}$$

where $L_x(t) = \frac{1}{2}x^T P^* x$. The derivative of the Lyapunov candidate is given by

$$\dot{L}(x, \tilde{\theta}) = \dot{L}_x(t) + \dot{L}_\theta(t). \tag{55}$$

Consider the first term and using the system dynamics to get

$$\dot{L}_x(t) = V_x^{*^T}\dot{x} = [V_x^{*^T}Ax + V_x^{*^T}Bu + V_x^{*^T}D\eta]. \tag{56}$$

Add and subtract $V_x^{*^T}Bu^* + V_x^{*^T}D\eta^*$ to get

$$\begin{aligned} &\dot{L}_x(t) = [V_x^{*^T}Ax + V_x^{*^T}Bu^* + V_x^{*^T}D\eta^*] \\ &+V_x^{*^T}B(u - u^*) + V_x^{*^T}D(\eta - \eta^*) \end{aligned}. \tag{57}$$

Using the Hamiltonian (GARE), $H(x, u^*, \eta^*)$, we have

$$-Q(x) - u^{*T}Ru^* + \sigma^2\eta^{*T}\eta^* = V_x^{*T}[Ax + Bu^* + D\eta^*]. \tag{58}$$

Substituting (58) in (57)

$$\begin{aligned} &\dot{L}_x = -Q(x) - u^{*T}Ru^* + \sigma^2\eta^{*T}\eta^* + V_x^{*^T}Bu(\breve{x}) \\ &-V_x^{*^T}Bu^* - V_x^{*^T}D\eta^* \end{aligned}. \tag{59}$$

Adding and subtracting $V_x^{*^T}Bu^*(\breve{x})$ in (59), we get

$$\begin{aligned} &\dot{L}_x(t) = -Q(x) - u^{*T}Ru^* + \sigma^2\eta^{*T}\eta^* - V_x^{*^T}D\eta^* \\ &+V_x^{*^T}B[u(\breve{x}) - u^*(\breve{x})] + V_x^{*^T}B[u^*(\breve{x}) - u^*(x)] \end{aligned}. \tag{60}$$

Now, define $\tilde{u} = u^*(\breve{x}) - u(\breve{x})$, and using the definition (32) and (10), we have $\tilde{u} = \frac{1}{2}R^{-1}\tilde{G}^{\mu\nu}\breve{x}$. Substitute $\tilde{u}$ in (59) to get

$$\begin{aligned} &\dot{L}_x(t) = -Q(x) - u^{*T}Ru^* + \sigma^2\eta^{*T}\eta^* - V_x^{*^T}D\eta^* \\ &+2u^{*T}R\tilde{G}^{\mu\nu}\breve{x} + V_x^{*^T}B[u^*(\breve{x}) - u^*(x)] \end{aligned}$$

Using the definition of η^*, we get the Lyapunov time-derivate

$$\dot{L}_x(t) = -Q(x) - u^{*T}Ru^* + \sigma^2\eta^{*T}\eta^* - 2u^{*T}R\tilde{G}^{\mu\nu}\breve{x} \tag{61}$$

Applying the norm operator (61) can be bounded as

$$\dot{L}_x(t) \leq -\bar{\delta}_x\|x\|^2 + 2\left\|u^{*T}R\right\|\left\|\tilde{G}^{\mu\nu}\breve{x}\right\| \tag{62}$$

where $\bar{\delta}_x = [Q + \frac{1}{4}P^*BR^{-1}B^TP^* - \frac{1}{4\sigma^2}P^*DD^TP^*]$. Using the YoungâĂŹs inequality [16], we get

$$\dot{L}_x(t) \leq -\bar{\delta}_x\|x\|^2 + 2\left\|u^{*T}\right\|^2 + \frac{1}{2}\|R\|^2\left\|\tilde{G}^{\mu\nu}\breve{x}\right\|^2.$$

Finally, application of the norm operator and using the fact that $\left\|\tilde{G}^{\bullet\nu}\breve{x}\right\| \leq \left\|\tilde{\theta}\Delta\phi(\Delta T)\right\|$, reveals

$$\dot{L}_x(t) \leq -(\bar{\delta}_x - 2K_M^2)\|x\|^2 + \frac{1}{2}\|R\|^2\left\|\tilde{\theta}\Delta\phi(\Delta T)\right\|^2. \tag{63}$$

Now, using (34), the estimation error dynamics is revealed as $\dot{\tilde{\theta}}(t) = -\dot{\hat{\theta}}(t)$ and

$$\dot{\tilde{\theta}} = \alpha\frac{[\Delta\phi(\Delta T)]}{1 + ([\Delta\phi(\Delta T)]^T[\Delta\phi(\Delta T)])^2}E_{k+1}^T(t). \tag{64}$$

Let $L_\theta(t) = \frac{1}{2}\tilde{\theta}^T\tilde{\theta}$, using (64) and (31), the Lyapunov time derivative is obtained as

$$\dot{L}_\theta(t) = -\alpha\frac{\tilde{\theta}^T\Delta\phi(\Delta T)[\Delta\phi(\Delta T)]^T\tilde{\theta}}{(1 + [\Delta\phi(\Delta T)]^T[\Delta\phi(\Delta T)])^2}. \tag{65}$$

Using (63) and (65) in (55), we get

$$\dot{L}(t) \leq -(\bar{\delta}_x - K_M^2)\|x\|^2 - (\frac{\alpha}{\rho} - \frac{1}{2}\|R\|^2)\left\|\tilde{\theta}\Delta\phi(\Delta T)\right\|^2 \tag{66}$$

where $\rho = (1 + [\Delta\phi(\Delta T)]^T[\Delta\phi(\Delta T)])^2$.

Proof of Lemma 3:

The optimal control input for the overall system is stabilizing. Therefore, the closed-loop system matrix $(A - BK^*)$ is Hurwitz. The Lyapunov equation is given by $(A - BK^*)^T\bar{P} + \bar{P}(A - BK^*) = -\bar{Q}$, has a positive definite solution $\bar{Q}$. The matrix $\bar{Q}$ can be chosen diagonal. Consider the Lyapunov function candidate $L(t) = x^T(t)\bar{P}x(t)$, with $\bar{P}$ being a positive definite matrix of appropriate dimension. The first derivative, along the overall system dynamics (36) can be expressed as

$$\begin{aligned}\dot{L}(t) &= \dot{x}^T(t)\bar{P}x(t) + x^T(t)\bar{P}\dot{x}(t)\\ &= x^T(t)[(A - BK)^T\bar{P} + \bar{P}(A - BK)]x(t) = -x^T(t)\bar{Q}x(t).\end{aligned}$$

Since, $\bar{Q}$ is a diagonal matrix, the first difference in terms of the subsystem state vector can be expressed as

$$\dot{L}(t) = -\sum\nolimits_{i=1}^{N} x_i^T(t)\bar{Q}_i x_i(t) \leq -\sum\nolimits_{i=1}^{N} \bar{q}_{\min}\|x_i(t)\|^2 < 0$$

where $\bar{q}_{\min}$ is the minimum singular value of $\bar{Q}_i$. This implies the subsystem (35) is asymptotic stable.

V. APPROXIMATE OPTIMAL EVENT-TRIGGERED CONTROL OF NONLINEAR SYSTEMS

ABSTRACT

In this paper, a novel approach is proposed for a nonlinear dynamical system using zero-sum game theory to optimize simultaneously both the event-triggering sampling instants and state feedback control policy. In the proposed scheme, the nonlinear control policy and the event-triggering sampling errors are considered as two non-cooperative players and a min-max optimization is devised to determine the optimal control policy and an event-triggering condition such that a balance between the frequency of feedback and system performance is achieved. First, a solution to this optimization problem is developed by assuming the system dynamics of the nonlinear system are known. Subsequently, an artificial neural network (NN) is employed to learn an approximate solution to the Hamilton-Jacobi-Isaacs (HJI) equation, in a forward-in-time and online manner, using a hybrid learning scheme. Next, NN identifiers are introduced to relax the requirement of the complete nonlinear dynamics and a model free approximate optimal event-triggered control scheme is proposed. Finally, the proposed approach is extended to the distributed approximate optimal control of nonlinear interconnected systems. The local ultimate boundedness of the resulting closed-loop nonlinear system is demonstrated. By using a numerical example, the performance of the near optimal design is evaluated through simulation studies.

1. INTRODUCTION

Traditional feedback controllers use instantaneous sensor measurements from the system as feedback signals to update the control input. With the advent of networked control systems (NCS), the feedback-loop in the modern control systems is closed via a communication network. The traditional feedback approach with a fixed sensor sampling rate is found to be expensive for the NCS due to communication overhead. Event-based sampling [1]-[3] and control, on the other hand, is increasingly gaining prominence among control researchers because of its computational and communication resource saving capability. In an event sampled framework, the sensor measurements are sampled based on certain state dependent criteria referred to as event-triggering condition. The controller is executed only at these aperiodic sampling instants. The event-triggering condition, in general, is designed by taking into account the stability, and, hence, proven to be advantageous [3] over its periodic counterpart.

For an event-triggered control, the system is required to be input-to-state stable (ISS) with respect to the measurement error. The event sampling instants are designed to reduce the frequency of feedback instants while guaranteeing the system stability. Here, the inter-event time intervals need to be lower bounded by non-zero positive constant to avoid accumulation point and zeno-behavior [3]. In line with these requirements, in the literature, two approaches are proposed for event-triggered control.

In the first approach, the sensor measurements and the control input are held between two consecutive events at the controller and actuator by using a zero order hold (ZOH), respectively. In contrast, the second approach uses a model of the system at the controller to provide the feedback information between event sampling instants [4]. A comprehensive survey on different event-triggered control approaches and their benefits are presented in [5]-[7]. It should be noted that the majority of the event-triggered techniques [1]-[4] are designed for stabilization without any performance criterion under the assumption that the system dynamics are known.

Optimal control [8], on the other hand, not only stabilizes the system but also optimizes the performance based on a performance function. Optimal control of nonlinear dynamic systems in continuous-time is a challenging problem due to the difficulty involved in obtaining a closed-form solution or value function to the Hamilton-Jacobi-Bellman equation. Adaptive dynamic programming (ADP) techniques [9]-[16], are used to solve the optimal control of such nonlinear systems online by finding an approximated value function.

Among the earlier works on ADP-based optimal control [12]-[14], the reinforcement learning technique using dynamic programming is combined with the adaptive control theory and a neural network (NN) based framework, to generate an online yet approximate solution to the optimal control without needing the knowledge of system dynamics. Later, online policy iteration schemes [15] are introduced to obtain the solution of HJB equation and attain optimality. In addition, an alternate single NN-based ADP approach is presented in [11] for an affine nonlinear continuous-time system without using an iterative technique. The NN weights are tuned online and periodically to achieve near optimality.

Recently, event-triggered optimal controllers are developed for a nonlinear system using NN based online approximators in [9],[10] wherein the event-triggering instants are designed to maintain system stability alone. In contrast, the authors in [17]-[18] proposed an optimal event-triggering mechanism by formulating a cost function that penalizes the number of events for a linear system with known dynamics.

In summary, the optimal event-triggering instants and controller design are, in general, considered as mutually exclusive problems. To the best knowledge of the authors, a simultaneous optimal co-design of the controller and the event-triggering sampling mechanism is not attempted in the literature for nonlinear systems with known and uncertain system dynamics. The benefit of such a control scheme lies in the fact that the control inputs to the system explicitly take into account the effects of aperiodic event sampled feedback and hence, the system performance is optimized. Moreover, the design parameters can be chosen directly to meet a predefined performance measure.

Motivated by the above facts, in this paper, a novel approximate optimal event-triggered control design scheme for nonlinear continuous-time systems is presented. First, the design of both control policy and the event-triggering mechanism is formulated as a two-player zero-sum game with known system dynamics. Here, a novel cost function is introduced as a function of state vector, control policy and the measurement/event-triggering error. The control policy and the measurement error due to event-triggered feedback will be considered as two non-cooperative players. The saddle point solution to this min-max problem results in the minimization of the control policy while maximizing the measurement error.

The resulting measurement error from this min max optimization problem is utilized as the dynamic threshold in an event-trigger condition to determine the sampling instants. Since the control policy explicitly accounts for the worst-case event-triggering error, the stability and the performance of the system is preserved. Moreover, since the inter-event time is directly proportional to the event-triggering error and utilizing the maximum event trigger error as a dynamic threshold results in optimizing the inter-event time. This net result is an optimal event-triggered controller which explicitly takes into account both generation of event-triggered sampling instants and control policy.

Next, an approximate solution to the simultaneous optimization of event sampling instants and control co-design problem is proposed when the system internal dynamics are considered uncertain. Here an artificial neural network (NN) is employed to learn the approximate optimal value function and to determine the optimal control policy while maximizing the event-triggering intervals in a forward-in-time manner by using a hybrid learning scheme [10]. Next, NN identifiers are introduced to relax the requirement of accurate knowledge of the internal dynamics and the input gain function to obtain the saddle point solution for the min-max optimization problem online. The Lyapunov stability analysis is used to guarantee local ultimate boundedness of the state vector and the NN weight estimation errors. Finally, the proposed approximate optimal event-triggered control scheme is extended for distributed control of interconnected system.

The contributions of the paper include: 1) a novel optimal event sampling instant and controller co-design using zero-sum game formulation for affine nonlinear systems; 2) development of an online NN learning scheme for generating optimal control and event-triggering policies when the system dynamics are uncertain; 3) extension of the approximate optimal event-triggered design to the distributed control of interconnected systems; 4) derivation of inter-event time or event triggered sampling instants for the cases of known and uncertain system dynamics; 5) Lyapunov stability analysis and verification of the proposed design using numerical examples via simulation.

The paper is organized as follows. In Section II, the system nonlinear system dynamics are introduced and the problem statement is presented. In Section III, the main results are presented for the case when the system dynamics are known. In Section IV, a NN based hybrid learning approach is proposed to solve the optimization problem forward-in-time, online, when the system internal dynamics are uncertain and the NN identifier based design is introduced to relax the requirement of both the internal dynamics and input gain function. In Section V, the extension of this approach to the distributed approximate optimal control of interconnected system is presented. Finally, simulation results are provided to show the effectiveness of the controller designed in Section VI. Conclusions follow in Section VII.

In this paper, $\Re$ denotes the set of all real numbers; denotes the set of all natural numbers. Euclidean norm is used for vectors and Frobenius norm [27]-[28] is used for matrices.The next section presents a brief background on the system dynamics and the problem statement.

2. BACKGROUND AND PROBLEM STATEMENT

2.1 System Description

Consider the nonlinear dynamical system represented by

$$\dot{x}(t) = f(x) + g(x)u(t), x(0) = x_0 \tag{1}$$

where $x \in \Omega \subset \Re^n$ is the state vector of the system, Ω is a compact set in the n-dimensional Euclidean space; $u(t)$ is the control input, $f : \Omega \to \Re^n$, and$g : \Omega \to \Re^{n\times m}$ are nonlinear maps representing internal dynamics and input gain function. The function f satisfies $f(0) = 0$ and the function $\|g(x)\| = 0$ if and only if $x = 0$. The control input for (1) is of the form

$$u(t) = \mu(x(t)) \tag{2}$$

where $\mu : \Omega_u \to \Re^m$ is a nonlinear map satisfying $\mu(0) = 0$ and Ω_u is a compact subset of Ω.

In the traditional periodic/continuous feedback framework, the control policy, μ, is continuously implemented using the current feedback signal $x(t)$. In contrast, in the event-triggered control framework $x(t)$ is available only at certain aperiodic event-based sampling instants. These time instants can be represented using the sequence $\{t_k\}_{k\in\{0,\mathbb{N}\}} \subseteq t$, such that $0 = t_0 < t_1 < \ldots$ The control policy will be held at the actuator using a zero-order hold circuit and satisfies $u(t) = \mu(\breve{x}(t))$ wherein $\breve{x}(t) = x(t_k), \ \forall t \in [t_k, t_{k+1})$. Hence, the control signals are piecewise continuous.

The discrete aperiodic sampling instants can be determined dynamically by using an event-triggering mechanism. Note that due to the difference between $x(t)$ and $\breve{x}(t)$, there will be an event-triggering error (or measurement error), which is defined as

$$e(t) = \breve{x}(t) - x(t), \quad \forall t \in [t_k, t_{k+1}). \tag{3}$$

At the sampling instants, the error (3) is reset to zero or $e(t_k) = 0$. Define the difference between continuously updated control (2) and the event sampled control policy as

$$\eta(t) = u(\breve{x}(t)) - u(x(t)). \tag{4}$$

In the rest of the paper η in (4) is referred as control sampling error policy.

Assumption 1: The computational delay is considered negligible and the sensors are assumed to be noise free.

Remark 1: For a linear system, the control policy is represented as $u(t) = Dx(t)$, where D is a control gain matrix. Due to the linearity, (4) will be a represented as $\eta(t) = De(t)$ for a linear system. However, because of the nonlinear policy (2), a linear relationship between $\eta(t)$ and the event-triggering error does not exist. Therefore, for simplicity, the difference in the event-sampled control and continuous control policy, $\eta(t)$, is defined as control sampling error policy.

Next, the problem of designing an event-triggering condition to determine $\{t_k\}$ and the optimal control policy (2) co-design are defined.

2.2 Problem Statement

Consider the nonlinear dynamical system given by (1). Let the control policy (2) be implemented with event-triggered feedback. Then, the system dynamics (1) can be re-written as

$$\dot{x}(t) = f(x) + g(x)u(\breve{x}). \tag{5}$$

Define the performance measure for the system (1) to be

$$\|\zeta(t)\|^2 = Q(x(t)) + u^T(t)Ru(t) \tag{6}$$

where Q is a positive definite function satisfying $Q(0) = 0$ and R is a positive definite matrix. The functions $Q,\ R$ penalize the states and the control policy, respectively.

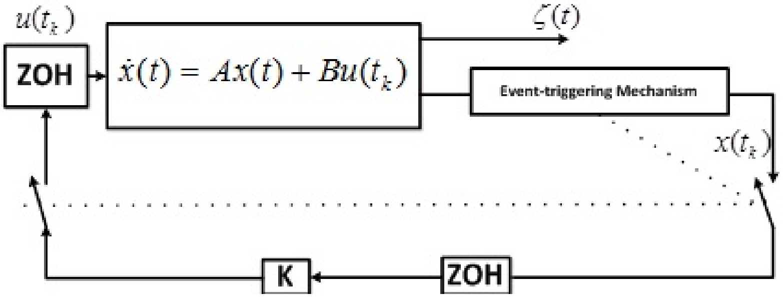


Fig. 6.1. Networked control system and event-triggered feedback.

A block diagram representation of the control architecture for event-triggered implementation of state feedback control is given in Fig. 6.1. The event-triggering mechanism monitors the sensor measurements and dynamically determines the time-instants $\{t_k\}$ to close the feedback loop. Here, the control policy in (2) are generated by solving a minimization problem associated with the performance measure (5) and the event-execution rule to determine the sequence $\{t_k\}$ is obtained by designing an upper bound for the measurement error (3) based on the stability of the controlled system.

In this paper, the objective is to develop an optimal control policy which minimizes (6) while simultaneously maximizing inter event sampling interval and meeting the performance given by (6).

Remark 2: To determine the event-triggering instants, the state vector is sampled as a function of the event-triggering error by using the stability criterion [3]. For example, consider the system (5); an ISS Lyapunov function, $L(x)$ is chosen such that its time derivative is represented as $\dot{L} = -\bar{\alpha}(\|x\|)+\bar{\gamma}(\|e\|)$, where $\bar{\alpha}, \bar{\gamma}$ are positive definite functions. Using the Lyapunov time-derivative, the event-triggering condition is chosen as $\|e\| \leq \sigma\bar{\gamma}^{-1}(\bar{\alpha}(\|x\|))$, for some positive constant σ. The bound on the event-triggering error $\sigma\bar{\gamma}^{-1}(\bar{\alpha}(\|x\|))$ ensures that the negative term in the Lyapunov time derivative, $\dot{L}(t)$ is dominant and hence, ensures stable operation of the system. However, this does not provide any information about the system performance during the inter-event period.

In the next section, a zero-sum game based event-triggered control scheme is presented which satisfies the objectives defined in this section. The resulting event-triggering mechanism increases the time between successive events while the optimal control policy ensures satisfactory system performance.

3. PROPOSED METHODOLOGY

In this section, the control and the event trigger sampling interval policies will be considered as two non-cooperative players applied to the system. A cost function is defined as a function of system state vector, control input vector and the sampling error policy. It will be demonstrated that the objectives listed in Section 2 will be achieved by determining a saddle point solution to the optimization problem associated with the cost function subject to the dynamic constraint (5).

The maximizing solution to the optimization problem will act as the threshold to generate events while the optimal minimizing control policy will be applied to the system with the feedback generated at these event-triggered sampling instants. Existence of such saddle point solution to the min-max optimization problem depends on certain properties of the system which are discussed in Remark 3 [8],[26].

Utilizing the system dynamics (5) and the definition of the sampling error policy (4), we can rewrite the system dynamics (5) by adding and subtracting $u(x(t))$ as

$$\dot{x}(t) = f(x) + g(x)u(x(t)) + h(x)\eta(t) \tag{7}$$

where $h(x) = g(x)$ and $\eta(t)$ is defined as in (4). Now define the infinite horizon cost function using the performance measure (6) as

$$J(x, \eta, u) = \int_t^{\infty} [\|\zeta(t)\|^2 - \sigma^2 \eta^T \eta] d\tau. \tag{8}$$

where $\sigma > 0$ represents the attenuation constant. The objective is to find an optimal saddle-point solution (u^*, η^*) so that the optimal value function satisfies

$$V^*(x(t)) = \min_u \max_\eta J(u, \eta) = \max_\eta \min_u J(u, \eta). \tag{9}$$

Using the infinitesimal version of the cost function (8) and the system dynamics (7), the Hamiltonian function can be defined as

$$H(x,u,\eta) = Q(x) + u^T Ru - \sigma^2 \eta^T \eta + V_x^T [f + gu(t) + h\eta(t)] \tag{10}$$

where $V_x = \partial V/\partial x$ with $V(x)$ being the value-function defined using the integral expression in (8). The optimal policies are obtained as [25]-[26]

$$u(x, V_x^*) = -\frac{1}{2} R^{-1} g^T(x) V_x^* \tag{11}$$

$$\eta(x, V_x^*) = \frac{1}{2\sigma^2} h^T(x) V_x^* = \eta^*(x, V_x^*) \tag{12}$$

where V_x^* is the gradient of the optimal value function along the state trajectory. Substituting optimal policies in the Hamiltonian will result in the continuous-time Hamilton-Jacobi-Isaacs (HJI) equation

$$H = Q(x) + V_x^{*T} f - \frac{1}{4} V_x^{*T} g R^{-1} g^T V_x^* + \frac{1}{4\sigma^2} V_x^{*T} h h^T V_x^*. \tag{13}$$

Assumption 2: The control policies are Lipschitz continuous over compact sets and satisfy $\left\| u(\breve{x}) - u(x) \right\| \leq L_u \left\| e \right\|$ where $L_u > 0$ being the Lipschitz constant [27].

Remark 3: Consider the infinite horizon cost function (8) and the nonlinear system dynamics (7). Let the system be reachable and zero-state observable with $Q(x) = C(x)^T C(x)$, with a nonlinear map C. Then, there exists a minimum positive definite solution for the Hamilton Jacobi Isaacs (HJI) equation ([26],[8]) when $\sigma > \sigma^*$, where σ^* is the H_∞ gain of the system.

Remark 4: Note that if there is a positive definite solution to the HJI equation, then the optimal cost function is finite and the control policy asymptotically stabilizes the system. Further, the optimal policies (11) and (12) are functions of the optimal value function V^*. To determine the optimal value function, solution to the HJI equation is required which is non-trivial. Therefore, function approximators are utilized to generate an estimated optimal value function which is utilized in the optimal policies.

Remark 5: If the dynamics f, g are linear maps represented by A, B, respectively, the HJI equation becomes the game algebraic Riccati equation (GARE) [8]. The optimal value function, $V^*(x)$, for the GARE exists if the system is controllable and the $(A, \sqrt{Q})$ is observable. The optimal value function then is given by $V^*(x) = x^T(t)P^*x(t)$, where P^* is the positive definite solution to GARE.

Lemma 1: Consider the infinite horizon cost function (8) and the nonlinear input affine system dynamics (7). Let be the positive definite solution for the HJI equation (13), then the optimal policy

$$u(x) = -\frac{1}{2}R^{-1}g^T(x)V_x^*(t) \tag{14}$$

generates a local ISS Lyapunov function for (7) with respect to the measurement error $e(t)$.

Proof: See Appendix.

Remark 6: The smooth function $L(x) = V^*(x)$ satisfies, $\bar{\alpha}(V^*(x)) \leq V^*(x) \leq \bar{\bar{\alpha}}(V^*(x))$, where $\bar{\bar{\alpha}}$ and $\bar{\alpha}$ represent the positive definite functions. Further, from the proof of Lemma 1, we have $\|x\| \geq \gamma \|e\|$ implies $\dot{L}_x < 0$, where $\gamma = \|L_v L_u h\| / \left\|\delta_{x,m}\right\|$, $\delta_{x,m}, L_v$ are positive constants defined in the proof of Lemma 1 using Q and R. Thus, it can be concluded that $L(x)$ is a local ISS Lyapunov function [27].

Next, the main results of this section are presented.

Theorem 1 (Case of Known System Dynamics): Consider the infinite horizon cost function (8) and the affine nonlinear system dynamics (7). Let be the positive definite solution for the HJI equation (13), and the optimal policy given by (14) be applied to the system with the following event-triggering condition given by

$$\|e(t)\| \leq \frac{\|\eta^*(t)\|}{L_u}, \quad t \in [t_k, t_{k+1}), \forall k \in \{0, \mathbb{N}\}. \tag{15}$$

Then, the closed-loop system is asymptotically stable when Q, R, σ are selected such that $\delta_{x,m} > 0$, where $\delta_{x,m}$ is a function of Q, R, σ. In addition, a positive minimum inter-event time, τ_m, exists such that

$$\tau_m \geq \frac{1}{\|hL_u\|}\log(\frac{\|hL_u\|}{X_m}\|e^*\| + 1) \tag{16}$$

where X_m is a positive constant defined in the proof.

Proof: See Appendix.

Remark 7: The proposed event-trigger condition (15) allows the measurement error to increase until the system performance defined by (6) is not deteriorated. This increases the inter-event time (Proof of Theorem 1).

Remark 8: The expression for the inter-event sampling interval obtained in the proof of Theorem 1 can be utilized to generate events automatically without using the event-triggering mechanism. Such a scheme is known as self-triggering scheme and it obviates the computation required by the event-triggering mechanism to determine t_k.

Remark 9: Note that the proposed event-triggering condition (15) is a function of $\eta^*(t)$. In contrast to the traditional event-triggering conditions [3], the event-triggering error in the proposed scheme is bounded by the worst-case difference between the continuous and event-sampled control policy η^*. Thus, the inter-event sampling time obtained using the proposed condition (15) specifies the maximum time for which the control policy is not required to be updated with the latest sensor feedback information. Also, note that all the signals required to check the event-triggering condition $(x(t), x(t_k), \eta^*)$ are available at the event-triggering mechanism.

In the next section, an optimal adaptive event-triggered control design using NN based hybrid learning approach is presented.

4. NEURAL NETWORK CONTROLLER DESIGN

The control policy and the event-triggering condition require the optimal value function which is incidentally the solution to the HJI equation. The HJI equation does not have a closed-form solution [12],[15]. Therefore, numerical solutions are constructed by using reinforcement learning techniques [24] and approximate optimal control solutions are obtained. The saddle point solution for the proposed min-max problem formulated for the event-triggered control design can be learned online, in a forward-in-time manner using a

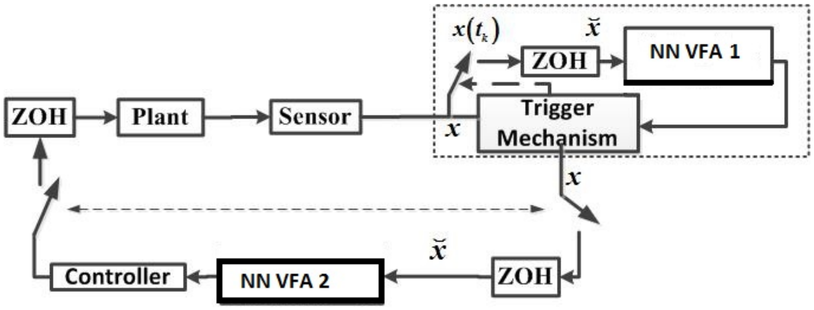


Fig. 6.2. Approximate optimal event sampled control system.

NN based value function approximator. First, a solution is proposed to relax the requirement of accurate knowledge of the system internal dynamics $f(x)$. Then an NN identifier based design is introduced to relax the accurate knowledge of both $f(x)$ and $g(x)$.

4.1 Case Of Known Control Coefficient Matrix

A block diagram of the proposed learning scheme is given in Fig. 6.2. It can be observed that in order to learn the optimal control policy and the event-triggering threshold, two value-function approximators (VFA) are required one at the controller and the other at the event-triggering mechanism. Both the value-function approximators learn the optimal value function corresponding to the HJI equation (13) and the initial values for the both the NN weights are same.

Using the infinitesimal version of the cost function (8), we have

$$\dot{V} = -Q(x) - u^T Ru + \sigma^2 \eta^T \eta \tag{17}$$

Integrating both sides of (17) in the interval $[t_k, t_{k+1})$, reveals

$$V^*(t_{k+1}) - V^*(t_k) = \int_{t_k}^{t_{k+1}} (-Q(x) - u^T Ru + \sigma^2 \eta^T \eta) d\tau. \tag{18}$$

Equation (18) is called the Bellman equation [15]. The Bellman equation is a fixed point equation with the optimal value function being the fixed point solution. Therefore, if the optimal value function and control policies in (18) are replaced by the estimated quantities, there would be an error, referred to as the Bellman error/temporal difference (TD) error [10], [24].

Define the estimate of the optimal value function as $\hat{V}$. Now replacing the optimal value function in (18) with the estimated optimal value function reveals [11]

$$\chi_{k+1}(t) = \int_{t_k}^{t_{k+1}} (x^T Qx + u^T Ru - \sigma^2 \eta^T \eta) d\tau + \hat{V}(t_{k+1}) - \hat{V}(t_k) \tag{19}$$

where χ_{k+1} is the Bellman residual error/ temporal difference error calculated at the occurrence of $k + 1$ event.

Assuming that the solution to the HJI equation is a smooth function, the approximation of the optimal value function can be represented in parametric form using artificial neural networks as

$$V^*(x) = W^T \phi(\omega^T x) + \varepsilon(x) \tag{20}$$

where W is the target NN weights, $\phi(x)$ is the smooth activation function satisfying $\phi(0) = 0$, $\varepsilon(x)$ is the reconstruction error and ω is the weights of the first layer which are randomly chosen, held constant to form a stochastic basis and will not be explicitly written henceforth [28].

Assumption 3: The target weight vector $W \in \Omega_W \subset \Re^{N_o \times 1}$ satisfies the bound $\|W\| \leq W_M$. Let the number of hidden layer neurons be denoted as N_o. The set of activation functions $[\phi_1 \; \phi_2 \ldots \phi_{N_o}]^T$ form a basis on the compact set Ω with $\|\phi(x)\| \leq N_o$. The reconstruction error satisfies $\|\varepsilon(x)\| \leq \varepsilon_M$.

Next, define the estimated NN weights, $\hat{W}$ and the estimated approximate optimal value function

$$\hat{V}(x) = \hat{W}^T \phi(x) \tag{21}$$

Substituting the estimate of the approximated optimal value function from (21) in (19) yields

$$\chi_{k+1}(t) = \int_0^T (Q(x) + u^T Ru - \sigma^2 \eta^T \eta) d\tau + \hat{W}^T \Delta\phi(\tau) \tag{22}$$

where $\Delta\phi(\tau) = \phi(t_{k+1}) - \phi(t_k)$ and $\chi_{k+1}(t)$ is the residual error calculated at the event-sampling instant t_{k+1}. Similarly, using (20) in (18) yields

$$W^T \Delta\phi(\tau) + \Delta\varepsilon(\tau) = \int_{t_k}^{t_{k+1}} (-Q(x) - u^T Ru + \sigma^2 \eta^T \eta) d\tau \tag{23}$$

where $\Delta\varepsilon(\tau) = \varepsilon(x(t_{k+1})) - \varepsilon(x(t_k))$. Define weight estimation error as $\tilde{W} = W - \hat{W}$. Then, substituting for the right hand side of (23) in (22) to get

$$-\chi_{k+1}(t) = W^T(t)\Delta\phi(\tau) - \hat{W}^T(t)\Delta\phi(\tau) + \Delta\varepsilon = \tilde{W}^T(t)\Delta\phi(\tau) + \Delta\varepsilon. \tag{24}$$

Equation (24) provides the relationship between TD error and the weight estimation error. With this relationship, the main results for this section are presented in the following theorem.

Theorem 2 (Case of Known Control Coefficient Matrix): Consider the infinite horizon cost function (8) and the nonlinear input affine system dynamics (7). Let W be a bounded and constant target NN weights for the value function approximator and $\hat{W}(0)$ be the initial estimated NN weights defined in a compact set Ω_W. Let the control policy given by

$$u(x) = -\frac{1}{2} R^{-1} g^T(x) \hat{V}_x, \tag{25}$$

be applied to the system with $\hat{V}_x$ being the gradient of the value function (21) with respect to the state vector. Choose an initial stabilizing control policy such that the resulting cost is finite. Further, let the event-triggering condition satisfying

$$\|e(t)\| \le \frac{\|\hat{\eta}(t)\|}{L_u}, \quad t \in [t_k, t_{k+1}), \forall k \in \{0, \mathbb{N}\}, \tag{26}$$

be used where $\hat{\eta} = \frac{1}{2\sigma^2} h^T(x)\hat{V}_x(t)$. Consider the value-function NN weight adaptation rule given by

$$\dot{\hat{W}} = \begin{cases} -\alpha \frac{[\Delta\phi(\tau)]}{(1+[\Delta\phi(\tau)]^T[\Delta\phi(\tau)])^2} \chi_k^T(t), \;\; t = t_k \\ -\alpha \frac{[\Delta\phi(\tau)]}{(1+[\Delta\phi(\tau)]^T[\Delta\phi(\tau)])^2} \chi^T(t_k), \;\; t \in (t_k, t_{k+1}). \end{cases} \tag{27}$$

Then, the state vector and the NN weight estimation error converges locally and becomes ultimately bounded with the event-triggering instants $k \longrightarrow \infty$, provided the design parameters α, Q, R, σ are chosen provided: $\bar{\delta}_x > 2L_u^2$, $\frac{\alpha}{\rho} > \frac{1}{2}\|R\|^2$ where $\rho = (1 + [\Delta\phi(\Delta T)]^T[\Delta\phi(\Delta T)])^2$, $\bar{\delta}_x = [Q(x) + u^{*T}Ru^* - \sigma^2\eta^{*T}\eta]$, where $\alpha > 0$, is the learning step. The bounds are defined as $\frac{1}{2}g_M^2\nabla\varepsilon_M^2 + \alpha^2\varepsilon_M^2$, where $\|g(x)\| \leq g_M$, $\|\nabla\varepsilon\| \leq \nabla\varepsilon_M$, $\|\varepsilon\| \leq \varepsilon_M$ with $\nabla\varepsilon_M, g_M, \varepsilon_M$ are positive constants.

Proof: See Appendix.

Remark 10: In contrast to the traditional policy-iteration scheme [24], the parameter tuning proposed in (27) is a hybrid learning scheme [10] which can be implemented online. The HJI residual error, χ_k, is calculated at every event-triggering instant, t_k, and the parameters are updated continuously using (27) both at the event sampling and inter-event intervals. The update rule utilizes the new information obtained at the event-triggering instant to calculate the Bellman error, $\chi_k(t)$, and the updates in the inter-event period, $[t_k, t_{k+1})$ tries to reduce the HJI residual error calculated at the last event-triggering instant, t_k. It is demonstrated in [10] that such a hybrid learning scheme improves the learning efficiency in the event-triggered feedback framework.

Remark 11: If the system dynamics are linear, the solution to the GARE is required instead of the solution to the HJI equation. Therefore, the approximation to the solution for the GARE can be represented as $V^*(x) = W^T\phi(x)$, where $\phi(x)$ being a regression function obtained by using the Kronecker product of $x^T(t) \otimes x(t)$ and W is obtained by representing the matrix P^* in a vector form [8].

Remark 12: Note that the bounds are obtained in Theorem 2 as a function of the NN reconstruction error. It has been demonstrated that as the number of hidden layer neurons are increased the reconstruction error converges to zero [28]. In this special case, by appropriate design of the NN approximator, the state vector and the NN weight estimation error converge to zero asymptotically.

4.2 Unknown Control Coefficient Matrix Using Identifier

Note that the control policy (25) and the event-triggering condition (26) still require the knowledge of the nonlinear function $g(x)$. To relax this requirement, consider the NN identifier as

$$\dot{x} = \hat{f}(\hat{x}) + \hat{g}(\hat{x})u(t) - A\tilde{x}(t) \tag{28}$$

where $\hat{f}, \hat{g}$ are the approximated functions of the nonlinear dynamics f, g. The forcing function, $\tilde{x} = x - \hat{x}$, is the state estimation error, and $A > 0$ is a linear map which stabilizes the NN identifier during the learning phase. Using NN approximation, the parametric equations for the nonlinear functions in (28) are $g(\bullet) = W_g \zeta_g(\bullet) + \varepsilon_g(\bullet)$, $f(\bullet) = W_f \zeta_f(\bullet) + \varepsilon_f(\bullet)$ where $W_\bullet$ denotes the target NN weights, $\zeta_\bullet$ denotes the bounded NN activation functions and $\varepsilon_\bullet$ denotes the bounded reconstruction errors. Using the estimate of the NN weights, $\hat{W}_\bullet$, define $\hat{f}(\bullet) = \hat{W}_f \zeta_f(\bullet)$ and $\hat{g}(\bullet) = \hat{W}_g \zeta_g(\bullet)$. Now to analyze the stability of (28), using (28) and (1), the dynamic equation describing the evolution of the state estimation error, $\tilde{x}(t)$ is revealed as

$$\begin{aligned}\dot{\tilde{x}} = \tilde{W}_f \zeta_f + W_f \tilde{\zeta}_f - \tilde{W}_f \tilde{\zeta}_f + [\tilde{W}_g \zeta_g + W_g \tilde{\zeta}_g - \\ \tilde{W}_g \tilde{\zeta}_g]u + \varepsilon_g u + \varepsilon_f + A\tilde{x}\end{aligned} \tag{29}$$

with $\tilde{\zeta}_\bullet = \zeta_\bullet(x) - \zeta_\bullet(\hat{x})$, $\tilde{W}_\bullet = W_\bullet - \hat{W}_\bullet$. The local bounded regulation of $\tilde{x}(t), \tilde{W}_\bullet(t)$ is observed when (29) is injected with a non-zero bounded input $e(t)$ and this result is summarized next.

Lemma 2: ([9][10]) Consider the identifier dynamics (28). Using the estimation error, $\tilde{x}(t)$, as a forcing function, define NN weight tuning using the Levenberg-Marquardt scheme with sigma modification term to avoid parameter drift as

$$\dot{\hat{W}}_f = \frac{\alpha_f \zeta_f \tilde{x}^T}{c_f + \left\|\tilde{x}^T\right\|^2} - \kappa_f \hat{W}_f, \;\; \dot{\hat{W}}_g = \frac{\alpha_g \zeta_g u \tilde{x}^T}{c_g + \left\|\tilde{x}^T\right\|^2 \left\|u^T\right\|^2} - \kappa_g \hat{W}_g \tag{30}$$

where $\alpha_f, \alpha_g, \kappa_f, \kappa_g, c_f, c_g$ are positive design constants. The error dynamics using (30) are obtained as

$$\dot{\tilde{W}}_f = \frac{-\alpha_f \zeta_f \tilde{x}^T}{c_f + \left\|\tilde{x}^T\right\|^2} + \kappa_f \hat{W}_f, \;\; \dot{\tilde{W}}_g = \frac{-\alpha_g \zeta_g u \tilde{x}^T}{c_g + \left\|\tilde{x}^T\right\|^2 \left\|u^T\right\|^2} + \kappa_g \hat{W}_g. \tag{31}$$

If $u(t)$ is stabilizing, then there exists $\alpha_\bullet, \kappa_\bullet, A_i > 0$ such that (29) and (31) are stable and $\tilde{x}(t), \tilde{W}_\bullet(t)$ are locally ultimately bounded. The bounds are functions of the reconstruction error and the sigma modification gain $\kappa_\bullet$.

Proof: See Appendix.

Remark 13: Using the NN identifier, the approximate optimal event-based control scheme can be developed by relaxing the requirement of the complete knowledge of the nonlinear dynamics. The stabilization of control policy requirement is not needed in the corollary.

Corollary 1: Consider the infinite horizon cost function (8) and the nonlinear input affine system dynamics (7). Let W be the a constant target weight matrix for the value function estimator and $\hat{W}(0)$ be the initial estimated weight matrix in Ω_W. Use the NN identifier (28) to obtain the approximation of the nonlinear system dynamics. Tune the NN identifier using the weight update rule (30). Let the control policy given by

$$u(x) = -\frac{1}{2} R^{-1} \hat{g}^T(t) \hat{V}_x, \tag{32}$$

be asserted on the system with an initial stabilizing policy. Let the event-triggering condition be

$$\|e(t)\| \leq \frac{\|\hat{\eta}(t)\|}{L_u}, \;\; t \in [t_k, t_{k+1}), \forall k \in \{0, \mathbb{N}\} \tag{33}$$

where $\hat{\eta} = \frac{1}{2\sigma^2}\hat{h}^T(x)\hat{V}_x(t)$. Next, consider the value-function NN hybrid weight adaptation rule given by

$$\dot{\hat{W}} = \begin{cases} -\alpha \frac{[\Delta\phi(\tau)]}{(1+[\Delta\phi(\tau)]^T[\Delta\phi(\tau)])^2}\chi_k^T(t), \ t = t_k \\ -\alpha \frac{[\Delta\phi(\tau)]}{(1+[\Delta\phi(\tau)]^T[\Delta\phi(\tau)])^2}\chi^T(t_k), \ t \in (t_k, t_{k+1}). \end{cases} \tag{34}$$

Then, the states of the system, NN identifier and the NN weight estimation errors converge locally to a bound with the event-triggering instants $k \longrightarrow \infty$.

Proof: See Appendix.

Remark 14: Using the NN identifier along with the value function approximator provides an additional benefit when compared to the traditional actor-critic architecture [14]-[15], which is an alternate approach to design approximate optimal controllers. The advantage of this approach is that the NN identifier can be used for online exploration [10] and the identifier state vector can be substituted for actual state vector to mimic a model-based event-triggering scheme which reduces the effects of network delays and packet losses [4]. *Remark 15:* Once the states reach their bounds, a dead-zone operator can be used to stop the event-triggering mechanism to generate redundant events [9],[22].

In the next section, the hybrid-learning based zeroâĂŞsum game theoretic formulation is extended to the distributed control using decentralized event-triggering conditions for interconnected systems.

5. EXTENSION TO DISTRIBUTED APPROXIMATE OPTIMAL CONTROL

Consider a nonlinear input-affine system composed of N interconnected subsystems, each of the form

$$\dot{x}_i = f_i(x_i) + g_i(x_i)u_i + \sum_{\substack{j=1 \\ j\neq i}}^{N} \Delta_{ij}(x_i, x_j), \quad x_i(0) = x_{i0} \tag{35}$$

where $x_i(t) \in \Omega_i \subseteq \Re^{n_i\times 1}$ represents the state vector of the i^{th} subsystem; $\dot{x}_i(t)$ its time derivative; Ω_i is a compact set; $u_i(t) \in \Re^{m_i}$ is the control input; f_i, g_i, are uncertain nonlinear maps and Δ_{ij} is the uncertain nonlinear interconnection between i^{th} and j^{th}

subsystem. Let x_{i0} be the given initial subsystem state. Using the subsystem dynamics, the augmented system dynamics can be represented as

$$\dot{x} = f(x) + g(x)u, x(0) = x_0 \tag{36}$$

(36) where $f = [(f_1 + \sum_{j=2}^{N} \Delta_{1j})^T, ., (f_N + \sum_{j=1}^{N-1} \Delta_{Nj})^T]^T$, $x = [x_1^T, ., x_N^T]^T \in \Omega \subseteq \Re^n$, $n = \sum_{i=1}^{N} n_i$, $u = [u_1^T, .., u_N^T]^T \in \Re^m$, $m = \sum_{i=1}^{N} m_i$, $g = diag([g_1(x_1).., g_N(x_N)])$, Ω is obtained as a finite union of Ω_i.

Remark 16: ([10]) Using the performance measure for the augmented system (36), the cost function for the individual subsystems can be represented using the relation $V(x) = \sum_{i=1}^{N} V_i(x)$ and the resulting control policy $u_i(x)$ is a distributed control policy as it is defined as a function of the local states and the states of the neighboring interconnected subsystems.

Assumption 3: The dynamics (35) and (36) are stabilizable with origin as the equilibrium point. Full state measurements are available for control. The communication network which facilitates information sharing among subsystems is lossless.

Corollary 2: Consider the infinite horizon cost function (8) and the nonlinear system dynamics (36). Use the NN identifier at each subsystem defined by (28) and the identifier NN weights be updated using (30). Let W_i be the constant, bounded, target parameter vector for the NN-function approximator at the i^{th} subsystem. Let the control policy

$$u_i(t) = -\frac{1}{2}[R_i^{-1}\hat{g}_i(x)\hat{V}_{ix}(t)], \tag{37}$$

be applied to the subsystem, where R_i is a positive definite matrix, $\hat{V}_{ix}$ is the gradient of the estimated optimal cost function of the i^{th} subsystem with respect to the states, x_i and let the decentralized event-triggering condition satisfies

$$\|e_i(t)\| \leq \frac{1}{L_{u_i}} \|\hat{\eta}_i(t)\|, \quad t \in [t_k, t_{k+1}), \forall k \in \{0, \mathbb{N}\} \tag{38}$$

where $R = diag(R_i)$, $\hat{\eta} = [\hat{\eta}_1^T \ \hat{\eta}_2^T \ .. \ \hat{\eta}_N^T]^T, L_{u_i} > 0$. Let the value-function NN weights at each subsystem be updated using

$$\dot{\hat{W}}_i = \begin{cases} -\alpha_i \frac{[\Delta\phi_i(\tau)]}{(1+[\Delta\phi_i(\tau)]^T[\Delta\phi_i(\tau)])^2} \chi_{i,k}^T(t), \ t = t_k \\ -\alpha_i \frac{[\Delta\phi_i(\tau)]}{(1+[\Delta\phi_i(\tau)]^T[\Delta\phi_i(\tau)])^2} \chi_i^T(t_k), \ t \in (t_k, t_{k+1}). \end{cases} \tag{39}$$

where $\hat{W}_i$ is the estimate of W_i, α_i is the learning rate and $\Delta\phi_i$ is defined similar to (34) at the i^{th} subsystem. Then, the state vector and the parameter estimation error converges to an ultimate bound which is a function of the reconstruction error, $\kappa_{\bullet}$ the sigma modification term in the identifier weight tuning law, as the event-triggering instants $k \to \infty$.

Proof: The proof of this Corollary follows a similar line of argument as Theorem 2. To avoid redundancy, the detailed derivations are omitted. In the next section, simulation results are provided to verify the theoretical claims.

6. SIMULATION RESULTS

For the simulation results, first, consider the unstable nonlinear dynamics in continuous-time [11] as $\dot{x} = f(x) + g(x)u$, with $x = [x_1 x_2]^T$ and $u = [u_1 u_2]^T$. The nonlinear dynamics are $\dot{x}_1 = -(29x_1 + 87x_1x_2^2)/8 - (2x_2 + 3x_2x_1^2)/4 + u_1$ and $\dot{x}_2 = -(x_1 + 3x_1x_2^2)/4 + 3u_2$. To verify the advantages of the proposed method, the results of our approach is compared with that of an event-triggered optimal approximate controller with the event-triggering condition of the form [22].

For the case of uncertain system dynamics $(f, g$ unknown), the proposed NN approximate-learning based control approach is compared with the traditional event-triggered NN approximate optimal control scheme (Sahoo et. al, [17]). The considered numerical example is studied in H_∞ control schemes and the analytical solution to the HJI equation (optimal value function) is calculated as $V^*(x) = x_1^2 + 2x_2^2 + 3x_1x_2$ [11],[15].

The simulation analyses carried out with different design parameters and the values of the penalizing matrices Q, R are taken as $10I, 0.2I$. The value of σ is taken as 0.7. These values are chosen as they resulted in comparable control effort and state trajectories. Both the VFA NN weights are initialized with the same random values from the interval [-2,2].

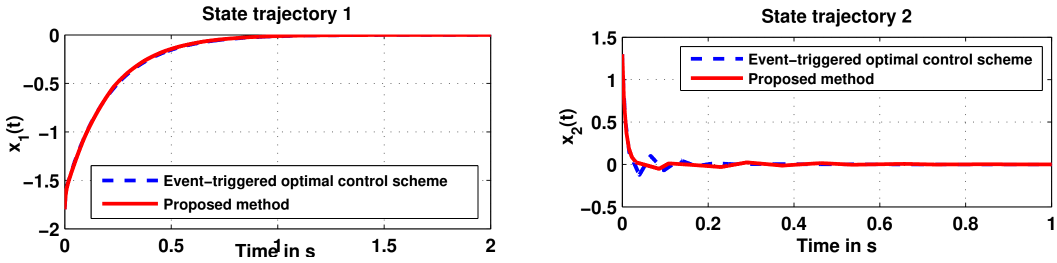


Fig. 6.3. Comparison of state trajectories x_1 and x_2.

Fig. 6.3 depicts the convergence of the closed-loop system state vector, and Fig. 6.4 depicts the comparison plots of the control inputs. The proposed method is contrasted with the event-triggered implementation of approximate optimal controller. It can be observed that the state trajectories are satisfactory with similar control effort.

Further, the lower bound on the inter-event times is observed to be 1 ms. It is clear from Fig. 6.5 that the event-triggering threshold with the proposed approach is considerably higher than the traditional approach. This elongated the inter-event time, reducing the resource utilization which is one of the primary objectives of the design. Fig. 6.5 shows the inter-event times.

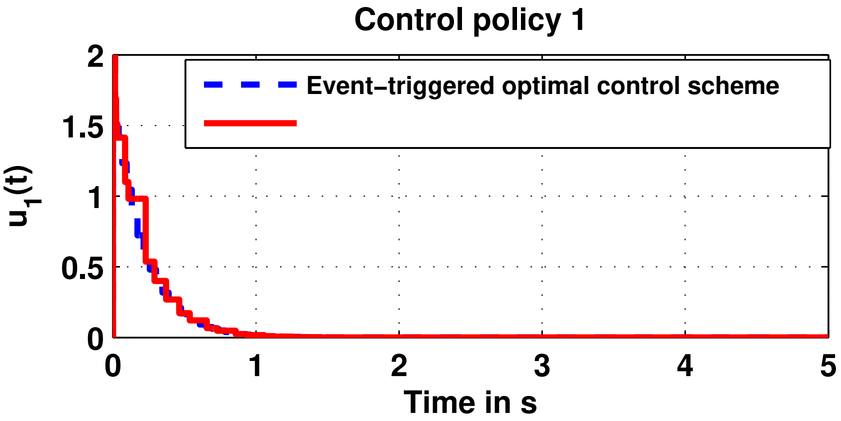


Fig. 6.4. Comparison of control policy u_1.

Table 6.1. Analysis of approximate optimal control scheme.

σ	Avg. IET in s		Cumulative cost		Events	
	ET-NN	P.M	ET-NN	P.M	ET-NN	P.M
0.90	0.0350	0. 0685	2.074e+5	1.442e+5	290	120
0.925	0.0357	0.0675	2.085e+5	1.530e+5	280	130
0.95	0.0362	0.0744	2.100e+5	1.602e+5	270	140
0.975	0.0369	0.0583	2.110e+5	1.654e+5	260	160
1	0.0604	0.0562	2.114e+5	1.704e+5	260	170

The comparison of inter-event time in Fig. 6.5 and the cumulative cost function in Fig. 6.6 reveals the benefit of using the proposed scheme. It is observed that the average inter-event time is increased considerably and the cumulative cost is reduced with the proposed approach. Due to space consideration, all the simulation figures are not included and the results are summarized in Table 6.1.

In Table 6.1, proposed method is abbreviated as PM, event-triggered NN based approximate optimal control [17] is abbreviated as ET-NN.

7. CONCLUSIONS

This paper proposes a novel approach for simultaneously optimizing both the event-triggering sampling instants and state feedback controller using zero-sum game formulation for a class of nonlinear system. The proposed design scheme provides a tractable trade-off between the frequency of events and the system performance cost by utilizing the min-

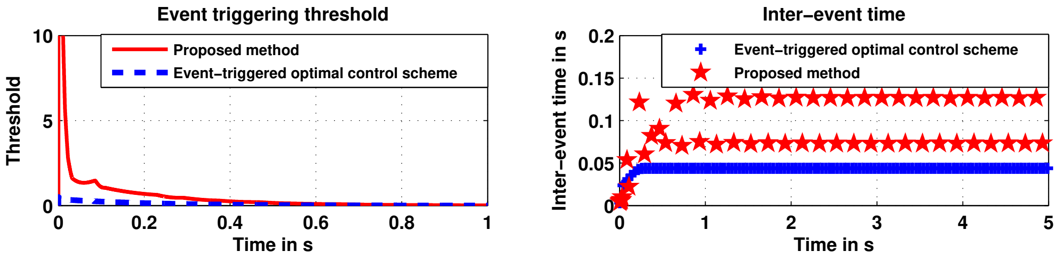


Fig. 6.5. Comparison of the performance of event-triggering mechanism.

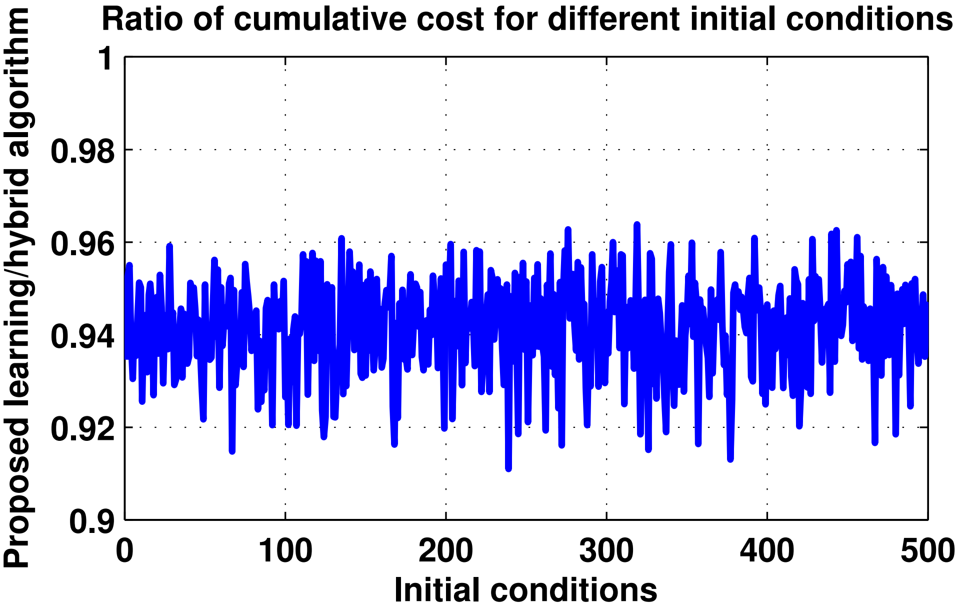


Fig. 6.6. Comparison of cumulative cost.

max optimization of the cost function. The inter-event time interval increases with the proposed event-triggering condition which considerably reduces the communication cost when compared to the traditional event-triggering schemes. The approximation based NN-learning scheme generates the optimal control policy and the event-triggering condition forward-in-time by obviating the curse-of-dimensionality. The NN identifiers relaxed the requirement of the accurate knowledge of the system dynamics to implement the proposed scheme.

REFERENCES

[1] R. Dorf, M. Farren, and C. Phillips, “Adaptive sampling frequency for sampled-data control systems,” *IRE Transactions on Automatic Control*, vol. 7, no. 1, pp. 38–47, Jan 1962.

[2] K. J. Astrom and B. M. Bernhardsson, “Comparison of Riemann and Lebesgue sampling for first order stochastic systems,” in *Proceedings of the 41st IEEE Conference on Decision and Control, 2002.*, vol. 2, Dec 2002, pp. 2011–2016 vol.2.

[3] P. Tabuada, “Event-triggered real-time scheduling of stabilizing control tasks,” *IEEE Transactions on Automatic Control*, vol. 52, no. 9, pp. 1680–1685, Sept 2007.

[4] W. Heemels and M. Donkers, “Model-based periodic event-triggered control for linear systems,” *Automatica*, vol. 49, no. 3, pp. 698–711, 2013.

[5] X. Meng and T. Chen, “Optimal sampling and performance comparison of periodic and event based impulse control,” *IEEE Transactions on Automatic Control*, vol. 57, no. 12, pp. 3252–3259, Dec 2012.

[6] M. Jost, M. S. Darup, and M. MÃűnnigmann, “Optimal and suboptimal event-triggering in linear model predictive control,” in *2015 European Control Conference (ECC)*, July 2015, pp. 1153–1158.

[7] A. Girard, “Dynamic triggering mechanisms for event-triggered control,” *IEEE Transactions on Automatic Control*, vol. 60, no. 7, pp. 1992–1997, July 2015.

[8] F. L. Lewis and V. L. Syrmos, *Optimal control.* John Wiley & Sons, 1995.

[9] 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.

[10] V. Narayanan and S. Jagannathan, “Approximate optimal distributed control of uncertain nonlinear interconnected systems with event-sampled feedback,” in *2016 IEEE 55th Conference on Decision and Control (CDC)*, Dec 2016, pp. 5827–5832.

[11] T. Dierks and S. Jagannathan, “Optimal control of affine nonlinear continuous-time systems using an online hamilton-jacobi-isaacs formulation,” in *49th IEEE Conference on Decision and Control (CDC)*, Dec 2010, pp. 3048–3053.

[12] D. P. Bertsekas, D. P. Bertsekas, D. P. Bertsekas, and D. P. Bertsekas, *Dynamic programming and optimal control*. Athena Scientific Belmont, MA, 1995, vol. 1, no. 2.

[13] Z. Chen and S. Jagannathan, “Generalized hamilton jacobi bellman formulation -based neural network control of affine nonlinear discrete-time systems,” *IEEE Transactions on Neural Networks*, vol. 19, no. 1, pp. 90–106, Jan 2008.

[14] J. Si, *Handbook of learning and approximate dynamic programming*. John Wiley & Sons, 2004, vol. 2.

[15] 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, 2012.

[16] Y. Jiang and Z. P. Jiang, “Global adaptive dynamic programming for continuous-time nonlinear systems,” *IEEE Transactions on Automatic Control*, vol. 60, no. 11, pp. 2917–2929, Nov 2015.

[17] A. Molin and S. Hirche, “On the optimality of certainty equivalence for event-triggered control systems,” *IEEE Transactions on Automatic Control*, vol. 58, no. 2, pp. 470–474, Feb 2013.

[18] A. Molin, “Optimal event-triggered control with communication constraints,” *Technische Universität München, München*, 2014.

[19] J. Lunze and D. Lehmann, “A state-feedback approach to event-based control,” *Automatica*, vol. 46, no. 1, pp. 211–215, 2010.

[20] M. Mazo and P. Tabuada, “Decentralized event-triggered control over wireless sensor/actuator networks,” *IEEE Transactions on Automatic Control*, vol. 56, no. 10, pp. 2456–2461, Oct 2011.

[21] 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.

[22] P. Tallapragada and N. Chopra, "On event triggered tracking for nonlinear systems," *IEEE Transactions on Automatic Control*, vol. 58, no. 9, pp. 2343–2348, Sept 2013.

[23] C. De Persis, R. Sailer, and F. Wirth, "Parsimonious event-triggered distributed control: A zeno free approach," *Automatica*, vol. 49, no. 7, pp. 2116–2124, 2013.

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

[25] T. Basar and P. Bernhard, "H infinity optimal control and related minimax design problems," *Birkhaüser, Boston, Massachusetts*, 1991.

[26] A. J. van der Schaft, "L2-gain analysis of nonlinear systems and nonlinear state-feedback h infin; control," *IEEE Transactions on Automatic Control*, vol. 37, no. 6, pp. 770–784, Jun 1992.

[27] H. Khalil, *Nonlinear Systems*, ser. Pearson Education. Prentice Hall, 2002.

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

[29] W. Rudin, *Principles of mathematical analysis*, ser. International series in pure and applied mathematics. McGraw-Hill, 1964.

APPENDIX

Proof of Lemma 1:

Choose the positive definite Lyapunov function $L_x(x) = V^*(x)$. The time derivative along the system dynamics can be obtained as

$$\dot{L}_x(t) = V_x^{*^T} \dot{x} = [V_x^{*^T} f + V_x^{*^T} gu + V_x^{*^T} h\eta]. \tag{40}$$

Add and subtract $V_x^{*^T} gu^* + V_x^{*^T} h\eta^*$ to get

$$\begin{aligned} \dot{L}_x(t) &= [V_x^{*^T} f + V_x^{*^T} gu^* + V_x^{*^T} h\eta^*] \\ &+V_x^{*^T} g(u - u^*) + V_x^{*^T} h(\eta - \eta^*) \end{aligned} . \tag{41}$$

From (10), with (u^*, η^*, V^*), we have $H(x, u^*, \eta^*) = 0$ and

$$-Q(x) - u^{*T} Ru^* + \sigma^2 \eta^{*T} \eta^* = V_x^{*T} [f + gu^* + h\eta^*]. \tag{42}$$

Using the right hand side of (42) in (41) results in

$$\begin{aligned} \dot{L}_x(t) &= -Q(x) - u^{*T} Ru^* + \sigma^2 \eta^{*T} \eta^* \\ &+V_x^{*^T} g(u - u^*) + V_x^{*^T} h(\eta - \eta^*). \end{aligned} \tag{43}$$

Using the definition, $\eta^* = \frac{1}{2\sigma^2} h^T V_x^*$, and substituting the control policy u^* from (14) in (43), we get $\dot{L}_x(t) = -Q(x) - u^{*T} Ru^* - \sigma^2 \eta^{*T} \eta^* + 2\sigma^2 \eta^{*T} \eta$. By using the definition of optimal policies (11), and (12), we have

$$\dot{L}_x(t) \leq -x^T \delta_x x + V_x^{*T} h\eta \tag{44}$$

where $\delta_x > 0$. Applying norm operator to (44) reveals

$$\dot{L}_x(t) \leq -\delta_{x,m} \|x\|^2 + L_v L_u \, \|h\| \, \|x\| \, \|e\| \tag{45}$$

where L_u, L_v are Lipschitz constants. From (45) it can be concluded that the closed-loop system is input-to-state stable as Q, R, σ are chosen as positive definite functions resulting in $\delta_{x,m} > 0$.

Proof of Theorem 1:

With the Lyapunov function candidate chosen similar to that in Lemma 2, from (44), we have

$$\dot{L}_x(t) = -x^T \delta_x x + V_x^{*T} h\eta. \tag{46}$$

Applying norm operator and using the event-triggering condition (15)

$$\dot{L}_x(t) \leq -\delta_{x,m} \|x\|^2 + \left\|V_x^{*T}\right\| \|x\| \|\eta^*\| . \tag{47}$$

Using the definition of η^* and (12), we get

$$\dot{L}_x(t) \leq -(\delta_{x,m} - \frac{1}{2\sigma^2}\left\|V_x^{*T}\right\|^2)\|x\|^2. \tag{48}$$

From the definition of $\delta_{x,m}$ in (44), note that $\delta_{x,m} > \frac{1}{2\sigma^2}\left\|V_x^{*T}\right\|^2$. Now to derive the positive inter-event time, use (3) and taking the time-derivative reveals

$$\dot{e}(t) = \dot{\breve{x}}(t) - \dot{x}(t), \;\; t \in [t_k, t_{k+1}). \tag{49}$$

Noting that $\breve{x}$ is a constant in the inter-event period, we have $\dot{\breve{x}} = 0$. Taking the norm operator and substituting the system dynamics reveals

$$\|\dot{e}(t)\| = \|\dot{x}(t)\| = \|f + gu^* + h\eta\| . \tag{50}$$

From (48), there exists a $X_m > 0$ such that $\|f + gu^*\| \leq X_m$. Using this relation in (50) yields

$$\|\dot{e}(t)\| \leq \|hL_u\| \, \|e\| + X_m. \tag{51}$$

Integrating using the comparison lemma [22], reveals

$$\|e(t)\| \leq \frac{X_m}{\|hL_u\|}(e^{\|hL_u\|(t-t_k)} - 1), \quad t \geq t_k. \tag{52}$$

Substituting $t = t_{k+1}$ reveals the minimum positive inter-event time

$$(t_{k+1} - t_k)_{\min} \geq \frac{1}{\|hL_u\|} \log(\frac{\|hL_u\|}{X_m} \, \|e^*\| + 1). \tag{53}$$

Proof of Theorem 2:

Consider the Lyapunov candidate

$$L(x, \tilde{W}) = L_x(t) + L_W(t) \tag{54}$$

where $L_x(t) = V^*(x)$ and $L_W(t) = \frac{1}{2}\tilde{W}^T\tilde{W}$. The derivative of the Lyapunov candidate is given by

$$\dot{L}(x,\tilde{W}) = \dot{L}_x(t) + \dot{L}_W(t). \tag{55}$$

Consider the first term and using the system dynamics to get

$$\dot{L}_x(t) = V_x^{*^T}\dot{x} = [V_x^{*^T} f + V_x^{*^T} gu + V_x^{*^T} h\eta]. \tag{56}$$

Add and subtract $V_x^{*^T} gu^* + V_x^{*^T} h\eta^*$ to get

$$\begin{aligned}&\dot{L}_x(t) = [V_x^{*^T} f + V_x^{*^T} gu^* + V_x^{*^T} h\eta^*]\\ &+V_x^{*^T} g(u-u^*) + V_x^{*^T} h(\eta-\eta^*)\end{aligned}. \tag{57}$$

Using the Hamiltonian (GARE), $H(x,u^*,\eta^*)$, we have

$$-Q(x) - u^{*T}Ru^* + \sigma^2\eta^{*T}\eta^* = V_x^{*T}[f + gu^* + h\eta^*]. \tag{58}$$

Substituting (58) in (57)

$$\begin{aligned}&\dot{L}_x = -Q(x) - u^{*T}Ru^* + \sigma^2\eta^{*T}\eta^* + V_x^{*^T} gu(\breve{x})\\ &-V_x^{*^T} gu^* - V_x^{*^T} h\eta^*\end{aligned}. \tag{59}$$

Adding and subtracting $V_x^{*^T} gu^*(\breve{x})$ in (59), we get

$$\begin{aligned}&\dot{L}_x(t) = -Q(x) - u^{*T}Ru^* + \sigma^2\eta^{*T}\eta^* - V_x^{*^T} h\eta^*\\ &+V_x^{*^T} g[u(\breve{x}) - u^*(\breve{x})] + V_x^{*^T} g[u^*(\breve{x}) - u^*(x)]\end{aligned}. \tag{60}$$

Now, define $\tilde{u} = u^*(\breve{x}) - u(\breve{x})$, using the definition (25) and (11), we have $\tilde{u} = -\frac{1}{2}R^{-1}g(\breve{x})\nabla\varepsilon(\breve{x}) - \frac{1}{2}R^{-1}g(\breve{x})\nabla\phi^T(\breve{x})\tilde{W}$. Substitute $\tilde{u}$ in (60) to get

$$\begin{aligned}&\dot{L}_x(t) = -Q(x) - u^{*T}Ru^* + \sigma^2\eta^{*T}\eta^* - V_x^{*^T} h\eta^*\\ &-2u^{*T}R\tilde{u} + V_x^{*^T} g[u^*(\breve{x}) - u^*(x)]\end{aligned}$$

Using the definition of η^*, we get the Lyapunov time-derivate

$$\dot{L}_x(t) = -Q(x) - u^{*T}Ru^* + \sigma^2\eta^{*T}\eta^* - 2u^{*T}R\tilde{u} \tag{61}$$

Applying the norm operator in (61) to get

$$\dot{L}_x(t) \leq -\bar{\delta}_x\|x\|^2 + \left\|u^{*T}\right\|\left\|g(\breve{x})\nabla\varepsilon(\breve{x})\right\| + \left\|u^{*T}\right\|\left\|g(\breve{x})\nabla\phi^T(\breve{x})\tilde{W}\right\| \tag{62}$$

where $\bar{\delta}_x = [Q + \frac{1}{4}P^*BR^{-1}B^TP^* - \frac{1}{4\sigma^2}P^*DD^TP^*]$. Using the Youngs inequality, we get

$$\dot{L}_x(t) \leq -\bar{\delta}_x\|x\|^2 + \left\|u^{*T}\right\|^2 + \frac{1}{2}g_M^2\left\|\nabla\phi^T(\breve{x})\tilde{W}\right\|^2 + \frac{1}{2}g_M^2\nabla\varepsilon_M^2.$$

Therefore, using the definition of u^* , (64) is simplified as

$$\dot{L}_x(t) \leq -(\bar{\delta}_x - L_u^2)\|x\|^2 + \frac{1}{2}g_M^2\left\|\nabla\phi^T(\breve{x})\tilde{W}\right\|^2 + \frac{1}{2}g_M^2\nabla\varepsilon_M^2. \tag{63}$$

Now consider the second term in the Lyapunov function. The estimation error dynamics is revealed as $\dot{\tilde{\theta}}(t) = -\dot{\hat{\theta}}(t)$ and

$$\dot{\tilde{W}} = \alpha\frac{[\Delta\phi(\Delta\tau)]}{1 + ([\Delta\phi(\Delta\tau)]^T[\Delta\phi(\Delta\tau)])^2}\chi_{k+1}^T(t). \tag{64}$$

Let $L_W(t) = \frac{1}{2}\tilde{W}^T\tilde{W}$, using (64) and (31), the Lyapunov time derivative is obtained as

$$\dot{L}_W(t) = -\alpha\frac{\tilde{W}^T\Delta\phi(\Delta\tau)[\Delta\phi(\tau)^T\tilde{W} + \Delta\varepsilon]}{(1 + [\Delta\phi(\Delta\tau)]^T[\Delta\phi(\Delta\tau)])^2}. \tag{65}$$

Using (63) and (65) in (55), we get

$$\begin{aligned}\dot{L}(t) &\leq -(\bar{\delta}_x - L_u^2)\|x\|^2 - (\tfrac{\alpha}{\rho} - \tfrac{g_M^2}{2} - \tfrac{1}{2})\left\|\tilde{W}\Delta\phi(\tau)\right\|^2 \\ &+\tfrac{1}{2}g_M^2\nabla\varepsilon_M^2 + \alpha^2\varepsilon_M^2\end{aligned} \tag{66}$$

where $\rho = (1 + [\Delta\phi(\Delta\tau)]^T[\Delta\phi(\Delta\tau)])^2$. From (66), it can be seen that the state vector and the NN weight estimation errors converge to their bounds in the inter-event period. Thus, the closed loop system is input to state stable [27] in the presence of measurement error and from the event-sampling condition, the measurement error remain bounded in the inter-event period. Therefore, it can be concluded that the closed-loop system is ultimately bounded similar to [22].

Alternatively, consider the event-triggering sampling instants. Choose the Lyapunov candidate function (54). First, note that the Lyapunov function is continuous because of the fact that the state vector and the weight estimation errors are continuous [29]. For the NN based learning scheme, define the closed-loop state vector $\xi = [x^T\ \tilde{W}^T]^T$. From (66), $\xi(t)$ is converging to its ultimate bound in the inter-event period. Hence, there exists a positive minimum inter event time (53) as long as the state vector and the weight estimation errors are outside their ultimate bounds (In this case, the variable X_M in equation (51) should be modified as $X_M + B$, where B is the bound for the state vector derived in (66)). This implies that the sequence of discrete sampling instants do not have an accumulation point [9],[22]-[23]. Therefore, it can be concluded that the set of points corresponding to the event-triggering sampling instants $\{t_k\}$ is countable with Lebesgue measure of zero [2]-[3],[29]. Thus, at the inter-event time interval and at the event-triggering sampling instants, the time derivative of the Lyapunov function is less than zero and the closed-loop state $\xi(t)$ locally converges to the ultimate bound defined in (66).

Proof of Lemma 2:

Consider the Lyapunov function candidate $L_I(\tilde{x}, \tilde{W}_f, \tilde{W}_g) = L_{\tilde{x}} + L_{\tilde{f}} + L_{\tilde{g}}$, with $L_{\tilde{x}} = \frac{\mu_{i1}}{2}\tilde{x}^T\Pi\tilde{x}$, $L_{\tilde{f}} = \frac{\mu_2}{2}\tilde{W}_f^T\tilde{W}_f + \frac{\mu_4}{4}(\tilde{W}_f^T\tilde{W}_f)^2$, and $L_{\tilde{g}} = \frac{\mu_3}{2}\tilde{W}_g^T\tilde{W}_g + \frac{\mu_5}{4}(\tilde{W}_g^T\tilde{W}_g)^2$, where $\mu_\bullet, \Pi$, are positive constants of appropriate dimensions. Consider the first term in the Lyapunov function. Taking the derivative and substituting the estimation error dynamics (29) yields

$$\dot{L}_{\tilde{x}} = \mu_1\tilde{x}^T\Pi(\tilde{W}_f\zeta_f + W_f\tilde{\zeta}_f - \tilde{W}_f\tilde{\zeta}_f + [\tilde{W}_g\zeta_g + W_g\tilde{\zeta}_g \\ -\tilde{W}_g\tilde{\zeta}_g]u + \varepsilon_g u + \varepsilon_f + A\tilde{x})$$

$$\dot{L}_{\tilde{x}} = \mu_1\tilde{x}^T\Pi A\tilde{x} + \mu_1\tilde{x}^T\Pi\tilde{W}_f\zeta_f + \mu_1\tilde{x}^T\Pi W_f\tilde{\zeta}_f - \mu_1\tilde{x}^T\Pi\tilde{W}_f\tilde{\zeta}_f \\ +\mu_1\tilde{x}^T\Pi[\tilde{W}_g\zeta_g + W_g\tilde{\zeta}_g - \tilde{W}_g\tilde{\zeta}_g]u + \mu_1\tilde{x}^T\Pi[\varepsilon_g u + \varepsilon_f].$$

Apply the norm operator and by choosing the matrix A such that the minimum singular value of A is given as $-\lambda_{\min}(\bar{q})$ reveals

$$\dot{L}_{\tilde{x}} \leq -(\lambda_{\min}(\mu_1\Pi\bar{q}) - \tfrac{7}{2})\|\tilde{x}\|^2 + \tfrac{1}{2}\|\mu_1\|^2\left\|\Pi\tilde{W}_f\zeta_f\right\|^2 \\ +\tfrac{1}{2}\|\mu_1\|^2\left\|\Pi[\varepsilon_f + \varepsilon_g u]\right\|^2 + \tfrac{1}{2}\|\Pi\|^2\|\mu_1\|^2 W_{fM}^2\left\|\tilde{\zeta}_f\right\|^2 \\ +\tfrac{1}{2}\|\Pi\|^2\|\mu_1\|^2\left\|\tilde{W}_f\right\|^2\left\|\tilde{\zeta}_f\right\|^2 + \tfrac{1}{2}\|\mu_1\|^2\left\|\Pi\tilde{W}_g\zeta_g u\right\|^2 \\ +\tfrac{1}{2}\|\Pi\|^2\|\mu_1\|^2\left\|\tilde{W}_g\right\|^2\left\|\tilde{\zeta}_g\right\|^2\|u\|^2 + \tfrac{1}{2}\|\Pi\|^2\|\mu_1\|^2 W_{gM}^2\left\|\tilde{\zeta}_g\right\|^2\|u\|^2.$$

Using the YoungâĂŹs inequality [29] and grouping similar terms yield

$$\dot{L}_{\tilde{x}} \leq -(\lambda_{\min}(\mu_1\Pi\bar{q}) - \frac{7}{2})\left\|\tilde{x}^T\right\|^2 + \frac{1}{2}\left\|\tilde{W}_g\right\|^4 + \frac{1}{2}\left\|\tilde{W}_f\right\|^4 + \eta_{o\tilde{x}B}$$

where

$$\eta_{o\tilde{x}B} = \tfrac{1}{2}\|\mu_1\|^2\|\Pi\|^2((\varepsilon_{fM}^2 + \varepsilon_{gM}^2 B_u^2) + \|\mu_1\|^2\|\Pi\|^2 N_{of}^2 \\ +N_{og}^2 B_u^4\|\mu_1\|^2\|\Pi\|^2 + W_{gM}^2 N_{og} B_u^2 + W_{fM}^2 N_{of}).$$

Now consider the second term in the Lyapunov candidate function. Taking the derivative and using the weight estimation error dynamics (31) reveals

$$\begin{aligned}\dot{L}_{\tilde{f}} = &-\frac{\mu_2 \tilde{W}_f^T \alpha_f \zeta_f \tilde{x}^T}{c_f + \tilde{x}^T \tilde{x}} + \mu_2 \tilde{W}_f^T \kappa_f \hat{W}_f \\ &-\frac{\mu_4 (\tilde{W}_f^T \tilde{W}_f) \tilde{W}_f^T \alpha_f \zeta_f \tilde{x}^T}{c_f + \tilde{x}^T \tilde{x}} + \mu_4 (\tilde{W}_f^T \tilde{W}_f) \tilde{W}_f^T \kappa_f \hat{W}_f\end{aligned}$$

Apply the norm operator and using the fact that the activation functions are bounded such that $\|\zeta_\bullet\| \leq N_{o\bullet}$, reveals

$$\begin{aligned}\dot{L}_{\tilde{f}} \leq &-(\lambda_{\min}(\mu_2 \kappa_f) - 1)\left\|\tilde{W}_f\right\|^2 - (\lambda_{\min}(\mu_4 \kappa_f) - 2)\left\|\tilde{W}_f\right\|^4 \\ &+\tfrac{1}{2}\|\mu_2\|^2 N_{of}^2 \left\|\alpha_f\right\|^2 + \tfrac{1}{8}\|\mu_4\|^4 N_{of}^4 \left\|\alpha_f\right\|^4 + \tfrac{1}{2}\|\mu_2\|^2 \left\|\kappa_f W_f\right\|^2 \\ &+\tfrac{1}{8}\|\mu_4\|^4 \left\|\kappa_f W_f\right\|^4 \\ \leq &-(\lambda_{\min}(\mu_2 \kappa_f) - 1)\left\|\tilde{W}_f\right\|^2 - (\lambda_{\min}(\mu_4 \kappa_f) - 2)\left\|\tilde{W}_f\right\|^4 + \eta_{ofB}\end{aligned}$$

where the bound η_{ofB} is defined as

$$\eta_{ofB} = \tfrac{1}{2}\|\mu_2\|^2 \left\|\alpha_f\right\|^2 N_{of}^2 + \tfrac{1}{8}\|\mu_4\|^4 \left\|\alpha_f\right\|^4 N_{of}^4 + \tfrac{1}{2}\|\mu_2\|^2 \left\|\kappa_f\right\|^2 W_{fM}^2 + \tfrac{1}{8}\|\mu_4\|^4 \left\|\kappa_f\right\|^4 W_{fM}^4,$$

and $\|W_\bullet\| \leq W_{\bullet M}$. Finally, consider the last term in the Lyapunov candidate function. Taking the derivative and substituting the weight estimation error dynamics (31) yields

$$\dot{L}_{\tilde{g}} = \mu_3 \tilde{W}_g^T \left(-\frac{\alpha_g \zeta_g u \tilde{x}^T}{c_g + \tilde{x}^T \tilde{x} u^T u} + \kappa_g \hat{W}_g\right) + \mu_5 (\tilde{W}_g^T \tilde{W}_g)(\tilde{W}_g^T (-\frac{\alpha_g \zeta_g u \tilde{x}^T}{c_g + \tilde{x}^T \tilde{x} u^T u} + \kappa_g \hat{W}_g))$$

On simplification we get

$$\dot{L}_{\tilde{g}} \leq -(\lambda_{\min}(\mu_3 \kappa_g) - 1)\left\|\tilde{W}_g\right\|^2 - (\lambda_{\min}(\mu_5 \kappa_g) - 2)\left\|\tilde{W}_g\right\|^4 + \eta_{oBg}$$

with the bound $\begin{aligned}\eta_{oBg} = &\tfrac{1}{2}\|\mu_3\|^2 \left\|\alpha_g\right\|^2 N_{og} + \tfrac{1}{8}\|\mu_5\|^4 \left\|\alpha_g\right\|^4 N_{og}^2 \\ &+\tfrac{1}{2}\|\mu_3\|^2 \left\|\kappa_g\right\|^2 W_{gM}^2 + \tfrac{1}{8}\|\mu_5\|^4 \left\|\kappa_g\right\|^4 W_{gM}^4.\end{aligned}$ Combining all the terms Lyapunov time derivative terms to obtain the first derivative of the Lyapunov function as

$$\begin{aligned}\dot{L}_I \leq &-(\lambda_{\min}(\mu_1 \bar{\Pi}) - \tfrac{7}{2})\left\|\tilde{x}^T\right\|^2 - (\lambda_{\min}(\mu_2 \kappa_f) - 1)\left\|\tilde{W}_f\right\|^2 \\ &-(\lambda_{\min}(\mu_4 \kappa_f) - \tfrac{7}{4})\left\|\tilde{W}_f\right\|^4 - (\lambda_{\min}(\mu_3 \kappa_g) - 1)\left\|\tilde{W}_g\right\|^2 \\ &-(\lambda_{\min}(\mu_5 \kappa_g) - \tfrac{7}{4})\left\|\tilde{W}_g\right\|^4 + \eta_{oB}\end{aligned}$$

with $\eta_{oB} = \eta_{oBg} + \eta_{ofB} + \eta_{o\tilde{x}B}$. This reveals that the identification and the NN weight estimation errors of the identifiers at each subsystem is locally ultimately bounded.

Proof of Corollary 1:

Similar to the proof of Theorem 2, we can consider two cases. First, consider the inter-event triggering interval. Let the Lyapunov candidate function $L_c = L_I + L$ where the individual terms are defined as

$$L_I(\tilde{x}, \tilde{W}_f, \tilde{W}_g) = L_{\tilde{x}} + L_{\tilde{f}} + L_{\tilde{g}},\ \ L(x, \tilde{W}) = L_x(t) + L_W(t).$$

Consider the second term. Taking the Lyapunov time-derivative and using the result from Theorem 2, we have

$$\dot{L}_x(t) = -Q(x) - u^{*T}Ru^{*} + \sigma^2\eta^{*T}\eta^{*} - 2u^{*T}R\tilde{u}$$

where $\begin{array}{l} \tilde{u} = -\frac{1}{2}R^{-1}g(\breve{x})\nabla\varepsilon(\breve{x}) - \frac{1}{2}R^{-1}g(\breve{x})\nabla\phi^{T}(\breve{x})\tilde{W} \\ -\frac{1}{2}R^{-1}\tilde{g}(\breve{x})\nabla\phi^{T}(\breve{x})W + \frac{1}{2}R^{-1}\tilde{g}(\breve{x})\nabla\phi^{T}(\breve{x})\tilde{W} \end{array}$ and $\tilde{g} = g - \hat{g}$. Similar to the simplification procedure of the proof of Theorem 2, using the YoungâĂŹs inequality, we get

$$\begin{array}{l} \dot{L}(t) \leq -(\bar{\delta}_x - L_u^2)\|x\|^2 - (\frac{\alpha}{\rho} - \frac{g_M^2}{2} - \frac{1}{2})\left\|\tilde{W}\Delta\phi(\tau)\right\|^2 \\ +\frac{1}{2}g_M^2\nabla\varepsilon_M^2 + \alpha^2\varepsilon_M^2 + \frac{1}{2}\left\|\tilde{W}_{\tilde{g}}\right\|^4 + \frac{1}{2}B(g_M^2\nabla\varepsilon_M^2) \end{array}$$

where B is a bounded term which is a function of the reconstruction error due to the NN approximation of the optimal value function and the nonlinear function $g(x)$. Now consider the first term Lyapunov function term corresponding to the NN identifier, using the Lyapunov derivative and using the simplification procedure similar to the proof of Lemma 2, we get the time-derivative $\dot{L}_I$. Combining the derivatives of the Lyapunov function corresponding to the system, identifiers, we get

$$\begin{array}{l} \dot{L}_I \leq -(\bar{\delta}_x - L_u^2)\|x\|^2 - (\frac{\alpha}{\rho} - \frac{g_M^2}{2} - \frac{1}{2})\left\|\tilde{W}\Delta\phi(\tau)\right\|^2 \\ -(\lambda_{\min}(\mu_1\bar{\Pi}) - \frac{7}{2})\left\|\tilde{x}^T\right\|^2 - (\lambda_{\min}(\mu_2\kappa_f) - 1)\left\|\tilde{W}_f\right\|^2 \\ -(\lambda_{\min}(\mu_4\kappa_f) - \frac{7}{4})\left\|\tilde{W}_f\right\|^4 - (\lambda_{\min}(\mu_3\kappa_g) - 1)\left\|\tilde{W}_g\right\|^2 \\ -(\lambda_{\min}(\mu_5\kappa_g) - \frac{9}{4})\left\|\tilde{W}_g\right\|^4 + \eta_{oB} + \frac{1}{2}B(g_M^2\nabla\varepsilon_M^2) + \alpha^2\varepsilon_M^2. \end{array}$$

Finally, similar arguments at the end of Theorem 2 can be used to demonstrate that the number of event-triggered sampling instants is countable and hence the closed-loop system which includes the NN identifier and the NN value function approximator are locally ultimately bounded for all time.

SECTION

2. CONCLUSIONS AND FUTURE WORK

In this dissertation, event-sampled stochastic Q-learning and adaptive dynamic programming techniques are developed for adaptive near optimal distributed control of linear and a class of nonlinear interconnected systems. The event sampled approximation is used to estimate the system dynamics, value function and optimal control input for uncertain interconnected systems. The event sampled approximation property of the neural network (NN) is revisited and the restrictive coupling between frequency of events and convergence of learning scheme is relaxed by introducing a hybrid learning scheme. The aperiodic transmission and controller execution instants are determined by designing novel event sampling conditions which optimize the frequency of feedback instants and system performance. The event sampling conditions orchestrated the sampling and transmission instants to achieve the accuracy in estimation/approximation and control performance with effective resource utilization.

2.1. CONCLUSIONS

In the first paper, an event driven hybrid Q-learning technique was developed to design the optimal control policies. The designed event sampled optimal adaptive control policies were able to regulate the system states with a reduced number of controller executions. Instead of generating more events to facilitate learning, the hybrid learning scheme was able to accelerate the learning and improved the estimate of the optimal value function during the learning phase which in turn improved the system performance.

On the other hand, for the case of nonlinear interconnected systems, in Paper II, the event sampled NN based approximation and hybrid weight update scheme approximated the unknown optimal value functions with a small bounded error. These results are validated with the numerical examples. The introduction of distributed observers to relax the need for internal state measurements increases the computation when compared to state feedback schemes whereas it is found to be more practical. However, the hybrid learning scheme performed consistently better that the traditional TD learning. Further, it was observed that the change in the NN weight initialization and learning gains for the weight update schemes affect the number of controller update.

An event sampled near optimal adaptive regulator was proposed in Paper III for uncertain nonlinear interconnected systems. The reinforcement learning frame work, to solve the infinite horizon distributed optimal control problem in a forward in time manner, is redesigned with event-sampled feedback information; leading to an event-driven hybrid adaptive dynamic programming. Near optimality was achieved with complete unknown system dynamics. The novel distributed NN identifier structure proposed to approximate the system dynamics with intermittent update at the event sampled instants performed satisfactorily. The aperiodic update scheme at the event sampled instants determined by the adaptive event sampling condition drove the NN weight estimation errors within a small bound. A novel event-sampled learning scheme was also developed to overcome the drawbacks in the hybrid learning. To take advantage of the available inter-event time, an exploration strategy using NN identifiers was proposed and it minimized the cost further, off-setting the effects of initial NN weights at the expense of additional computations.

The fourth paper proposed a novel approach for simultaneously optimizing both the event-triggering sampling instants and control policy using zero-sum game formulation. The proposed design scheme provides a tractable trade-off between the frequency of events and the system performance cost by utilizing the min-max optimization of the cost function. The inter-event time interval increases with the proposed event-triggering condition which considerably reduces the communication cost when compared to the traditional event-triggering schemes. This inter-event time expression can be used not only to find the successive event-triggering instants but also to develop an optimal self-triggering control scheme. The model-free hybrid Q-learning scheme extends these results when the linear system dynamics are uncertain. Finally, the decentralized event-triggering condition enables distributed control of interconnected systems confirming the generic nature of the proposed design.

The fifth paper proposed a novel approach for simultaneously optimizing both the event-triggering sampling instants and state feedback controller using zero-sum game formulation for a class of nonlinear system. The inter-event time interval increases with the proposed event-triggering condition which considerably reduces the communication cost when compared to the traditional event-triggering schemes. The approximation based NN-learning scheme generates the approximate optimal control policy and the event-triggering condition forward-in-time by obviating the curse-of-dimensionality. The NN identifiers relaxed the requirement of the accurate knowledge of the system dynamics to implement the proposed scheme.

2.2. FUTURE WORK

As part of the future work, the controllers proposed for the interconnected systems can be made immune to the cyber-attacks. This needs a redefinition of the performance index by taking into account the effects of attack inputs which is not considered yet. This will increase the reliability and resilience of the networked control systems. Extending the ideas of distributed model-free control to develop controllers for large-scale networks could be an area of future research. Since many practical systems can be modelled as complex-networks composed of component subsystems, developing controllers for such complex systems is a challenging problem.

Exploration in the online reinforcement learning framework offers several challenges, especially in designing exploration policy for Markov decision processes with higher dimensional state, action space. Further, the deep NN architecture for learning and control of complex large-scale systems is a potential future direction of research. Due to the approximation accuracy that can be achieved by the deep NNs, they offer a huge scope for improvement in decision making processes involving complex tasks and in optimal control of large-scale systems.

BIBLIOGRAPHY

[1] Jamshidi, M.: 'Large-scale systems: modeling, control, and fuzzy logic' (Prentice-Hall, Inc., 1996).

[2] Azwirman Gusrialdi. *Performance-oriented distributed control design for interconnected systems*. PhD thesis, Universitätsbibliothek der TU München, 2012.

[3] Anuradha M Annaswamy, Ahmad R Malekpour, and Stefanos Baros. Emerging research topics in control for smart infrastructures. *Annual Reviews in Control*, 42:259–270, 2016.

[4] Christopher John Cornish Hellaby Watkins. *Learning from delayed rewards*. PhD thesis, University of Cambridge England, 1989.

[5] Danil V Prokhorov and Donald C Wunsch. Adaptive critic designs. *IEEE transactions on Neural Networks*, 8(5):997–1007, 1997.

[6] Peter C Frumhoff and Jayne Baker. A genetic component to division of labour within honey bee colonies. *Nature*, 333(6171):358–361, 1988.

[7] Florian T Muijres and Michael H Dickinson. Bird flight: fly with a little flap from your friends. *Nature*, 505(7483):295–296, 2014.

[8] James F Kennedy, James Kennedy, Russell C Eberhart, and Yuhui Shi. *Swarm intelligence*. Morgan Kaufmann, 2001.

[9] Jeremy Hsu. Ibm's new brain [news]. *IEEE Spectrum*, 51(10):17–19, 2014.

[10] Dragoslav D Šiljak. *Large-scale dynamic systems: stability and structure*, volume 2. North Holland, 1978.

[11] Lubomir Bakule. Decentralized control: An overview. *Annual reviews in control*, 32(1):87–98, 2008.

[12] Gianluca Antonelli. Interconnected dynamic systems: An overview on distributed control. *IEEE Control Systems*, 33(1):76–88, 2013.

[13] Lucian Busoniu, Robert Babuska, and Bart De Schutter. A comprehensive survey of multiagent reinforcement learning. *IEEE Transactions on Systems Man and Cybernetics Part C Applications and Reviews*, 38(2):156, 2008.

[14] Karl Johan Astrom and Bo M Bernhardsson. Comparison of riemann and lebesgue sampling for first order stochastic systems. In *Decision and Control, 2002, Proceedings of the 41st IEEE Conference on*, volume 2, pages 2011–2016. IEEE, 2002.

[15] Dmitris Hristu-Varsakelis and Panganamala R Kumar. Interrupt-based feedback control over a shared communication medium. In *Decision and Control, 2002, Proceedings of the 41st IEEE Conference on*, volume 3, pages 3223–3228. IEEE, 2002.

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

[17] Manuel Mazo and Paulo Tabuada. Decentralized event-triggered control over wireless sensor/actuator networks. *IEEE Transactions on Automatic Control*, 56(10):2456–2461, 2011.

[18] Randy Cogill. Event-based control using quadratic approximate value functions. In *Decision and Control, 2009 held jointly with the 2009 28th Chinese Control Conference. CDC/CCC 2009. Proceedings of the 48th IEEE Conference on*, pages 5883–5888. IEEE, 2009.

[19] Adam Molin and Sandra Hirche. On the optimality of certainty equivalence for event-triggered control systems. *IEEE Transactions on Automatic Control*, 58(2):470–474, 2013.

[20] Orhan C Imer and Tamer Basar. To measure or to control: optimal control with scheduled measurements and controls. In *American Control Conference, 2006*, pages 6–pp. IEEE, 2006.

[21] Eloy Garcia and Panos J Antsaklis. Model-based event-triggered control for systems with quantization and time-varying network delays. *IEEE Transactions on Automatic Control*, 58(2):422–434, 2013.

[22] Avimanyu Sahoo. *Event sampled optimal adaptive regulation of linear and a class of nonlinear systems*. PhD thesis, Missouri University of Science and Technology, 2015.

[23] Pavankumar Tallapragada and Nikhil Chopra. On event triggered tracking for nonlinear systems. *IEEE Transactions on Automatic Control*, 58(9):2343–2348, 2013.

[24] WPMH Heemels and MCF Donkers. Model-based periodic event-triggered control for linear systems. *Automatica*, 49(3):698–711, 2013.

[25] Frank L Lewis, Draguna Vrabie, and Kyriakos G Vamvoudakis. Reinforcement learning and feedback control: Using natural decision methods to design optimal adaptive controllers. *IEEE Control Systems*, 32(6):76–105, 2012.

[26] Zheng Chen and Sarangapani Jagannathan. Generalized hamilton–jacobi–bellman formulation-based neural network control of affine nonlinear discrete-time systems. *IEEE Transactions on Neural Networks*, 19(1):90–106, 2008.

[27] Xiaoli Wang, Yiguang Hong, Jie Huang, and Zhong-Ping Jiang. A distributed control approach to a robust output regulation problem for multi-agent linear systems. *IEEE Transactions on Automatic control*, 55(12):2891–2895, 2010.

[28] Kumpati S Narendra and Snehasis Mukhopadhyay. To communicate or not to communicate: A decision-theoretic approach to decentralized adaptive control. In *American Control Conference (ACC), 2010*, pages 6369–6376. IEEE, 2010.

[29] Dimitri P Bertsekas, Dimitri P Bertsekas, Dimitri P Bertsekas, and Dimitri P Bertsekas. *Dynamic programming and optimal control*, volume 1. Athena Scientific Belmont, MA, 1995.

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

[31] 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.

VITA

Vignesh Narayanan (S'14) received the B.Tech. degree in electrical and electronics engineering from SASTRA University, Thanjavur, India, and M.Tech. degree in electrical engineering from the National Institute of Technology, Kurukshetra, India, in 2012 and 2014, respectively. He received his Ph.D. degree in electrical engineering from Missouri University of Science and Technology, Rolla, MO, USA, in July 2017.