Fall 2012

Lyapunov based optimal control of a class of nonlinear systems

Hassan Zargarzadeh

Department: Electrical and Computer Engineering

LYAPUNOV BASED OPTIMAL CONTROL OF A CLASS OF NONLINEAR SYSTEMS

by

HASSAN ZARGARZADEH

A DISSERTATION

Presented to the Faculty of the Graduate School of the

MISSOURI UNIVERSITY OF SCIENCE AND TECHNOLOGY

In Partial Fulfillment of the Requirements for the Degree

DOCTOR OF PHILOSOPHY

in

ELECTRICAL ENGINEERING

2012

Approved

Jagannathan Sarangapani, Advisor
J. A. Drallmeier
S. N. Balakrishnan
Cihan H. Dagli
Kelvin T. Erickson

PUBLICATION DISSERTATION OPTION

This dissertation consist the following five articles that have been submitted or published as follows:

Paper I, pages 17-57, H. Zargarzadeh, S. Jagannathan and J. Drallmeier, "Robust Optimal Control of Uncertain Nonaffine Multi-Input and Multi-Output Nonlinear Discrete-time Systems with Application to HCCI Engines", Published in, Int. J. of Adaptive Control and Signal processing (Invited paper for the special issue).

Paper II, pages 58-94, H. Zargarzadeh, S. Jagannathan, "A Discrete-Time Extremum Seeking Method with Application to Efficiency Optimization of HCCI Engines," To be submitted, *IEEE Transaction on Control System Technology*.

Paper III, pages 95-135, H. Zargarzadeh, T. Dierks, S. Jagannathan, "Adaptive Neural Network-based Optimal Control of Nonlinear Continuous-time Systems in Strict Feedback Form," Under review, Int. J. of Adaptive Control and Signal Processing (invited paper for the special issue).

Paper IV, pages 136-176, D. Nodland, H. Zargarzadeh, S. Jagannathan, "Neural Network-based Optimal Adaptive Output Feedback Control of a Helicopter UAV", Under review, IEEE Transaction on Neural Networks.

Paper V, pages 177-238, H. Zargarzadeh, S. Jagannathan, "Optimal Adaptive Control of Nonlinear Continuous-time Systems in Strict Feedback Form", Under review, IEEE Transaction on Neural Networks.

Other Articles:

David Nodland, H. Zargarzadeh, A. Ghosh, and S. Jagannathan, Neuro-optimal control of an unmanned helicopter", Journal of Defense Modeling and Simulation, Accepted for publication, August 2011. (invited paper)

ABSTRACT

Optimal control of nonlinear systems is in fact difficult since it requires the solution to the Hamilton-Jacobi-Bellman (HJB) equation which has no closed-form solution. In contrast to offline and/or online iterative schemes for optimal control, this dissertation in the form of five papers focuses on the design of iteration free, online optimal adaptive controllers for nonlinear discrete and continuous-time systems whose dynamics are completely or partially unknown even when the states not measurable.

Thus, in Paper I, motivated by homogeneous charge compression ignition (HCCI) engine dynamics, a neural network-based infinite horizon robust optimal controller is introduced for uncertain nonaffine nonlinear discrete-time systems. First, the nonaffine system is transformed into an affine-like representation while the resulting higher order terms are mitigated by using a robust term. The optimal adaptive controller for the affine-like system solves HJB equation and identifies the system dynamics provided a target set point is given. Since it is difficult to define the set point a priori in Paper II, an extremum seeking control loop is designed while maximizing an uncertain output function.

On the other hand, Paper III focuses on the infinite horizon online optimal tracking control of known nonlinear continuous-time systems in strict feedback form by using state and output feedback by relaxing the initial admissible controller requirement. Paper IV applies the optimal controller from Paper III to an underactuated helicopter attitude and position tracking problem. In Paper V, the optimal control of nonlinear continuous-time systems in strict feedback form from Paper III is revisited by using state and output feedback when the internal dynamics are unknown. Closed-loop stability is demonstrated for all the controller designs developed in this dissertation by using Lyapunov analysis.

ACKNOWLEDGMENTS

I would like to thank my mentor, Prof. Jagannathan Sarangapani, for his guidance, generous support, and patience over the last three years. I also would like to thank Dr. Drallmeier for his cooperation constructive suggestion over the dissertation. I also would like to thank Dr. Cihan Dagli, Dr. Kelvin Erickson, and Dr. S. N. Balakrishnan for serving on my doctoral committee. In addition, I would like to thank the National Science Foundation for providing financial support.

I also thank my wife, Simin, my parents, Tooba and Mirza-Ali, as well as the rest of my family for their love and support. Their encouragement and sympathety along with my father's financial support has truly been a blessing. Additionally, I have several golleagues and friends who have played vital rules in my educational progress and family matters. Among those I should name and specially thank Dr. Mehdi Ferdowsi, Dr. Shahab Mehraeen, Dr. Behdis Eslamnoor, Reza Ahmadi, and Hasan Ferdowsi.

TABLE OF CONTENTS

SECTION

LIST OF ILLUSTRATIONS

PAPER V

LIST OF TABLES

1. INTRODUCTION

1.1.BACKGROUND

Optimal control of nonlinear continuous and discrete-time systems is a subject of research for the past couple of decades [1][2]. Unlike Riccati-based solution to the linear systems, optimal control of nonlinear systems is a challenging problem since it involves the solution to the Hamilton- Jacobi-Bellman (HJB) equation, which does not have a closed-form solution [1][3].

Several methods are introduced for the problem of nonlinear optimal control that can be categorized as offline [4] and online [5]. In the offline schemes, the controller is tuned a priori whereas the online approaches try to approximate the value function by using the Bellman equation while simultaneously guaranteeing the stability of the closed-loop system. This dissertation aims to establish novel online optimal adaptive schemes for certain classes of nonlinear discrete-time and continuous-time systems whose dynamics are completely or partially unknown.

In the literature, numerous methodologies are employed to find a control scheme that minimizes the Bellman error. For the first time, the idea of searching a compact set for the best possible trajectory and the corresponding optimal controller was introduced in [2]. In [6], it is shown that a nonlinear controller is able to optimize a particular cost function that yields an inverse optimal control. Model predictive control is a different method to obtain finite-horizon optimal control [7]. Another approach that extends the results of linear optimal control theory to nonlinear systems is the state dependent Riccati equation (SDRE) [13]. However, the SDRE yields a sub-optimal solution.

By contrast, adaptive dynamic programming (ADP) is an approach where the solution to the HJB equation is found for generating the optimal control input in an approximate manner. These infinite horizon ADP schemes are based on either policy or value iteration. These schemes form the core of a methodology known by various names, such as approximate dynamic programming, neuro-dynamic programming, or reinforcement learning [8]. However the common theme among these methods is the iterative methodology for generating the policy or value function.

In the policy iteration method, given an initial admissible control input and within a sampling interval, the objective is to iterate the policy until a solution is found that minimizes the cost function [10]. By contrast, in the value iteration-based schemes, an initial admissible control input is not needed [9]. On the other hand, Q-learning is a reinforcement learning technique that works by learning an action-value function that gives the expected utility of taking a given action in a given state and following a fixed policy thereafter. One of the advantages of Q-learning is that it is able to compare the expected utility of the available actions without requiring a model of the environment [11].

A class of reinforcement learning methods is based on the actor-critic structure [12], where an actor component applies an action or control policy while a critic component assesses the value of that action. Based on this assessment of the value, various schemes may be used to modify or improve the action in the sense that the new policy yields a value that is improved over the previous one.

All of the above ADP methodologies are based on an iterative solution of either the policy or the value function. In the iteration based ADP schemes, it is normally

assumed that sufficient number of iterations can be executed within a sampling interval for the purpose of the convergence of the value function or policy. An insufficient number of iterations within a sampling interval will result in the instability of these schemes. Therefore, recently a new ADP framework is introduced without using an iterative approach for affine nonlinear discrete-time system [14] and later extended in [15] to affine nonlinear continuous-time systems.

Beside online optimal control of systems, online optimization of them is also an interesting research topic [16]. In this case, instead of minimizing a cost function along the system trajectories (in optimal control systems), it is desired to maximize or minimize a performance function that potentially can impose some state constraints [17]. To this end, extremum seeking was introduced several decades ago and it is able to find the extremum of a unknown performance function in an adaptive manner [16]. The proof of the stability has been shown in the literature recently for different type of systems including linear, nonlinear, discrete, and continuous-time systems [18][19]. Nonetheless, there are some remaining systems on which the stability of the extremum seeking method is not yet examined.

The next subsection presents an overview of the available online optimal approaches in the literature since the current dissertation establishes novel approaches in the field of online optimal adaptive control of nonlinear systems.

1.2. OVERVIEW OF ONLINE OPTIMAL CONTROL METODOLOGIES

It is well-known that for finding an optimal control scheme for nonlinear systems, a solution to the HJB equation based on the boundary conditions is desirable. This requires solving differential equations of the HJB equation by using the system dynamics

backward-in-time [1][3] which is very difficult to do so. For the case of linear systems, the HJB equation becomes the Riccati equation (RE) that is relatively much easier to solve. Instead of the Riccati equation, the algebraic Riccati equation (ARE) determines the solution to the infinite-horizon time-invariant linear quadratic regulator (LQR) as well as that of the infinite horizon time-invariant linear quadratic Gaussian control (LQG). Compared with the solution of the RE backward in time, the solution of the ARE is less time consuming and easier to solve in a forward-in-time manner, although it offers a suboptimal solution [3]. In addition, the ARE cannot be used for linear systems when the system dynamics are uncertain. Nonetheless, solving either the RE or HJB equation in real-time is still a very difficult problem to the control researchers.

The suboptimal control of nonlinear systems can be achieved by assuming that the nonlinear dynamics has a linear state dependent representation given by $\dot{x} = A(x)x + B(x)u$ where x is the state vector and u is the control input [26]. Then one can solve the state-dependent RE (SDRE). The iteration-based solution to the SDRE is shown to converge [27] and the existence of the solution is studied in [28].

In contrast to dynamic programming approaches that tend to solve the RE or HJB equation backward in time, reinforcement learning schemes try to find forward in time optimal solutions without needing the system dynamics [28]. Online policy iteration is a technique that uses an adaptive (or neural network) estimate of the value function and then iterates the applied policy on the system until it converges to a policy that minimizes the Bellman error.

By using an initial admissible controller, policy iteration schemes evaluate the cost function by using the current value function and updates the value function until it

minimizes the cost function, then improves the policy [10]. By contrast, online value iteration schemes do not require an initial admissible controller. For approximating the policy or value functions, onlinear approximators such as neural networks (NN) can be used. First for a given control policy, a least squares solution is obtained as the process of identifying the NN weights of the value function by iteration, and eventually the best solution is applied to generate the policy [9]. Both policy and value iteration based schemes yield optimal adaptive controllers. Moreover, such policy or value iteration based schemes can be implemented at two different time-scales since the control action in an inner loop occurs once a sampling interval whereas the performance is evaluated in an outer loop over a longer horizon, corresponding to the convergence time needed for the least square computation.

In value function learning, one requires knowledge of the system dynamics. At a minimum, the input coupling matrix $g(x)$ for nonlinear affine system ($\dot{x} = f(x) + g(x)u$) or input matrix B in the case of linear system ($\dot{x} = Ax + Bu$) is required. To avoid needing the system dynamics, Q-learning [29] based scheme can be applied. In this approach, by learning a quality function based on the input-output information of the system, the need for the system dynamics is relaxed. To identify the quality function, iterative-based schemes are also proposed [30].

As mentioned above, all the iterative optimal schemes have an inner and outer loop. The inner loop is a fast loop that yields the policy, value, or Q function while the outer loop by using the information of the inner loop stabilizes the closed loop system. These schemes have to be implemented as a two-time scale problem. In fact, the stability

of the iterative framework relies on the assumption that sufficient number of iterations can be performed within an sampling interval in the outer loop.

In summary, original iteration-based online optimal methods are intended to solve the HJB equation given the knowledge of the system dynamics. When the system dynamics are uncertain, it is impossible to find the solution to the HJB equation. However, an alternative approach is to solve the value function directly as an unknown function in the HJB equation by using adaptive methods. This way, while the controller is generating a stabilizing control law it also adapts itself such that the applied control law solves the HJB equation simultaneously. While these approximate and iterative schemes generate optimal control in the forward in time manner, a large of iterations is needed for convergence of the approximate value function which is a major drawback.

In [14], for discrete-time nonlinear systems, an online approximator is proposed to solve the HJB equation without policy iteration while the overall stability of the system is guaranteed. The same scheme is proposed for continuous time systems in [15] where a single online approximator (SOLA) is used to minimize the cost function estimation error while guaranteeing the overall stability of the closed loop system.

Since the main theme of the work [14][15] is based on standard adaptive methods with iteration-free update laws, they can be easily developed to the cases where the system dynamics is unknown. This has been the motivation of the current dissertation.

1.3.OVERVIEW OF EXTREMUM SEEKING OF NONLINEAR SYSTEMS

In some control problems it is difficult to determine the best operating point that satisfies a predefined set of constraints and/or optimizes a performance function. In fact, stabilization of a system is not sufficient for a closed loop system when the performance

of the system varies in different operating points. Therefore, after stabilization, it might be desirable to have a method that is able to find an operating point that guarantees the best performance of the system.

In a wide class of control problems, the operating point that optimizes the plant performance is unknown and requires to be found. On the other hand, the extremum point of the performance function may be uncertain due to the uncertainty in the plant parameters. Therefore, in the literature, self-optimization, extremum control, or extremum seeking approaches [16] are utilized for this purpose.

Extremum seeking approach has been coined in 1922, a few decades before the introduction of linear adaptive methods. Since extremum seeking methods are adaptive against the performance function uncertainties, authors [18] tend to introduce them as the first adaptive control methods reported in the literature. This method has been widely applied to engineering systems as photovoltaic systems [20], soft landing of electromagnetic actuators [21], and PID controller tuning [22].

Extremum seeking can be applied to HCCI engines control system where it is required to maximize an efficiency function and keep the pressure rise rate (PRR) of the cylinder constrained. Since the engine dynamics are defined in a discrete-time manner [23], this dissertation is motivated to focus on extremum seeking stability of nonlinear discrete-time systems. The author in [19], considered a nonlinear plant represented as a cascade combination of linear dynamics and a static nonlinearity. Since this representation is not the case of engine dynamics representation, in this dissertation, we consider a nonlinear dynamical system with a nonlinear state-to-output performance mapping.

1.4.ORGANIZATION OF THE DISSERTATION

In this dissertation, novel online optimal adaptive schemes are developed to control nonlinear discrete/continuous systems in an optimal manner while the knowledge of the system dynamics is fully or partially unknown. Moreover, due to application being the HCCI engines, a novel extremum seeking method is developed that is able to drive an optimally discrete-time nonlinear stabilized system to its optimized operating point.

This dissertation is presented in the form of five papers and their relationship to one another is illustrated in the Figure 1.1 The common theme in the four of five papers is optimal adaptive control of nonlinear systems in affine or strict-feedback form, whose dynamics are not necessarily known while a persistent excitation condition is needed in order to learn the unknown cost function. The second chapter's theme as mentioned above, is based on optimization of the system performance by seeking the best operating point. This is necessary for many optimal adaptive control problems when the operating point is not defined a priori.

The first paper extends the iteration-free optimal adaptive control of affine work in [14] to solve the online optimal problem for the nonaffine nonlinear discrete time systems in input-output form with completely unknown system dynamics. First the nonaffine nonlinear system is transformed to an affine-like equivalent system in input-output form with higher order terms. The NN identifier identifies the system dynamics of the affine nonlinear discrete-time system. The optimal control scheme subsequently provides an online optimal control law for the affine part of the system provided an initial admissible controller is given. In addition, a robust term is employed to mitigate the higher order terms. Finally, the control scheme is applied to a homogeneous charge compression ignition (HCCI) engine dynamics whose dynamics are represented as a

nonaffine nonlinear discrete-time system. The dynamics of the HCCI engine changes with the fuel type. Lyapunov based uniformly ultimately bounded (UUB) stability of the overall closed-loop system is demonstrated.

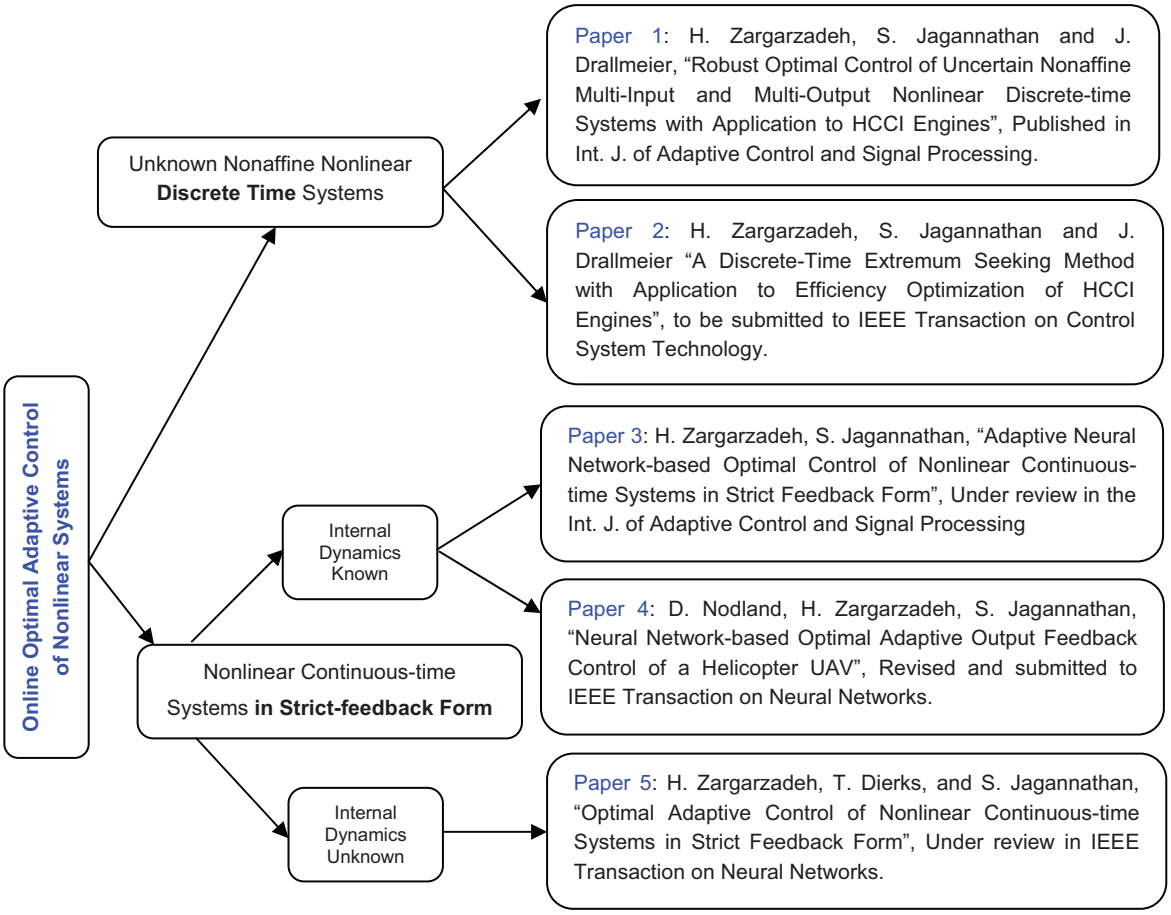

Fig 1.1 Dissertation outline

By having the results of Paper I, we are able to stabilize an unknown nonafine system in optimal manner on any feasible operating point in an optimal manner. Nonetheless, this is not enough in several applications including the HCCI engine case where we require the engine to works in the most efficient operating point. With this motivation, Paper II takes the closed-loop dynamics from the first paper adds an outer loop that is able to find the extremum of a predefined performance function. In this

fashion, the closed loop scheme takes a multiloop representation whose inner loop is the optimal adaptive NN-based control and the outer loop is the extremum seeking scheme that generates the setpoint for the inner loop. Since such systems representations render a singularly perturbed dynamics, the traditional proof of the stability is done in two steps. The first step is to show the stability of the outer loop using averaging analysis [24] with the assumption that the inner loop is fast enough to follow any desired set point with no delay. With the assumption that the state-to-output map has a unique extremum, we show that the proposed method is able to locally converge to the extremum point. In the second step, the stability of the overall dynamical system is examined using singular perturbation method for discrete-time nonlinear systems [25]. It is shown that the overall system is UUB stable in a neighborhood that can be arbitrarly small by choosing a proper extremum seeking parameters and large enough number of NN basis function vector.

On the other hand, Paper III extends the work of [15] where an online optimal adaptive scheme is established to control multi-input and multi-output (MIMO) nonlinear continuous-time systems in strict feedback form with known dynamics, and without using policy/value iterations and initial admissible controller. Here, it was shown that the tracking problem of MIMO strict feedback systems can be solved as the optimal stabilization of the corresponding error dynamics if a proper feedforward term can be designed. Subsequently, state feedback control scheme is developed for the affine nonlinear continuous-time system that is expressed with tracking error. Next, the state feedback-based optimal adaptive control scheme is extended by using output feedback. Lyapunov analysis is utilized to demonstrate the UUB stability of the overall closed loop system.

Paper IV is an application of Paper III to an unmanned helicopter with underactuated dynamics in order to optimally track a desired position and orientation. Because of underactuated nonlinear dynamics, high-performance controller design for unmanned helicopters is a challenging problem. This paper introduces a NN based optimal controller by using output feedback for trajectory tracking with out using an initial admissible controller but considering the dynamics are known. The output-feedback control system employs the backstepping methodology, using kinematic and dynamic controllers and a NN observer to generate the tracking control law based on output measurements. The online approximator-based dynamic controller learns the infinite-horizon cost function in continuous time and calculates the corresponding optimal control input in order to stabilize the corresponding error dynamics. A UUB stability is included based on Lyapunov approach.

In Paper V, the internal dynamics of the nonlinear continuous-time systems in strict feedback form are considered unknown and an adaptive scheme is utilized not only to approximate the cost function but also the unknown internal dynamics. The Lyapunov based stability analysis indicates that the tracking error converges to zero using the proposed optimal adaptive controller provided the cost function and the known dynamics are represented as linear in the unknown parameters, The unknown parameters of the cost function and the internal dynamics converge to their true values under a persistency of excitation condition on the input signals. Finally, the results are extended to the output feedback control of strict feedback systems.

1.5.CONTRIBUTION OF THE DISSERTATION

The main objective of this dissertation is to develop a suite of novel optimal adaptive control schemes for a class of nonlinear discrete/continuous-time systems when

the system dynamics are unknown or the system states are not necessarily measurable. Therefore, the proposed optimal adaptive control schemes can not only maintain the overall system stability but also they can adaptively learn the solution of the HJB equation. Once the nonlinear optimal controller is designed the system can be stabilized on any desired setpoint, although we may require the setpoint to be the optimum operating point in some applications. Therefore, motivated by HCCI engine performance optimization problem, an extremum seeking method is developed for nonlinear discrete-time systems with unknown output functions.

Optimal adaptive control for affine discrete-time systems with unknown internal dynamics is previously studied in [14], whereas the current work considers an unknown nonaffine discrete-time system in input output form with uncertain dynamics. Moreover, for the case of continuous-time systems, optimal adaptive control is derived when the internal dynamics are unknown in contrast to [31] where an iterative optimal approach and an initial admissible controller are needed. In contrast, the proposed schemes deal with partially unknown nonlinear continuous-time systems without using an iterative solution.

The contributions of the Paper I include providing a suitable representation of unknown nonaffine systems that can be identified online using a single NN identifier. Then, an online optimal adaptive controller is introduced to control the affine part of the identified system dynamics. Since the bounds of the higher order residual terms of the nonaffine system are unknown, a novel robust auxiliary controller is introduced using singularly perturbation system theory in order to mitigate them and an overall boundedness of the closed-loop system is demonstrated. Finally, the control scheme is

applied to the HCCI engine to show a significant performance comparison with other traditional controllers.

To the best knowledge of the author, the extremum seeking method is not studied in the literature in a generic case. In constrast to [19] where a nonlinear plant represented as a cascade combination of linear dynamics and a static nonlinearity, Paper II proposes a novel averaging methodology to prove the stability of the reduced model [24] of the extremum seeking scheme in a general case. Then, a discrete-time version of the sigularly pertubed system theory from [25] is employed to prove the overall stability.

In Paper III, to the best knowledge of the authors, this was the first time an optimal adaptive control of strict-feedback nonlinear continuous-time systems was considered by using the state and output feedback control. A neural network-based controller is shown to provide a UUB of the closed-loop system while the actual adaptive control input approaches the optimal one. An initial stabilizing controller is not required and value and policy iterations are not employed. Paper IV provides an application of optimal adaptive control of strict feedback control systems to an unmanned aerial vehicle (UAV) helicopter with underactuated dynamics. This paper introduces an optimal controller design via output feedback for trajectory tracking of a helicopter UAV using a NN observer while demonstrating the boundedness of the overall closed-loop system. Simulation results show the effectiveness of the proposed control design for trajectory tracking. The adaptive optimal control of nonlinear strict feedback continuous-time systems is examined by using state and output feedback in Paper V while the internal dynamics is unknown. Here value or policy iterations are not utilized and an initial admissible control is not required.

REFERENCES

[1] D. E. Kirk, Optimal Control: An Introduction, Prentice-Hall, 1970.

[2] Beard R, Saridis G, Wen J. Improving the performance of stabilizing controls for nonlinear systems. IEEE Control Systems Magazine 1996; 16(5):27 – 35.

[3] F.L. Lewis, and V.L. Syrmos, Optimal Control, 2nd Edition, Wiley, New York.

[4] M. Abu-Khalaf, F. L. Lewis, "Nearly optimal control laws for nonlinear systems with saturating actuators using a neural network HJB approach," *Automatica*, vol. 41, no. 5, pp. 779-791, 2005.

[5] K. G. Vamvoudakis, F. L. Lewis, "Online actor_critic algorithm to solve the continuous-time infinite horizon optimal control problem," Automatica, vol. 46, no. 5, pp. 878-888, 2010.

[6] P. Moylan, B. Anderson, "Nonlinear regulator theory and an inverse optimal control problem," *Automatic Control, IEEE Transactions on* , vol.18, no.5, pp. 460- 465, 1973.

[7] L. Grüne, J. Pannek, *Nonlinear Model Predictive Control: Theory and Algorithms*, Springer, 2011.

[8] D. P. Bertesekas, *Dynamic Programming and Optimal Control*, Athena Scientific, c2005-2007.

[9] A. Al-Tamimi, F. L. Lewis, and M. Abu-Khalaf, "Discrete-Time Nonlinear HJB Solution Using Approximate Dynamic Programming: Convergence Proof," *Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on* , vol.38, no.4, pp.943-949, 2008.

[10] G. A. Hewer, "An iterative technique for the computation of the steadystate gains for the discrete optimal regulator," *IEEE Tra*ns. Autom. Control, vol. AC-16, no. 4, pp. 382–384, 1971.

[11] A. L. Strehl, L. Li, E. Wiewiora, J. Langford, and M. L. Littman. "Pac model-free reinforcement learning," *In Proc. 23nd ICML 2006*, pages 881–888, 2006.

[12] A. G. Barto, R. S. Sutton, and C. Anderson, "Neuron-like adaptive elements that can solve difficult learning control problems," *IEEE Trans. Syst., Man, Cybern.*, vol. SMC-13, pp. 834–846, 1983.

[13] J.S Shamma, and J.R. Cloutier , "Existence of SDRE stabilizing feedback," Automatic Control, IEEE Transactions on , vol.48, no.3, pp. 513- 517, March 2003.

[14] T. Dierks, and S. Jagannthan, "Optimal control of affine nonlinear discrete-time systems," *Control and Automation, 2009. MED '09. 17th Mediterranean Conference on* , pp.1390-1395, 24-26 June 2009.

[15] T. Dierks and S. Jagannathan, "Optimal control of affine nonlinear continuous-time systems using an online Hamilton-Jacobi-Isaacs formulation," *Decision and Control (CDC), 2010 49th IEEE Conference on* , pp.3048-3053, 15-17 Dec. 2010.

[16] C. S. Drapper and Y. T. Li, "Principles of optimalizing control systems and an application to the internal combustion engine," ASME, vol. 160, pp. 1-16, 1951.

[17] Li Yaoyu, M. A. Rotea, G.T.-C. Chiu, L. G. Mongeau, In-Su Paek, "Extremum seeking control of a tunable thermoacoustic cooler," Control Systems Technology, IEEE Transactions on , vol.13, no.4, pp. 527- 536, July 2005.

[18] M. Krstic´ and H.-H. Wang, "Stability of extremum seeking feedback for general nonlinear dynamic systems," *Automatica*, vol. 36, pp. 595–601, 2000.

[19] C. J. Young, M. Krstic, and K.B. Ariyur, J. S. Lee, "Extremum seeking control for discrete-time systems," Automatic Control, IEEE Transactions on , vol.47, no.2, pp.318-323, Feb 2002.

[20] R. Leyva, C. Alonso, I. Queinnec, A. Cid-Pastor, D. Lagrange, and L. Martinez-Salamero, "MPPT of photovoltaic systems using extremum—Seeking control," IEEE Trans. Aerosp. Electron. Syst., vol. 42, no. 1, pp. 249–258, Jan. 2006.

[21] K. S. Peterson and A. G. Stefanopoulou, "Extremum seeking control for soft landing of an electromechanical valve actuator," Automatica, vol. 40, no. 6, pp. 1063–1069, 2004.

[22] N. J. Killingsworth and M. Krstic´, "PID tuning using extremum seeking: Online, model-free performance optimization," IEEE Control Syst. Mag., vol. 26, no. 2, pp. 70–79, Feb. 2006.

[23] J. B. Bettis, J. A. Massey, J. A. Drallmeier, S. Jagannathan, "A thermodynamics-based homogeneous charge compression ignition engine model for adaptive nonlinear controller development," Journal of Automobile Engineering, 2012.

[24] H. K. Khalil, Nonlinear Systems, 3rd ed. Upper Saddle River, NJ: Prentice-Hall, 2002.

[25] R. B. Bouyekhf and A. El Moudni, "On analysis of discrete singularly perturbed nonlinear systems: Application to the study of stability properties," J. Franklin Inst., vol. 334B, no. 2, pp. 199–212, 1997.

[26] S. C. Beeler, H. T. Tran and H. T. Banks, "Feedback control methodologies for nonlinear systems," *Journal of Optimization Theory and Application*, vol. 107, no. 1, 2000.

[27] T. Cimen and S. P. Banks, “Global optimal feedback control for general nonlinear systems with nonquadratic performance criteria,” Systems and Control Letters, vol. 53, no. 5, pp. 327-346, 2004.

[28] A. Al-Tamimi and F. L. Lewis, “Discrete-time nonlinear HJB solution using approximate dynamic programming: convergence proof,” *Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on* , vol.38, no.4, pp.943-949, 2008.

[29] C. Watkins, “Learning from delayed rewards,” Ph.D. thesis, Cambridge Univ., Cambridge, England, 1989.

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

[31] D. Vrabie, M. Abu-Khalaf, F.L. Lewis, and Y. Wang; , "Continuous-Time ADP for linear systems with partially unknown dynamics," *Approximate Dynamic Programming and Reinforcement Learning*, 2007. ADPRL 2007.

[32] D. Vrabie and F. L. Lewis, “Neural network approach to continuous-time direct adaptive optimal control for partially unknown nonlinear systems” *Neural Networks*, vol. 22, no. 3, 2009.

PAPER

I. ROBUST OPTIMAL CONTROL OF UNCERTAIN NONAFFINE MULTI-INPUT AND MULTI-OUTPUT NONLINEAR DISCRETE SYSTEMS WITH APPLICATION TO HCCI ENGINES

H. Zargarzadeh, S. Jagannathan, and J. Drallmeier

Abstract — Multi-input and multi-output (MIMO) optimal control of unknown nonaffine nonlinear discrete-time systems is a challenging problem due to the presence of control inputs inside the unknown nonlinearity. In this paper, the nonaffine nonlinear discrete-time system is transformed to an affine-like equivalent nonlinear discrete-time system in the input-output form. Next, a forward-in-time Hamilton-Jacobi-Bellman (HJB) equation-based optimal approach, without using value and policy iterations, is developed to control the affine-like nonlinear discrete-time system by using both neural networks (NN) as an online approximator and output measurements alone. To overcome the need to know the control gain matrix in the optimal controller, a new online discrete-time NN identifier is introduced. The robustness of the overall closed loop system is shown via singularly perturbation analysis by using an additional auxiliary term to mitigate the higher-order terms. Lyapunov stability of the overall system, which includes the online identifier and robust control term, demonstrates that the closed-loop signals are bounded and the approximate control input approaches the optimal control signal with a bounded error. The proposed optimal control approach is applied to a cycle-by-cycle discrete-time representation of an experimentally validated homogeneous charge compression ignition (HCCI) fuel-flexible engine whose dynamics are modeled as

uncertain nonlinear, nonaffine, and MIMO discrete-time system. Simulation results are included to demonstrate the efficacy of the approach in presence of actuator disturbances.

I. INTRODUCTION

Online optimal control of uncertain nonlinear systems is a challenging problem due to the difficulty of solving the Hamilton-Jacobi-Bellman (HJB) equation which does not have a closed-form solution. In addition, controlling a nonaffine nonlinear discrete-time system in general is a major challenge due to the coupled nonlinear relationship between the states and the control input within the unknown nonlinearity. Recently, neural networks (NN) as online approximators have been successfully applied to learn the uncertain nonlinear system dynamics in an online fashion because of their universal function approximation property.

The NN-based optimal control of affine nonlinear systems is now available in the literature either in continuous or discrete-time systems [1]-[4] by using HJB equation in forward-in-time manner via value and policy iterations. While [1] and [1] present offline based schemes, others [3] address optimal control in an online manner for affine nonlinear discrete-time systems.

In [1] and [3] the input gain matrix[1] (IGM) of the affine system is considered known while the internal system dynamics are considered unknown. The work in [5] introduces an adaptive dynamic programming (ADP)-based scheme for optimal control of unknown affine systems. The authors in [1] and [6] deal with online optimal control of affine nonlinear system whose input gain matrix (IGM) is considered known. Here in these works [1] and [6], the cost function is estimated through the HJB equation offline,

---

[1] In a general form of discrete time affine systems i.e. $x_{k+1} = f(x(k)) + g(x(k))u_k$, $f(x(k))$ and $g(x(k))$ are considered as internal dynamics and input gain matrix respectively.

whereas the work in [3] estimates the cost function with an online NN based estimator while proving the overall convergence of the NN based controller. In [6], convergence of the heuristic dynamic programming algorithm (HDP) via value and policy iterations is demonstrated and closed-loop stability is not shown. It is found that an insufficient number of iterations in the value and policy iteration-based optimal control schemes [3,6] will not only cause convergence issues but also instability. Therefore the optimal controller in [3] is developed without using value and policy iterations and closed-loop stability analysis is demonstrated. However, all these methods [1-6] assume that the states of the system are measurable. Unfortunately, in many practical applications, such as the proposed control of HCCI engines, states are not available which necessitates an output feedback based optimal control scheme.

Therefore, this paper addresses forward-in-time based optimal control of unknown nonaffine MIMO discrete-time systems by transforming the nonaffine nonlinear discrete-time system into an affine-like equivalent system in the *input-output* form with higher order terms. The input-output form relaxes the need for state availability. Next, a NN identifier is proposed to learn the unknown IGM matrix online whose estimation is required in the optimal controller design. Next, in order to mitigate the modeling errors due to higher order terms, an auxiliary term is designed via fast dynamic inversion technique. The fast dynamic solver, along with the closed loop system, forms a singularly perturbed system whose stability is shown to be guaranteed. Thus this auxiliary term ensures robustness against modeling errors and reduces the ultimate bounds of the closed-loop system by mitigating the effect of higher order terms.

Subsequently, the forward-in-time approach similar to [3] is introduced to the generic unknown affine-like equivalent system by using output feedback without using value and policy iterations. Here, the value function and the control input are updated once a sampling interval. Using an initial stabilizing control, a NN online approximator (OLA) is tuned to learn the cost function which is subsequently utilized along with the estimated IGM to generate the optimal control input. The nonaffine nature of the system, online identifier and lack of system states complicate the stability analysis whereas the boundedness of all the closed-loop signals and the actual control input to the optimal value are demonstrated. The net result is the output feedback-based robust optimal controller design for nonaffine MIMO nonlinear discrete-time systems.

Finally, the proposed optimal controller is applied to a homogeneous charge compression ignition (HCCI) engine which is a practical example of an unknown nonlinear MIMO discrete-time system with structural uncertainties due to variations in the fuel type or ambient operating conditions [12]. Compared to regular spark ignition (SI) engines, the HCCI engines have the advantage of increased thermal efficiency and low nitrous oxides, NO_x, and particulate matter emissions [8]-[10]. The HCCI engines do not have an ignition system, and managing the combustion appears to be a challenging control problem. In other words, achieving and maintaining HCCI mode of operation in diverse operating situations requires an appropriate *closed loop control* strategy. The control approach should be optimal under a variety of fuel types which in turn imposes a variety of combustion chemical kinetics and that being unknown, necessitates an online learning feature for the controller.

Due to complex engine dynamics [11] and the presence of uncertainties, the HCCI engine dynamics in nonaffine form are transformed into the input-output form for control purposes. Numerical results show that the proposed NN-based robust optimal control scheme can successfully control the engine dynamics and is able to adaptively tune the initial admissible controller to attain an optimal controller online. The paper is organized as follows.

The state space representation of the MIMO system dynamics are given in Section II. Section III introduces a system identification technique while Section IV establishes an overall robust online optimal control approach where the identified information of the system dynamics is used. Here the robust term is utilized first to mitigate the higher order terms that appear as the result of transformation from nonaffine to affine nonlinear discrete-time system. Next it is shown, in this section, that any initial admissible controller can be tuned to an optimal online controller that minimizes a desired performance index. Section V introduces the experimentally validated representation of a HCCI engine and performance of the proposed online optimal controller.

II. STATE SPACE REPRESENTATION OF THE OUTPUT CONTROL OF NONAFFINE SYSTEMS

Consider a generic form of nonaffine system described by

$$x_{k+1} = f(x_k, u_k) \tag{1}$$

$$y_k = g(x_k, u_k) \tag{2}$$

where $u_k \in E_u \subset \mathbb{R}^m$, $x_k \in E_x \subset \mathbb{R}^n$, and $y_k \in E_y \subset \mathbb{R}^\ell$ represent the system input, states, and the outputs respectively, and $f(\bullet)$ and $g(\bullet)$ are assumed to be unknown continuous nonlinear functions with respect to x_k and u_k. The system representation (1)

indicates a multi-input and multi-output (MIMO) nonlinear discrete-time system whose output y_k is controlled through the input u_k. Next, the following assumption is needed before we proceed.

Assumption 1. The output function is a diffeomorphism with respect to a given u_k u_k**.** In other words, $g(x_k, u_k)$ is a one-to-one mapping between y_k and x_k for a given u_k. Then, there exists a one-to-one function g^{-1} such that

$$x_k = g^{-1}(y_k, u_k) \tag{3}$$

Using (1) and (2), we

$$\begin{aligned} &y_{k+1} = g(x_{k+1}, u_{k+1}) = g(f(x_k, u_k), u_{k+1}) = h(x_k, u_k, u_{k+1}) \\ &= h(g^{-1}(y_k, u_k), u_k, u_{k+1}) \end{aligned} \tag{4}$$

In other words, $g(x_k, u_k)$ is a one-to-one mapping between y_k and x_k for a g u_k iven. Then, there exists a one-to-one function g^{-1} such that

$$x_k = g^{-1}(y_k, u_k) \tag{5}$$

Using (1) and (2), we have

$$y_{k+1} = g(x_{k+1}, u_{k+1}) = g(f(x_k, u_k), u_{k+1}) = h(x_k, u_k, u_{k+1}) = h(g^{-1}(y_k, u_k), u_k, u_{k+1}) \tag{6}$$

Now, assume that the overall controller is designed such that the system input increment, denoted by Δu_k, is generated. This implies that the system input and the controller output are connected such that $u_{k+1} = u_k + \Delta u_k$. Using (1) and (3) we can express the system dynamics as

$$y_{k+1} = h(y_k, u_k, \Delta u_k)$$

or

$$X_{k+1} = \left(\left(u_k + \Delta u_k \right)^T \quad h(g^{-1}(y_k, u_k), u_k, u_k + \Delta u_k)^T \right)^T = H(X_k, \Delta u_k) \tag{7}$$

where $X_{k+1} = (u_k \quad y_k)^T \subset \mathbb{R}^{m+\ell}$. By applying the Taylor series expansion, equation (4) can be expanded as [12]

$$\equiv \sum_{i=0}^{q} F_i(X_k) \Delta u_k^i + O(X_k, \Delta u_k), \;\; q > 1$$

$$\begin{aligned} X_{k+1} = H(X_k, \Delta u_k) &= H(X_k) + \frac{\partial}{\partial \Delta u_k} H(X_k, \Delta u_k)\,|_{\Delta u_k = 0}\; \Delta u_k \\ &+ \frac{1}{2} \frac{\partial}{\partial \Delta u_k} \left(\frac{\partial}{\partial \Delta u_k} H(X_k, \Delta u_k) \Delta u_k \right)|_{\Delta u_k = 0}\; \Delta u_k + \cdots \end{aligned} \tag{8}$$

where

$$O(X_k, \Delta u_k)\; F_0(X_k) = H(X_k)\;,\; F_1(X_k) = \frac{\partial}{\partial \Delta u_k} H(X_k, \Delta u_k)\,|_{\Delta u_k = 0}\;, \ldots$$

Moreover, represents the higher order terms of the Taylor series expansion, , and $q \in \mathbb{N}$ denoting the number of terms considered in the control design. Also, it is clear that $F_{i=0}(X_k)$ is denoted as $H(X_k)$ in (8). Moreover, Δu_k^i denotes a vector whose elements include all possible *i*th multiplication of the elements of Δu_k. For example

$$\Delta u_k^2 = \begin{pmatrix} \Delta u_k(1)\Delta u_k(1), \Delta u_k(1)\Delta u_k(2), ..., \Delta u_k(1)\Delta u_k(m), \Delta u_k(2)\Delta u_k(2) \\ , \Delta u_k(2)\Delta u_k(3), ..., \Delta u_k(2)\Delta u_k(m), ..., \Delta u_k(m)\Delta u_k(m) \end{pmatrix}, \tag{9}$$

with $\Delta u_k^2 \in \mathbb{R}^{m \times (m+1)/2}$. Typically, for practical systems, the higher order terms, $O(X_k, \Delta u_k)$, can be considered small and negligible due to uniform convergence of Taylor series when $H(X_k, \Delta u_k)$ is differentiable [12]. In the sequel, we denote $O(X_k, \Delta u_k)$ by O_k for the sake of simplicity. Therefore, the unknown affine-like system representation of (1) takes the following input-output form as

$$X_{k+1} = F(X_k) + G(X_k)\Delta u_k + O_k, \tag{10}$$

where $F(X_k) = F_0(X_k)$ is the internal dynamics and $G(X_k) = F_1(X_k)$ being the input gain matrix (IGM).

In this paper, the robust optimal control scheme for the affine-like system (10) in the input-output form is introduced which requires $G(X_k)$ and the higher order terms O_k to be known while the information on internal dynamics, $F(X_k)$, is not required [3] [5]. In order to overcome the need for the IGM and the higher order terms, a new online NN identifier is introduced in the next section. Subsequently, the robust optimal control scheme is designed by using two terms one for robustness and the second for optimality as presented in Section IV.

III. ONLINE NN-BASED IDENTIFIER

The objective of this section is to introduce a novel NN-based identifier for (5) when the system inputs and outputs are given. The NN-based identifier is novel in the sense that it can directly identify $F(X_k)$, IGM, and O_k by using a single NN and without using system states. There are several approaches for identification of either affine or nonaffine nonlinear systems by using offline methods [1],[7],[13],[22],[23] whereas the proposed identifier works online.

Assume that a stabilizing input is applied to the system(10), the following expression can be used for approximation of the system (10) at the instant *k* in a compact set as

$$X_k = \Psi(X_{k-1}, \Delta u_{k-1})^T W \Delta U_{k-1} + \overline{\varepsilon}_{k-1}. \tag{11}$$

where $\Psi(X_{k-1},\Delta u_{k-1})\in\mathbb{R}^{p\times(m+\ell)}$, $W\in\mathbb{R}^{p\times(m+\ell)}$, $\Delta U_k=(1\quad\cdots\quad 1\quad \Delta u_k^T)^T\in\mathbb{R}^{(m+\ell)}$, $\overline{\varepsilon}_k\in\mathbb{R}^{m+\ell}$, $\left\|\Psi(X_{k-1},\Delta u_{k-1})\right\|\le\Psi_M$, $\left\|\Psi(X_{k-1},\Delta u_{k-1})U_{k-1}\right\|\le\overline{\Psi}_M$ being the bounded NN activation function and $\overline{\varepsilon}_k$ is the estimation error satisfying $\left\|\overline{\varepsilon}_k\right\|<\overline{\varepsilon}_M$ [17]. Moreover, the following bound $U_m\le\left\|\Delta U_k\right\|$ holds due to the presence of constant values in the input vector in fact $U_m=\left\|\Delta U_k\right\|_{\Delta u_k=0}$.

It should be noted that the identification scheme has the advantage of identifying $F(X_k)$, $G(X_k)$, and O_k separately and online. By identifying separately, we mean that the identifier has only one NN, whereas the identification is complete, $F(X_k)$, $G(X_k)$, and O_k can be distinguished without any further algebraic operations. Select $\Psi(X_k,\Delta u_k)^T=\left(\Psi_F(X_k)^T,\Psi_O(X_k,\Delta u_k)^T,\Psi_G(X_k)^T\right)$ and $W=\left(W_F^T,W_O^T,W_G^T\right)^T$ for convenience. Therefore, from (10) and (11) we have

Now, by considering ΔU_k in (10) we have

$$X_{k+1}=W_F^T\Psi_F(X_k)+W_G^T\Psi_G(X_k)\Delta u_k+W_O^T\Psi_O(X_k,\Delta u_k)+\overline{\varepsilon}_k. \tag{12}$$

Our goal next is to identify the NN weight matrix W denoted here as $\hat{W}_k=\left(\hat{W}_F^T,\hat{W}_O^T,\hat{W}_G^T\right)^T$ by estimating the state vector X_k with $\hat{X}_k$ where

$$\hat{X}_k=\Psi(X_{k-1},\Delta u_{k-1})^T\hat{W}\Delta U_{k-1}\equiv\hat{W}_F^T\Psi_F(X_k)+\hat{W}_G^T\Psi_G(X_k)\Delta u_k+\hat{W}_O^T\Psi_O(X_k,\Delta u_k),$$

$$\equiv\hat{F}(X_k)+\hat{G}(X_k)\Delta u_k+\hat{O}(X_k,\Delta u_k) \tag{13}$$

Therefore,

$$X_{k+1}=\hat{F}(X_k)+\hat{G}(X_k)\Delta u_k+\hat{G}(X_k)\Delta\rho_k+\hat{O}_k+\tilde{e}_{k+1}. \tag{14}$$

where the identification error is defined as

$$\tilde{e}_k = X_k - \hat{X}_k = \Psi(X_{k-1}, \Delta u_{k-1})^T (W - \hat{W}_{k-1}) \Delta U_{k-1} + \bar{\varepsilon}_{k-1}$$

$$= \Psi(X_{k-1}, \Delta u_{k-1})^T \tilde{W}_{k-1} \Delta U_{k-1} + \bar{\varepsilon}_{k-1} \tag{15}$$

Define the update law for the actual NN weights $\hat{W}_k$ as

$$\hat{W}_k = \hat{W}_{k-1} + \alpha \tilde{E}_k \Psi(X_{k-1}, \Delta u_{k-1}) \Delta U_{k-1}^T \Big/ \left(\left\| \Delta U_{k-1} \right\|^2 + 1 \right), \tag{16}$$

where

$$\tilde{E}_k = diag(\tilde{e}_k) \in \mathbb{R}^{(m+\ell)\times(m+\ell)}, \tag{17}$$

with $\alpha > 0$ being the design parameter or NN learning rate and $\hat{W}_k = 0$. The error dynamics in weight update law are written as

$$\tilde{W}_k = \tilde{W}_{k-1} - \alpha \tilde{E}_k \Psi(X_{k-1}, \Delta u_{k-1}) \Delta U_{k-1}^T \Big/ \left(\left\| \Delta U_{k-1} \right\|^2 + 1 \right). \tag{18}$$

Now, by considering ΔU_k in (10) we have

$$X_{k+1} = W_F^T \Psi_F(X_k) + W_G^T \Psi_G(X_k) \Delta u_k + W_O^T \Psi_O(X_k, \Delta u_k) + \bar{\varepsilon}_k. \tag{19}$$

It should be noted that if the identification error $\left\| \tilde{e}_k \right\|$ and weight estimation error $\left\| \tilde{W}_k \right\|_F$ converges, the identification of $F(X_k)$, $G(X_k)$, and O_k is complete. Therefore, the following theorem can be stated to show the boundedness of the identification and the NN weight estimation errors. Here we will use Ψ_k instead of $\Psi(X_k, \Delta u_k)$ for the sake of brevity.

Theorem 1: Assume that the proposed identifier in (11) with the update law in (18) is used to identify the system (16) and the design parameter is chosen as $0 < \alpha < 32 U_{\mathrm{m}}^2 \Psi_m^2 \Big/ \left(4\bar{\Psi}_M^2 + \bar{\Psi}_M^2 \Psi_M^2 \right)$. Then the identification error $\tilde{e}_k$ and the NN weight

estimation errors $\tilde{W}_k$ are *uniformly ultimately bounded* (UUB) [7] and converge to a basin of attraction with the bounds given by $\left\|\tilde{W}_k\right\| \le W_B$ or $\left\|\tilde{e}_k\right\| \le e_B$ where

$$e_B \triangleq \bar{\varepsilon}_M \sqrt{1+\frac{\Psi_M^2}{4}+\frac{\left(\Psi_M^2+2\alpha\bar{\Psi}_M^2\right)^2}{\alpha\Psi_M^2\left(8U_{\mathrm{m}}^2\Psi_m^2-\alpha\Psi_M^2U_M^2\left\{4+\Psi_M^2\right\}\right)}} \tag{20}$$

or

$$W_B \triangleq 2\bar{\varepsilon}_M\left(\Psi_M^2+2\alpha\bar{\Psi}_M^2\right)\Big/\left\{8U_{\mathrm{m}}^2\Psi_m^2\Psi_M-\alpha\bar{\Psi}_M^2\left(4\Psi_M+\Psi_M^3\right)\right\}$$

$$+\bar{\varepsilon}_M\sqrt{1+\frac{\Psi_M^2}{4}+\frac{\left(\Psi_M^2+2\alpha\bar{\Psi}_M^2\right)^2}{\alpha\Psi_M^2\left(8U_{\mathrm{m}}^2\Psi_m^2-\alpha\bar{\Psi}_M^2\left(4+\Psi_M^2\right)\right)}}\,. \tag{21}$$

Proof. Define $V_I=\alpha^2\tilde{e}_k^T\tilde{e}_k+tr\left\{\tilde{W}_k^T\tilde{W}_k\right\}$ as a Lyapunov candidate function. The first difference can be written as

$$\Delta V_I=\alpha^2\tilde{e}_{k+1}^T\tilde{e}_{k+1}-\alpha^2\tilde{e}_k^T\tilde{e}_k+tr\left\{\tilde{W}_{k+1}^T\tilde{W}_{k+1}\right\}-tr\left\{\tilde{W}_k^T\tilde{W}_k\right\}$$

$$=-\alpha^2\tilde{e}_k^T\tilde{e}_k+\alpha^2\tilde{e}_{k+1}^T\tilde{e}_{k+1}-2tr\left\{\alpha\Delta U_k\Psi(X_k)^T\tilde{E}_{k+1}\tilde{W}_k\Big/\left(\left\|\Delta U_k\right\|^2+1\right)\right\}$$

$$+tr\{\alpha^2\Delta U_k\Psi(X_k)^T\tilde{E}_{k+1}^T\tilde{E}_{k+1}\Psi(X_k)\Delta U_k^T\}\ . \tag{22}$$

Using the fact that $tr(ABC)=tr(CAB)$, (22) can be simplified as

$$\Delta V_I=-\alpha^2\tilde{e}_k^T\tilde{e}_k+\alpha^2\tilde{e}_{k+1}^T\tilde{e}_{k+1}-2\alpha\,tr\left\{\Psi(X_k)^T\tilde{E}_{k+1}\tilde{W}_k\Delta U_k\right\}\Big/\left(\left\|\Delta U_k\right\|^2+1\right)$$

$$+\alpha^2\left\|\tilde{E}_{k+1}\right\|_F^2\left\|\Psi(X_k)\right\|_F^2\left\|\Delta U_k\right\|^2\Big/\left(\left\|\Delta U_k\right\|^2+1\right)^2$$

From (17) and using the equation (15) we can write

$$\Delta V_I=-\alpha^2\tilde{e}_k^T\tilde{e}_k+\alpha^2\left(\Psi(X_k)^T\tilde{W}_k\Delta U_k+\bar{\varepsilon}_k\right)^T\times$$

$$\left(\Psi(X_k)^T\tilde{W}_k\Delta U_k+O_k+\bar{\varepsilon}_k\right)\left\{1+\left\|\Psi(X_k)\right\|^2\left\|\Delta U_k\right\|^2\Big/\left(\left\|\Delta U_k\right\|^2+1\right)^2\right\}$$

$$-2\alpha tr\left\{\Psi(X_k)^T\bar{\varepsilon}_k\tilde{W}_k\,\Delta U_k\Big/\left(\|\Delta U_k\|^2+1\right)\right\}$$

$$-2\alpha\|\Psi(X_k)\|^2\left\|\tilde{W}_k\right\|^2\|\Delta U_k\|^2\Big/\left(\|\Delta U_k\|^2+1\right). \tag{23}$$

Now, after simplifying the inequality(23), the first difference can be expressed as

$$\Delta V_I \le -\alpha^2\tilde{e}_k^T\tilde{e}_k - 2\alpha\|\Psi(X_k)\|^2\left\|\tilde{W}_k\right\|^2\|\Delta U_k\|^2\Big/\left(\|\Delta U_k\|^2+1\right)$$

$$+\alpha^2\begin{pmatrix}\|\Psi(X_k)\|^2\left\|\tilde{W}_k\right\|^2\|\Delta U_k\|^2+\|\bar{\varepsilon}_k\|^2\\ +2\|\Psi(X_k)\|\left\|\tilde{W}_k\right\|\|\bar{\varepsilon}_k\|\|\Delta U_k\|\end{pmatrix}\left\{1+\frac{\|\Psi(X_k)\|^2\|\Delta U_k\|^2}{\left(\|\Delta U_k\|^2+1\right)^2}\right\}$$

$$+2\alpha\|\Psi(X_k)\|\left\|\tilde{W}_k\right\|\|\Delta U_k\|\|\bar{\varepsilon}_k\|\Big/\left(\|\Delta U_k\|^2+1\right). \tag{24}$$

Using the fact that $\|\Delta U_k\|^2\Big/\left(\|\Delta U_k\|^2+1\right)^2 \le 1/4$, $\Psi_m \le \|\Psi(X_k)\| \le \Psi_M$, and $U_m \le \|U(X_k)\|$, we can simplify the above inequality to the following form given by

$$\Delta V_I \le -\alpha^2\tilde{e}_k^T\tilde{e}_k + \alpha\left(-2U_{\mathrm{m}}^2\Psi_m^2+\alpha\bar{\Psi}_M^2\left\{1+\Psi_M^2/4\right\}\right)\left\|\tilde{W}_k\right\|^2$$

$$+\left(\alpha+2\alpha^2\bar{\Psi}_M\right)\Psi_M\bar{\varepsilon}_M\left\|\tilde{W}_k\right\|+\alpha^2\bar{\varepsilon}_M^2\left\{1+\Psi_M^2/4\right\}. \tag{25}$$

Here, in order for the first difference to be less than zero, design parameter α has to be selected satisfying $-2U_{\mathrm{m}}^2\Psi_m^2+\alpha\Psi_M^2\left\{1+\Psi_M^2/4\right\}<0$ which implies that

$$0<\alpha<32\Psi_m^2U_{\mathrm{m}}^2\Big/\left(4+\Psi_M^2\right)\bar{\Psi}_M^2. \tag{26}$$

Now, completing the squares in (25) we get

$$\Delta V_I \le -\alpha^2\tilde{e}_k^T\tilde{e}_k - \{\sqrt{\alpha\left(2U_{\mathrm{m}}^2\Psi_m^2-\alpha\bar{\Psi}_M^2\left\{1+\Psi_M^2/4\right\}\right)}\left\|\tilde{W}_k\right\|$$
$$-\bar{\varepsilon}_M\sqrt{\alpha\left(1+2\alpha\bar{\Psi}_M^2\right)^2\Psi_M^2\Big/\{4\left(2U_{\mathrm{m}}^2\Psi_m^2-\alpha\bar{\Psi}_M^2\left\{1+\Psi_M^2/4\right\}\right)\}}\}^2$$

$$+\{\alpha\left(\Psi_M^2+2\alpha\bar{\Psi}_M^2\right)^2\Big/4\Psi_M^2\left(2U_{\mathrm{m}}^2\Psi_m^2-\alpha\Psi_M^2\bar{\Psi}_M^2\left\{1+\Psi_M^2/4\right\}\right)$$

$$+\alpha^2\left\{1+\Psi_M^2/4\right\}\}\bar{\varepsilon}_M^2. \tag{27}$$

Now, $\Delta V_I < 0$ implies that $\|\tilde{e}_k\| > e_B$ or $\|\tilde{W}_k\| > W_B$ which demonstrates UUB of the identifier with the bounds given by (14) and (15). ■

IV. NN-BASED FORWARD IN TIME OPTIMAL REGULATOR

After identifying the IGM and higher order terms online, the next step is to optimally stabilize the affine-like system by using the identifier and to guarantee that the closed-loop remains stable in the neighborhood around the origin while ensuring that the actual control input is bounded and close to the optimal control input.

To this end, the affine representation of (5) is considered with the cost function $J(k)$ such that

$$J(X_k) \equiv \sum_{i=0}^{\infty} r(k+i) = r(X_k, \Delta u_k) + J(X_{k+1}) \tag{28}$$

where $r(X_k, \Delta u_k) = Q(X_k) + \Delta u_k^T R \Delta u_k$ with $Q(X_k) \geq 0$, $Q(0) = 0$, and $R \in \mathbb{R}^{m\times m}$ as a positive definite matrix. In the sequel, we will denote $J(X_k)$ by J_k for the sake of simplicity. Next the following definition is needed in order to proceed.

Definition 1: [14] A control Δu_k is admissible with respect to the infinite horizon cost function (28) on a compact set Σ provided the control action Δu_k is a) continuous on Σ, b) stabilizes **(1)** on Σ with $\Delta u_k|_{X_k=0} = 0$, and c) makes $J(X_0)$ finite (upper bounded) for all $X_0 \in \Sigma$.

The objective is to minimize J_k by starting with an admissible control law and modifying it with respect to the system dynamics so that the estimated cost function and

control input converge to the optimal cost function J_k^* and control law Δu_k^* respectively. By applying the stationarity condition [14] to (28) we have

$$\frac{\partial J_k}{\partial \Delta u_k} = \frac{\partial r(X_k, \Delta u_k)}{\partial \Delta u_k} + \frac{\partial J_{k+1}}{\partial \Delta u_k} = 2R\Delta u_k + \left(\frac{\partial X_{k+1}}{\partial \Delta u_k}\right)^T \frac{\partial J_{k+1}}{\partial X_{k+1}} = 0 \tag{29}$$

Using (29) and (14) we can write

$$\Delta u_k^* = -\frac{1}{2} R^{-1} \hat{G}^T(X_k) \frac{\partial J_{k+1}^*}{\partial X_{k+1}} - \frac{1}{2} R^{-1} \left(\frac{\partial \hat{O}_k}{\partial \Delta u_k} + \frac{\partial \tilde{e}_{k+1}}{\partial \Delta u_k}\right)^T \frac{\partial J_{k+1}^*}{\partial X_{k+1}} \tag{30}$$

Remark 1: Solving (30) is not possible in general since the second term in (30) on the right hand side is also a function of Δu_k. Therefore, we need to mitigate this term before trying to solve the optimal controller. Moreover, in the previous section, we have proven that the NN identifier will keep the identification error bounded regardless of Δu_k. This implies that $\tilde{e}_{k+1}$, in the system dynamics (14), plays the role of a bounded disturbance that eventually converges to a reasonable bound which could be small due to evolution of the identifier over time. Therefore, by temporarily ignoring the identification error, in the next subsection we will try to mitigate the higher order terms, $\hat{O}_k$, by designing a robust term. Subsequently, by assuming that O_k is mitigated, we will show in Subsection B that optimal controller $\Delta u_k^* = -0.5R^{-1}G^T(X_k)\partial J_{k+1}^* / \partial X_{k+1}$ of the affine system $X_{k+1} = F(X_k) + G(X_k)\Delta u_k$ can stabilize (10).

Thus in order to mitigate the higher order term, the design of an auxiliary term, $\Delta \rho_k$, is required for robustness in addition to the optimal controller term

$$\Delta u_k = \Delta \hat{u}_k + \Delta \rho_k . \tag{31}$$

where $\Delta\hat{u}_k$ is a NN estimation of Δu_k^*. The next subsections are dedicated to the design the update laws of $\Delta\rho_k$ and $\Delta\hat{u}_k$ along with the assessment of the overall system stability. First the design of the robust optimal controller is introduced.

A. Affinization of Nonaffine Systems Using Singular Perturbation

In order to minimize the effect of the higher order terms O_k and to ensure robustness, we design an auxiliary term. Nonetheless, in equation (10), $F(X_k)$, $G(X_k)$, $O(X_k)$ are not known. However, using (15), we have $X_{k+1} = \hat{X}_{k+1} + \tilde{e}_{k+1}$. Now, using (14) and rewriting the system dynamics in terms of identified variables as

$$X_{k+1} = \hat{F}(X_k) + \hat{G}(X_k)\Delta u_k + \hat{O}_k + \tilde{e}_{k+1}. \tag{32}$$

Using (31) the above equation can be expressed as

$$X_{k+1} = \hat{F}(X_k) + \hat{G}(X_k)\Delta\hat{u}_k + \hat{G}(X_k)\Delta\rho_k + \hat{O}_k + \tilde{e}_{k+1}. \tag{33}$$

Now our aim, is to design Δu_k such that the affine system $X_{k+1} = \hat{F}(X_k) + \hat{G}(X_k)\Delta\hat{u}_k$ minimizes the cost function (28) while the robust term $\Delta\rho_k$ mitigates the higher order term such that $\hat{G}(X_k)\Delta\rho_k + \hat{O}_k = 0$. In other words, the design of the robust optimal controller for the system (8) is equivalent to designing a controller for (30). Therefore, this subsection is dedicated to derive $\Delta\rho_k$ whereas the next subsection will deal with the optimal controller design.

It is desired to find a solution to $\Delta\rho_k$ namely $\Delta\overline{\rho}_k$ that solves the following equation

$$\hat{G}(X_k)\Delta\overline{\rho}_k + \hat{O}_k\left(X_k, \Delta\hat{u}_k + \Delta\overline{\rho}_k\right) = 0, \tag{34}$$

provided a solution exists for (34). Assume that $\Delta\rho_k$ is going to be updated such that it converges to the solution of (34) i.e. $\Delta\bar{\rho}_k$. To this effect we define an error dynamics as

$$\zeta_k = \hat{G}(X_k)\Delta\rho_k + \hat{O}_k\left(X_k, \Delta\hat{u}_k + \Delta\rho_k\right), \tag{35}$$

where ζ_k is the estimation error in equation (34) caused by $\Delta\rho_k$. Our aim is to design the dynamics of the robust controller term $\Delta\rho_k$ such that $\left\|\zeta_{k+1}\right\|^2 \le \alpha_\zeta^2 \left\|\zeta_k\right\|^2$ for $\alpha_\zeta < 1$ that means $\Delta\rho_k$ converge to $\Delta\bar{\rho}_k$. In other words, ζ_k has to be exponentially stable. Moreover, it should converge faster than the optimal controller for the sake of robustness. Now, consider the following update law for $\Delta\rho_k$ as

$$\Delta\rho_{k+1} = \Delta\rho_k + \delta\rho_k, \tag{36}$$

which makes (35) to have the following form

$$\hat{G}(X_{k+1})\Delta\rho_{k+1} + \hat{O}\left(X_{k+1}, \Delta\hat{u}_{k+1} + \Delta\rho_{k+1}\right) =$$

$$\hat{G}(X_{k+1})\left(\Delta\rho_k + \delta\rho_k\right) + \hat{O}\left(X_{k+1}, \Delta\hat{u}_{k+1} + \Delta\rho_k + \delta\rho_k\right) = \zeta_{k+1}. \tag{37}$$

By using Taylor series first order approximation we can write

$$\hat{O}\left(X_{k+1}, \Delta\hat{u}_{k+1} + \Delta\rho_k + \delta\rho_k\right) \cong$$

$$\hat{O}\left(X_{k+1}, \Delta\hat{u}_{k+1} + \Delta\rho_k\right) + \frac{\partial}{\partial\Delta\hat{u}_k}\hat{O}\left(X_{k+1}, \Delta\hat{u}_{k+1} + \Delta\rho_k\right)\delta\rho_k. \tag{38}$$

Equation (38) is valid since the convergence of the outputs to a certain neighborhood of the operation point, which is the origin, is guaranteed by the proposed optimal controller which is presented in the next subsection. Equation (37) can be written as

$$\left\{\hat{G}(X_{k+1})+\partial\hat{O}\left(X_{k+1},\Delta\hat{u}_{k+1}+\Delta\rho_k\right)/\partial\Delta\hat{u}_k\right\}\delta\rho_k$$

$$+\hat{G}(X_{k+1})\Delta\rho_k+\hat{O}\left(X_{k+1},\Delta\hat{u}_{k+1}+\Delta\rho_k\right)=\zeta_{k+1}. \tag{39}$$

We desire to have the left hand side of the equation (39) equal to $\alpha_\zeta\zeta_k$, where α_ζ is the rate of convergence of the fast dynamics which is known as ε in continuous-time singularly perturbed systems. Therefore, using pseudo-inversion to calculate the desired $\delta\rho_k$ yields

$$\delta\rho_k=-A_\zeta^{-1}\left(k\right)B_\zeta\left(k\right), \tag{40}$$

with

$$A_\zeta\left(k\right)=\left\{\hat{G}(X_{k+1})+\partial\hat{O}\left(X_{k+1},\Delta\hat{u}_{k+1}+\Delta\rho_k\right)/\partial\Delta\hat{u}_k\right\},$$

$$B_\zeta\left(k\right)=\left\{\hat{G}(X_{k+1})\Delta\rho_k+\hat{O}\left(X_{k+1},\Delta\hat{u}_{k+1}+\Delta\rho_k\right)-\alpha_\zeta\zeta_k\right\},$$

And

$$0<\alpha_\zeta<1. \tag{41}$$

Remark 2: It is clear that by having an invertible A_ζ, $\left\|\zeta_{k+1}\right\|=\alpha_\zeta\left\|\zeta_k\right\|$ holds. However, in the case when A_ζ is not invertible, $\left\|\zeta_{k+1}\right\|<\left\|\zeta_k\right\|$ may not hold and therefore, the robust term should be chosen as $\delta\rho_k=0$ which means a solution may not exist and converting the system into affine form fails. In the sequel, we assume A_ζ is invertible and therefore there exists $L_\zeta<1$ such that

$$\left\|\zeta_{k+1}\right\|-\left\|\zeta_k\right\|\le-L_\zeta\left\|\zeta_k\right\|. \tag{42}$$

Remark 3: When (34) holds, the identified higher order term $\hat{O}_k$ gets mitigated and therefore the overall system behaves as an affine system. For guaranteeing the

existence of solution for $\Delta\rho_k$ the *implicit function theorem* requires $\hat{G}(X_{k+1}) + \partial\hat{O}\left(X_{k+1}, \Delta\hat{u}_{k+1} + \Delta\rho_k\right)/\partial\Delta\hat{u}_k$ to be invertible. Therefore, the invertability of A_ζ is an inevitable condition. To the best knowledge of the authors, the only work on converting a nonaffine system [21] into an affine-like equivalent, requires the same assumption. Consequently, under the locally invertible assumption of A_ζ, the robust controller (40) improves the convergence by mitigating the modeling errors due to higher order terms of the Taylor series expansion so that the optimal controller can be used for the affine system.

Therefore, we consider the following representation of a discrete singularly perturbed robust controller

$$\Delta\rho_{k+1} = \Delta\rho_k + \delta\rho_k = \Delta\rho_k - A_\zeta^{-1}(k) B_\zeta(k), \tag{43}$$

for which an online robust optimal controller will be designed in the sequel.

Remark 4: For convergence of $\Delta\rho_k$ to the solution of (34) ($\Delta\rho_k \to \Delta\bar{\rho}_k$), it is required that system dynamics (33) be stable. Therefore, as will be mentioned later, it is necessary that Δu_k to be initialized to an admissible controller and remain admissible while we are update it towards an optimal solution.

Now, the system dynamics represented in (33) can be represented by using (35) as

$$X_{k+1} = \hat{F}(X_k) + \hat{G}(X_k)\Delta\hat{u}_k + \zeta_k + \tilde{e}_{k+1}, \tag{44}$$

where $\Delta\hat{u}_k$ is going to be designed in the next subsections B and C. So far, we demonstrated the stability of ζ_k and $\tilde{e}_k$. In the next subsections it will be observed that the overall system X_k, ζ_k, and $\tilde{e}_k$ will also be stable in the closed loop form. Moreover,

simultaneously, the estimation of the cost function $J(X_k)$ will converge close to its optimal value forcing $\Delta\hat{u}_k$ to converge to the optimal solution ($\Delta\hat{u}_k \to \Delta u_k^*$).

B. Cost Function Approximation for Optimal Regulator Design

The objective of the optimal control law is to stabilize the system (10) while minimizing the cost-function (28). The cost function (28) will be approximated by an OLA and written as

$$\hat{J}(k) = \hat{J}(X_k) = \hat{\Phi}_k^T \sigma(X_k) = \hat{\Phi}_k^T \sigma(k) \ , \tag{45}$$

where $\hat{J}(k)$ represents an approximated value of the original cost function $J(k)$, $\hat{\Phi}_k$ is the vector of actual parameter vector for the target OLA parameter vector, Φ , and $\sigma(k) = \{\sigma_\ell(k)\}_1^{L:}$ is set of activation functions which are each chosen to be basis sets and thus are linearly independent. The basis function should satisfy $\|\sigma(0)\| = 0$ for $\|x\| = 0$ with $x \in \Re^n$. Selection of $\sigma(\bullet)$ in this way ensures $J(0) = 0$ can be satisfied [16]. For convenience, define the error in the cost function as

$$e_c(k) = r(k-1) + \hat{\Phi}_k^T \sigma(k) - \hat{\Phi}_k^T \sigma(k-1) \tag{46}$$

whose dynamics are given by

$$e_c(k+1) = r(k) + \hat{\Phi}_{k+1}^T \left(\sigma(k+1) - \sigma(k)\right). \tag{47}$$

Next, we define an auxiliary cost error vector as

$$E_c(k) = Y(k-1) + \hat{\Phi}_k^T \overline{X}(k-1) \in \Re^{1\times(1+j)} \tag{48}$$

where $Y(k-1) = [r(k-1) \quad r(k-2) \ldots r(k-1-j)]$ and $\overline{X}(k-1) = [\Delta\sigma(k) \quad \Delta\sigma(k-1) \ldots \Delta\sigma(k-j)]$ with $\Delta\sigma(k) = \sigma(k) - \sigma(k-1)$, $0 < j < k-1 \in \mathbf{N}$ and $\mathbf{N}$ being the set of natural real numbers. It is useful to observe that (48) can be rewritten

as $E_c(k) = [e_c(k \mid k) \;\; e_c(k \mid k-1) \cdots e_c(k \mid k-j)]$ where the notation $e_c(k \mid k-1)$ means the cost error $e_c(k-1)$ re-evaluated at time .. using the actual cost parameter matrix $\hat{\Phi}_k^T$. The dynamics of the auxiliary vector (48) are formed similar to (47) and revealed to be

$$E_c^T(k+1) = Y^T(k) + \bar{X}^T(k)\hat{\Phi}_{k+1} \tag{49}$$

Examining the error dynamics in (49), it is observed that they closely resemble a nonlinear discrete-time system with $\hat{\Phi}_{k+1}$ being the control input, and $Y(k)$ and $X(k)$ being nonlinear vector fields. To proceed, the following technical results are needed.

Now define the cost function OLA parameter update to be

$$\hat{\Phi}_{k+1} = \bar{X}(k)\left(\bar{X}^T(k)\bar{X}(k)\right)^{-1}\left(\alpha_c E_c^T(k) - Y^T(k)\right) \tag{50}$$

where $0 < \alpha_c < 1$, and substituting (50) into (49) reveals

$$E_c^T(k+1) = \alpha_c E_c^T(k). \tag{51}$$

Remark 5: It is interesting to observe that the parameter update law (50) resembles the least squares update rule commonly used in offline ADP [6] and [15]; however, instead of summing over a mesh of training points, the update (50) represents a sum over the system's time history stored in $E_c(k)$. Thus, the update (50) uses data collected in real time instead of data formed offline.

Remark 6: As a result of *Lemma 1*, the matrix $\bar{X}^T(k)\bar{X}(k)$ is invertible provided $X(k) \neq 0$ which can be viewed as a persistency of excitation (PE) condition. Observing the definition of the cost function (28) and OLA approximation (45), it is evident that both become zero only when $X_k = 0$. To ensure the PE condition, an output measurement noise will be added to X_k.

As a final step in the cost function OLA design, we define the parameter estimation error to be $\tilde{\Phi}_k = \Phi - \hat{\Phi}_k$, and rewrite (28) using the ideal OLA representation

$$J(X_k) = \Phi^T \sigma(X_k) + \varepsilon_c \tag{52}$$

revealing $\Phi^T \sigma(X_{k+1}) + \varepsilon_c(k) = r(k) + \Phi^T \sigma(X_k) + \varepsilon_c(k+1)$ which can be rewritten as $r(k) = -\Phi^T \Delta\sigma(X_k) - \Delta\varepsilon_c(k)$ where $\Delta\varepsilon_c(k) = \varepsilon_c(k+1) - \varepsilon_c(k)$ and $\Delta\sigma(X_k) = \sigma(X_{k+1}) - \sigma(X_k)$. Substituting $r(k)$ into (47) as well as utilizing (46) and $e_c(k+1) = \alpha_c e_c(k)$ from (51) yields

$$\tilde{\Phi}_{k+1}^T \Delta\sigma(X_k) = -\alpha_c (r(k-1) + \hat{\Phi}_k^T \Delta\sigma(k-1)) - \Delta\varepsilon_c(k). \tag{53}$$

In a similar manner as $r(k)$, we now form $r(k-1) = -\Phi^T \Delta\sigma(X_{k-1}) - \Delta\varepsilon_c(k-1)$ and substitute this expression into (53), revealing $\Delta\sigma^T(X_k)\tilde{\Phi}_{k+1} = \alpha_c \Delta\sigma^T(X_k)\tilde{\Phi}_k + \alpha_c \Delta\varepsilon_c(k-1) - \Delta\varepsilon_c(k)$, and the OLA parameter estimation error dynamics are revealed to be

$$\tilde{\Phi}_{k+1}^T = \alpha_c \Delta\sigma(X_k)\left(\Delta\sigma^T(X_k)\Delta\sigma(X_k)\right)^{-1} \Delta\sigma(X_{k-1})\tilde{\Phi}_k + \Delta\sigma(X_k)$$

$$\left(\Delta\sigma^T(X_k)\Delta\sigma(X_k)\right)^{-1} (\alpha_c \Delta\varepsilon_c(k-1) - \Delta\varepsilon_c(k)). \tag{54}$$

Next, the boundedness of the cost function error (46) and OLA estimation error (54) is demonstrated in the following theorem. In order to proceed, the following definition is needed.

Theorem 2 [3]: *(Boundedness of the Cost OLA Errors).* Let $u_0(x_k)$ be a fixed admissible control policy for the controllable system (10), and let the cost NN weight update law be given by (50). Then, given a positive constant δ_c satisfying $0 < \delta_c < 1/2$,

there exists positive constant, α_c given by $\alpha_c = \sqrt{1/2 - \delta_c}$ such that the critic NN weight estimation error (14) is uniformly ultimately bounded (UUB) with ultimate bounds given by $\left\|\tilde{\Phi}_k\right\|_F \le b'_\Phi$ for a positive constant b'_Φ. ■

Remark 7: The results of *Theorem 2* are drawn under the assumption of a fixed control policy or when the control input is not updated with time, and these results will aid in the proof of *Theorem 2* where only an initial admissible control is required. Moreover, the estimation of the cost function presented in the above is a generic approach that does not depend on the system structure. This approach was used in [3] for an affine system while here it is employed for a nonaffine system.

C. Estimation of the Optimal Feedback Control Signal

The objective of this section is to find the control policy which minimizes the approximated cost function (45). To begin the development of the feedback control policy, we define a NN to estimate (30) as

$$\Delta u^*(k) = \Delta u^*(X_k) = \Theta_k^T \vartheta(X_k) + \varepsilon_A . \tag{55}$$

Therefore, define an OLA approximation of (55) to be

$$\Delta \hat{u}(k) = \Delta \hat{u}(X_k) = \hat{\Theta}_k^T \vartheta(X_k) \tag{56}$$

where $\Delta \hat{u}(k)$ is introduced in (31) and $\hat{\Theta}_k$ is the estimated value of the ideal parameter matrix Θ and $\vartheta(\bullet)$ denotes the linearly independent basis function.

Next, the optimal control signal error is defined to be the difference between the feedback control applied to (10) and the optimal control signal, as

$$e_a(k) = \hat{\Theta}_k^T \vartheta(X_k) + R^{-1} \hat{G}^T(k) \left(\partial \sigma(X_k)/\partial X_k\right)^T \hat{\Phi}_k / 2 \text{ , and} \tag{57}$$

$$e_a(k+1) = \hat{\Theta}_{k+1}^T \vartheta(X_k) + \frac{1}{2} R^{-1} \hat{G}^T(k+1) \left(\frac{\partial \sigma(X_{k+2})}{\partial X_{k+2}} \right)^T \hat{\Phi}_{k+1} \,, \tag{58}$$

where, as it is mentioned before, the $\tilde{e}_k$ and ζ_k are ignored in the equation (44). One should notice that the identifier error and the robust term errors are previously shown to be bounded and play the role of disturbance in the closed loop system. This is the reason that they are ignored in the design of the estimation error (57). Nonetheless, we will see that $\tilde{e}_k$ and ζ_k will appear in the overall stability proof of the system.

The proposed control OLA parameter update is defined to be

$$\hat{\Theta}_{k+1} = \hat{\Theta}_k - \alpha_a \vartheta(k) e_a^T(k) \big/ (\vartheta^T(k)\vartheta(k) + 1) \tag{59}$$

where $0 < \alpha_a < 1$ is a positive design parameter. Substituting the parameter update (59) into (58) yields

$$e_a(k+1) = \hat{\Theta}_k^T \vartheta\left(X_{k+1}\right) - \alpha_a e_a(k) \vartheta^T(k) \vartheta(X_{k+1}) \Big/ \left(\vartheta^T(k)\vartheta(k) + 1 \right)$$

$$+ \frac{R^{-1} \hat{G}^T(k+1)}{2} \left(\frac{\partial \sigma(X_{k+2})}{\partial X_{k+2}} \right)^T \hat{\Phi}_c(k+1) \,. \tag{60}$$

Since the control policy $u(x(k))$ in (56) minimizes the cost function (28), from (30) we can write

$$0 = \varepsilon_A(k+1) + \frac{1}{2} R^{-1} \hat{G}^T(k+1) \frac{\partial \varepsilon_c\left(k+2\right)^T}{\partial X_{k+2}} +$$

$$\Theta^T \vartheta\left(k+1\right) + \frac{1}{2} R^{-1} \hat{G}^T(k+1) \frac{\partial \sigma(X_{k+2})^T}{\partial X_{k+2}} \Phi \,. \tag{61}$$

Subtracting (61) from (60) along with defining the control OLA parameter estimation error as $\tilde{\Theta}_k = \Theta - \hat{\Theta}_k$ while recalling $\tilde{\Theta}_{k+1} = \Theta - \hat{\Theta}_{k+1}$ yields

$$e_a(k+1) = -\alpha_a e_a(k)\,\vartheta^T(k)\vartheta(X_{k+1})\Big/\left(\vartheta^T(k)\vartheta(k)+1\right)$$

$$-\varepsilon_A(k+1) - \frac{1}{2}R^{-1}\hat{G}^T(k+1)\frac{\partial \varepsilon_c(k+2)}{\partial X_{k+2}}$$

$$-\frac{R^{-1}\hat{G}^T(k+1)}{2}\frac{\partial \sigma(X_{k+2})}{\partial X_{k+2}}^T \tilde{\Phi}_{k+1} - \tilde{\Theta}_k^T\vartheta(X_{k+1}) \tag{62}$$

As a final step, we form the parameter estimation error dynamics as

$$\tilde{\Theta}_{k+1} = \tilde{\Theta}_k + \alpha_a \vartheta(k)\frac{e_a^T(k)}{\vartheta^T(k)\vartheta(k)+1}. \tag{63}$$

Remark 8: As mentioned before, in order to calculate $e_a(k)$ in (57) and implement the OLA parameter update(59), knowledge of IGM is required while that of $F(k)$ is not.

In the following theorem, it will be shown that by starting with an initial stabilizing control, the control OLA update (59)ensures all future control inputs are also admissible. The following corollary illustrates that for any admissible control policy, the system states as well as the cost and control approximator basis functions are bounded.

Corollary 1:(Boundedness of OLA Basis Functions) [3]. Let $\mu(k)$ be any admissible control for the controllable system (10). Then, there exists a positive constant $X_0 = X_{k=0}$ such that $X_0 \geq \|X_k\|$ for all $k > 0$. Moreover, there exists positive constants $\vartheta_M = \|\vartheta(x_0)\|$ and $\sigma_M = \|\sigma(X_o)\|$ such that $\vartheta_M \geq \|\vartheta(X_k)\|$ and $\sigma_M \geq \|\sigma(X_k)\|$ for all $k > 0$.

Now, we are ready to propose the main result which is the stability analysis of the optimal NN-based controller.

D. Convergence Proof

Theorem 3 (*Overall Stability Proof of the Robust Optimal Controller*): Consider the nonlinear singularly perturbed discrete-time system rewritten as (44) and (43). Assume that the proposed online NN identifier is given by (11) with the weight update law shown in (16) is used to identify the affine-like system(10). Let the optimal control law proposed in (56) by using a NN is applied to the affine like representation of the nonaffine system (10) when $\alpha_u/2 < \delta_J < 1/2$. Let the NN (45) is applied to estimate the cost function. Assume that there exist $L_\zeta < 1$ so that (42) holds. Then the overall system (44) and (43) along with (36) is uniformly ultimate bounded and $\Delta\rho_k$ asymptotically exponentially converges to $\Delta\bar{\rho}_k$, where $\Delta\bar{\rho}_k$ is the solution of (34). In addition, the robust controller $\Delta\hat{u}_k + \Delta\rho_k$ converges to the robust optimal controller $\Delta u_k^* + \Delta\bar{\rho}_k$ in (30) with an arbitrarily small error. Moreover, the state error X_k, the weight estimation error $\tilde{\Theta}_k$, the cost function weight estimation error $\tilde{\Phi}_k$, the identification error $\tilde{e}_k$, and NN weight estimation error $\tilde{W}_k$ for identifier are uniformly ultimately bounded i.e. $\left\|\tilde{W}_k\right\| \le b_W$, $\left\|X_k\right\| \le b_X$, $\left\|\Xi_{Ak}\right\| \le b_\Xi$, $\left\|\tilde{e}_k\right\| \le b_e$, $\left\|\tilde{\Phi}_k\right\| \le b_\Phi$, $\left\|\zeta_k\right\| \le b_\zeta$.

Proof- Consider the following Lyapunov function candidate

$$V_k = V_D(X_k) + V_u(\tilde{\Theta}_k) + V_J(\tilde{\Phi}_k) + V_I(\tilde{W}_k, \tilde{e}_k) + V_\rho(\zeta_k)$$

$$= K_D\left\|X_k\right\| + tr\left\{\tilde{\Theta}_k^T\tilde{\Theta}_k\right\} + tr\left\{\tilde{W}_k^T\tilde{W}_k\right\} + \left(\tilde{\Phi}_k^T\Delta\sigma\left(X_{k-1}\right)\right)^2 + \alpha^2\tilde{e}_k^T\tilde{e}_k + \left\|\zeta_k\right\|, \qquad (64)$$

where $K_D \in \mathbb{R}^+$ and $\zeta_k = \Delta\rho_k - \Delta\bar{\rho}_k$. Therefore, the first difference of the Lyapunov function can be denoted as follows

$$\Delta V_k = \Delta V_D(X_k) + \Delta V_u(\tilde{\Theta}_k) + \Delta V_J(\tilde{\Phi}_k) + \Delta V_I(\tilde{W}_k) + \Delta V_\rho(\zeta_k)\ . \qquad (65)$$

Due to complexity of ΔV_k, we investigate the first difference of individual terms $\Delta V_D(X_k)$, $\Delta V_u(\tilde{\Theta}_k)$, $\Delta V_J(\tilde{\Phi}_k)$, $\Delta V_I(\tilde{W}_k)$, and $\Delta V_\rho(\zeta_k)$. Then, the bounds that make $\Delta V_k < 0$ will be calculated.

ΔV_D**:** Consider the positive definite and radially unbounded function $V_D(k) = K_D \|X_k\|$ and its first difference:

$$\Delta V_D = V_D(k+1) - V_D(k) = K_D \left\| F_k + G_k \{\Delta u_k + \Delta \rho_k\} + O_k \right\| - K_D \|X_k\|$$

$$= K_D \left\| \hat{F}_k + \hat{G}_k \hat{\Theta}_{k-1}^T \vartheta_k + \alpha_u \hat{G}_k \|\vartheta_k\|^2 \Delta \tilde{u}_{k-1}^T \Big/ \left(\|\vartheta_k\|^2 + 1 \right) + \zeta_k + \tilde{e}_{k+1} \right\| - K_D \|X_k\|, \quad (66)$$

Considering that $\hat{\Theta}_{k-1}^T \vartheta_k$ resembles an admissible controller (if $\hat{\Theta}_{k-1}$ is not updated but renders the input remains admissible), (66) can be rewritten as follows

$$\Delta V_D \le -K_D(1-L)\|X_k\|$$

$$+K_D \alpha_u \left(\|\vartheta_k\|^2 \Big/ \left(\|\vartheta_k\|^2 + 1 \right) \right) \left\| \hat{G}_k \left(\Xi_{Ak} + \varepsilon_{uk} \right) \right\| + K_D \|\zeta_k\| + K_D \|\tilde{e}_{k+1}\|, \quad (67)$$

where $\Xi_{Ak} = \tilde{\Theta}_{k-1}^T \vartheta_k$ and it is assumed that the following holds due to the initial input admissibility.

$$\left\| \hat{F}_k + \hat{G}_k \{\Delta u_k + \Delta \rho_k\} + \hat{O}_k + \tilde{e}_{k+1} \right\| = \|X_{k+1}\| \le L \|X_k\| \quad (68)$$

for the initial step with $L < 1$. Therefore, with the assumption $\left\|\varepsilon_k^o\right\| \le \varepsilon_M^o$, using (67) the first difference can be written as

$$\Delta V_D \le -K_D(1-L)\|X_k\| + K_D \alpha_u G_M \|\Xi_{Ak}\|$$

$$+K_D \alpha_u G_M \|\varepsilon_{uM}\| + K_D \|\zeta_k\| + K_D(1+2\alpha)\|\tilde{e}_{k+1}\|, \quad (69)$$

where the following fact is used

$$\left\|\hat{G}_k \Delta\tilde{u}_k\right\| = \left\|\left(G_k + \tilde{G}_k\right)\Delta\tilde{u}_k\right\| \le \left\|G_k \Delta\tilde{u}_k\right\| + \left\|\tilde{G}_k \Delta\tilde{u}_k\right\| \le K_D \alpha_u G_M \left\|\Xi_{Ak}\right\| + K_D \alpha_u G_M \left\|\varepsilon_{uM}\right\| + 2\left\|\tilde{e}_{k+1}\right\|$$

, and assuming G_M being the upper bound on the gain matrix G_k.

$\Delta V_u(\tilde{\Theta}_k)$: Consider $V_u(\tilde{\Theta}_k) = tr\left\{\tilde{\Theta}_k^T \tilde{\Theta}_k\right\}$ as a positive definite and radially unbounded function. The first difference of $V_u(\tilde{\Theta}_k)$ can be expressed as follows

$$\Delta V_u(\tilde{\Theta}_k) = tr\{(\tilde{\Theta}_k + \alpha_u \frac{\vartheta_k e_a^T(k)}{\left\|\vartheta_k\right\|^2 + 1})^T (\tilde{\Theta}_k + \alpha_u \frac{\vartheta_k e_a^T(k)}{\left\|\vartheta_k\right\|^2 + 1})\} - tr\{\tilde{\Theta}_k^T \tilde{\Theta}_k\}$$

$$= \alpha_u^2 \frac{\left\|\vartheta_k\right\|}{\left\|\vartheta_k\right\|^2 + 1} e_a^T(k) e_a(k) + 2\alpha_u \frac{1}{\left\|\vartheta_k\right\|^2 + 1} e_a^T(k) \tilde{\Theta}_k^T \vartheta_k, \tag{70}$$

where

$$\Delta\tilde{u}_k = \hat{\Theta}_k^T \vartheta_k + \frac{1}{2} \hat{G}_k^T \frac{\partial \sigma_{k+1}^*}{\partial X_{k+1}} \hat{\Phi}_k + \varepsilon_{uk} =$$

$$\Theta_k^T \vartheta_k - \tilde{\Theta}_k^T \vartheta_k + \frac{1}{2}\left(G_K^T - \tilde{G}_K^T\right) \frac{\partial \sigma_{k+1}}{\partial X_{k+1}}^T (\Phi - \tilde{\Phi}_k) + \varepsilon_{uk}, \tag{71}$$

with $G_k - \hat{G}_k = \tilde{G}_k$. Moreover, it is clear that

$$\Theta_k^T \vartheta_k + \frac{1}{2} G_k^T \frac{\partial \sigma_{k+1}}{\partial X_{k+1}}^T \Phi_k = \varepsilon_{uk}^* \tag{72}$$

$$\Delta\tilde{u}_k = -\tilde{\Theta}_k^T \vartheta_k - \frac{1}{2} G_K^T \frac{\partial \sigma_{k+1}}{\partial X_{k+1}}^T \tilde{\Phi}_k - \frac{1}{2} \tilde{G}_K^T \frac{\partial \sigma_{k+1}}{\partial X_{k+1}}^T \Phi_k + \frac{1}{2} \tilde{G}_K^T \frac{\partial \sigma_{k+1}}{\partial X_{k+1}}^T \tilde{\Phi}_k + \varepsilon_{uk} + \varepsilon_{uk}^*$$

where ε_{uk}^* is the estimation error of the desired optimal control law. It is shown in [3] that

$$\Delta V_u(\tilde{\Theta}_k) \le -\frac{\alpha_u}{2(\vartheta_M^2 + 1)}\left(3 - 3\alpha_u\right)\left\|\Xi_{Ak}\right\|^2 + \alpha_u \Pi_A \left\|\tilde{\Phi}_k\right\|^2 + \alpha_u \Pi_G \left\|\tilde{G}_k\right\|^2$$

$$+\left(\alpha_u\left(5\alpha_u+4\right)+\alpha_u^2\left(\sigma'_M G_M \lambda_{\max}\left(R^{-1}\right)\right)^2\right)\left(\varepsilon_{uk}+\varepsilon_{uk}^*\right), \tag{73}$$

where

$$\Pi_A=\left(\left(1+5\alpha_u/4\right)\left(\sigma'_M G_M \lambda_{\max}\left(R^{-1}\right)\right)^2+\alpha_u/2\right)$$

$$\Pi_G=\left(\left(1+5\alpha_u/4\right)\left(\sigma'_M \Phi_M \lambda_{\max}\left(R^{-1}\right)\right)^2+\alpha_u/2\right)$$

$\Delta V_J(\tilde{\Phi}_k)$: In the work [3], it is shown that with $V_J(\tilde{\Phi}_k)=\left(\tilde{\Phi}_k^T\Delta\sigma\left(X_{k-1}\right)\right)^2$ and the chosen update law we have

$$\Delta V_J(\tilde{\Phi}_k)\le -2\delta_c\Delta_{\min}^2\left\|\tilde{\Phi}_k\right\|^2+2\Delta\bar{\varepsilon}_{IM}^2. \tag{74}$$

$\Delta V_I(\tilde{W}_k)$: Negativeness of this term of (65) is already investigated in Theorem 1. With the fact that

$$\left\|\tilde{G}_k\right\|^2\le\left\|\Psi_M\tilde{W}_k\right\|^2\le\left\|\Psi_M\right\|^2\left\|\tilde{W}_k\right\|^2 \tag{75}$$

$\Delta V_\rho(\zeta_k)$: As we designed this term, for each time step, it is guaranteed that (42) holds

Now, we are ready to find the overall stability bounds. (27), (42), (69), (73), and (74), can be used to write

$$\Delta V_k\le-\alpha^2\tilde{e}_k^T\tilde{e}_k+\left\{\alpha_u\Pi_G\Psi_M^2+\alpha\left(-2U_{\mathrm{m}}^2\Psi_m^2+\alpha\bar{\Psi}_M^2\left\{1+\Psi_M^2/4\right\}\right)\right\}\left\|\tilde{W}_k\right\|^2$$

$$+\left(\alpha+2\alpha^2\bar{\Psi}_M\right)\Psi_M\bar{\varepsilon}_M\left\|\tilde{W}_k\right\|-K_D(1-L)\left\|X_k\right\|+K_D\left\|\zeta_k\right\|-L_\zeta\left\|\zeta_k\right\|$$

$$-\frac{\alpha_u}{2(\vartheta_M^2+1)}\left(3-3\alpha_u\right)\left\|\Xi_{Ak}\right\|^2+K_D\alpha_u G_M\left\|\Xi_{Ak}\right\|+\left(-2\delta_J\Delta_{\min}^2+\alpha_u\Pi_A\right)\left\|\tilde{\Phi}_k\right\|^2+\varepsilon_{SM} \tag{76}$$

With

$$\varepsilon_{SM}=\left(\alpha_u\left(5\alpha_u+4\right)+\alpha_u^2\left(\sigma_M'G_M\lambda_{\max}\left(R^{-1}\right)\right)^2\right)\left(\varepsilon_{uk}+\varepsilon_{uk}^*\right)$$

$$+K_D\alpha_uG_M\left\|\varepsilon_{uM}\right\|+\alpha^2\bar{\varepsilon}_M^2\left\{1+\Psi_M^2/4\right\}+2\Delta\bar{\varepsilon}_{IM}^2+(1+2\alpha)K_De_B \tag{77}$$

It is obvious that K_D should be chosen such that $K_D<L_\zeta<1$. Therefore, ΔV_k is negative as long as the following is satisfied

$$\left\|\tilde{W}_k\right\|>\frac{\bar{\varepsilon}_M\left(\Psi_M+2\alpha\dfrac{\bar{\Psi}_M^2}{\Psi_M}\right)+\sqrt{\begin{array}{l}\alpha\left(-8U_{\mathrm{m}}^2\Psi_m^2+\alpha\bar{\Psi}_M^2\left\{4+\Psi_M^2\right\}\right)\varepsilon_{SM}+\\4\varepsilon_{SM}\alpha_u\Pi_G\Psi_M^2-\left(\alpha+2\alpha^2\bar{\Psi}_M\right)^2\Psi_M^2\bar{\varepsilon}_M^2\end{array}}}{2\alpha_u\Pi_G\Psi_M^2+2\alpha\left(-2U_{\mathrm{m}}^2\Psi_m^2+\alpha\bar{\Psi}_M^2\left\{1+\Psi_M^2/4\right\}\right)}\equiv b_W,\text{or} \tag{78}$$

$$\left\|X_k\right\|>\varepsilon_{SM}/(1-L)\equiv b_X,\text{ or} \tag{79}$$

$$\left\|\Xi_{Ak}\right\|>\frac{K_D\alpha_uG_M+\sqrt{K_D^2\alpha_u^2G_M^2-\dfrac{4\varepsilon_{SM}\alpha_u}{2(\vartheta_M^2+1)}\left(3-3\alpha_u\right)}}{\dfrac{\alpha_u}{(\vartheta_M^2+1)}\left(3-3\alpha_u\right)}\equiv b_\Xi,\text{ or} \tag{80}$$

$$\left\|\tilde{e}_k\right\|>\sqrt{\varepsilon_{SM}}/\alpha\equiv b_e,\text{ or} \tag{81}$$

$$\left\|\tilde{\Phi}_k\right\|>\sqrt{\varepsilon_{SM}/\left(\alpha_u\Pi_A-2\delta_c\Delta_{\min}^2\right)}\equiv b_\Phi,\text{ or} \tag{82}$$

$$\left\|\zeta_k\right\|>\frac{\varepsilon_{SM}}{L_\zeta-K_D}. \tag{83}$$

Therefore, the closed loop system is UUB with the bounds given by (78)-(83). One can observe the dependence of the convergence of the closed-loop system on the NNs learning rate parameters α_u, α, and α_c as well as the higher order terms. Although the convergence is guaranteed, finding proper learning rates for the best convergence rate is a challenge. ■

Next, Figure 1 depicts the overall controller block diagram representation. Although the internal dynamics $F(X_k)$ is not required, the NN identifier generates an estimation of $G(X_k)$ and O_k. As mentioned before, the robust controller block in Fig. 1 is used to mitigate the effects of modeling errors in the form of higher order terms and to represent the closed-loop system as an affine system. The rest of the controller block diagram represents the proposed online optimal controller. Here value and policy iterations are not utilized. Instead, the cost function is updated once every sampling interval. The only requirement for having the overall closed loop system to be stable is an initial admissible controller which is shown in the block diagram as the initial value of $\hat{\Theta}_k$. After tuning, the initial admissible controller will become robust optimal controller over time.

V. APPLICATION TO THE HCCI ENGINE

Low temperature combustion modes, such as Homogeneous Charge Compression Ignition (HCCI), represent a promising means to increase the efficiency and significantly reduce the emissions of internal combustion engines. Implementation and control are difficult, however, due to the dependence of the combustion event on chemical kinetics rather than an external trigger. In [11], the author outlined a nonlinear control-oriented model of a single cylinder HCCI engine, which is physically based on a five state which utilizes fully vaporized gasoline-type fuels, exhaust gas recirculation and intake air heating in order to achieve HCCI operation. The onset of combustion, which is vital for control, is modeled using an Arrhenius Reaction Rate expression which relates the combustion timing to both charge dilution and temperature.

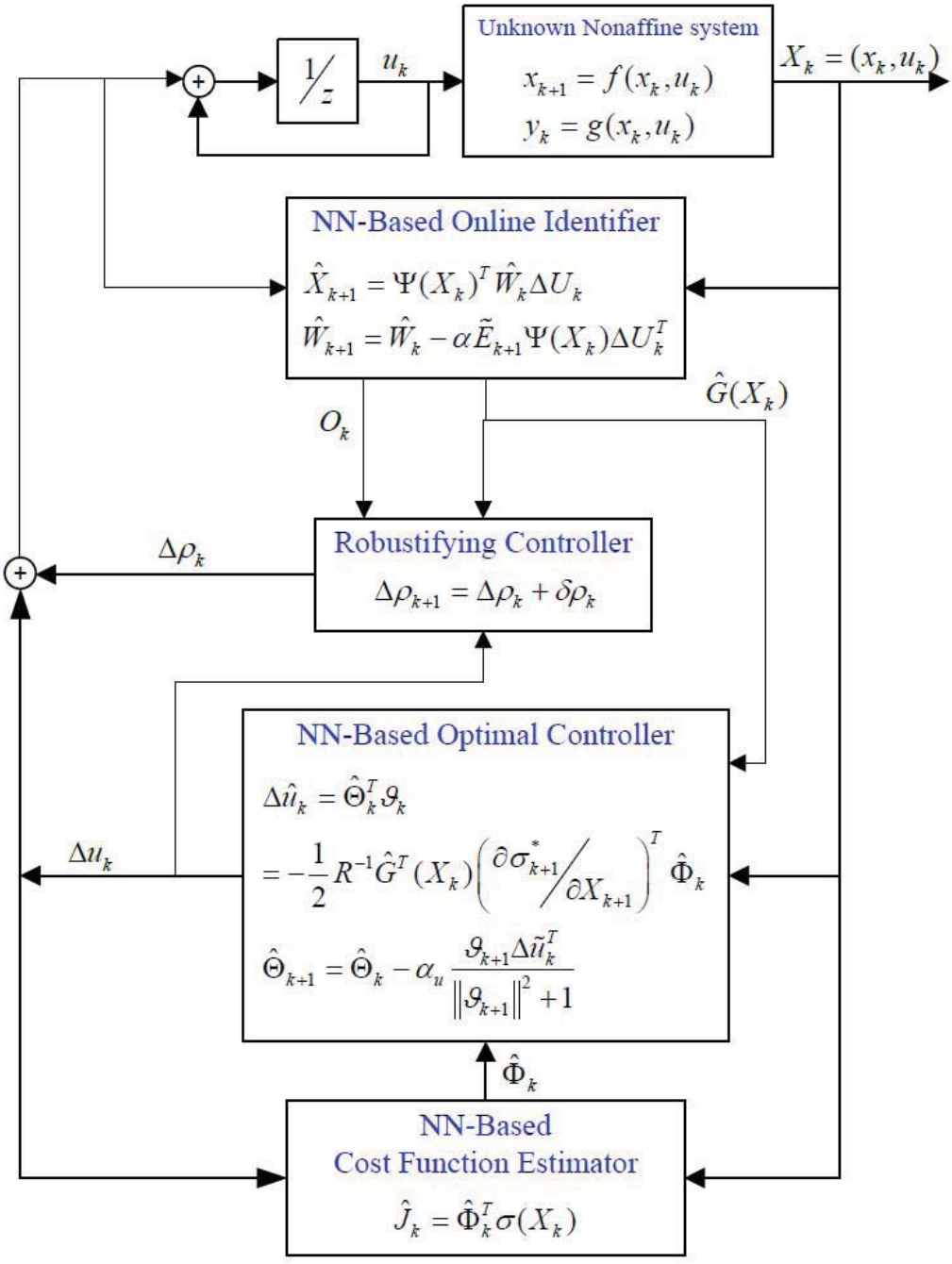

Figure 1. The proposed controller block diagram representation.

This model is aimed at capturing the behavior of an engine thermodynamic cycle. The model is validated against experimental data from a single cylinder CI engine operating under HCCI conditions at two different fueling rates. Predicted combustion timing and peak in-cylinder pressure values from simulation agree very well with the experiment at both operating conditions. Once validated, trends in the model dynamics are investigated. The result is a discrete-time nonlinear control model which provides a platform for developing and validating various nonlinear control strategies. Figure 2 depicts the laboratory HCCI engine whose operating data are used for deriving the model. HCCI engine dynamics in discrete-time is well-developed in [7] and it is given in (84) through (105) as a MIMO nonaffine nonlinear discrete-time. The system dynamics are derived from cycle-by-cycle information of the thermo-dynamical behavior of the engine. The model is represented as follows:

$$\alpha_{i,k+1} = 0.091255843297\, T_{in,k}$$

$$\times\left[\left(c_{2,k-1}\alpha_{i,k} + c_{egr,k-1}\alpha_{e,k-1} + c_{4,k-1} - R_{u1}N_{3,k}\right)T_{1,k}V_{23,k} \Big/ \left(Z_1\, d\, V_1\right)\right]^{1-\lambda/\gamma}$$

$$\times\left[c_{2,k-1}\alpha_{i,k} + c_{egr,k-1}\alpha_{e,k-1} + c_{4,k-1} - R_{u1}N_{3,k}\right]\Big/ d \tag{84}$$

$$T_{1,k+1} = \Big\{c_{1,k}T_{in,k} + c_{egr,k}\alpha_{e,k}T_{egr} + c_{2,k}\alpha_{i,k+1}\chi \times$$

$$\left[\left(c_{2,k-1}\alpha_{i,k} + c_{egr,k-1}\alpha_{e,k-1} + c_{4,k-1} - R_{u1}N_{3,k}\right)T_{1,k}V_{23,k} \div (Z_1 dV_1)\right]^{\gamma-1/\gamma} \times$$

$$\left[d / (c_{2,k-1}\alpha_{i,k} + c_{egr,k-1}\alpha_{e,k-1} + c_{4,k-1} - R_{u1}N_{3,k})\right]\Big\} \div \left(c_{1,k} + c_{2,k}\alpha_{i,k+1} + c_{egr,k}\alpha_{e,k}\right) \tag{85}$$

$$\Delta\theta_{k+1} = 0.0174533 \times [2.0677 \times 10^{-18}\left(0.0000351555\right)^{\varphi_k} \times$$

$$\left(0.992961373\right)^{\left(V_1/V_{SOC,k+1}\right)^{\gamma-1}T_{1,k+1}}\left(1.16093521\right)^{\theta_{SOC,k+1}\left(180/\pi\right)}] \tag{86}$$

$$\alpha_{i,k+1} = 0.091255843297\, T_{in,k}$$

$$\times\left[\left(c_{2,k-1}\alpha_{i,k} + c_{egr,k-1}\alpha_{e,k-1} + c_{4,k-1} - R_{u1}N_{3,k}\right)T_{1,k}V_{23,k} \Big/ \left(Z_1\, d\, V_1\right)\right]^{1-\lambda/\gamma}$$

$$\times\left[c_{2,k-1}\alpha_{i,k} + c_{egr,k-1}\alpha_{e,k-1} + c_{4,k-1} - R_{u1}N_{3,k}\right]\Big/ d \tag{87}$$

$$T_{1,k+1} = \left\{c_{1,k}T_{in,k} + c_{egr,k}\alpha_{e,k}T_{egr} + c_{2,k}\alpha_{i,k+1}\chi \times\right.$$

$$\left[\left(c_{2,k-1}\alpha_{i,k} + c_{egr,k-1}\alpha_{e,k-1} + c_{4,k-1} - R_{u1}N_{3,k}\right)T_{1,k}V_{23,k} \div (Z_1 d V_1)\right]^{\gamma-1/\gamma} \times$$

$$\left.\left[d / (c_{2,k-1}\alpha_{i,k} + c_{egr,k-1}\alpha_{e,k-1} + c_{4,k-1} - R_{u1}N_{3,k})\right]\right\} \div \left(c_{1,k} + c_{2,k}\alpha_{i,k+1} + c_{egr,k}\alpha_{e,k}\right) \tag{88}$$

$$\Delta\theta_{k+1} = 0.0174533 \times [2.0677 \times 10^{-18}\left(0.0000351555\right)^{\varphi_k} \times$$

$$\left(0.992961373\right)^{\left(V_1/V_{SOC,k+1}\right)^{\gamma-1}T_{1,k+1}}\left(1.16093521\right)^{\theta_{SOC,k+1}\left(180/\pi\right)}] \tag{89}$$

$$\theta_{23,k+1} = \left(\frac{\mathrm{K_{th}}\,\omega}{\mathrm{A}\,\varphi_k{}^{\mathrm{a}}\left[11\left(\alpha_{\mathrm{i},k+1}\left(1-\varphi_{k-1}\right)+1\right)\right]^{\mathrm{b}}}\right)\left(\frac{\mathrm{V_c R_{u\,2}}}{\mathrm{P_{atm}}}\right)^{\mathrm{a+b}} \times$$

$$\left(\left[\varphi_k + 52.36 + \alpha_{i,k+1}\left(4\varphi_{k-1} + 52.36\right) + \alpha_{e,k}\left(\varphi_k + 52.36\right)\right]\frac{T_{1,k+1}}{\mathrm{V_1}}\right)^{\mathrm{a+b}} \times$$

$$\exp\left[\mathrm{E_a}\Big/\left(\mathrm{V_1}/\mathrm{V_c}\right)^{\gamma-1}\mathrm{T_{1,k+1}}\right] + \Delta\theta_{k+1} + \theta_{IVC} + \theta_{offset} \tag{90}$$

$$\begin{aligned} d = {} & (c_{3,k-1} + \left(c_{1,k-1} + c_{2,k-1}\alpha_{i,k} + c_{egr,k-1}\alpha_{e,k-1} - R_{u1}N_{2,k}\right)\left(V_1/V_{23,k}\right)^{\gamma-1}T_{1,k} \\ & - \left(c_{1,k-1} - c_{4,k-1}\right)T_{ref}) \end{aligned} \tag{91}$$

$$Z_1 = \frac{N_{3,k}}{N_{2,k}} = \frac{4\left(\varphi_{k-1} + \varphi_{k-2}\alpha_{i,k}\right) + 52.36\left(1 + \alpha_{i,k}\right) + \alpha_{e,k-1}\left(\varphi_{k-1} + 52.36\right)}{\varphi_{k-1} + 52.36 + \alpha_{i,k}\left(4\varphi_{k-2} + 52.36\right) + \alpha_{e,k-1}\left(\varphi_{k-1} + 52.36\right)} \tag{92}$$

$$c_{1,k} = \varphi_k\,\overline{c}_{p\,C8H18,R} + 11\overline{c}_{p\,O2,R} + 41.36\,\overline{c}_{p\,N2,R} \tag{93}$$

$$c_{2,k} = 7\varphi_{k-1}\,\overline{c}_{p\,CO2,P} + 8\varphi_{k-1}\,\overline{c}_{p\,H2O,P} + 41.36\,\overline{c}_{p\,N2,P} + 11\left(1-\varphi_{k-1}\right)\overline{c}_{p\,O2,P} \tag{94}$$

$$c_{egr,k} = \overline{c_{p\,N2,E}}\left(\varphi_k + 52.36\right) \tag{95}$$

$$c_{3,k} = \varphi_k LHV_{C8H18}\left(1-\varepsilon\right) + 11\left(1-\varphi_k\right)\overline{c_{p\,O2,P}} \tag{96}$$

$$c_{4,k} = 7\varphi_k \overline{c_{p\,CO2,P}} + 8\varphi_k \overline{c_{p\,H2O,P}} + 41.36\,\overline{c_{p\,N2,P}} \tag{97}$$

$$V_{23,k} = V_c[1 + 0.5\left(r_c - 1\right)\times\left(R + 1 - \cos\left(\vartheta_{23,k}\right) - \sqrt{R^2 - \sin^2\left(\vartheta_{23,k}\right)}\right)] \tag{98}$$

$$N_{t,k} = P_{in}\left(V_d + V_c\right)/R_{u2}T_{in,k}\;\; N_{f,k} = gpm_k\,4\pi/60\,\omega\,MW_f \tag{99}$$

$$X_{egr,k} = \alpha_{e,k}\left(\varphi_{k-1} + 52.36\right)\div$$

$$\{\left(\varphi_{k-1} + 52.36\right)\left(1+\alpha_{e,k}\right) + \alpha_{i,k+1}\left(15\varphi_{k-1} + 41.36 + 11\left(1-\varphi_{k-1}\right)\right)\} \tag{100}$$

$$N_{egr,k} = X_{egr,k}N_{t,k}\;\;,\;\; N_{iegr,k} = \alpha_{i,k+1}N_{t,k}\;\;\text{, and}\;\; N_{a,k} = N_{t,k} - N_{egr,k} - N_{iegr,k}$$

$$, N_{fs,k} = N_{a,k}MW_aFA_s/\left(MW_f\right) \tag{101}$$

$$\varphi_k = N_{f,k} \div N_{fs,k} \tag{102}$$

$$\theta_{\mathrm{SOC,k+1}} = \theta_{IVC} + \theta_{offset} + \exp\left[\frac{\mathrm{E_a}}{\mathrm{T_{1,k+1}}}\left(\frac{\mathrm{V}_c}{\mathrm{V_1}}\right)^{\gamma-1}\right]\left(\frac{\mathrm{K_{th}}\,\omega}{\mathrm{A}\,\varphi_k{}^{\mathrm{a}}\left[11\left(\alpha_{\mathrm{i,k+1}}\left(1-\varphi_{k-1}\right)+1\right)\right]^{\mathrm{b}}}\right)\times$$

$$\left(\mathrm{V_c R_{u\,2}/P_{atm}}\right)^{\mathrm{a+b}}\times$$

$$\left(\left\{\left[\varphi_k + 52.36 + \alpha_{i,k+1}\left(4\varphi_{k-1} + 52.36\right) + \alpha_{e,k}\left(\varphi_k + 52.36\right)\right]T_{1,k+1}/\mathrm{V_1}\right\}\right)^{\mathrm{a+b}} \tag{103}$$

$$V_{SOC,k+1} = V_c[1 + 0.5\left(r_c - 1\right)\times\left(R + 1 - \cos\left(\vartheta_{SOC,k+1}\right) - \sqrt{R^2 - \sin^2\left(\vartheta_{SOC,k+1}\right)}\right)] \tag{104}$$

$$P_{SOC,k} = P_{in}\left(V_1/V_{soc,k}\right)^{\gamma} \tag{105}$$

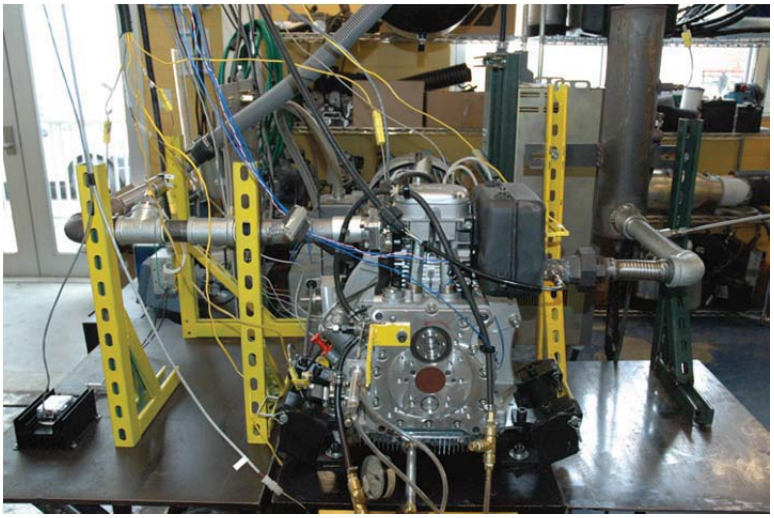

Figure 2. Laboratory version of the HCCI engine.

All the parameters are introduced in Table 1. In this model, it is assumed that C_7H_{16} is used as the engine fuel. The HCCI model inputs are the intake air-fuel mixture temperature $T_{in}(k)$ and the fuel rate $gpm(k)$. The system outputs are assumed to be crank angle $\theta_{23,k}$ and $P_{3,k}$ peak pressure. The objective of controlling HCCI engines is usually to regulate the output while minimizing a given value function.

By observing the system dynamics (84)-(105), it is clear that the system is nonaffine, MIMO and has uncertain dyanamics. Therefore, the proposed controller from the previous sections can be applied to the HCCI engine representation. In this section, some simulation results are provided to show the significance of the proposed approach. It is required to have an admissible controller Δu_k^o. Then, the update law (16), (36), (50), and (59) will converge to a optimal control signal Δu_k^*. The dispersion noise is injected to the system dynamics (as actuator and sensor noise) to provide persistence of excitation for the system identification and cost function estimation.

Now, we consider two different cases to observe the improvement made by the online near optimal controller. The first case is when (16), (36), (50), and (59) are not updated; and the second case is when they are updated and Δu_k^o converges to Δu_k^*. Figure 3 illustrates the convergence of the closed loop system with the two above cases when the set point is $(\theta_{23,k}, P_3) = (365CAD, 0.55KN/cm^2)$. Moreover, we investigate the index function (28) where $Q = 1$ and $R = 1$. This figure clearly shows that the trajectory selected by the online optimal controller is much shorter.

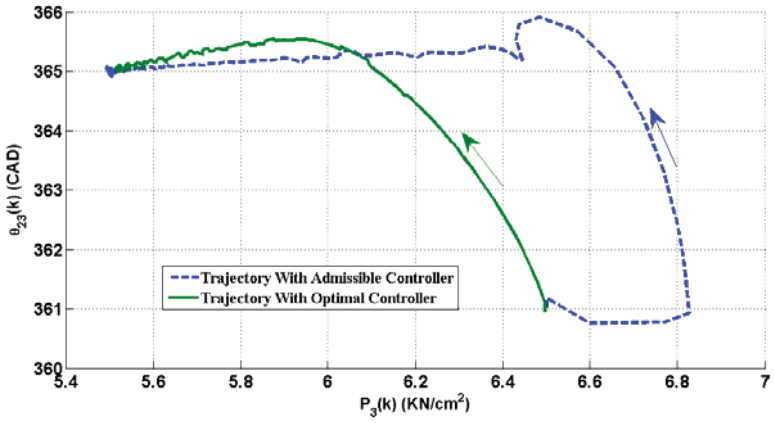

Figure 3. Convergence of the closed loop system for $(\theta_{23,k}, P_3) = (365CAD, 0.55KN/cm^2)$ with the initial admissable and suboptimal controllers.

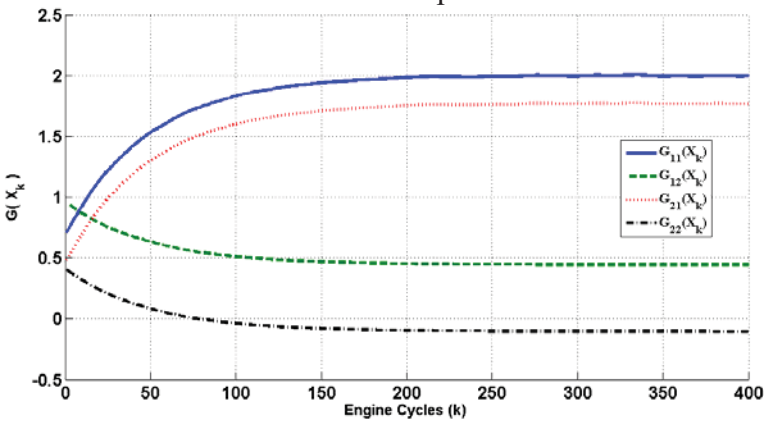

Figure 4.Convergence of $G(X_k)$

Assume that both of the controllers have been operating at steady state mode. Under this scenario, the estimation of $G(X_k)$ is illustrated in Figure 4. Figure 5 makes a

comparison between the control inputs applied by the initial admissible and optimal controllers. Figure 5 shows that the optimal controller starts with a same behavior as that of the admissible controller, whereas after about 200 engine cycles, the control effort begins to decrease in magnitude when compared to the original admissible controller. This figure shows that the optimal controller spends significantly smaller control effort while keeping the system stable.

The injected process and sensor noise will not allow the system to converge to its set point whereas the output variation at this set point can be viewed as the dispersion reduction ability. Figure 6 compares two controllers in terms of reduction in dispersion where $P_{3,k}$ is plotted versus $\theta_{23,k}$. Compared with the initial admissible controller, the optimal controller, after a transient behavior, shows significant smaller variation around the equilibrium point.

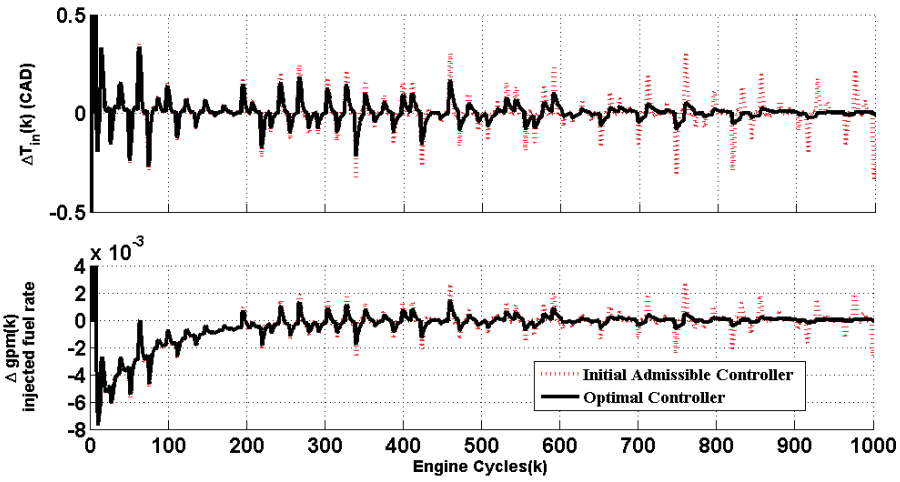

Figure 5. Performance comparison of the initial admissible and the optimal controller.

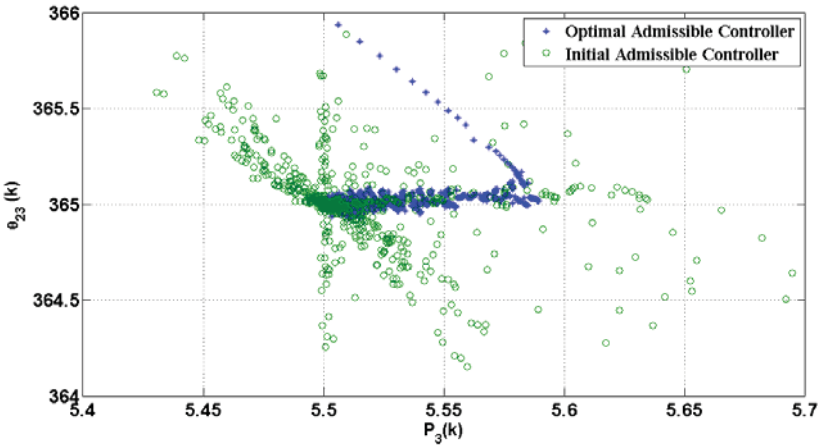

Figure 6. Comparison of initial admissible and the suboptimal controllers.

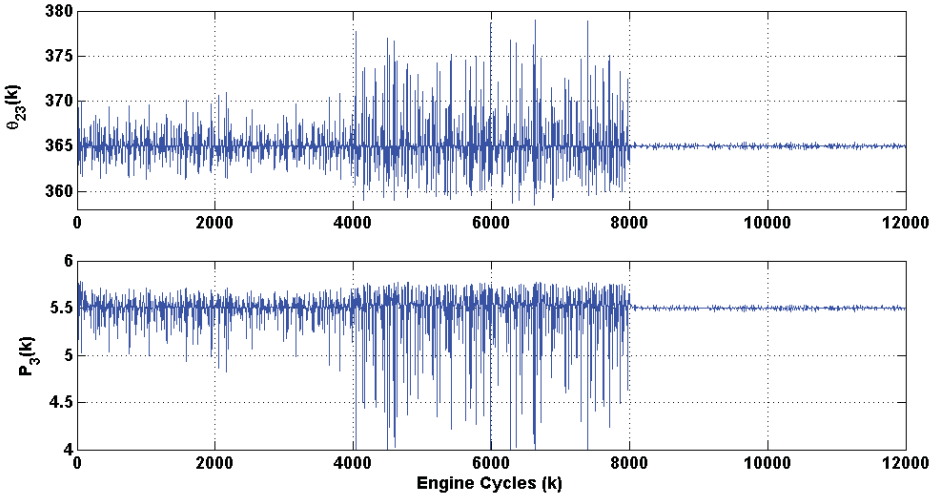

Figure 7. Comparison among open loop, admissible, and the sub-optimal controllers when the setpoint is ($P_3 = 0.55$, $\theta_{23} = 370$): the controller switches from open-loop to admissible at k=400; then, to the sub-optimal controller at k=800.

In another numerical experiment, we compare open loop, admissible, and the optimal controllers in terms of their cyclic dispersion reduction abilities. Over all numerical examinations k=0.008333s (the engine speed which is 188.49 rpm), $\alpha = 0.005$, $\alpha_u = 0.001$, $\sigma(X_k) = \{x_1^2, x_1x_2, x_2^2, x_1^4, x_1^3x_2, ..., x_2^6\}$, $\Psi(X_k) = \{1, ..., 1, \sin(X), \sin(2X), ..., \tanh(X), \tanh(2X), ...\} \in \mathbb{R}^{16\times 32}$, $\vartheta = \{\tanh(X), \tanh(2x), ...\} \in \mathbb{R}^{16\times 32}$, $\alpha_\zeta = 0.1$, and the initial admissible controller is chosen as $-0.1 \times [(\theta_{23} - \theta_{23d}), (P_3 - P_{3d})]^T$.

To this end, we assume that initially the engine is considered to be running in the open-loop mode for 400 engine cycles, then it switches to an admissible controller for the same duration of run, and finally, to the optimal controller. Figure 7 shows the result of this experiment for set point $(P_3 = 0.55\,KN/cm^2, \theta_{23} = 365CAD)$. It is obvious that the optimal controller significantly reduces the dispersion when compared with the open-loop and the admissible controllers. We conclude from Figure 7 that the proposed controller shows the best results in terms of output dispersion reduction.

VI. CONCLUSIONS

This paper presents an online robust optimal control of unknown nonaffine nonlinear discrete-time systems by using inputs and outputs. The stability of the closed loop system is shown under the assumption that an admissible controller is available and it is updated until it converges to an optimal controller. The optimal controller includes three separate neural networks: 1) the cost approximation NN; 2) the optimal controller NN; 3) the NN identifier. Moreover, it is shown that we can mitigate the modeling errors due to higher order terms by using a robust term. The net result is the design of a robust optimal controller by using output feedback. As an application, the approach is applied to the MIMO nonaffine representation of an HCCI engine that is validated experimentally. The simulation results show a significant reduction in the cyclic dispersion when the robust optimal controller is utilized.

REFERENCES

[1] Dierks T, Thumati B T, Jagannathan S. Optimal control of unknown affine nonlinear discrete-time systems using offline-trained neural networks with proof of convergence. Neural Networks 2008; 22(5-6):851-860.

[2] Beard R, Saridis G, Wen J. Improving the performance of stabilizing controls for nonlinear systems. IEEE Control Systems Magazine 1996; 16(5):27 – 35.

[3] Dierks T, Jagannathan S. Optimal control of affine nonlinear discrete-time systems with unknown internal dynamics using online approximation. in Proc. Joint 48th IEEE Conference on Decision and Control and 28th Chinese Control Conference 2009; 6750-6755.

[4] Prokhorov D, Wunsch D, Adaptive critic design. IEEE Transactions on Neural Networks 1997; 8(5): 997-1007.

[5] Murray J J, Cox C J, Lendaris G G. Adaptive dynamic programming. IEEE Transaction on Systems, Man, and Cybernetics 2002; 32(2): 140-153.

[6] Al-Tamimi A, Lewis F L, Abu-Khalaf M. Discrete-time nonlinear HJB solution using approximate dynamic programming: convergence proof. IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics 2008. 38(4): 943-949.

[7] Jagannathan S. Neural network control on nonlinear discrete-time systems. CRC press, 2006.

[8] Yao M, Zhaolei Z, Haifeng L. Progress and recent trends in homogeneous charge compression ignition (HCCI) engines. Progress in Energy and Combustion Science 2009; 35(5): 398-437.

[9] Stanglmaier R, Charles R. Homogeneous charge compression ignition (HCCI): benefits, compromises, and future engine applications. SAE paper 1999.

[10] Juttu S, Thipse S, Marathe N, Gajendra B M. Homogeneous Charge Compression Ignition (HCCI): a new concept for near zero NO_x and particulate matter from diesel engine combustion. SAE paper 2007.

[11] Bettis J, Massey J, Drallmeier J, Jagannathan S. Thermodynamic based modeling for nonlinear control of combustion phasing in HCCI. Technical Meeting Champaign 2010; 21-23.

[12] Hartman P. Ordinary differential equations 2nd edition. SIAM, 2002.

[13] Lewis F L, Jagannathan S, Yeşildirek A. Neural network control of robot manipulators and nonlinear systems. CRC Press, 1999.

[14] Beard R W, McLain T W. Successive Garlerkin approximation algorithms for nonlinear optimal and robust control. Int. J. Control 1998; 71(5): 717-743.

[15] Chen Z, Jagannathan S. Generalized Hamilton-Jacobi-Bellman formulation based neural network control of affine nonlinear discrete-time systems. IEEE Trans. On Neural Networks 2008; 19(3): 90-106.

[16] Lewis F L, Syrmos V L. Optimal Control 2nd ed. Hoboken, NJ: Wiley, 1995.

[17] W. Rudin, Principles of Mathematical Analysis, 3rd ed. McGraw-Hill: New York, 1976.

[18] Hovakimyan N, Lavretsky E, Sasane A. Dynamic inversion for nonaffine-in-control systems via time-scale separation. Part I-II. Journal of Dynamical and Control Systems 2007; 13(4): 451-465.

[19] Park K, Lim J. Time-scale separation of nonlinear singularly perturbed discrete systems. Control Automation and Systems (ICCAS) 2010; 892-895.

[20] Juang J N, Phan M Q. Identification and Control of Mechanical Systems. Cambridge University Press, 2001.

[21] Stefanovski J. Feedback affinization of nonlinear control systems. Systems and Control Letters 2002; 46: 207-217.

[22] Khalil H. Nonlinear systems. Prentice-Hall, 2002.

[23] Boutalis Y, Theodoridis D C, Christodoulou M A. A New Neuro-FDS Definition for Indirect Adaptive Control of Unknown Nonlinear Systems Using a Method of Parameter Hopping. Neural Networks, IEEE Transactions on 2009; 20(4): 609-625.

[24] Theodoridis D, Boutalis Y, Christodoulou M. A new adaptive Neuro-Fuzzy controller for trajectory tracking of robot manipulators. International Journal of Robotics and Automation 2011; 26(1):1-12.

II. A DISCRETE-TIME EXTREMUM SEEKING METHOD COUPLED WITH OPTIMAL ADAPTIVE CONTROLLER FOR NONLINEAR DISCRETE TIME SYSTEMS WITH APPLICATION TO EFFICIENCY OPTIMIZATION OF HCCI ENGINES

H. Zargarzadeh, *Student Member, IEEE*, S. Jagannathan, *Senior Member, IEEE, and* J. A. Drallmeier

SUMMARY– Identifying an optimal set point is important for a number of control applications once a system is stabilized. This optimal set point is determined by finding the extremum of an output or performance function of an unknown nonlinear system. Therefore, this paper introduces an extremum seeking method with an optimal adaptive stabilizing controller for nonlinear nonaffine discrete-time systems. First, a novel averaging method is used for the nonlinear discrete-time system to show that the unique extremum points are stable equilibrium points. Then, a singular perturbation method in discrete-time is employed to show that the overall closed-loop system will dynamically converge to the extremum. Finally, as an example, the proposed approach is applied to identify a set point which maximizes the performance for a generic linear multivariable system and in terms of efficiency of the HCCI engines which are represented by non-affine dynamics with an uncertain output function. This approach is able to find an operating point that not only maximizes efficiency, but also minimizes the pressure rise rates of the cylinder due to engine operating constraints.

1. INTRODUCTION

Identifying a suitable operating point by maximizing an output function, which is a function of system states or parameters, is of great importance for a nonlinear system besides controlling the system around it in a stable manner. Given an output function and any associated constraints, it is always desirable to optimize the system performance by

choosing the best operating point. It is well known in the literature that optimal control is able to minimize a cost function in either direct [1] or inverse [2] manner, where a desired set point or trajectory is given beforehand. Nonetheless, in a wide class of control problems, the operating point that optimizes the plant performance is unknown and need to be identified online. In many cases, the extremum point of the output function of a nonlinear function is utilized to select this operating point and it may be unknown. Therefore, in the literature, self-optimization, extremum control, or extremum seeking approaches [16] are utilized for identifying a suitable operating set point.

Extremum seeking word is coined in 1922, a few decades before the introduction of linear adaptive control methods. Since extremum seeking methods are adaptive against the performance function uncertainties, authors in [18] tend to introduce them as the first adaptive control method reported in the literature. Since then, extremum seeking has been widely applied to engineering systems. For example, maximum power point tracking of photovoltaic systems [20], soft landing of electromagnetic actuators [8], PID tuning [9], thermoacoustic coolers [17] and so on are derived by using extremum seeking techniques or its variants. Moreover, extremum seeking has been considered in automotive applications such as antilock breaking system [11], combustion instability [12], and optimization of variable cam timing engine operation [14].

This paper considers the problem of extremum seeking of nonlinear discrete-time systems whose output function is uncertain. In order to address this issue, in [19], a nonlinear plant represented as a cascade combination of linear dynamics and a static nonlinearity is considered. In contrast, in this paper, a nonlinear dynamical system with a nonlinear state-to-output mapping is considered with an optimal adaptive stabilizing

controller in the inner loop. It is assumed that the closed loop system has a faster inner loop and a slower outer loop. The nonlinear discrete-time plant is stabilized by using the fast inner loop, whereas a slow extremum seeking outer loop is employed to find the optimum set point.

We first show the stability of the outer loop using averaging analysis [15][24] provided the inner loop is fast enough to follow any desired set point which is considered a mild and customary assumption for extremum seeking methods [18] due to employing singular perturbation method [24] for the stability proof . By viewing the state-to-output map as a output or performance function, which has a unique extremum, it was shown that the proposed method is able to locally converge to this set point. The stability of the overall dynamical system is examined by using singular perturbation method for nonlinear discrete-time systems [25]. It is shown that the overall closed loop method converge to the optimum set point with a uniformly ultimately bounded (UUB) stability.

To solve the problem, a nonlinear discrete-time system is considered which can be stabilized at any desired set point with UUB stability. This can be guaranteed by any controller e.g. a neural network (NN) based robust optimal adaptive controller [6]. Then the objective is to design an outer loop that is able to find the set point which renders an extremum for a predefined yet unknown function of the system states. It is shown that the set of inner and the outer loops will form a singularly perturbed system. The proof of the stability of the singular perturbed discrete-time systems is provided in [25] for asymptotically stable systems. In this paper, these results are extended to show that the overall nonlinear discrete-time system will remain UUB when the proposed nonlinear

extremum seeking method is applied to a system that is UUB. The proposed approach is verified initially on a nonlinear system.

Low temperature combustion mode engines, such as Homogeneous Charge Compression Ignition (HCCI) engines, represent a promising trend to increase the efficiency and significantly reduce the emissions of internal combustion engines [2]. HCCI engines are one of the complicated mechanical nonlinear systems, due to the dependence of the combustion event on chemical kinetics rather than an external trigger and control of such systems becomes a challenge [6] due to uncertain engine dynamics.

In [6], the authors considered the problem of optimally controlling the engine while the system dynamics is fully unknown, although the problem of choosing the best operating point that maximizes the performance is predetermined. In fact, it is desired to find a suitable crank angle that maximizes a predefined performance function while the cylinder pressure rise rare (PRR) is kept below a safe threshold. By defining a suitable performance objective, the HCCI engine becomes a practical example for implementation of the extremum approach in discrete-time which is developed in this paper. Numerical results are shown on an experimentally validated engine model to verify the theoretical claims.

In fact, we assume that an internal control loop is used from [6] to provide the ability of driving the nonlinear system to any desired setpoint. Then the proposed method is used to design an external control loop by extremum seeking in order to find the best set point that maximizes the performance function.

The paper is organized as the following. After the introduction, Section II is dedicated to demonstrate the theoretical results to the stability of the proposed extremum

seeking method. Section III introduces the linear multivariable systems and HCCI engines and the problem of maximization of performance. Finally, Section IV provides results to numerically verify the results in the previous sections.

2. A DISCRETE TIME EXTREMUM SEEKING METHOD FOR NONLINEAR SYSTEMS

Consider a general class of nonlinear discrete-time systems in the following representation provided by

$$x_{k+1} = f(x_k, u_k)$$

$$y_k = h(x_k), \tag{1}$$

where $u_k \in E_u \subset \mathbb{R}^m$, $x_k \in E_x \subset \mathbb{R}^n$, and $y_k \in E_y \subset \mathbb{R}$ represent the system input, states, and the outputs respectively, and $f(\bullet)$ and $h(\bullet)$ are assumed to be unknown continuous nonlinear functions with respect to x_k and u_k. We assume that there exist a stabilizing controller $u_k = \alpha(x_k, \vartheta_k) \in E_u \subset \mathbb{R}^m$ such that

$$x_{k+1} = f(x_k, \alpha(x_k, \vartheta_k)), \tag{2}$$

$$y_k = h(x_k) \tag{3}$$

is uniformly ultimately bounded (UUB) with respect to the equilibrium point $x_k = \ell(\vartheta_k)$ where $\vartheta_k \in \mathbb{R}$ is an adjustable or a design parameter. Without loss of generality, we assume that the state to output mapping can be represented as

$$h(x_k) = h^* - (x_k - x^*)^2, \tag{4}$$

where $h^* \in E_y$ is a unique extremum of the output function and $x^* \in E_x$ is the value for the system state that results in $h(x^*) = h^*$. The objective is to find ϑ^* such that $x^* \equiv \ell(\vartheta^*)$. The following assumptions are necessary in order to proceed.

Assumption 1 [4]: There exists a smooth function $\ell : \mathbb{R} \to \mathbb{R}^n$ such that $f(x_k, \alpha(x_k, \vartheta_k)) = 0$ if and only if $x_k \equiv \ell(\vartheta)$, that is, a unique equilibrium point can be found depending upon ϑ.

Assumption 2: For each $\vartheta \in \mathbb{R}$, the equilibrium point $x \equiv \ell(\vartheta)$ is UUB.

Assumption 3: There exist $\vartheta^* \in \mathbb{R}$ such that $\left(h \circ \ell\right)'(\vartheta^*) = 0$ and $(h \circ \ell)''(\vartheta^*) < 0$, where the $h \circ \ell(\bullet) \triangleq h(\ell(\bullet))$ denotes the composition of $h(\bullet)$ and $\ell(\bullet)$.

In Assumption 1, uniqueness of $x_k \equiv \ell(\vartheta)$ is required to guarantee the uniqueness of the system trajectory [18]. Assumption 2 provides the stability of the equilibrium point that is required since the objective is to find the equilibrium point which maximizes the output performance function. Assumption 3 is a standard assumption that guarantees the uniqueness of the extremum point of the output function. It should be noted that the extremum seeking method solely deals with the systems having unique extremum points. The property of (4) also justifies this assumption [18][19].

The block diagram of the proposed extremum seeking approach is illustrated in Figure 1. To start, we write the system dynamics as

$$x_{k+1} = f(x_k, \alpha(x_k, \hat{\vartheta}_k + a\sin(\omega k))) \tag{5}$$

$$\hat{\vartheta}_{k+1} = \hat{\vartheta}_k - \gamma b\left(\eta_k + y_k\right)\sin(\omega k) \tag{6}$$

$$\eta_{k+1} = -\varpi\eta_k - y_k(1+\varpi) \tag{7}$$

Now define $\tilde{\vartheta}_k = \hat{\vartheta}_k - \vartheta^*$ and $\tilde{\eta}_k = \eta_k + h \circ \ell(\vartheta^*)$. The system dynamics are rewritten as

$$x_{k+1} = f(x_k, \alpha(x_k, \tilde{\vartheta}_k + \vartheta^* + a\sin(\omega k))) \tag{8}$$

$$\tilde{\vartheta}_{k+1} = \tilde{\vartheta}_k - \gamma b\left(\tilde{\eta}_k - h \circ \ell(\theta^*) + h(x_k)\right)\sin(\omega k) \quad (9)$$

$$\tilde{\eta}_{k+1} - h \circ \ell(\theta^*) = -\varpi(\tilde{\eta}_k - h \circ \ell(\vartheta^*)) - h(x_k)(1+\varpi) \quad (10)$$

or

$$x_{k+1} = f(x_k, \alpha(x_k, \tilde{\vartheta}_k + \vartheta^* + a\sin(\omega k))) \quad (11)$$

$$\tilde{\vartheta}_{k+1} = \tilde{\vartheta}_k - \gamma b\left(\tilde{\eta}_k + h(x_k) - h \circ \ell(\vartheta^*)\right)\sin(\omega k) \quad (12)$$

$$\tilde{\eta}_{k+1} = -\varpi\tilde{\eta}_k - \left(h(x_k) - h \circ \ell(\vartheta^*)\right)(1+\varpi)\,. \quad (13)$$

The overall feedback system (11)-(13) has two time scales. The stabilized plant (11) dynamics are at a *faster time scale*, while the periodic perturbation and filter dynamics respectively in extremum seeking (12)(13) are at a *slower time-scale* [19][25][24].

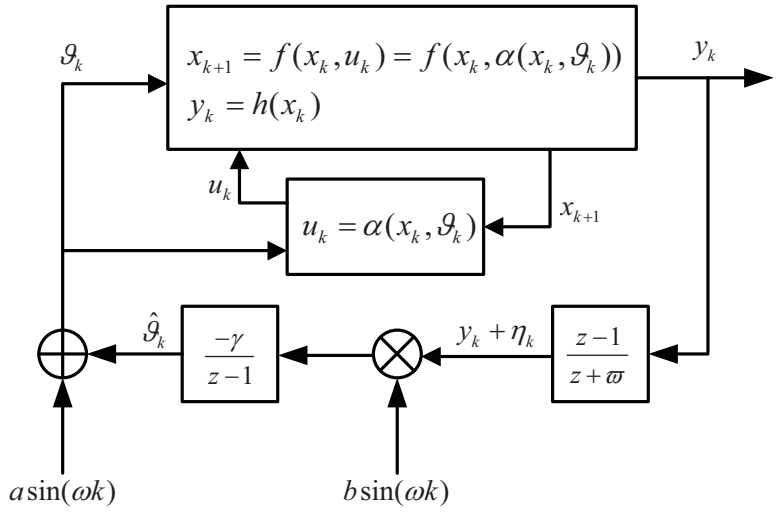

Figure. 1. Block diagram representation of the proposed extremum seeking scheme.

The set of equations (12) and (13) constitute a dynamic system due to the extremum seeking outer loop. Now, using *averaging* analysis, it is necessary to show that the closed loop system is stable around the unique extremum point of (4) which is also a

stable equilibrium point of (2) when ϑ^* is applied. The second aspect is to show that the closed loop is able to converge to the neighborhood close to the unique extremum using *singular perturbation* method [25][24].

In singularly perturbed systems, it is not always easy to determine whether or not the fast subsystem is fast enough or the slow subsystem is slow enough (pp.424, [24]). In fact, it is not always clear how to pick the design parameters for the closed loop in order to be considered slow or fast. However, in many applications, our knowledge of the system dynamics can help in the right choice if design parameters. Methods of selecting parameters are shown in Khalil (pp.423-429, [24]).

2.1. Averaging Analysis

Now, considering the state x_k at its equilibrium point $\ell(\vartheta^* + \tilde{\vartheta}_k + a\sin(\omega k))$ and substituting it in the slower dynamics given by (12)-(13) which forms the *reduced system* [24] as

$$\tilde{\vartheta}_{k+1} = \tilde{\vartheta}_k - \gamma b\Big(\tilde{\eta}_k + \upsilon(\tilde{\theta}_r + a\sin(\tau))\Big)\sin(\tau) \tag{14}$$

$$\tilde{\eta}_{k+1} = -\varpi\tilde{\eta}_k - \upsilon(\tilde{\theta}_r + a\sin(\tau))(1+\varpi), \tag{15}$$

where

$$\upsilon(\tilde{\theta}_r + a\sin(\tau)) = h\circ\ell(\vartheta^* + \tilde{\vartheta}_k + a\sin(\tau)) - h\circ\ell(\vartheta^*). \tag{16}$$

Here, in the light of Assumption 3, it is obvious that $\upsilon(0) = 0$ and $\upsilon'(0) = 0$. Without loss of generality ω can be represented as $\omega = 2\pi/n$ with $n \in \mathbb{N}$ being a natural number. Then, by averaging (14) and (15) over n sampling instants to get

$$\tilde{\vartheta}^a_{k+1} = \tilde{\vartheta}^a_k - \frac{\gamma b}{n}\Big(\sum\nolimits_{i=1}^{n}\upsilon(\tilde{\theta}^a_k + a\sin(\omega i))\sin(\omega i) + \tilde{\eta}_k\sum\nolimits_{i=1}^{n}\sin(\omega i)\Big)$$

$$= \tilde{\vartheta}_k^a - \frac{\gamma b}{n}\sum_{i=1}^{n} \upsilon(\tilde{\theta}_k^a + a\sin(\omega i))\sin(\omega i) \tag{17}$$

$$\tilde{\eta}_{k+1}^a = -\varpi\tilde{\eta}_k^a - \frac{(1+\varpi)}{n}\sum_{i=1}^{n} \upsilon(\tilde{\theta}_k^a + a\sin(\omega i)) \ . \tag{18}$$

Here, and in the sequel, the superscript "a" represents the parameter after averaging process. Obviously, an *averaging equilibrium point* of (17) and (18) can be represented as $\tilde{\vartheta}_k^{a,e} \equiv \tilde{\vartheta}_{k+1}^a = \tilde{\vartheta}_k^a$ and $\tilde{\eta}_k^{a,e} \equiv \tilde{\eta}_{k+1}^a = \tilde{\eta}_k^a$. Obviously, from (17) and (18), the stability of this equilibrium point is guaranteed when the following conditions are satisfied:

$$0 \equiv -\gamma b\sum_{i=1}^{n} \upsilon(\tilde{\vartheta}_k^{a,e} + a\sin(\omega i))\sin(\omega i) \tag{19}$$

$$\tilde{\eta}_k^{a,e} \equiv -\frac{1}{n}\sum_{i=1}^{n} \upsilon(\tilde{\vartheta}_k^{a,e} + a\sin(\omega i)) \tag{20}$$

Now writing (19) and (20) to get

$$0 = \sum_{i=1}^{n} \upsilon(\tilde{\vartheta}_k^{a,e} + a\sin(\frac{2\pi}{n} i))\sin(\frac{2\pi}{n} i) \tag{21}$$

$$\tilde{\eta}_k^{a,e} = -\frac{1}{n}\sum_{i=1}^{n} \upsilon(\tilde{\vartheta}_k^{a,e} + a\sin(\frac{2\pi}{n} i)) \tag{22}$$

It is obvious that the solution of (21) and (22) are only a function of a as the magnitude of the sinusoidal perturbation $a\sin(\omega k)$ in Figure 1. Therefore, by using Taylor series expansion, $\tilde{\vartheta}_k^{a,e}$ can be expressed as

$$\tilde{\vartheta}_k^{a,e} = b_1 a + b_2 a^2 + b_3 a^3 + O(a^4) \tag{23}$$

Therefore,

$$\upsilon(\tilde{\vartheta}_k^{a,e} + a\sin(\frac{2\pi}{n}i)) = \upsilon(b_1 a + b_2 a^2 + O(a^3) + a\sin(\frac{2\pi}{n}i))$$

$$= \upsilon(0) + \upsilon'(0)(b_1 a + b_2 a^2 + O(a^3) + a\sin(\frac{2\pi}{n}i))$$

$$+\frac{1}{2}\upsilon''(0)(b_1 a + b_2 a^2 + O(a^3) + a\sin(\frac{2\pi}{n}i))^2 + O(a^3)$$

$$= \frac{1}{2}\upsilon''(0)(b_1 a + b_2 a^2 + O(a^3) + a\sin(\frac{2\pi}{n}i))^2 + O(a^3)\ . \tag{24}$$

Now, from (21) one can get

$$0 = \sum_{i=1}^{n}\begin{bmatrix}\upsilon''(0)(b_1 a + b_2 a^2 + b_3 a^3 + O(a^3)) \\ +a\sin(\frac{2\pi i}{n}))^2\sin(\frac{2\pi i}{n})\end{bmatrix} + \sum_{i=1}^{n} O_i(a^3)\sin(\frac{2\pi}{n}i)$$

$$= \upsilon''(0)b_1 a^2\sum_{i=1}^{n}\sin^2(\frac{2\pi}{n}i) + \frac{1}{2}\upsilon''(0)a^2\sum_{i=1}^{n}\sin^3(\frac{2\pi}{n}i)$$

$$+\frac{1}{2}\upsilon''(0)b_1^2 a^2\sum_{i=1}^{n}\sin(\frac{2\pi}{n}i)$$

$$+\frac{1}{6}\upsilon'''(0)\sum_{i=1}^{n}\left[b_1 a + b_2 a^2 + b_3 a^3 + O(a^3) + a\sin(\frac{2\pi}{n}i)\right]^3\sin(\frac{2\pi}{n}i)$$

$$+\sum_{i=1}^{n} O_i(a^4)\sin(\frac{2\pi}{n}i) \tag{25}$$

or

$$0 = \upsilon''(0)b_1\sum_{i=1}^{n}\sin^2(\frac{2\pi}{n}i) + \frac{1}{2}\upsilon''(0)\sum_{i=1}^{n}\sin^3(\frac{2\pi}{n}i)$$

$$+\frac{1}{2}\upsilon''(0)b_1^2\sum_{i=1}^{n}\sin(\frac{2\pi}{n}i) + \frac{1}{6}a\upsilon'''(0)\sum_{i=1}^{n}[b_1 + b_2 a + b_3 a^2 + O(a^3)$$

$$+\sin(\frac{2\pi i}{n})]^3\sin(\frac{2\pi i}{n}) + \sum_{i=1}^{n} O(a^4)\sin(\frac{2\pi}{n}i) \tag{26}$$

Obviously, $O(a^4) \cong 0$ when a is chosen small enough, and it is necessary that $b_1 = b_2 = b_3 = 0$ in order to ensure (26) holds. Therefore, we conclude that the averaging equilibrium point for the system (17)-(18) can be represented as

$$\tilde{\theta}_k^{a,e} = O(a^4) \cong 0 \tag{27}$$

$$\tilde{\eta}_k^{a,e} = -\frac{1}{n}\sum_{i=1}^{n} \upsilon(\tilde{\vartheta}_k^{a,e} + a\sin(\frac{2\pi}{n}i))$$

$$= -\frac{1}{2n}\sum_{i=1}^{n} \upsilon''(0)(O(a^4) + a\sin(\frac{2\pi}{n}i))^2 - O(a^3)$$

$$= -\frac{a^2}{2n}\upsilon''(0)\sum_{i=1}^{n} \sin^2(\frac{2\pi}{n}i) + O(a^3) \ . \tag{28}$$

With $n \geq 3$, we can write

$$\frac{1}{n}\sum_{i=1}^{n} \sin^2(2\pi i / n) = \frac{1}{n}\sum_{i=1}^{n} [\frac{1}{2} - \frac{1}{2}\cos(4\pi i / n)] = \frac{1}{2}. \tag{29}$$

Similarly we have

$$\frac{1}{n}\sum_{i=1}^{n} \sin^4(2\pi i / n) = \frac{1}{4n}\sum_{i=1}^{n} [1 - \cos(4\pi i / n)]^2 = \frac{1}{4} \times \frac{3}{2} = \frac{3}{8}. \tag{30}$$

Therefore, (27) and (28) imply that the equilibrium point of the system (17) and (18) can be represented as

$$\tilde{\theta}_k^{a,e} = O(a^4) \cong 0 \tag{31}$$

and

$$\tilde{\eta}_k^{a,e} = -\frac{a^2}{4}\upsilon''(0) + O(a^3) = O(a^2) \cong 0. \tag{32}$$

By choosing a properly small enough value for the amplitude a, the above analysis presents that the closed-loop reduced order system has a unique equilibrium

point arbitrarily close to the origin. The next step is to study the stability of this equilibrium point. Now, the Jacobian of (17) and (18) can be written as

$$J_r^a = \begin{bmatrix} 1 & \frac{\gamma b}{n}\sum_{i=1}^{n} \upsilon(a\sin(\frac{2\pi}{n}i))\sin(\frac{2\pi}{n}i) \\ -\frac{(1+\varpi)}{n}\sum_{i=1}^{n} \upsilon(a\sin(\frac{2\pi}{n}i)) & -\varpi \end{bmatrix}$$ which in

turn can be expressed as

$$J_r^a \cong \begin{bmatrix} 1 & \xi_1 \\ -\xi_2 & -\varpi \end{bmatrix} \tag{33}$$

where

$$\xi_1 \equiv \frac{\gamma a^3 b}{6n}\upsilon'''(0)\sum_{i=1}^{n}\sin^4(\frac{2\pi}{n}i) \tag{34}$$

$$\xi_2 \equiv \frac{(1+\varpi)}{n}a^2\upsilon''(0)\sum_{i=1}^{n}\sin^2(\frac{2\pi}{n}i)\ . \tag{35}$$

Using (29) and (30), (34) and (35) can be rewritten as

$$\xi_1 \cong \gamma a^3 b\upsilon'''(0)/16 \tag{36}$$

and

$$\xi_2 \cong (1+\varpi)a^2\upsilon''(0)/2\ . \tag{37}$$

Hence, by using Jury's method [18], the stability of (33) requires the following conditions to hold

$$\left|\xi_1\xi_2 - \varpi\right| < 1,\ \xi_1\xi_2 > 0,\ and\ \ \xi_1\xi_2 > 2(\varpi - 1) \tag{38}$$

Observing these conditions, the first condition can be satisfied by choosing small enough a and b. Moreover, since $\varpi - 1 < 0$ due to the stability of the linear filters in the outer loop of the block diagram shown in Figure 1, $\xi_1\xi_2 > 0$ is sufficient to satisfy the

second and the third conditions from (38). Now, using (36), the following theorem is proven.

Theorem 1. Consider the system described by (17) and (18) with the assumptions $\upsilon(0)=0$ and $\upsilon'(0)=0$. Then there exist $\overline{a},\underline{a}\in\mathbb{R}$ and $\overline{b},\underline{b}\in\mathbb{R}$ such that $a\in(\underline{a},\overline{a})$ and $b\in(\underline{b},\overline{b})$ makes conditions in (38) hold and the system (17)-(18) has a stable solution $\tilde{\vartheta}_k^{a,e}$ and $\tilde{\eta}_k^{a,e}$ with period 2π. In addition, this solution, from (27) and (28), satisfies

$$\| [\tilde{\vartheta}_k^{a,e}(\tau) \quad \tilde{\eta}_k^{a,e}(\tau)]^T \| \leq O(a^2) . \tag{39}$$

Proof. It has been shown that (17) and (18) are stable under the condition (33) provided (38) holds with an appropriate choice of a and b. The bounds $(\underline{b},\overline{b})$ and $(\underline{a},\overline{a})$ can be determined by expanding the left condition of (38) as

$$\frac{\varpi-1}{\varpi+1} < \frac{1}{32}\gamma a^5 b\upsilon'''(0)\upsilon''(0) < 1 \quad . \tag{40}$$

Finally, (39) holds due to (31) and (32). ■

This result shows that, under the assumption of the fast inner loop, the system is able to reach the desired point and the proposed extremum seeking method makes the closed loop system to converge to a neighborhood which is bounded by $O(a^2)$. Moreover, this bound can be arbitrarily approach the origin by choosing a to be small enough.

Once we demonstrated that the equilibrium point is stable, we need to show that the closed-loop system will converge to this point from any initial condition. To this end, we will use singular perturbation analysis in the next subsection.

2.2. Singular Perturbation Analysis

Now, consider the full system depicted in Figure 1 whose state space representation is given in (11)-(13). Without loss of generality, we assume that $b=\sigma a$

with $\sigma \in \mathbb{R}$. This shows that, by choosing $a = 0$, (11) will represent the *boundary layer* [24] and (12) and (13) will represent the reduced model. For the sake of brevity, we denote $\tau = \omega k$ and represent the system (12)-(13) as

$$z_{k+1} = G(\tau, x_k, z_k), \tag{41}$$

with $z_k = (\tilde{\vartheta}_k, \tilde{\eta}_k)$. Theorem 1 shows that there exists a periodic stable solution $z_r^{2\pi}$ such that

$$z_r^{2\pi}{}_{k+1}(\tau) = G(\tau, L(\tau, z_r^{2\pi}{}_k), z_r^{2\pi}{}_k(\tau)) \tag{42}$$

is stable, where $L(\tau, z_r^{2\pi}{}_k) = \ell\left(\theta^* + \tilde{\theta} + a\sin\tau\right)$. In order to make the system as a singularly perturbed representation, we shift the state z as

$$\tilde{z}_k = z_k - z_r^{2\pi}{}_k(\tau), \tag{43}$$

By using (39), can represent the singular perturbed representation as

$$\tilde{z}_{k+1} = \tilde{G}(\tau, x_k, \tilde{z}_k) \tag{44}$$

$$x_{k+1} = \tilde{F}(x_k, \alpha(\tau, x_k, \tilde{z})) \tag{45}$$

where

$$\tilde{G}(\tau, x_k, \tilde{z}_k) = G(\tau, x_k, z_k) - G(\tau, L(\tau, z_r^{2\pi}{}_k), z_r^{2\pi}{}_k(\tau)) \tag{46}$$

$$\tilde{F}(\tau, x_k, \tilde{z}) = f(x_k, \alpha(x_k, \tilde{\vartheta}_k + \underbrace{\vartheta^* - \tilde{\vartheta}_r^{2\pi}(\tau)}_{\tilde{z}} + \tilde{\vartheta}_r^{2\pi}(\tau) + a\sin(\tau))). \tag{47}$$

In *Theorem 1,* it was shown that this equilibrium point is stable with properly chosen value for a and $b = \sigma a$. Moreover, as demonstrated previously if $a \to 0$ then the solution to the fast dynamics (47) will converge to a dynamics that is called *quasi-steady state model* represented by

$$x_k = L(\tau, \tilde{z} + z_r^{2\pi}(\tau))\Big|_{a=0}. \tag{48}$$

and the reduced model [24]

$$\tilde{z}_{r,k+1}(\tau) = \tilde{G}(\tau, L(\tau, \tilde{z}_r + z_r^{2\pi}(\tau)), \tilde{z}_r + z_r^{2\pi}(\tau)) \tag{49}$$

has an equilibrium point at the origin. Now consider the *boundary layer model* [19][24] as

$$x_{b,k+1}(\tau) = \tilde{F}(\tau, x_{b,k} + L(\tau, \tilde{z}_r + z_r^{2\pi}(\tau)), \tilde{z}) =$$

$$f(x_{b,k} + \ell(\theta), \alpha(x_{b,k} + \ell(\theta), \theta)) \ , \tag{50}$$

where $\theta = \theta^* + \tilde{\theta} + a\sin\tau$ should be considered as a parameter independent of k. Since we assumed that $x = \ell(\theta)$ is a locally stable point, x_b is also stable. The following assumptions are needed in order to proceed.

Assumption 4: Assume that there exists a $V : \mathbb{R}^n \to \mathbb{R}^+$, which is positive definite on $D_x \subseteq \mathbb{R}^n$, and in addition, radially unbounded when $D_x = \mathbb{R}^n$. Moreover, there exists a function $\varphi(x) : \mathbb{R}^n \to \mathbb{R}^+$ that is *locally* positive definite on D_x such that

$$V\left(\tilde{F}(x_k, \alpha(\tau, x_k, \tilde{z}))\right)\Big|_{a=0} - V(x_k) \le -\varphi(x_k) \qquad \forall x_k \in D_x . \tag{51}$$

Assumption 5: Assume that there exists a $W : \mathbb{R}^2 \to \mathbb{R}^+$ that is a positive definite on $D_z \subseteq \mathbb{R}^2$, and in addition, radially unbounded for $D_z = \mathbb{R}^2$. Moreover, there exists $\psi(z) : \mathbb{R}^2 \to \mathbb{R}^+$ that is *locally* positive definite on D_z such that

$$V\left(\tilde{z}_{k+1}^f\right) - V\left(\tilde{z}_k^f\right) \le -\psi(\tilde{z}_k) \qquad \forall x \in D_x , \tag{52}$$

where the superscript (f) signifies the fast part of the dynamics.

Existence of V and W are not strong assumptions due to the stability of the reduced (12)-(13) and boundary systems (11) by using the converse Lyapunov theorems [19].

Assumption 6: Consider the reduced model (49) and boundary layer (50). Assume that there exists real numbers λ_i, $i = \{1,2,...,6\}$ satisfying the following inequalities

$$\text{(a) } V\left(\tilde{F}(x_k,\alpha(\tau,x_k,\tilde{z}))\right) - V\left(\tilde{F}(x_k,\alpha(\tau,x_k,\tilde{z}))\Big|_{a=0}\right) \le \lambda_1\varphi(x_k) + \lambda_2 a\psi(\tilde{z}_k) \tag{53}$$

$$\text{(b) } W\left(\tilde{G}(\tau,x_k,\tilde{z}_k)\right) - W\left(\tilde{z}_{k+1}^f\right) \le \lambda_3\varphi(x_k) + \lambda_4 a\psi(\tilde{z}_k) \tag{54}$$

$$\text{(c) } W\left(\tilde{z}_k^f\right) - W\left(z_k\right) \le \lambda_5\varphi(x_k) + \lambda_6\psi(\tilde{z}_k)\,. \tag{55}$$

The above inequalities determine the permissible interaction between the slow and fast dynamics. Next the following theorem is stated.

Theorem 2. Consider the singularly perturbed nonlinear system whose boundary layer is represented in (45) when $a = 0$ and the reduced model is given by (44) with the Assumptions 1-6 hold. Assume that the boundary layer (45) with $a = 0$ is locally uniformly ultimately bounded stable for any $\hat{\theta} \in \mathbb{R}$ i.e. it is guaranteed that, with $a = 0$, x_k uniformly converges to a bound $b_x(\hat{\theta})$ in a neighborhood D_x. Moreover, let the fast part of the dynamics (44) represented by $\tilde{z}_k^f$ be stable under the Assumption 5 as proven in Theorem 1. Then, when $\lambda_1 + \lambda_3 + \lambda_5 \le 1$, the overall extremum seeking system (5)-(7) is UUB such that x_k converges to x^* and $\hat{\theta}$ converges to θ^* with bounds being b_x and b_θ respectively.

Proof. It is enough to choose $v = V + W$ as the Lyapunov candidate that is a positive definite function on $D = D_x \times D_z$ based on the Assumptions 4 and 5. Therefore, the forward difference can be written as

$$\Delta v = \Delta V + \Delta W$$

$$= v[\tilde{F}(x_k,\alpha(\tau,x_k,\tilde{z})),\tilde{G}(\tau,x_k,\tilde{z}_k),a] - v[x_k,\tilde{z}_k,a] \tag{56}$$

By some simple computation and under Assumptions 4-6, (56) along the closed loop system trajectory can be written as

$$\Delta v = V\left(\tilde{F}(x_k, \alpha(\tau, x_k, \tilde{z}))\right) - V\left(\tilde{F}(x_k, \alpha(\tau, x_k, \tilde{z}))\Big|_{a=0}\right)$$
$$+W\left(\tilde{G}(\tau, x_k, \tilde{z}_k)\right) - W\left(\tilde{z}_{k+1}^f\right) + W\left(\tilde{z}_k^f\right) - W\left(z_k\right)$$

Now using (53)-(55) we have

$$\Delta v \le -(1-\lambda_1-\lambda_3-\lambda_5)\varphi(x_k) - (1-\lambda_6 - a\lambda_2 - a\lambda_4)\psi(\tilde{z}_k) \le 0$$

$$\forall (x_k, \tilde{z}_k) \in D_x \times D_z \text{ and } a \le a_m \triangleq (1-\lambda_6)/(\lambda_2+\lambda_4). \quad (57)$$

The inequality (57) implies that there exists a neighborhood $b_x \times b_z \subseteq D_x \times D_z$ such that for $\forall (x_k, \tilde{z}_k) \in b_x \times b_z$ the inequality holds. Therefore, x_k will converge to the bound of b_x and $\tilde{z}_k$ will converge to the bound of b_z which means $\tilde{\theta}$ will be upper bounded by b_θ. ■

2.3. Inner Loop Stabilizer Design

As mentioned in (2)-(3), u_k is stabilizing and we relied on this assumption throughout the subsections A and B. Now, the question might be how to design such a controller for nonaffine systems. Depending upon the application, different classes of controller are proposed in the literature [20][20] for nonaffine systems. In the work [6], a robust optimal adaptive controller is proposed for controlling nonaffine systems.

In paper [6], the nonaffine nonlinear discrete-time system is transformed to an affine-like equivalent nonlinear discrete-time system in the input-output form. Next, a forward-in-time Hamilton-Jacobi-Bellman (HJB) equation-based optimal approach, without using value and policy iterations, is developed to control the affine-like nonlinear discrete-time system by using both neural networks (NN) as an online approximator and

output measurements alone. To overcome the need to know the control gain matrix in the optimal controller, a new online discrete-time NN identifier is introduced. The robustness of the overall closed loop system is shown via singularly perturbation analysis by using an additional auxiliary term to mitigate the higher-order terms. Lyapunov stability of the overall system, which includes the online identifier and robust control term, demonstrates that the closed-loop signals are bounded and the approximate control input approaches the optimal control signal with a bounded error. in the following, the steps the algorithm of computing u_k is provided, while the one can find the proof of the stability in [6].

Consider a generic form of nonaffine system (1). The system input and the controller output are connected such that $u_{k+1} = u_k + \Delta u_k$. Using (1) and (3) we can express the system dynamics as

$$y_{k+1} = h(y_k, u_k, \Delta u_k)$$

or

$$\begin{aligned} X_{k+1} &= \left(\left(u_k + \Delta u_k \right)^T \quad h(g^{-1}(y_k, u_k), u_k, u_k + \Delta u_k)^T \right)^T \\ &= H(X_k, \Delta u_k) \end{aligned} \tag{58}$$

where $X_{k+1} = (u_k \quad y_k)^T \subset \mathbb{R}^{m+\ell}$. By applying the Taylor series expansion, equation (58) can be expanded as

$$\begin{aligned} X_{k+1} = H(X_k, \Delta u_k) &= H(X_k) + \frac{\partial}{\partial \Delta u_k} H(X_k, \Delta u_k)\,|_{\Delta u_k = 0}\ \Delta u_k \\ &+ \frac{1}{2} \frac{\partial}{\partial \Delta u_k} \left(\frac{\partial}{\partial \Delta u_k} H(X_k, \Delta u_k) \Delta u_k \right)|_{\Delta u_k = 0}\ \Delta u_k + \cdots \\ &\equiv \sum_{i=0}^{q} F_i(X_k) \Delta u_k^i + O(X_k, \Delta u_k), \end{aligned} \tag{59}$$

where

$$F_0(X_k) = H(X_k),\ F_1(X_k) = \frac{\partial}{\partial \Delta u_k} H(X_k, \Delta u_k)\,|_{\Delta u_k = 0},\ \ldots$$

Therefore, the unknown affine-like system representation of (1) takes the following input-output form as

$$X_{k+1} = \bar{F}(X_k) + \bar{G}(X_k)\Delta u_k + \bar{O}_k, \tag{60}$$

where $\Psi(X_{k-1}, \Delta u_{k-1}) \in \mathbb{R}^{p\times(m+\ell)}$, $W \in \mathbb{R}^{p\times(m+\ell)}$, $\Delta U_k = (1 \quad \cdots \quad 1 \quad \Delta u_k^T)^T \in \mathbb{R}^{(m+\ell)}$, $\bar{\varepsilon}_k \in \mathbb{R}^{m+\ell}$, $\|\Psi(X_{k-1}, \Delta u_{k-1})\| \le \Psi_M$, $\|\Psi(X_{k-1}, \Delta u_{k-1})U_{k-1}\| \le \bar{\Psi}_M$ being the bounded NN activation function and $\bar{\varepsilon}_k$ is the estimation error satisfying $\|\bar{\varepsilon}_k\| < \bar{\varepsilon}_M$. Moreover, the following bound $U_m \le \|\Delta U_k\|$ holds due to the presence of constant values in the input vector in fact $U_m = \|\Delta U_k\|_{\Delta u_k = 0}$. In order identify the NN weight matrix W denoted here as $\hat{W}_k = \left(\hat{W}_{\bar{F}}^T, \hat{W}_{\bar{O}}^T, \hat{W}_{\bar{G}}^T\right)^T$ by estimating the state vector X_k with $\hat{X}_k$ where

$$\begin{aligned}\hat{X}_k &= \Psi(X_{k-1}, \Delta u_{k-1})^T \hat{W}\Delta U_{k-1} \equiv \hat{W}_{\bar{F}}^T \Psi_{\bar{F}}(X_k) \\ &+ \hat{W}_{\bar{G}}^T \Psi_{\bar{G}}(X_k)\Delta u_k + \hat{W}_{\bar{O}}^T \Psi_{\bar{O}}(X_k, \Delta u_k)\ , \\ &\equiv \hat{\bar{F}}(X_k) + \hat{\bar{G}}(X_k)\Delta u_k + \hat{\bar{O}}(X_k, \Delta u_k)\end{aligned} \tag{61}$$

Therefore,

$$\hat{X}_{k+1} = \hat{\bar{F}}(X_k) + \hat{\bar{G}}(X_k)\Delta u_k + \hat{\bar{G}}(X_k)\Delta \rho_k + \hat{\bar{O}}_k + \tilde{e}_{k+1}. \tag{62}$$

where the identification error is defined as

$$\begin{aligned}\tilde{e}_k &= X_k - \hat{X}_k = \Psi(X_{k-1}, \Delta u_{k-1})^T (W - \hat{W}_{k-1})\Delta U_{k-1} + \bar{\varepsilon}_{k-1} \\ &= \Psi(X_{k-1}, \Delta u_{k-1})^T \tilde{W}_{k-1}\Delta U_{k-1} + \bar{\varepsilon}_{k-1}\end{aligned} \tag{63}$$

Define the update law for the actual NN weights $\hat{W}_k$ as

$$\hat{W}_k = \hat{W}_{k-1} + \alpha \tilde{E}_k \Psi(X_{k-1}, \Delta u_{k-1}) \Delta U_{k-1}^T \Big/ \left(\left\| \Delta U_{k-1} \right\|^2 + 1 \right), \tag{64}$$

where

$$\tilde{E}_k = diag(\tilde{e}_k) \in \mathbb{R}^{(m+\ell)\times(m+\ell)}, \tag{65}$$

with $\alpha > 0$ is a design parameter or NN learning rate. After identification of the unknown dynamics we design the optimal controller. Consider the cost function $J(k)$ such that

$$J(X_k) \equiv \sum_{i=0}^{\infty} r(k+i) = r(X_k, \Delta u_k) + J(X_{k+1}) \tag{66}$$

where $r(X_k, \Delta u_k) = Q(X_k) + \Delta u_k^T R \Delta u_k$ with $Q(X_k) \geq 0$, $Q(0) = 0$, and $R \in \mathbb{R}^{m\times m}$ as a positive definite matrix. In the sequel, we will denote $J(X_k)$ by J_k for the sake of simplicity. Next the following definition is needed in order to proceed. The objective is to minimize J_k by starting with an admissible control law and modifying it with respect to the system dynamics so that the estimated cost function and control input converge to the optimal cost function J_k^* and control law Δu_k^* respectively. By applying the stationarity condition to (28) we have

$$\frac{\partial J_k}{\partial \Delta u_k} = \frac{\partial r(X_k, \Delta u_k)}{\partial \Delta u_k} + \frac{\partial J_{k+1}}{\partial \Delta u_k}$$

$$= 2R\Delta u_k + \left(\frac{\partial X_{k+1}}{\partial \Delta u_k} \right)^T \frac{\partial J_{k+1}}{\partial X_{k+1}} = 0 \tag{67}$$

Using (29) and (14) we can write

$$\Delta u_k^* = -\frac{1}{2} R^{-1} \hat{G}^T(X_k) \frac{\partial J_{k+1}^*}{\partial X_{k+1}} - \frac{1}{2} R^{-1} \left(\frac{\partial \hat{O}_k}{\partial \Delta u_k} + \frac{\partial \tilde{e}_{k+1}}{\partial \Delta u_k} \right)^T \frac{\partial J_{k+1}^*}{\partial X_{k+1}}. \tag{68}$$

In order to mitigate the higher order term, the design of an auxiliary term, $\Delta\rho_k$, is required for robustness in addition to the optimal controller term

$$\Delta u_k = \Delta\hat{u}_k + \Delta\rho_k . \tag{69}$$

update law for $\Delta\rho_k$ is chosen to be

$$\Delta\rho_{k+1} = \Delta\rho_k + \delta\rho_k \tag{70}$$

With the desired $\delta\rho_k$ as

$$\delta\rho_k = -A_\zeta^{-1}(k)\,B_\zeta(k), \tag{71}$$

with

$$A_\zeta(k) = \left\{\hat{G}(X_{k+1}) + \partial\hat{O}\left(X_{k+1}, \Delta\hat{u}_{k+1} + \Delta\rho_k\right)\Big/\partial\Delta\hat{u}_k\right\},$$

$$B_\zeta(k) = \left\{\hat{G}(X_{k+1})\Delta\rho_k + \hat{O}\left(X_{k+1}, \Delta\hat{u}_{k+1} + \Delta\rho_k\right) - \alpha_\zeta\zeta_k\right\},$$

and $0 < \alpha_\zeta < 1$. (72)

The cost function (28) will be approximated by an OLA and written as

$$\hat{J}(k) = \hat{J}(X_k) = \hat{\Phi}_k^T\sigma(X_k) = \hat{\Phi}_k^T\sigma(k)$$

where $\hat{J}(k)$ represents an approximated value of the original cost function $J(k)$, $\hat{\Phi}_k$ is the vector of actual parameter vector for the target OLA parameter vector, Φ , and $\sigma(k) = \{\sigma_\ell(k)\}_1^{L:}$ is set of activation functions which are each chosen to be basis sets and thus are linearly independent. Define the cost function OLA parameter update to be

$$\hat{\Phi}_{k+1} = \bar{X}(k)\left(\bar{X}^T(k)\bar{X}(k)\right)^{-1}\left(\alpha_c E_c^T(k) - Y^T(k)\right) \tag{73}$$

where $0 < \alpha_c < 1$. Finally, we need to define the optimal feedback optimal policy. we define a NN to estimate (30) as

$$\Delta u^*(k) = \Delta u^*(X_k) = \Theta_k^T \vartheta(X_k) + \varepsilon_A . \tag{74}$$

Therefore, define an OLA approximation of (55) to be

$$\Delta \hat{u}(k) = \Delta \hat{u}(X_k) = \hat{\Theta}_k^T \vartheta(X_k) \tag{75}$$

where $\Delta \hat{u}(k)$ is introduced in (31) and $\hat{\Theta}_k$ is the estimated value of the ideal parameter matrix Θ and $\vartheta(\bullet)$ denotes the linearly independent basis function.

Next, the optimal control signal error is defined to be the difference between the feedback control applied to (10) and the optimal control signal, as

$$e_a(k) = \hat{\Theta}_k^T \vartheta(X_k)$$

$$+R^{-1}\hat{G}^T(k)\left(\partial\sigma(X_k)/\partial X_k\right)^T \hat{\Phi}_k / 2 \text{ , and} \tag{76}$$

$$e_a(k+1) = \hat{\Theta}_{k+1}^T \vartheta(X_k) + \frac{1}{2} R^{-1}\hat{G}^T(k+1)\left(\frac{\partial\sigma(X_{k+2})}{\partial X_{k+2}}\right)^T \hat{\Phi}_{k+1} , \tag{77}$$

One should notice that the identifier error and the robust term errors are previously shown to be bounded and play the role of disturbance in the closed loop system. This is the reason that they are ignored in the design of the estimation error (57). Nonetheless, we will see that $\tilde{e}_k$ and ζ_k will appear in the overall stability proof of the system. The proposed control OLA parameter update is defined to be

$$\hat{\Theta}_{k+1} = \hat{\Theta}_k - \alpha_a \vartheta(k) e_a^T(k) / (\vartheta^T(k)\vartheta(k) + 1) \tag{78}$$

where $0 < \alpha_a < 1$ is a positive design parameter.

By now, the inner loop for generating the control law u_k is complete. As we mentioned earlier the proof of the stability is provided in [6] and is omitted here for the sake of brevity.

The next section is devoted to demonstrate one of the applications of the proposed extremum seeking method in discrete-time in conjunction with a NN based optimal controller. Two examples are included one a linear multivariable system to illustrate the proposed scheme while the second one is the HCCI engine.

3. SIMULATION RESULTS

In this subsection we consider both a linear multivariable system a nonlinear nonaffine system and propose a NN based optimal controller proposed by the authors in [6].

3.1. Application to Nonlinear Multivariable Systems

The system dynamics is represented as

$$\begin{bmatrix} x_1(k+1) \\ x_2(k+1) \end{bmatrix} = \begin{bmatrix} 0.2\sin x_1(k) + 0.6x_2(k) + u_1(k)(5+\cos(x_2(k)u_2(k))) \\ 0.1\cos x_2(k) + 0.6x_1(k) + u_2(k)(5+\sin(x_1(k)u_1(k))) \end{bmatrix}. \tag{79}$$

$$y_k = \begin{bmatrix} x_1(k) & x_2(k) \end{bmatrix}^T = x_k \tag{80}$$

Now, we define an unknown performance function

$$eff(k) = -(y_1 - 1)^2 - (y_2 - 2)^2, \tag{81}$$

to be maximized. Two separate extremum seeking controllers are used in order to generate the optimum elements of the setpoint $y^d = [y_1^d \quad y_2^d]$.

Figure 2 shows the block diagram for the extremum seeking closed loop block diagram representation. Each extremum seeking blocks has given the charge of maximizing the efficiency by seeking the optimum solution for y_1^d and y_2^d separately. In this figure, the initial admissible controller required in the proposed NN-based optimal controller [6] is chosen as $u(k) = 0.05\left\{y^d(k) - x(k) + u(k-1)\right\}$.

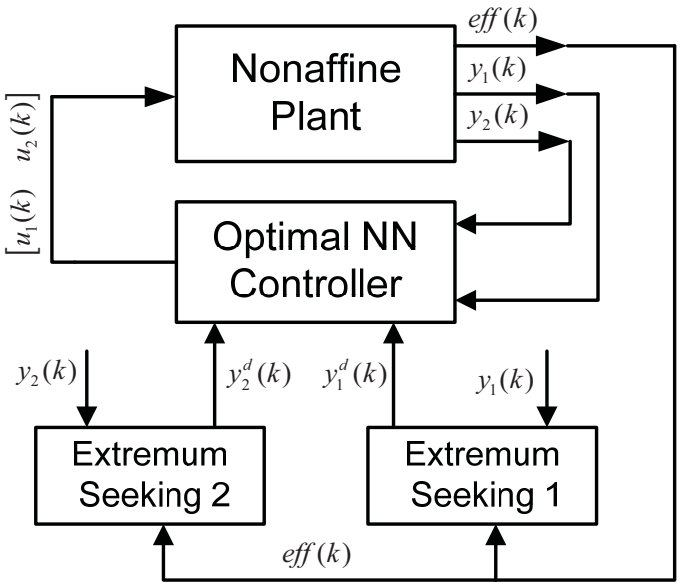

Figure. 2. The block diagram representation of the extremum seeking controller for the nonlinear MIMO Nonaffine system with the optimal controller being in the inner loop.

Figure 3 shows the convergence of the closed loop system while the state initial conditions are chosen to be $(y_1^d, y_2^d) = (0,0)$ and $(0.1, 0.1)$. This figure illustrates the efficiency with respect to the outputs trajectory in a 3D manner. The figure shows that how the extremum seeking is able to maximize the efficiency function (81) by finding the optimum values for y_1 and y_2 assuming that the function is unknown. Figure 4 also illustrates the output convergence with respect to time. In this case, the perturbation amplitude is taken as $a = b = 0.05$, the filter pole $\varpi = 0.1$, and $\omega = 2\pi/10$, and $\gamma = 10$.

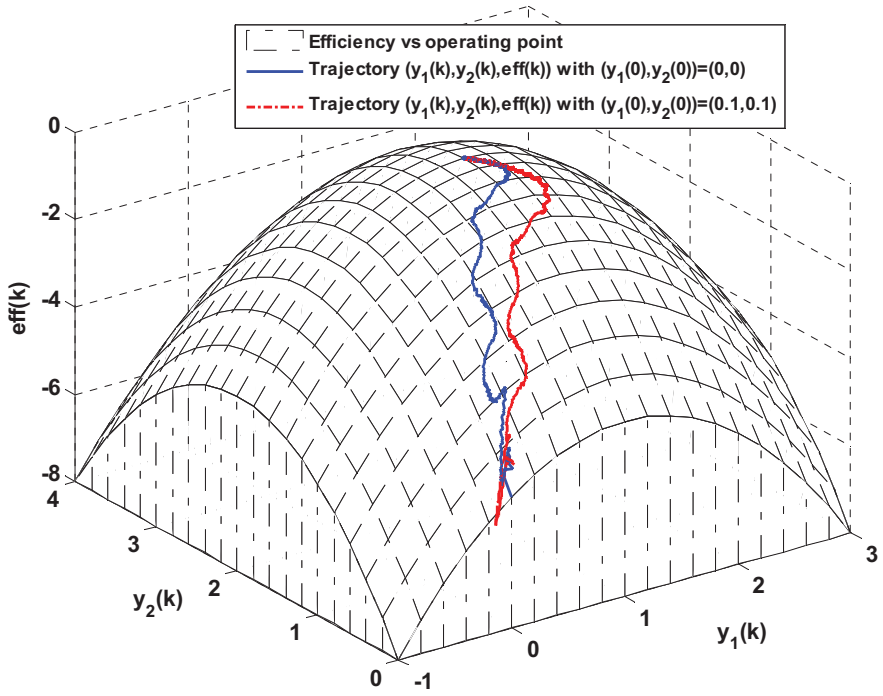

Figure. 3. The nonlinear system trajectory (starting from the origin) versus the efficiency while it converges to the extremum point $y^d = [1 \quad 2]^T$ for two different initial state conditions.

3.2. Application to HCCI Engine Performance Maximization

Low temperature combustion (LTC) modes in homogeneous charge compression ignition (HCCI) engines represent a promising means to increase the efficiency and significantly reduce the emissions of internal combustion (IC) engines. Controlling such engines is difficult due to the dependence of the combustion event on chemical kinetics rather than an external trigger. In [11], the author outlined a nonlinear control-oriented model of a single cylinder HCCI engine, which is physically based on a five state thermodynamic cycle. This model is aimed at capturing the behavior of an engine which utilizes fully vaporized gasoline-type fuels, exhaust gas recirculation and intake air heating in order to achieve HCCI operation. The onset of combustion, which is vital for control, is modeled using an Arrhenius reaction rate expression which relates the combustion timing to both charge dilution and temperature.

The model is validated against experimental data from a single cylinder combustion ignition (CI) engine operating under HCCI conditions at two different fueling rates. Predicted combustion timing and peak in-cylinder pressure values from simulation agree very well with the experiments at both operating conditions. Once validated, trends in the model dynamics are investigated. The result is a discrete-time nonlinear control representation which provides a platform for developing and validating various nonlinear control strategies.

Figure 5 depicts the block diagram representation of the HCCI engine whose operating data are obtained from the model. The system dynamics are derived from cycle-by-cycle information of the thermo-dynamical behavior of the engine and elaborately represented in [5]. The engine dynamics is represented as the nonaffine nonlinear systems, which makes the optimal control of the HCCI engine a challenging problem.

The authors in [6], has proposed an approach that provides a robust optimal controller that makes the closed loop system converge to an arbitrarily small bound in an optimal manner while this bound is a function of the reconstruction error of the neural networks. Therefore, $b_x(\hat{\theta})$ is provided by [6] and used here and we only require to choose a proper a for the purpose of extremum seeking.

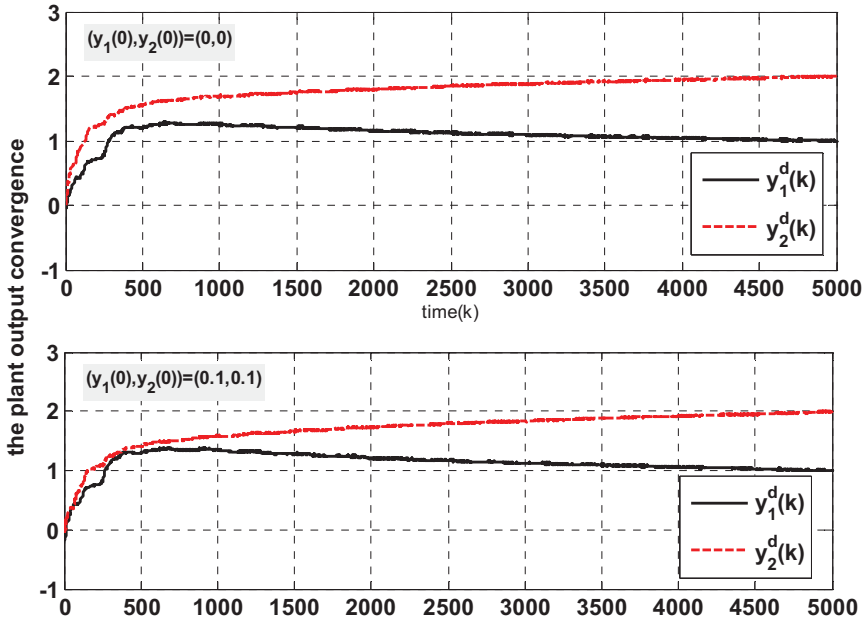

Figure. 4. The output convergence of the plant output to the optimum point for two different initial state condition.

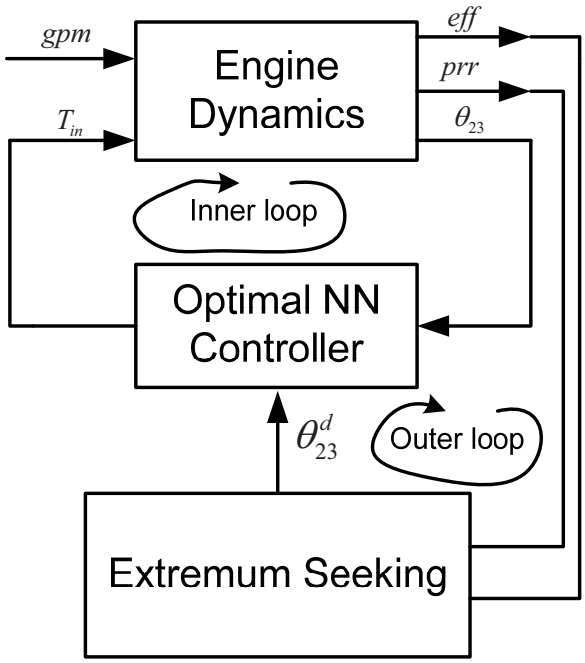

Figure. 5. The block diagram representation of the HCCI engine with the controller.

The Engines efficiency eff_k in Figure 5 has to be maximized by choosing a suitable crank angle θ_{23}^d which is the desired set point to be identified by the proposed

extremum seeking controller. Figure 6 shows that there exist a peak for the engine efficiency when the crank angle changes for three different fuel injection rates namely $gpm = \{6,9,11\}$ where gpm stands for grams per minute. On the other hand, it is desired that the peak pressure rise rate PRR be constrained to be less than 10bar/CAD due to the practical limitations of the engine where CAD is a unit (equal to one "ordinary" degree) used to measure the piston travel (position) e.g. to adjust ignition. When the piston is at its highest point, known as the top dead center, the crankshaft angle is at 0 crank angle degrees (CAD).

Thus we choose the following function to be maximized

$$y_k = -S_k (PRR_k - 8)^2 + eff_k \text{, where} \tag{82}$$

$$S_k = 1 \text{ if } PRR_k \geq 8 \tag{83}$$

$$S_k = 0 \text{ if } PRR_k < 8. \tag{84}$$

In order to maximize (82), the eff_k should be kept as close as possible to its peak while meeting the PRR_k constraint towards 8bar/CAD. The reason that we chose 8 instead of 10bar/CAD is to keep the engine in a safe margin from the dangerous values of the PRR. The proposed optimal adaptive NN controller along with extremum seeking feature is used to control the HCCI engine with the output function given by (58) through (60).

Figures 6 through 10 demonstrate the extremum seeking results for different values of the injected fuel rates {6,9,11} gpm while the fuel type is chosen to be UTG96. As it is obvious in these figures, the closed loop behavior is more scattered for the case of gmp=11 when compared to the other cases. The main reason is that the PRR is higher

than the threshold of 8bar/CAD. In fact, by $PRR_k \geq 8$, $S_k = 1$ holds which makes the magnitude of (82) to reduce which in turn increases the effect of the periodic signals injected for seeking the extremum.

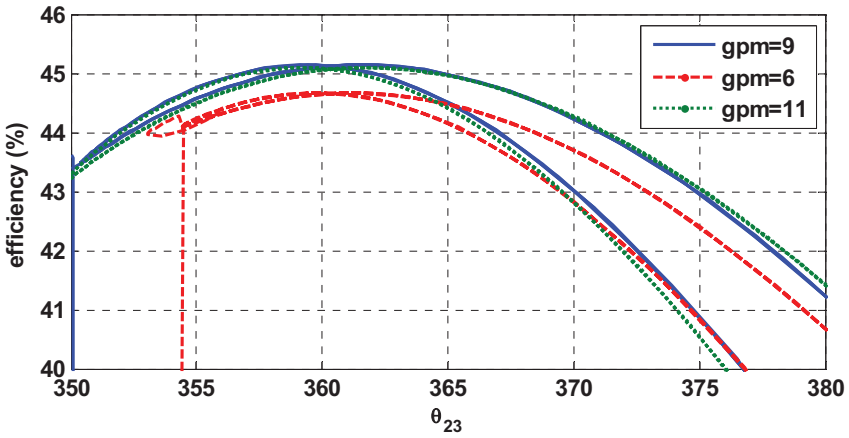

Figure. 6. Crank angle versus efficiency plot illustrating a peak with varying intake temperature.

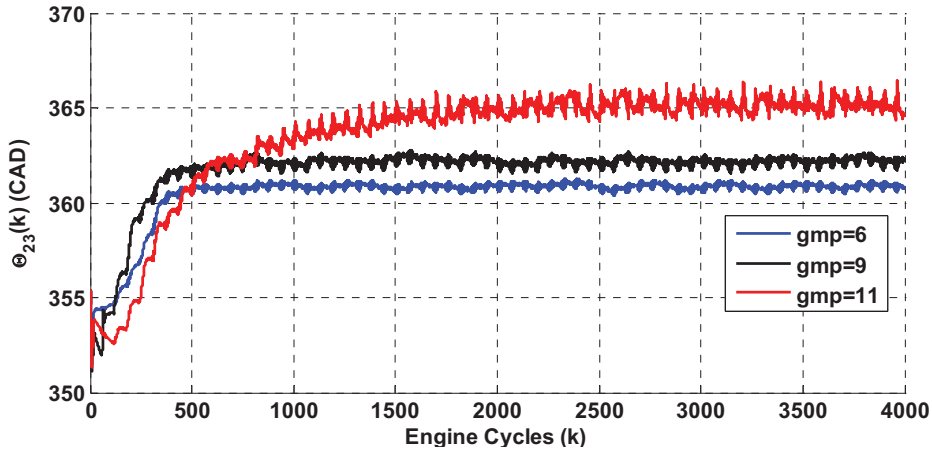

Figure. 7. The crank angle convergence to the optimum.

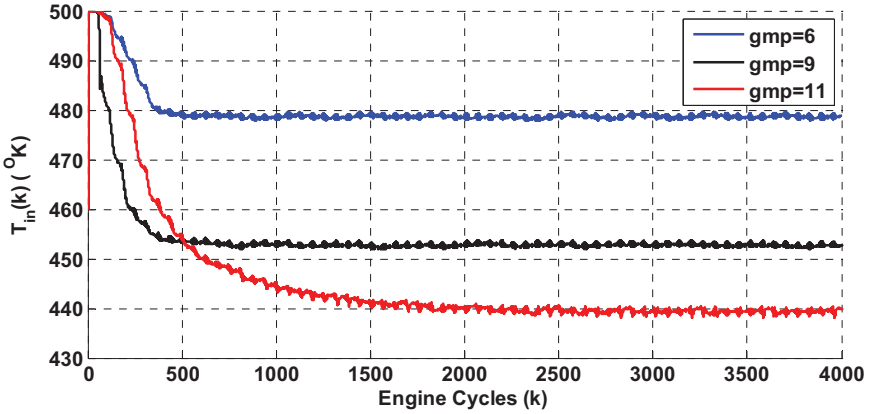

Figure. 8. The intake temperature as the system input.

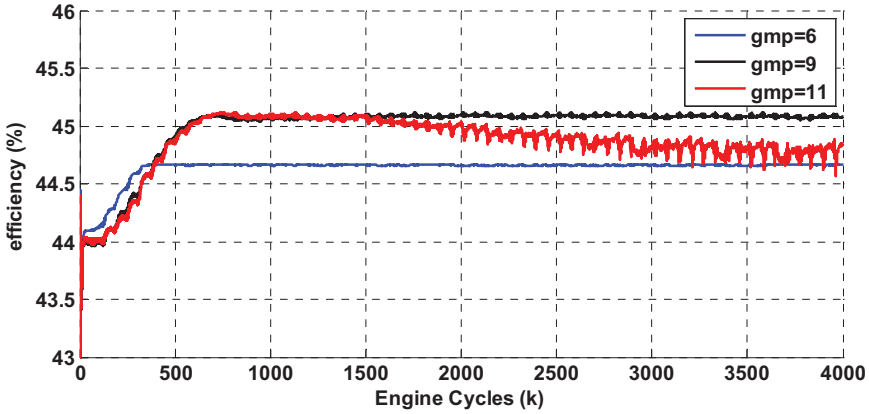

Figure. 9. Maximization of engine efficiency by using the extremum seeking control for different fixed fuel rates.

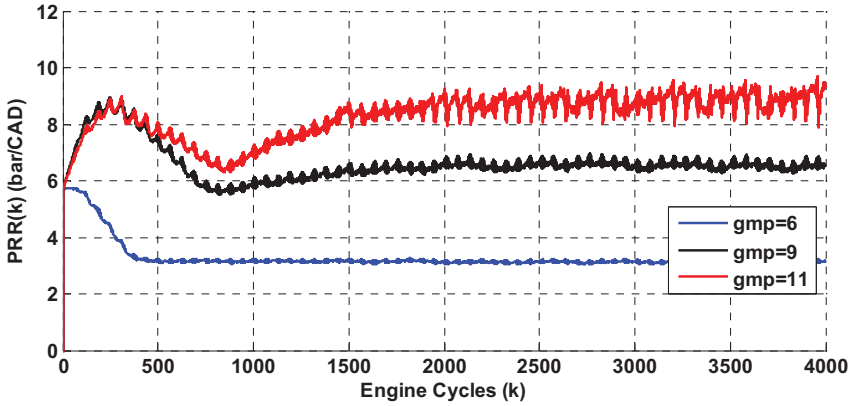

Figure. 10. The PRR within the safe margin.

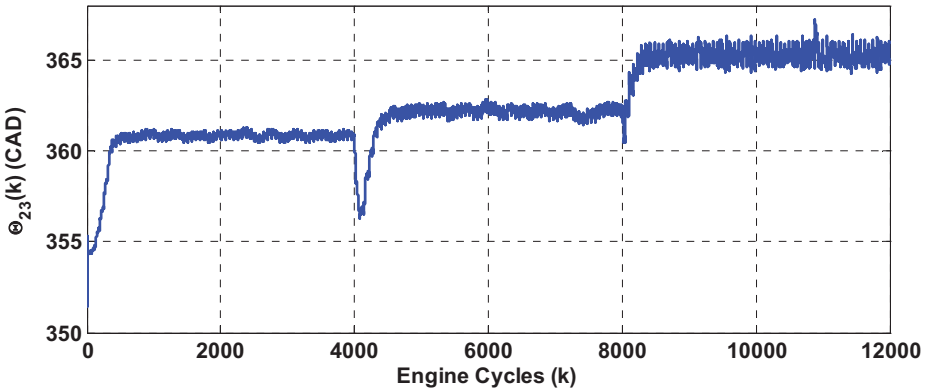

Figure. 11. The convergence of the crank angle to its optimal value by using fuel.

For the next few experiments shown in Figures 11 through 14, the fuel rate is changed online to show that the proposed extremum seeking controller is able to find the suitable operating point even when there is a change in the operating condition during the control process. Figure 13 shows that the efficiency reduces when the gpm increases from 11 to 13. The reason is that the PRR in this case tends to converge to a values higher than 8. This makes $S_k = 1$ which finally makes the extremum seeking to reduce the intake temperature and find a higher Θ_{23} (see Figure 11) that causes a lower efficiency value.

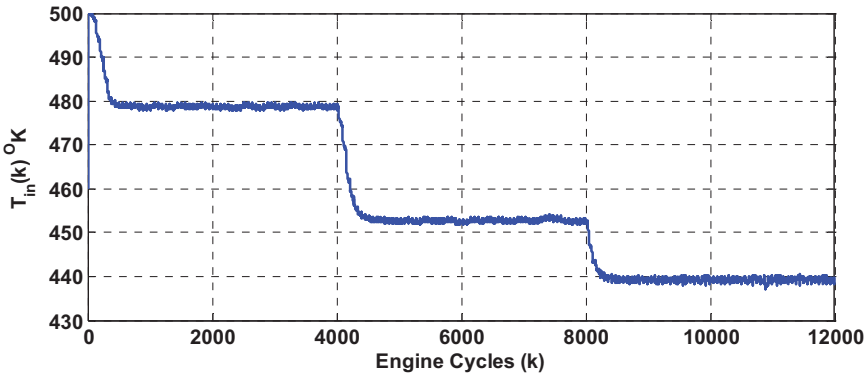

Figure. 12. The intake temperature applied to the inner control loop when the fuel rate changes from 6 to 9 and to 11 gpm once every 4000 cycles.

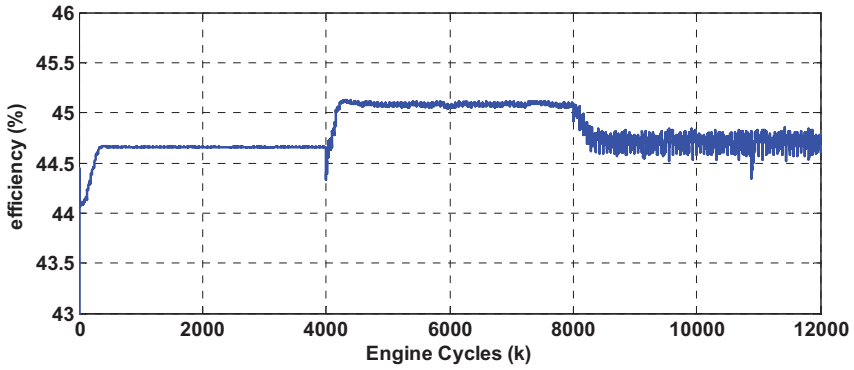

Figure. 13. Engine efficiency when the fuel rate changes from 6 to 9 and 11 gpm once every 4000 cycles.

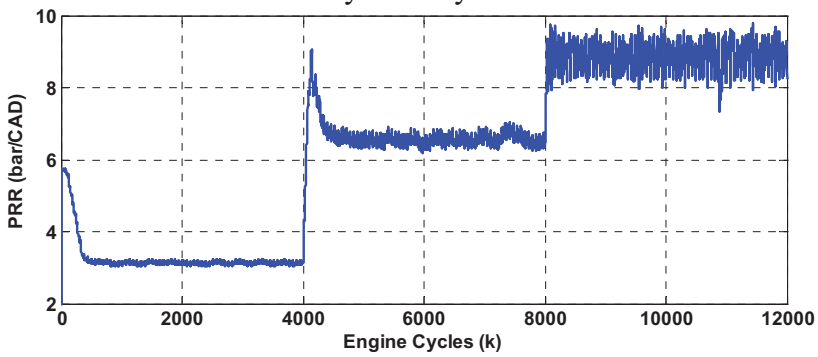

Figure. 14. The PRR convergence when the fuel rate changes sequentially with respect to Fig. 10.

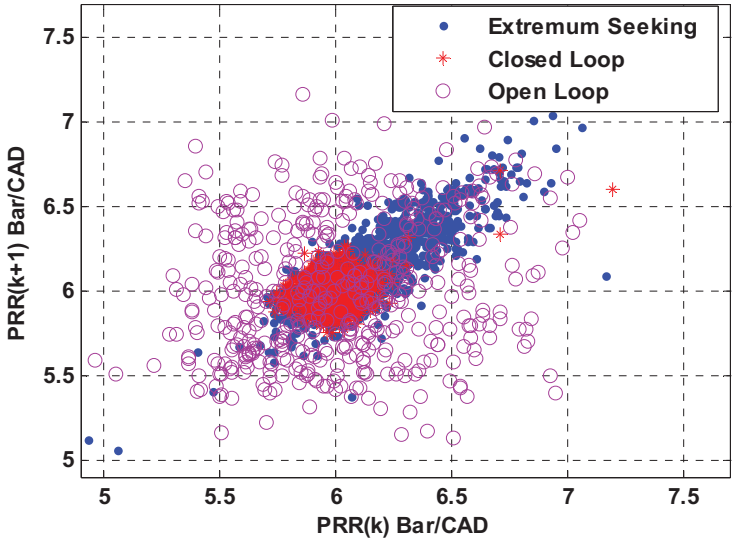

Figure. 15. Comparison of the return map of the pressure rise rate $PRR(k)$ for three cases: 1) extremum seeking (both loops closed) 2) closed loop (outer loop open and the NN-loop closed) 3) both loops open.

In another set of experiments, we study the cyclic dispersion resulting from controlling the HCCI engine by using open loop, closed loop without extremum seeking and with extremum seeking. In fact, as it is shown in Figure 5, that the proposed control approach for the closed loop system has an inner and outer loop. In the first case all the loops are closed which implies that extremum seeking is also considered. In the second case, the outer loop will be open which means that there is no extremum seeking and in the last case all the loops are open. The return map of the PRR and the crank angle $\Theta_{23}(k)$ of the engine are compared for the above three cases. A probing noise is added to the engine dynamics in order to compare the ability of the proposed controller to reduce the cyclic dispersion in any of the three cases. The fuel rate is chosen to be mpg=11. For the first case, the extremum seeking system will seek for the best operating point that maximizes the efficiency while keeping the pressure rise rate under 10bar/CAD.

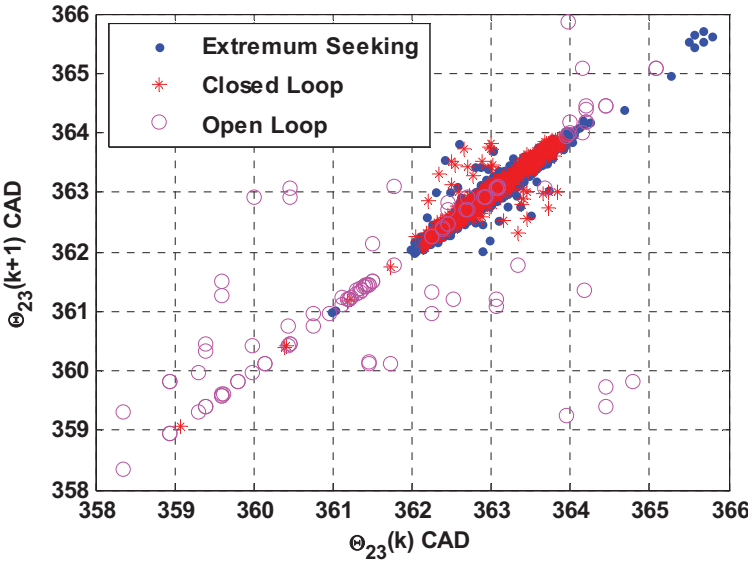

Figure. 16. Comparison of the return the map of the crank angle $\Theta_{23}(k)$ for three cases: 1) extremum seeking (both loops closed) 2) closed loop (outer loop open and the NN closed) 3) both loops open.

Figures 15 and 16 illustrate that the engine converges to an operating point due to the extremum seeking controller. This operating point is used as the target set point value in the following next two cases. The return map of $PRR(k)$ and Θ_{23} are provided. For the applied operating point of the second case the NN controller will find a set value in terms of the intake air temperature to provide the desired output. As the third case, the steady state value of the second case is applied to the engine in an open loop manner and the results are provided in Figures 15 and 16. From the figures, the reduction in cyclic dispersion for the closed-loop control is much better than the open loop case. On the other hand, the closed loop without the extremum control outer loop has a better performance when compared to the case that the extremum seeking algorithm acts as an outer loop.

The higher dispersion observed in the case of extremum seeking is due to the periodic injection of signals to the system dynamics in order to find the extremum operating point. In other words, the higher level of the cyclic dispersion is the price paid in order to find the extremum set point online. This aspect is observed when the fuel is varied from gpm=6 and gpm=11.

Table I and Table II compare the coefficient of variation (COV) of the PRR and the crank angle in the three cases of open loop, closed loop, and extremum seeking. As mentioned above, the closed loop case has the best COV among the other cases, although neither the closed loop nor the open loop are able to find the best operating point. Therefore, a reasonable increase in COV will be tolerable when an optimal operating point can be found. In both tables I and II, the columns for the extremum seeking and closed loop data has a number in percentage within brackets which indicates that the

percentage improvement in COV when compared to the COV for the open loop case. These results indeed concur that the closed-loop controller performs best while the addition of extremum seeking feature presents a tradeoff between identifying the setpoint with the increase in cyclic dispersion.

Table 1. Coefficient of variation for PRR and percentage of improvement comparing with the open loop case.

gpm	Extremum Seeking COV	Closed Loop COV	Open Loop COV
6	1.6604(-11%)	1.3860 (-25%)	1.8686
9	2.0325(-26%)	1.8491(-33%)	2.2770
11	2.2091(-1%)	1.9338(-13%)	2.2345

Table 2. Coefficient of variation for the crank angle and percentage of improvement comparing with the open loop case

gpm	Extremum Seeking COV	Closed Loop COV	Open Loop COV
6	0.0904(-11%)	0.0809(-21%)	0.10243
9	0.0890(-3%)	0.0741(-19%)	0.0915
11	0.0900(-2%)	0.08723(-5%)	0.09207

4. CONCLUSIONS

An extremum seeking method is introduced in this paper to find the extremum of a performance function for nonlinear discrete time systems. We first introduced that the extremum of the closed loop system is stable using an innovative discrete averaging method, and then by utilizing the singular perturbation method, the overall system is demonstrated to remain stable. The proposed method is based on a slow outer loop to a nonlinear system which already has fast closed loop controller in the inner loop. The stability analysis shows that if a fast stabilizing closed loop controller is provided for a plant, the proposed extremum seeking outer loop is able to converge to the optimum set point with an error defined by the UUB. The method is also applied to the HCCI engine dynamics to verify the theoretical result on a practical mechanical system. The results

show that the method can successfully maximize the efficiency while keeping the pressure rise rate within constraints.

REFERENCES

[1] T. Dierks and S. Jagannathan, "Online optimal control of affine nonlinear discrete-time systems with unknown internal dynamics by using time-based policy update," *IEEE Trans. on Neural Networks and Learning Systems*, vol. 23, no. 7, pp. 1118-1129, 2012.

[2] Z. Hua, and M. Krstic, "Optimal design of adaptive tracking controllers for non-linear systems," *Automatica*, vol. 33, no. 8, 1997.

[3] C. S. Drapper and Y. T. Li, "Principles of optimalizing control systems and an application to the internal combustion engine," *ASME*, vol. 160, pp. 1-16, 1951.

[4] M. Krstic´ and H.-H. Wang, "Stability of extremum seeking feedback for general nonlinear dynamic systems," *Automatica*, vol. 36, pp. 595–601, 2000.

[5] J. B. Bettis, J. A. Massey, J. A. Drallmeier, and S. Jagannathan, "A thermodynamics-based homogeneous charge compression ignition engine model for adaptive nonlinear controller development," *Journal of Automobile Engineering*, 2012.

[6] H. Zargarzadeh, S. Jagannathan, and J. Drallmeier, "Robust optimal control of uncertain nonaffine MIMO nonlinear discrete-time systems with application to HCCI engines," *Int. Journal of Adaptive Control and Signal Processing*, vol. 26, pp. 592-613, 2012.

[7] R. Leyva, C. Alonso, I. Queinnec, A. Cid-Pastor, D. Lagrange, and L. Martinez-Salamero, "MPPT of photovoltaic systems using extremum—Seeking control," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 42, no. 1, pp. 249–258, 2006.

[8] K. S. Peterson and A. G. Stefanopoulou, "Extremum seeking control for soft landing of an electromechanical valve actuator," *Automatica*, vol. 40, no. 6, pp. 1063–1069, 2004.

[9] N. J. Killingsworth and M. Krstic´, "PID tuning using extremum seeking: Online, model-free performance optimization," *IEEE Control Syst. Mag.*, vol. 26, no. 2, pp. 70–79, Feb. 2006.

[10] Li Yaoyu, M. A. Rotea, G.T.-C. Chiu, L. G. Mongeau, and In-Su Paek, "Extremum seeking control of a tunable thermoacoustic cooler," *Control Systems Technology, IEEE Transactions on* , vol.13, no.4, pp. 527- 536, July 2005.

[11] H. Yu and U. Ozguner, "Extremum-seeking control strategy for ABS system with time delay," *Proceedings of the American Control Conference*, vol. 5, pp. 3753-3758, 2002.

[12] A. Banaszuk and Y. Zhang, C. A. Jacobson, "Adaptive control of combustion instability using extremum-seeking ," *Proceedings of the American Control Conference*, pp.416-422, 2000.

[13] C. J. Young, M. Krstic, and K.B. Ariyur and J. S. Lee, "Extremum seeking control for discrete-time systems," *Automatic Control, IEEE Transactions on* , vol.47, no.2, pp.318-323, Feb 2002.

[14] D. Popovic, M. Jankovic, S. Magner and A. R. Teel, "Extremum seeking methods for optimization of variable cam timing engine operation," *IEEE Transactions on Control Systems Technology*, vol.14, no.3, pp. 398- 407, May 2006.

[15] E. W. Bai, L. C. Fu, and S. S. Sastry, "Averaging analysis for discrete time and sampled data adaptive systems," *IEEE Transactions on Circuits and Systems*, vol.35, no.2, pp.137-148, Feb 1988.

[16] R. B. Bouyekhf and A. El Moudni, "On analysis of discrete singularly perturbed nonlinear systems: Application to the study of stability properties," *J. Franklin Inst.*, vol. 334B, no. 2, pp. 199–212, 1997.

[17] H. K. Khalil, *Nonlinear Systems*, 3rd ed. Upper Saddle River, NJ: Prentice-Hall, 2002.

[18] K. Ogata, *Discrete-Time Control Systems*. Englewood Cliffs, NJ: Prentice Hall, 1995, ch. 4, pp. 185.

[19] Z. P. Jiang and Y. Wang, "A converse Lyapunov theorem for discrete-time systems with disturbances," *Systems & Control Letters*, vol. 45, no. 1, pp. 49-58, 2002.

[20] S. S. Ge, and J. Zhang, "Neural-network control of nonaffine nonlinear system with zero dynamics by state and output feedback," *Neural Networks, IEEE Transactions on* , vol.14, no.4, pp. 900- 918, 2003.

[21] L. Yih-Guang, W. Y. Wang, T. T. Lee, "Observer-based direct adaptive fuzzy-neural control for nonaffine nonlinear systems," *Neural Networks, IEEE Transactions on* , vol.16, no.4, pp.853-861, 2005.

III. ADAPTIVE NEURAL NETWORK-BASED OPTIMAL CONTROL OF NONLINEAR CONTINUOUS-TIME SYSTEM IN STRICT FEEDBACK FORM

Abstract— This paper focuses on neural network (NN) based optimal control of nonlinear continuous-time systems in strict feedback form. A single NN-based adaptive approach is designed to learn the infinite horizon continuous-time Hamilton-Jacobi-Bellman (HJB) equation while the corresponding optimal control input that minimizes the HJB equation is calculated in a forward-in-time manner without using value and policy iterations. First, the optimal control problem is solved for a generic multi-input and multi-output (MIMO) nonlinear system with a state feedback approach. Then, the approach is extended to single input and single output (SISO) nonlinear system by using output feedback via a nonlinear observer. Lyapunov techniques are used to show that all signals are uniformly ultimately bounded (UUB) and that the approximated control signals approach the optimal control inputs with small bounded error for both the state and output feedback-based controller designs. In the absence of NN reconstruction errors, asymptotic convergence to the optimal control is demonstrated. Finally, simulation examples are provided to validate the theoretical results.

Keywords- Online Nonlinear Optimal Control; Neural Network Control; Output Feedback Control; Strict Feedback Systems

I. INTRODUCTION

The stabilization of nonlinear systems is now an established field [1]-[4]. Many control techniques are available for stabilization of nonlinear systems such as feedback linearization [5][1], sliding mode scheme [1], backstepping [5], and online approximators (OLA's)-based methods [2]-[3] for both continuous and discrete-time

systems. However, it is desirable that the control law not only stabilizes the system, but also minimizes a pre-defined cost function [6],[7],[8]. In other words, optimal control of nonlinear systems is preferred over other techniques that guarantee stability alone.

It is well known that the optimal control of linear systems can be obtained by solving the well-known Riccati equation [8]. In contrast, the optimal control of nonlinear continuous or discrete-time systems is a much more challenging task that often requires solving the nonlinear Hamilton-Jacobi-Bellman (HJB) equation which does not have a closed-form solution.

Therefore nonlinear optimal control has been addressed initially with an off-line and backward-in-time approach [9] similar to Riccati equation in linear systems. Subsequently, different online approximator-based controller designs, often referred to as adaptive critic designs (ACD), are presented in [10]-[13] which evolve forward-in-time to overcome the iterative offline methodology. The central theme in these works is that the optimal control law and HJB function are approximated by online parametric structures, such as NN's in a forward-in-time manner using policy and value iterations. Although the techniques [10] are verified via simulations, the approximation errors are not considered and mathematical proofs of convergence are not offered.

Recently, several online methods are introduced to solve the optimal control via continuous and discrete time HJB and Hamilton-Jacobi-Isaacs (HJI) equations in [10]-[13]. In particular, [12], online policy iterations based on adaptive control and Q-learning [14] are developed to solve the continuous HJB and discrete HJI problems, respectively. Although, full knowledge of the system dynamics is not required, these methods [12] are applicable to linear systems and are based on value and policy iterations. While these

methods render stability, the number of iterations needed within a sampling interval for convergence is not known. In addition, these iterative schemes in general are not preferred for hardware implementation.

In contrast, in [6], a single online approximator-based ACD technique is introduced for continuous-time nonlinear system in affine form since traditional ACD schemes require two approximators or NNs. Lyapunov stability is included and policy and value iterations are not needed while computational complexity is reduced. Instead, value and policy are updated at each sampling interval thus making the scheme suitable for real-time control.

It is important to note that all the ACD techniques [6],[11] address either continuous-time or discrete-time nonlinear systems in affine form. To the best knowledge of the authors, no known adaptive critic-based optimal control scheme is available for nonlinear systems in strict feedback form. Different types of strict feedback systems are defined in the literature in [5]-[16] while control of such strict feedback nonlinear systems is considered by using backstepping scheme [5] without any optimality. Other papers focus on the unknown strict feedback system using NN-based schemes [15]-[16].

More recently, the inverse optimal control of strict feedback systems is introduced in [7] when the dynamics are assumed known. However, in the inverse optimal control problem, first the control law is designed, and then the associated cost function is identified for that control law in contrast with traditional optimal control where a control law is designed based on a given cost function.

Therefore, in this paper, a novel optimal control scheme is introduced for a nonlinear continuous-time system in strict-feedback form when the system dynamics are

considered known. The nonlinear system in strict feedback form is transformed into a nonlinear system in affine form by using the backstepping technique. Then a single online approximator (SOLA) is utilized to provide the cost function in forward-in-time manner. Lyapunov theory is utilized to demonstrate the convergence of this approximate optimal control scheme for the overall nonlinear system while explicitly considering the approximation errors resulting from the use of the online approximator (OLA) in the backstepping approach.

An initial stabilizing control is not required in contrast to [12] and the proposed scheme is developed forward-in-time in contrast with standard Riccati equation based backward-in-time solution. In addition, this scheme is developed without using value and/or policy iterations that are commonly found in ACD techniques [12], [13]. It is shown that the approximated control input approaches the optimal value over time. If the NN reconstruction errors become zero as in the case of traditional adaptive control, asymptotic stability is demonstrated. First, state-feedback based optimal design is considered and subsequently, an output feedback controller design is addressed.

The paper is organized as follows. Section II is dedicated to the optimal control of a class of strict feedback nonlinear continuous-time systems by transforming the system to an equivalent nonlinear system in affine form. Section III introduces an online optimal stabilization scheme for affine systems. Next, Section IV develops the results to an observer based output control approach where the states can are not measured. Finally, Section V provides numerical results for the proposed optimal controller.

In the next section, a solution to the optimal tracking control of nonlinear system in strict feedback form is introduced.

II. THE TRACKING PROBLEM FOR STRICT FEEDBACK SYSTEMS

Consider the multi-input multi-output (MIMO) nonlinear continuous-time system in the absence of disturbances described by

$$\dot{x}_i = f_i\left(x_1,...,x_i\right) + g_i\left(x_1,...,x_i\right)x_{i+1} \text{ for } \quad 1 \le i \le N \text{ and } N \ge 2 \tag{1}$$

$$\dot{x}_N = f_N\left(x_1,...,x_N\right) + g_N\left(x_1,...,x_N\right)u\,, \tag{2}$$

$$y = x_1 \tag{3}$$

where each $x_i \in \Re^m$ denotes a state vector, $u \in \Re^m$ represents the input vector with $f_i\left(x_1,...,x_i\right) \in \Re^m$, and $g_i\left(x_1,...,x_i\right) \in \Re^{m\times m}$ being nonlinear smooth functions. Here, for the system(1), the next system state is treated as the virtual control input. Nonetheless, the system is going to be controlled through the control input u. The following assumption is needed before we proceed.

Assumption 1. It is assumed that $g_i\left(x_1,...,x_i\right) \neq 0 \;\; (1 \le i \le N)$ belongs to $\Omega \in \Re^n$, and it is bounded above and below satisfying $g^i_{\min} \le \| g_i\left(x_1,...,x_i\right)\|_F \le g^i_{\max}$ when the Frobenius norm is applied and where $g^i_{\min}$ and $g^i_{\max}$ are positive constants. Besides, it is assumed that systems (1)-(2) is reachable.

Under the above conditions given in the Assumption 1, the optimal control input for the nonlinear system (1)-(2) can be calculated [8] through a backstepping approach.

In this case, the objective of our scheme is to design a controller u in order to have the output y to track a desired trajectory x_{1d} in an *optimal manner*. To this end, by applying the backstepping approach [5], the system given by (1)-(2) tracks a predesigned trajectory $\left(x_{2d},...,x_{Nd}\right)$. Now, we follow the steps in the standard backstepping scheme to attain the optimal scheme of the strict feedback systems.

To stabilize the tracking error, $e_1 = x_1 - x_{1d}$, the backstepping approach will use N steps [1] which are presented next.

Step.1: It is desired that x_1 follow the smooth desired trajectory x_{1d}. Therefore, the first system dynamics of (1) can be rewritten as

$$\begin{aligned} \dot{x}_1 - \dot{x}_{1d} = \dot{e}_1 &= -\dot{x}_{1d} + f_1(x_1) + g_1(x_1)x_{2d} + g_1(x_1)(x_2 - x_{2d}) \\ &= f_1(e_1) + g_1(x_1)x_{2d}^* + g_1(x_1)e_2 \ , \end{aligned} \tag{4}$$

where virtual control input x_{2d} is chosen such that $x_{2d} = x_{2d}^* + x_{2d}^a$ with the feedforward virtual control input x_{2d}^a selected by solving

$$-\dot{x}_{1d} + f_1(x_1) + g_1(x_1)x_{2d}^a = f_1(e_1). \tag{5}$$

Moreover, x_{2d}^* is going to be the optimal feedback control input. Section III is devoted to show the existence of x_{2d}^* and its design. Inevitably, e_2 cannot be zero due to dynamics of the second system of (1) and the desired output x_1 trajectory. Therefore, the next steps of the design procedure should handle this issue such that the last term of (4) gets cancelled by the next system dynamics in (1) by $i = 2$. Since the second step to the ($N-1$) step remains the same, we skip to the i*th* step.

Step. i: In this step, we need an optimal controller for the system (1)-(3) such that $e_i \to 0$. To this end, the system i in (1) can be rewritten as

$$\begin{aligned} &\dot{x}_i - \dot{x}_{id} = \dot{e}_i = -\dot{x}_{id} + f_i(x_1,...,x_i) + g_i(x_1,...,x_i)x_{(i+1)d} \\ &+ g_i(x_1,...,x_i)(x_{i+1} - x_{i+1d}) \end{aligned}$$

$$= f_i(e_1,...,e_i) + g_i(x_1,...,x_i)x_{(i+1)d}^* + g_i(x_1,...,x_i)e_{i+1} - g_{i-1}^T(x_1,...,x_{i-1})e_{i-1}, \tag{6}$$

where x_{id} is chosen such that $x_{(i+1)d} = x^*_{(i+1)d} + x^a_{(i+1)d}$, with the virtual control input $x^a_{(i+1)d}$ satisfying

$$-\dot{x}_{id} + f_i\left(x_1,...,x_i\right) + g_i\left(x_1,...,x_i\right)x^a_{(i+1)di} = f_i\left(e_1,...,e_i\right) - g^T_{i-1}\left(x_1,...,x_{i-1}\right)e_{i-1}.$$

(7)

As mentioned in the previous step, there exists an optimal solution for the virtual input $x^*_{(i+1)d}$ which will be designed in the next section. Moreover, the third term of (6) inevitably shows up due to the design procedure, while the fourth term is deliberately added due to stability considerations.

Step. N: In this step, similar to the previous steps, the system input will be designed. To this end, the system (2) can be rewritten as

$$\dot{x}_N - \dot{x}_{Nd} = \dot{e}_n = -\dot{x}_{Nd} + f_N\left(x_1,...,x_N\right) + g_N\left(x_1,...,x_i\right)u$$

$$= f_N\left(e_1,...,e_N\right) + g_N\left(x_1,...,x_N\right)u^* - g^T_{N-1}\left(x_1,...,x_{N-1}\right)e_N, \quad (8)$$

where x_{id} is chosen such that $u = u^* + u^a$, where the feedforward control input u^a is selected from

$$-\dot{x}_{Nd} + f_N\left(x_1,...,x_N\right) + g_N\left(x_1,...,x_N\right)u^a = f_N\left(e_1,...,e_N\right) - g^T_{N-1}\left(x_1,...,x_{N-1}\right)e_{N-1}, \quad (9)$$

As mentioned in the previous steps, there exists an optimal feedback control input u^* that will be designed in Section III. Now, we are ready to state the contribution of the current section in the following lemma.

Lemma 1. Consider the tracking dynamics defined in (4), (6), and (8). Assume that the virtual and real control input vector $U = \begin{bmatrix} x_{2d} & \cdots & x_{Nd} & u \end{bmatrix}$ is designed such that $U = U^a + U^*$ where $U^a = \begin{bmatrix} x^a_{2d} & \cdots & x^a_{Nd} & u^a \end{bmatrix}$ is the feedforward control input designed in

(5), (7), (9) and $U^* = \begin{bmatrix} x_{2d}^* & \cdots & x_{Nd}^* & u^* \end{bmatrix}$ represent the feedback control input which optimally stabilizes the following system

$$\begin{bmatrix} \dot{e}_1 \\ \vdots \\ \dot{e}_N \end{bmatrix} = \begin{bmatrix} f_1(e_1) \\ \vdots \\ f_N(e_1,...,e_N) \end{bmatrix} + \begin{bmatrix} g_1(x_1) & & 0 \\ & \ddots & \\ 0 & & g_N(x_1,...,x_N) \end{bmatrix} U^* \tag{10}$$

In this case, optimal control of (1) and (2) is equivalent to the optimal controller design for (10). In the other words, by applying $U = U^a + U^*$ to the system (1) and (2), the system dynamics (1) and (2) is transformed into the error dynamic system given by (10).

Proof. By choosing $J_1 = E^T E / 2$ with $E^T = \begin{bmatrix} e_1^T & \cdots & e_N^T \end{bmatrix}$ as the Lyapunov candidate and taking derivative through the system dynamics (4), (6), (8) we have

$$\dot{J}_1 = E^T \dot{E} = \sum_{i=1}^{N-1} e_i^T \left(f_i(e_1,...,e_i) + g_i(x_1,...,x_i) x_{(i+1)d}^* \right) + e_N^T \left(f_N(e_1,...,e_N) + g_N(x_1,...,x_N) u^* \right)$$

$$+ \sum_{i=1}^{N-1} e_i^T g_i(x_1,...,x_i) e_{i+1} - \sum_{i=2}^{N} e_i^T g_{i-1}^T(x_1,...,x_{i-1}) e_{i-1} . \tag{11}$$

One may easily recognize that the last two terms of (11) cancel each other. Therefore, the existence of an optimal controller to make the other terms negative is sufficient enough. On the other hand, equation (11) without the last two terms resembles the stability of the system (10) which proves the desired result.

III. OPTIMAL TRAJECTORY AND CONTROL INPUT DESIGN

Due to *Lemma 1*, the objective of this section to optimally stabilize the system (10). It is desired to design the optimal control vector defined by $\left[x_{2d}^*,...,x_{Nd}^*, u^* \right]$ such that the tracking error $(e_1,...,e_N)$ is stable while minimizing the cost function

$$V = \int_t^{\infty} r(E(\tau), U^*(\tau))d\tau\,, \tag{12}$$

where $E = \left[e_1,...,e_N\right]^T$, $U^* = \left[x_{2d}^*,...,x_{Nd}^*,u^*\right]^T$, and $\left[x_1,...,x_N\right] = X$. In(12), $r(E,U^*) = Q(E) + U^{*T}RU^*$, $Q(E) \geq 0$ is the positive semidefinite penalty on the states, and $R > 0 \in \Re^{M\times M}$ is a positive definite matrix with $M = mN$.

Equation (10) demonstrates that the optimal control of nonlinear system in strict feedback form can be transformed into solving optimal control of affine nonlinear system in the error domain. Now, consider the optimal stabilization problem for an affine type system in the error domain

$$\dot{E} = F(E) + G(X)U^*, \tag{13}$$

where $\left[f_1^T(e_1) \quad \cdots \quad f_N^T(e_1,...,e_N)\right]^T = F(E)$ and $G(X) = diag\left[g_1(x_1),...,g_N(x_1,...,x_N)\right]$. It is desired that E converges to zero while the cost function (12) is minimized.

Moving on, the control input U^* is required to be designed such that the cost function (12) is finite. We define the Hamiltonian for the cost function (11) with an associated admissible control input U to be [8]

$$H(E,U) = r(E,U) + V_E^T(E)\left(F(E) + G(X)U\right), \tag{14}$$

where $V_E(E)$ is the gradient of the $V(E)$ with respect to E. In the sequel, we will use the same terminology for denoting gradient of functions i.e. for any function $\Omega(\psi)$, $\Omega_\psi(\psi)$ means gradient of $\Omega(\psi)$ with respect to ψ. It is well-known that the optimal trajectory U^* that minimizes the cost function (12) also minimizes the Hamiltonian(14);

therefore, the optimal control is found by using the stationarity condition $\partial H(E,U)/\partial U = 0$ and revealed to be [8]

$$U^*(E) = -R^{-1}G(X)^T V_E^*(E)/2. \tag{15}$$

By substituting the optimal control (15) into the Hamiltonian (14) while observing $H(E,U) = 0$ reveals the HJB equation and the necessary and sufficient condition for optimal control to be [8]

$$Q(E) + V_E^T(E)F(X) - \frac{1}{4}V_E^T(E)G(X)R^{-1}G(X)^T V_E(E) = 0 \tag{16}$$

with $V^*(0) = 0$. For linear systems, equation (16) yields the standard algebraic Riccati equation (ARE) [8]. Before proceeding, the following technical lemma is required.

Lemma 2 [6]. Given the nonlinear system (13) with associated cost function (12) and optimal control (15), let $J(E)$ be a continuously differentiable, radially unbounded Lyapunov candidate such that $\dot{J}(E) = J_E^T(E)\dot{E} = J_E^T(E)\left(F(E) + G(X)U\right) < 0$ where $J_E^T(E)$ is the radially unbounded partial derivative of $J^T(E)$. Moreover, let $\bar{Q}(E)$ be a positive definite matrix satisfying $\left\|\bar{Q}(E)\right\| = 0$ only if $\|E\| = 0$ and $\bar{Q}_{\min} \le \left\|\bar{Q}(E)\right\| \le \bar{Q}_{\max}$ for $\chi_{\min} \le \|E\| \le \chi_{\max}$ for positive constants $\bar{Q}_{\min}, \bar{Q}_{\max}, \chi_{\min}$ and $\chi_{\max}$. In addition, let $\bar{Q}(E)$ satisfy $\lim_{E\to\infty}\bar{Q}(E) = \infty$ as well as

$$V_E^{*T}\bar{Q}(E)J_E = r(E,u^*) = Q(E) + U^{*T}RU^*. \tag{17}$$

Then, the following relation holds

$$J_E^T(F(E) + G(E)U^*) = -J_E^T\bar{Q}(E)J_E. \tag{18}$$

Proof: When the optimal control (15) is applied to the nonlinear system(13), the cost function (12) becomes a Lyapunov function rendering

$$\dot{V}^*(E) = V_E^{*T}(E)\dot{E} = V_E^{*T}(E)\left(F(E) + G(x)U^*\right) = -Q(E) - U^{*T}RU^* \tag{19}$$

From (15), after manipulation and substitution of (17), equation (19) is rewritten as

$$\begin{aligned} F(E) + G(X)U^* &= -(V_E^* V_E^{*T})^{-1} V_E^* (Q(E) + U^{*T}RU^*) \\ &= -(V_E^* V_E^{*T})^{-1} V_E^* V_E^{*T} \bar{Q}(E) J_E = -\bar{Q}(E) J_E \end{aligned} \tag{20}$$

Now, multiply both sides of (20) by J_E^T yields the desired relationship in(18).

■

In [13], the closed-loop dynamics $F(E) + G(E)U^*$ is required to satisfy a Lipschitz condition such that $\left\|F(E) + G(X)U^*\right\| \le K$ for a constant K. In contrast, the optimal closed loop dynamics are assumed to be upper bounded by a function of the system states in this work such that

$$\left\|F(E) + G(X)U^*\right\| \le \delta(E). \tag{21}$$

The generalized bound $\delta(E)$ is taken as $\delta(E) \equiv \sqrt[4]{K^* \left\|J_E\right\|}$ in this work where $\left\|J_E\right\|$ can be selected to satisfy general bounds and K^* is a constant. For example, if $\delta(E) = K_1 \left\|E\right\|$ for a constant K_1, then it can be shown that selecting $J(E) = (E^T E)^{(5/2)} / 5$ with $J_E(E) = (E^T E)^{(3/2)} E^T$ satisfies the bound. The assumption of a time-varying upper bound in (13) is a less stringent assumption than the constant upper bound required in [13]. The next section develops an approach for optimally stabilize the affine system which is required for optimal tracking of original strict feedback systems.

Moving on, we rewrite the cost function (11) using an OLA representation as

$$V(E) = \Theta^T \varphi(E) + \varepsilon(E), \tag{22}$$

where $\Theta \in \Re^L$ is the constant target OLA vector, $\varphi(E): \Re^n \to \Re^L$ is a linearly independent basis vector which satisfies $\varphi(0) = 0$, and $\varepsilon(E)$ is the OLA reconstruction error. The target OLA vector and reconstruction errors are assumed to be upper bounded according to $\|\Theta\| \le \Theta_M$ and $\|\varepsilon(E)\| \le \varepsilon_M$, respectively [3]. In addition, it will be assumed that the gradient of the OLA reconstruction error with respect to E is upper bounded according to $\|\partial\varepsilon(E)/\partial E\| = \|\nabla_E \varepsilon(E)\| \le \varepsilon'_M$. The gradient of the OLA cost function (22) is written as

$$\partial V(E)/\partial E = V_E(E) = \nabla_E^T \varphi(E)\Theta + \nabla_E \varepsilon(E). \tag{23}$$

Now, using (23), the optimal control (14) and HJB equation (16) are rewritten as

$$U^*(E) = -\frac{1}{2} R^{-1} G(E)^T \nabla_E^T \varphi(E)\Theta \; -\frac{1}{2} R^{-1} G(X)^T \nabla_E \varepsilon(E) \tag{24}$$

and

$$H^*(E,\Theta) = Q(E) + \Theta^T \nabla_E \varphi(E) F(E) \; -\frac{1}{4}\Theta^T \nabla_E \varphi(E) D \nabla_E^T \varphi(E)\Theta + \varepsilon_{HJB} = 0 \tag{25}$$

where $D = G(E)R^{-1}G(E)^T > 0$ is bounded such that $D_{\min} \le \|D\| \le D_{\max}$ for known constants $D_{\min}$ and $D_{\max}$, and

$$\varepsilon_{HJB} = \nabla_E \varepsilon^T \left(F(E) - \frac{1}{2} G(X) R^{-1} G(X)^T (\nabla_E^T \varphi(E)\Theta + \nabla_E \varepsilon)\right)$$

$$+\frac{1}{4}\nabla_E \varepsilon^T G(X) R^{-1} G(X)^T \nabla_E \varepsilon = \nabla_E \varepsilon^T (F(E) + G(X)U^*) + \frac{1}{4}\nabla_E \varepsilon^T D \nabla_E \varepsilon \tag{26}$$

is the residual error due to the OLA reconstruction error. Asserting the bounds for the optimal closed-loop dynamics (21) along with the boundedness of $G(X)$ and $\nabla_E \varepsilon$, the residual error ε_{HJB} is bounded above on a compact set according to $\left|\varepsilon_{HJB}\right| \le \varepsilon'_M \delta(E) + \varepsilon'^2_M D_{\max}$. In addition, it has been shown [13] that by increasing the dimension of the basis vector $\varphi(E)$ in the case of a single-layer NN, the OLA reconstruction error decreases.

Moving on, the OLA estimate of (11) is now written as

$$\hat{V}(E) = \hat{\Theta}^T \varphi(E) \tag{27}$$

where $\hat{\Theta}$ is the OLA estimate of the target parameter vector Θ. Similarly, the estimate of the optimal control (14) is written in terms of $\hat{\Theta}$ as

$$\hat{U}^* = -\frac{1}{2} R^{-1} G(X)^T \nabla_E^T \varphi(E) \hat{\Theta}. \tag{28}$$

It is shown [6] that an initial stabilizing control is not required to implement the proposed SOLA-based scheme in contrast to [11] and [13], which require initial control policies to be stabilizing. In fact, the proposed OLA parameter tuning law described next ensures that the system states remain bounded and that (28) will become admissible.

Now, using (22), the approximate Hamiltonian can be written as

$$\hat{H}(E, \hat{\Theta}) = Q(E) + \hat{\Theta}^T \nabla_E \varphi(E) F(E) - \frac{1}{4} \hat{\Theta}^T \nabla_E \varphi(E) D \nabla_E^T \varphi(E) \hat{\Theta}. \tag{29}$$

Observing the definition of the OLA approximation of the cost function (27) and the Hamiltonian function (29), it is evident that both become zero when $\|E\| = 0$. Thus, once the system states have converged to zero, the cost function approximation can no longer be updated. This can be viewed as a persistency of excitation (PE) requirement

for the inputs to the cost function OLA [11], [13]. That is, the system states must be persistently exiting long enough for the OLA to learn the optimal cost function.

Recalling the HJB equation shown in (16), the OLA estimate $\hat{\Theta}$ should be tuned to minimize $\hat{H}(E,\hat{\Theta})$. However, tuning to minimize $\hat{H}(E,\hat{\Theta})$ alone does not ensure the stability of the nonlinear system (13) during the OLA learning process. Therefore, the proposed OLA tuning algorithm is designed to minimize (29) while considering the stability of (13) and written as

$$\dot{\hat{\Theta}} = -\alpha_1 \frac{\hat{\sigma}}{(\hat{\sigma}^T\hat{\sigma}+1)^2}\left(Q(E)+\hat{\Theta}^T\nabla_E\varphi(E)F(E)-\frac{1}{4}\hat{\Theta}^T\nabla_E\varphi(E)D\nabla_E^T\varphi(E)\hat{\Theta}\right)$$

$$+\Sigma(E,\hat{u})\frac{\alpha_2}{2}\nabla_E\varphi(E)g(E)R^{-1}G(X)^T J_E(E) \tag{30}$$

where $\hat{\sigma} = \nabla_E\varphi F(E) - \nabla_E\varphi(E)D\nabla_E^T\varphi(E)\hat{\Theta}/2$, $\alpha_1 > 0$ and $\alpha_2 > 0$ are design constants, $J_E(E)$ is described in *Lemma 1*, and the operator $\Sigma(E,\hat{U})$ is given by

$$\Sigma(E,\hat{U}) = \begin{cases} 0 & \text{if } J_E^{\mathrm{T}}(E)\dot{E} = J_E^{\mathrm{T}}(E)(F(E) \\ & \quad -G(X)R^{-1}G(X)^T\nabla_E^T\varphi(E)\hat{\Theta}/2) < 0. \\ 1 & \text{otherwise} \end{cases} \tag{31}$$

The first term in (31) is the portion of the tuning law which seeks to minimize (29) and was derived using a normalized gradient descent scheme with the auxiliary HJB error defined as

$$E_{HJB} = \hat{H}(E,\hat{\Theta})^2/2. \tag{32}$$

Meanwhile, the second term in the OLA tuning law (30) is included to ensure the system states remain bounded while the SOLA scheme learns the optimal cost function. The form of the operator shown in (31) was selected based on the Lyapunov's sufficient

condition for stability (i.e. if $J(E) > 0$ and $\dot{J}(E) = J_E^T(E)\dot{E} < 0$, then the states E are stable). From the definition of the operator in (31), the second term in (30) is removed when the nonlinear system (13) exhibits stable behavior, and learning the HJB cost function becomes the primary objective of the OLA update (30). In contrast, when the system (13) exhibits signs of instability (i.e. $J_E^T(E)\dot{E} \geq 0$), the second term of (30) is activated and tunes the OLA parameter estimates until the nonlinear system (13) exhibits stable behavior.

Moving on, we now form the dynamics of the OLA parameter estimation error $\tilde{\Theta} = \Theta - \hat{\Theta}$. Observing $Q(E) = -\Theta^T \nabla_E \varphi(E) F(E) + \Theta^T \nabla_E \varphi(E) D \nabla_E^T \varphi(E) \Theta / 4 - \varepsilon_{HJB}$ from (24), the approximate HJB equation (29) can be rewritten in terms of $\tilde{\Theta}$ as

$$\hat{H}(E, \hat{\Theta}) = -\tilde{\Theta}^T \nabla_E \varphi(E) F(E) + \frac{1}{2} \tilde{\Theta}^T \nabla_E \varphi(E) D \nabla_E^T \varphi(E) \Theta$$

$$-\frac{1}{4} \tilde{\Theta}^T \nabla_E \varphi(E) D \nabla_E^T \varphi(E) \tilde{\Theta} - \varepsilon_{HJB}. \tag{33}$$

Next, observing $\dot{\tilde{\Theta}} = -\dot{\hat{\Theta}}$ and $\hat{\sigma} = \nabla_E \varphi(E)(\dot{E}^* + D \nabla_E \varepsilon / 2) + \nabla_E \varphi(E) D \nabla_E^T \varphi(E) \tilde{\Theta} / 2$ where $\dot{E}^* = F(E) + G(X) U^*$, the error dynamics of (20) are written as

$$\dot{\tilde{\Theta}} = -\frac{\alpha_1}{\rho^2} \left(\nabla_E \varphi(E) \left(\dot{E}^* + \frac{D \nabla_E \varepsilon}{2} \right) + \frac{\nabla_E \varphi(E) D \nabla_E^T \varphi(E) \tilde{\Theta}}{2} \right) \times$$

$$\left(\tilde{\Theta}^T \nabla_E \varphi(E) \left(\dot{E}^* + \frac{D \nabla_E \varepsilon}{2} \right) + \frac{1}{4} \tilde{\Theta}^T \nabla_E \varphi(E) D \nabla_E^T \varphi(E) \tilde{\Theta} + \varepsilon_{HJB} \right)$$

$$-\Sigma(E, \hat{U}) \frac{\alpha_2}{2} \nabla_E \varphi(E) G(X) R^{-1} G(X)^T J_E(E) \tag{34}$$

where $\rho = (\hat{\sigma}^T \hat{\sigma} + 1)$. Next, the stability of the SOLA-based adaptive scheme for optimal control is examined along with the stability of the nonlinear system (13).

Theorem 1: (SOLA-based Optimal Control Scheme). Given the nonlinear system (4), (6), and (8) with the target HJB equation (16), and let the tuning law for the SOLA be given by (30). Then, there exists computable positive constants b_{JE} and b_{Θ} such that the OLA approximation error $\tilde{\Theta}$ and $\|J_E(E)\|$ are uniformly ultimately bounded (*UUB)* [3] for all $t \geq t_0 + T$ with ultimate bounds given $\|J_E(E)\| \leq b_{JE}$ and $\left\|\tilde{\Theta}\right\| \leq b_{\Theta}$. Further, under OLA reconstruction errors, $\left\|V^* - \hat{V}\right\| \leq \varepsilon_{r1}$ and $\left\|U^* - \hat{U}\right\| \leq \varepsilon_{r2}$ for small positive constants ε_{r1} and ε_{r2}, respectively. Where, $b_{\Theta} \equiv \sqrt[4]{\eta(\varepsilon)/\beta_1}$ and $b_{JE} \equiv \alpha_1 \eta(\varepsilon) / \left(\alpha_2 \dot{x}_{\min} - \alpha_1 \beta_2 K^*\right)$. With β_2 chosen such that $\alpha_2 \dot{x}_{\min} - \alpha_1 \beta_2 K^* > 0$.

Proof. See appendix. ■

Next, the stability of the SOLA-based optimal control scheme can be examined when there are no OLA reconstruction errors as would be the case when standard adaptive control techniques [2] are utilized. In other words, when a NN is replaced with a standard linear in the unknown parameter (LIP) adaptive control, the parameter estimation errors and the states are globally asymptotically stable according to Corollary 1.

Corollary 1:(Ideal SOLA-based Optimal Control Scheme Convergence). Let the hypothesis of *Theorem 1* holds in the absence of OLA reconstruction errors. Then, the OLA approximation error vector $\tilde{\Theta}$ and system states E are globally asymptotically stable (GAS) and $\hat{V} \to V^*$ and $\hat{U} \to U^*$.

Proof: please refer to [6]. ■

Remark: It is shown in [17] that when the number of neurons in the hidden layer is chosen sufficiently large, the NN reconstruction error converges to zero. Alternatively, this error term is zero when a standard adaptive control is used.

The block diagram of the proposed state feedback-based optimal control scheme is shown in Fig. 1 where no value and policy iterations are utilized. Only the value function and control input are updated at the sampling interval.

IV. OBSERVATION BASED OUTPUT FEEDBACK CONTROL

Practically, the states are not measurable in a vast class of nonlinear systems. In this section, we consider the control problem of strict feedback control of the system (2)-(3) where $f_i(.)$ and $g_i(.)$ are known, whereas the state vector is not measured and only the output $y = h(x)$ is given. The multi-input multi-output (MIMO) feedback control of strict feedback systems will have to mitigate several challenges and will be relegated for a future publication. For example, selecting different outputs can change the relative degree of the system which in turn can complicate the process of the controller design. Therefore, we consider the system (1)-(3) to a single-input and single-output (SISO) case. This problem is still difficult as no known output feedback-based optimal control scheme is available in the forward-in-time manner for nonlinear systems, although recently for linear systems some results are achieved [17]. Now, assume that (1)-(3) is represented in

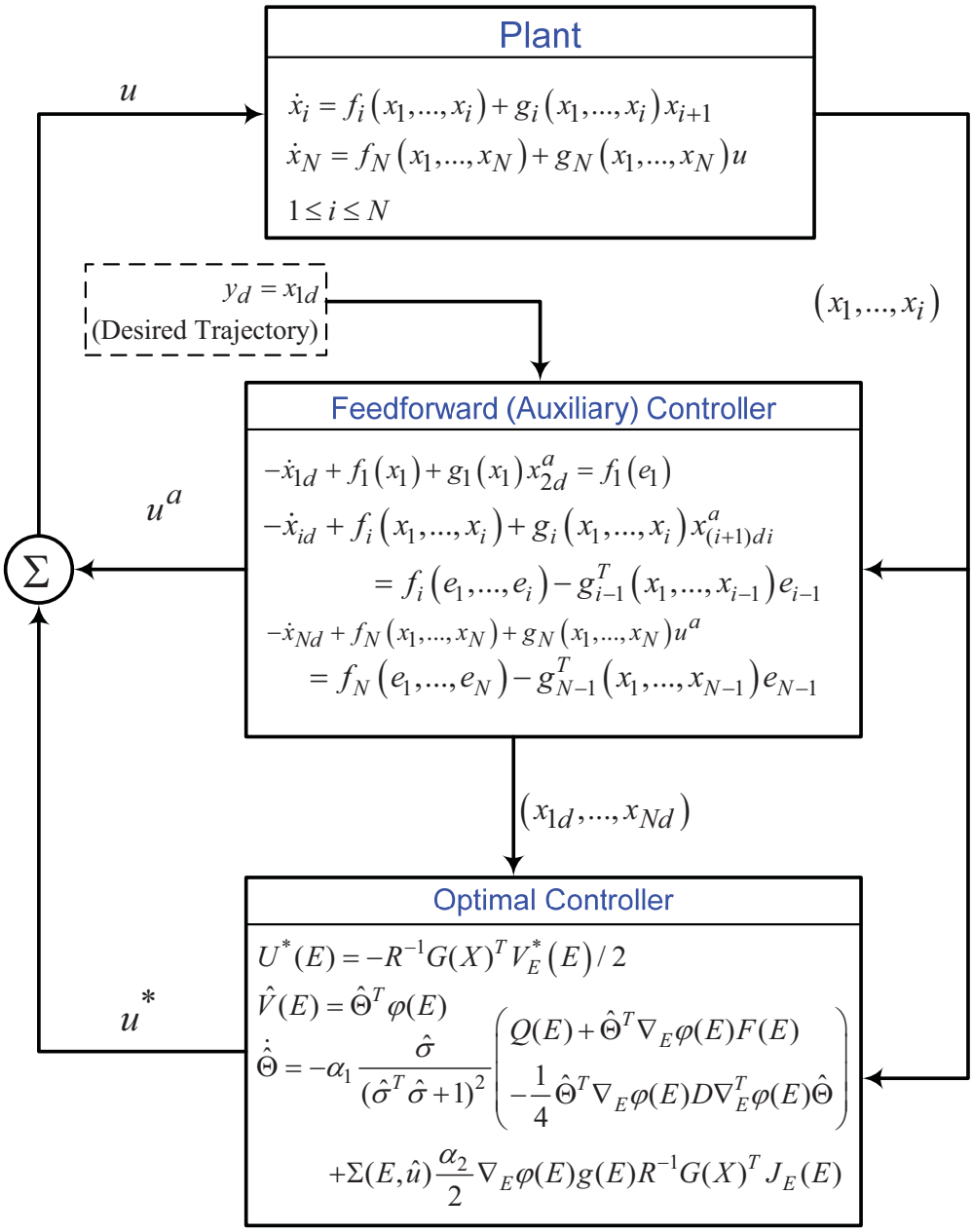

Figure 1. Block diagram of the state feedback-based optimal controller.

a SISO representation i.e. $x_i \in \Re$ and $u \in \Re$. It is shown [5] that, in this case, there exists a mapping $\zeta = (\zeta_1,...,\zeta_N) = \aleph(x_1,...,x_N)$ that transforms the system (1)-(3) into a new state space representation as

$$\dot{\zeta}_1 = \zeta_2 + \omega_1(y)$$

$$\dot{\zeta}_2 = \zeta_3 + \omega_2(y)$$

$$\vdots$$

$$\dot{\zeta}_{N-1} = \zeta_N + \omega_{N-1}(y)$$

$$\dot{\zeta}_N = \omega_N(y) + b\beta(y)u$$

$$y = \zeta_1 = h(x) \tag{35}$$

where $\omega_i(y) \in \Re$ are known functions of the output. The transformation $\aleph$ exists only when the relative degree of (1)-(3) (in SISO case) is equal to N. To overcome the need for states measurement, define the observer dynamics as

$$\dot{\hat{\zeta}} = A\hat{\zeta} + k(y - \hat{y}) + \omega(y) + b\beta(y)u$$

$$\hat{y} = c^T\hat{\zeta} \tag{36}$$

where

$$A = \begin{bmatrix} 0 & & I \\ \vdots & \ddots & \\ 0 & \cdots & 0 \end{bmatrix}, \quad b = \begin{bmatrix} 0 \\ \vdots \\ 0 \\ 1 \end{bmatrix}, \quad c = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \quad \omega(y) = \begin{bmatrix} \omega_1(y) \\ \vdots \\ \omega_n(y) \end{bmatrix}.$$

Therefore, with $A_o = A - kc^T$ being Hurwitz, we conclude that the closed-loop observer dynamics can decay exponentially to the origin. Therefore, by defining $\tilde{\zeta} = \zeta - \hat{\zeta}$ the observer error dynamics takes the following form.

$$\dot{\tilde{\zeta}} = A_o\tilde{\zeta}\,. \tag{37}$$

Now, we apply the same back stepping approach of previous section with the assumption that ζ_i for $i = 2,...,N$ are not measured but estimated using the observer(36). By following the Steps 1 through N, we get

$$\begin{bmatrix} \dot{\hat{e}}_1 \\ \vdots \\ \dot{\hat{e}}_N \end{bmatrix} = \begin{bmatrix} \omega_1(\hat{e}_1) \\ \vdots \\ \omega_N(\hat{e}_1) \end{bmatrix} + \begin{bmatrix} 1 & \cdots & & 0 \\ 0 & \ddots & & \vdots \\ \vdots & & 1 & 0 \\ 0 & \cdots & 0 & \beta(y) \end{bmatrix} U^* + A\tilde{\zeta}$$

$$\equiv \omega(\hat{e}_1) + \mathrm{B}(y)U^* + A\tilde{\zeta} \ , \tag{38}$$

with $\hat{e}_i = \hat{\zeta}_i - \zeta_{id}$ that implies $e_1 = \hat{e}_1 = y - y_d$ since $\hat{\zeta}_1 = \zeta_1 = y$. Moreover, the desired trajectory (feedforward controller) is designed as follows:

$$\begin{aligned} &-\dot{y}_{1d} + \varphi_1(y) + \hat{\zeta}^a_{2d} = \omega_1(\hat{e}_1) \\ &\vdots \\ &-\dot{\zeta}_{id} + \varphi_i(y) + \hat{\zeta}^a_{(i+1)di} = \omega_i(\hat{e}_i) - \hat{e}_{i-1} \\ &\vdots \\ &-\dot{\zeta}_{Nd} + \omega_N(y) + b\beta(y)u^a = \omega_N(\hat{e}_N) - \hat{e}_{N-1}. \end{aligned} \tag{39}$$

Equations (39) are identical to the equations derived for the steps 1 to N in Section II unless we used the system output y and the estimated states $\hat{\zeta}$ (by the observer) instead of the real value of ζ. This is the reason that an estimation error term $A\tilde{\zeta}$ appear in error dynamics (38). Moreover, in (39) $U_1^* = \left[\zeta_{2d}^* \quad \cdots \quad \zeta_{Nd}^* \quad u*\right]$. Theorem 2 will show that the state estimation error $\tilde{\zeta} = \zeta - \hat{\zeta}$ is guaranteed to be bounded which is necessary for the overall stability of the closed loop system. Using $r_1(\hat{E}, U^*) = Q_1(\hat{E}) + U_1^{*T} R_1 U_1^*$, where $\hat{E} = \left[\hat{e}_1, ..., \hat{e}_N\right]^T$ the target HJB equation takes the following form

$$Q_1(\hat{E}) + V_{1E}^T(\hat{E})\omega(e_1) - \frac{1}{4}V_{1\hat{E}}^T\left(\hat{E}\right)B(y)R_1^{-1}B(y)^T V_{1\hat{E}}\left(\hat{E}\right) = 0. \tag{40}$$

with $\hat{E} = \begin{bmatrix} e_1 & \cdots & e_N \end{bmatrix}^T$, $V_1 = \int_t^\infty r_1(\hat{E}(\tau), U^*(\tau))d\tau$ as the cost function, $R_1 \in \Re^{N\times N}$ with $R_1 > 0$ since $m = 1$, and $Q_1(\hat{E})$ is a positive semidefinite function of $\hat{E}$. Therefore, the optimal controller for this case can be represented as follows:

$$U_1^*(\hat{E}) = -\frac{1}{2} R_1^{-1} \mathrm{B}\left(y\right)^T V_{1\hat{E}}^*(\hat{E}). \tag{41}$$

Now, consider an OLA representation as

$$V_1(\hat{E}) = \Theta_1^T \varphi_1(\hat{E}) + \varepsilon_1(\hat{E}), \tag{42}$$

with $\varepsilon_1(\hat{E})$ as the estimation error and update law as

$$\dot{\hat{\Theta}}_1 = -\gamma_1 \frac{\hat{\sigma}_1}{(\hat{\sigma}_1^T \hat{\sigma}_1 + 1)^2} \left(Q_1(\hat{E}) + \hat{\Theta}_1^T \nabla_{\hat{E}} \varphi_1(\hat{E}) \omega(y) - \frac{1}{4} \hat{\Theta}^T \nabla_{\hat{E}} \varphi_1(\hat{E}) D_1 \nabla_{\hat{E}}^T \varphi_1(\hat{E}) \hat{\Theta}_1 \right)$$

$$+ \Sigma_1(\hat{E}, \hat{U}_1) \frac{\gamma_2}{2} \nabla_{\hat{E}} \varphi_1(\hat{E}) \mathrm{B}\left(y\right) R_1^{-1} \mathrm{B}\left(y\right)^T J_{1\hat{E}}(\hat{E}) \tag{43}$$

with

$$\Sigma_1(\hat{E}, \hat{U}) = \begin{cases} 0 & \text{if} \quad J_{1\hat{\mathrm{E}}}^{\mathrm{T}}(\hat{E})\dot{\hat{E}} = J_{1E}^{\mathrm{T}}(\hat{E})(\omega\left(\hat{e}_1\right) \\ & \quad - B(y)R^{-1}B(y)^T \nabla_{\hat{E}}^T \varphi_1(\hat{E})\hat{\Theta}_1 / 2) < 0. \\ 1 & \text{otherwise} \end{cases} \tag{44}$$

Here, $J_1(\hat{E})$ is a positive definite radially unbounded function of $\hat{E}$, $\gamma_1, \gamma_2 > 0$ are real design parameters, Θ_1 is the target parameter and $\varphi_1(\hat{E})$ the basis function for the estimation of $V_1(\hat{E})$. Moreover,

$$\hat{\sigma}_1 = \nabla_{\hat{E}} \varphi_1 \omega\left(\hat{e}_1\right) - \nabla_{\hat{E}} \varphi_1(\hat{E}) D_1 \nabla_{\hat{E}}^T \varphi_1(\hat{E}) \hat{\Theta}_1 / 2, \tag{45}$$

with $D_1 = B(y)R_1^{-1}B(y)^T > 0$ where $D_{1\min} < \| D_1 \| < D_{1\max}$. It is finally assumed that $\left\| \omega\left(\hat{e}_1\right) + B(y)U_1^* \right\| \le \delta_1(E) \equiv \sqrt[4]{K_1^* \| J_{\hat{E}} \|}$ with $K_1^* > 0$.

We can now introduce *Theorem 2* under the case where the states are not measured while the output is only available.

Theorem 2: (Output Feedback SOLA-based Optimal Control Scheme). Assume that the states of the nonlinear system (1) through (3) are not measurable while the output is only available with $m = 1$. Assume also that x_i are transformed using $\aleph[x_1,...,x_N]$ to ζ which forms the system dynamics into (35). Given the nonlinear system (35), the observer (36) and the target HJB equation (40), and let the tuning law for the SOLA be given by (43) with the cost function estimation $\hat{V}_1(\hat{E}) = \hat{\Theta}_1^T \varphi_1(\hat{E})$. Then, there exists computable positive constants b_{1JE} , $b_{1\Theta}$, and $b_{1\zeta}$ such that the OLA approximation error $\tilde{\Theta}_1 = \Theta_1 - \hat{\Theta}_1$, $\| J_{1\hat{E}}(\hat{E}) \|$, and $\tilde{\zeta}$ are *UUB* for all $t \geq t_0 + T$ with ultimate bounds given $\| J_{1\hat{E}}(\hat{E}) \| \leq b_{1JE}$, $\| \tilde{\Theta}_1 \| \leq b_{1\Theta}$, and $\| \tilde{\zeta} \| < b_{1\zeta}$. Further, under OLA reconstruction errors, $\| V_1^* - \hat{V}_1 \| \leq \overline{\varepsilon}_{r1}$ and $\| U_1^* - \hat{U}_1 \| \leq \overline{\varepsilon}_{r2}$ for small positive constants $\overline{\varepsilon}_{r1}$ and $\overline{\varepsilon}_{r2}$, respectively where $b_{1\Theta} \equiv \sqrt[4]{\eta_1(\varepsilon)/\tau_1}$, $b_{1JE} \equiv \gamma_1 \eta_1(\varepsilon) / \left(\gamma_2 \hat{\dot{E}}_{\min} - \gamma_1 \tau_2 K_1^* \right)$, and $b_{\zeta} \equiv \rho_1^{-1} \sqrt{\gamma_1 \eta_1(\varepsilon) / \lambda_{T\min}}$ provided $\gamma_2 / \gamma_1 > \tau_2 K_1^* / \hat{\dot{E}}_{\min}$ where $\rho_1 = (\hat{\sigma}_1^T \hat{\sigma}_1 + 1)$. In addition, $-T = A_o^T P + P A_o$ with P and T being an arbitrary positive definite matrix and $\lambda_{T\min}$ being the minimum eigenvalue of T.

Proof. See the appendix. ■

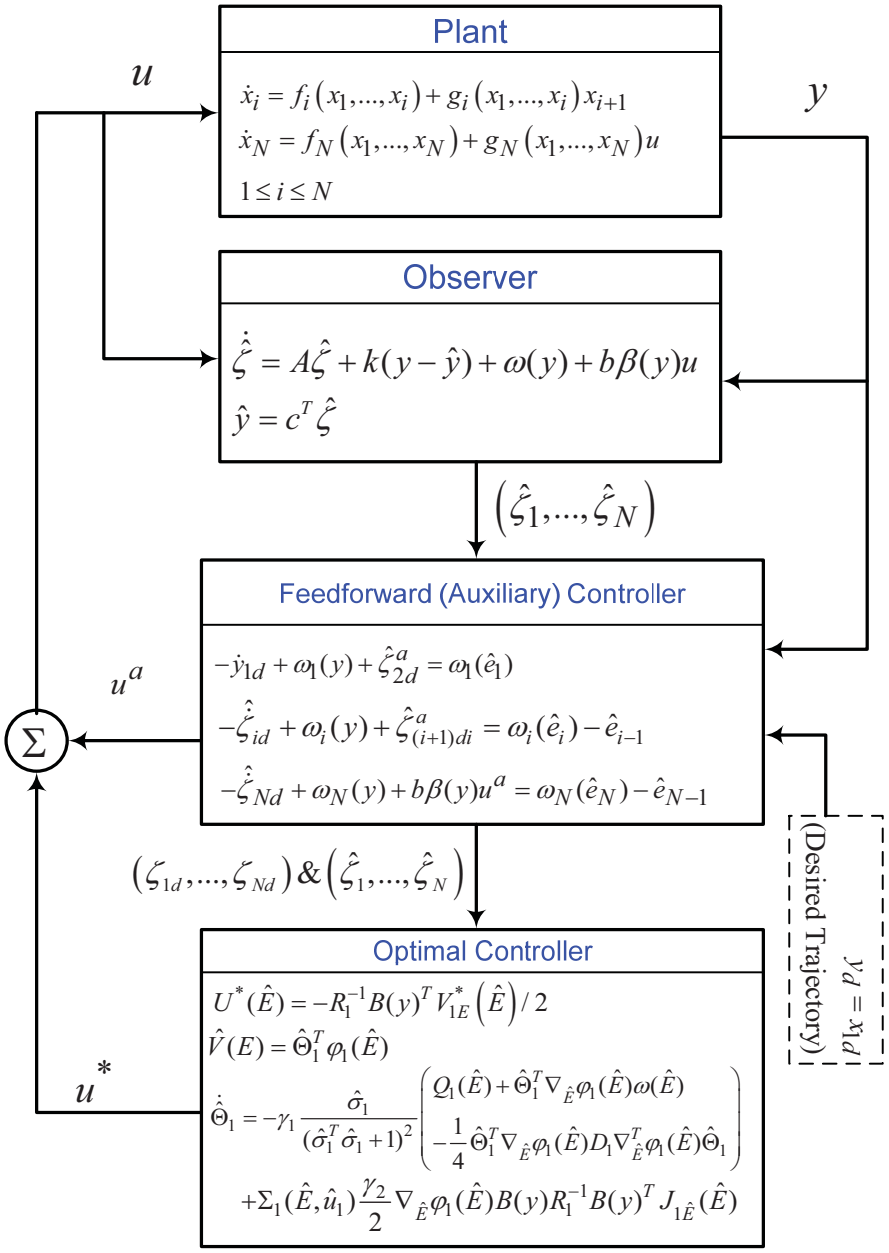

Figure 2. Block diagram the proposed output feedback controller.

The block diagram of the proposed output feedback-based optimal control scheme is shown below where no value and policy iterations are utilized. Only the value function and control input are updated at the sampling interval. The interesting point of this approach is now revealed in this diagram: the observer is observing the transformed

parameters ζ when it is applied to the real system. This means that by guaranteeing the existence of $\aleph(X)$, the user does not need the system model to the form of (35).

V. SIMULATION RESULTS

In this section, first a MIMO system is considered and a state feedback optimal approach is designed and verified in simulation. Subsequently, the output feedback-based optimal scheme is evaluated in another example.

A. *MIMO Online Optimal Control*

Consider the following nonlinear system in the form of (1)-(2) respectively as

$$\begin{bmatrix}\dot{x}_1\\ \dot{x}_2\end{bmatrix}=\begin{bmatrix}x_2\\ -x_1\left(\frac{\pi}{2}+\tan^{-1}(5x_1)\right)-\frac{5x_1^2}{2(1+25x_1^2)}+4x_2\end{bmatrix}+\begin{bmatrix}1 & 0\\ 0 & 3\end{bmatrix}\begin{pmatrix}z_1\\ z_2\end{pmatrix} \tag{46}$$

$$\begin{bmatrix}\dot{z}_1\\ \dot{z}_2\end{bmatrix}=\begin{bmatrix}-4z_1+x_2^2-2x_1\\ -3z_2+2x_2^2-x_1\end{bmatrix}+\begin{bmatrix}1+z_1^2 & 0\\ 0 & 1+\frac{1}{2}\cos(z_1+x_1)\end{bmatrix}u \tag{47}$$

Using the HJB cost function (3) with $Q(x)=E^TE$ and $R=1$, the basis vector for the SOLA-based scheme implementation was selected as $\varphi(E)=\begin{bmatrix}x_{e1} & x_{e2} & x_{e1}x_{e2} & x_{e1}^2 & x_{e2}^2 & x_{e1}^2\tan^{-1}(5x_{e1}) & x_{e1}^3\\ z_{e1} & z_{e2} & z_{e1}z_{e2} & z_{e1}^2 & z_{e2}^2 & z_{e1}^2\tan^{-1}(5z_{e1}) & z_{e1}^3\end{bmatrix}^T$ while the tuning parameters were selected as $\alpha_1=200$ and $\alpha_2=0.01$. Moreover, $x_{e1}=x_1-x_{1d}$, $x_{e2}=x_2-x_{2d}$, $z_{e1}=z_1-z_{1d}$, and $z_{e2}=z_2-z_{2d}$. The initial conditions of the system states were taken as $\begin{bmatrix}x_1 & x_2 & z_1 & z_2\end{bmatrix}^T=[2 \quad -2 \quad 2 \quad 2]^T$ while all NN weights were initialized to zero. That is, no initial stabilizing control was utilized for implementation of this online design for

the nonlinear system. Moreover, it is desired that the output track $X_d = \left[\sin(t/50) \quad \sin(t/40)\right]^T$ as the desired trajectory.

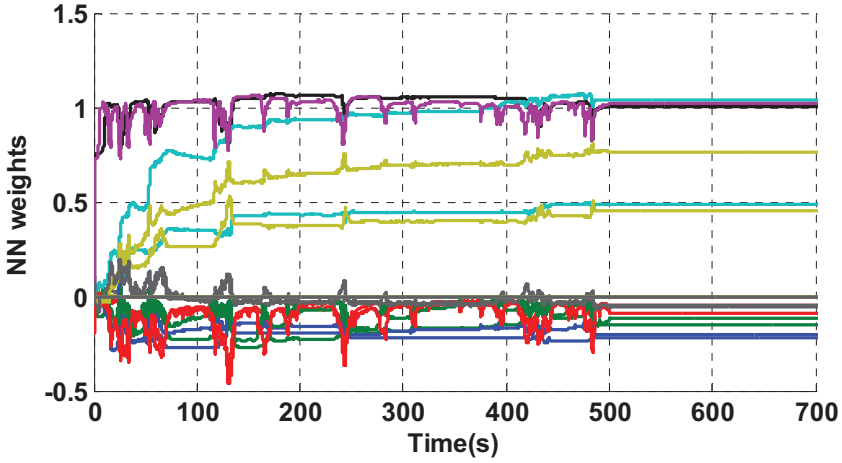

Figure 3. The evolution of NN weights with time.

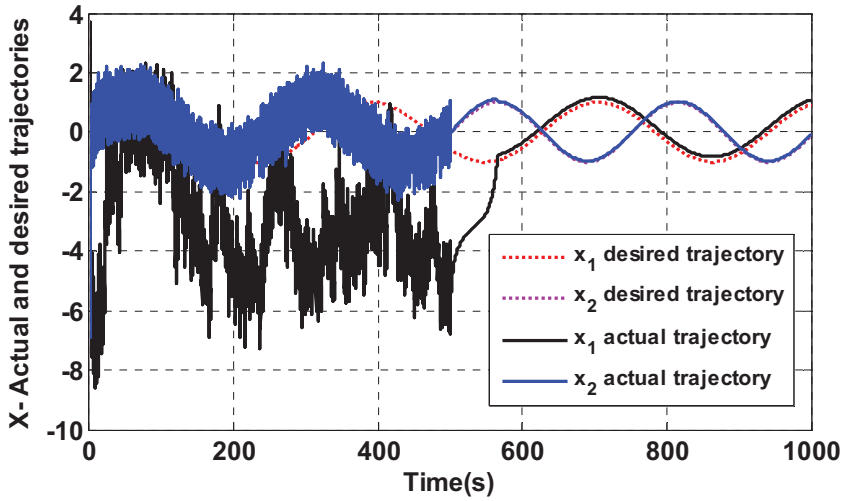

Figure 4. The convergence of system outputs to the desired trajectory.

Fig 3 depicts the evolution of the OLA weights during the online learning. Starting from zero, the weights of the online OLA are tuned to learn the optimal cost function. The system output ($X = [x_1, x_2]^T$) are shown in Fig. 4, and noise is added to each state to ensure the persistency of excitation condition (PE) condition is satisfied. After 150 seconds, the PE condition was no longer required and was thus removed. Fig. 5

depicts the stability of the internal system states ($z = [z_1, z_2]^T$).Fig. 6 shows the control input to the system $\hat{U}^*$.

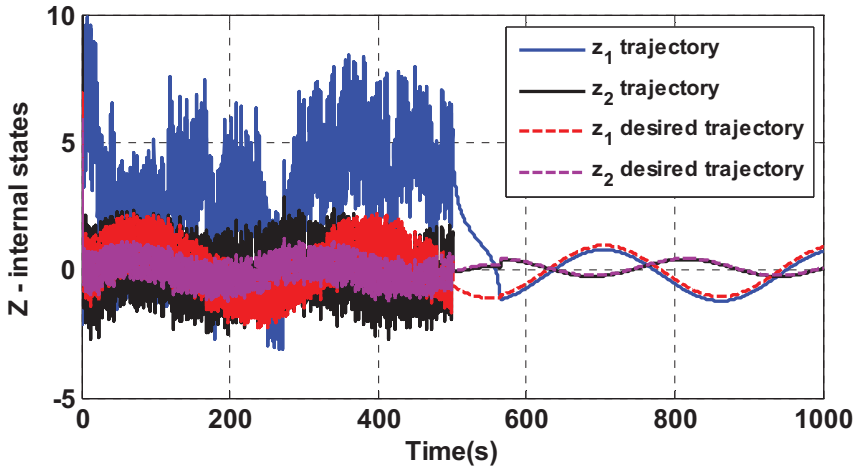

Figure 5. The convergence of the internal system states.

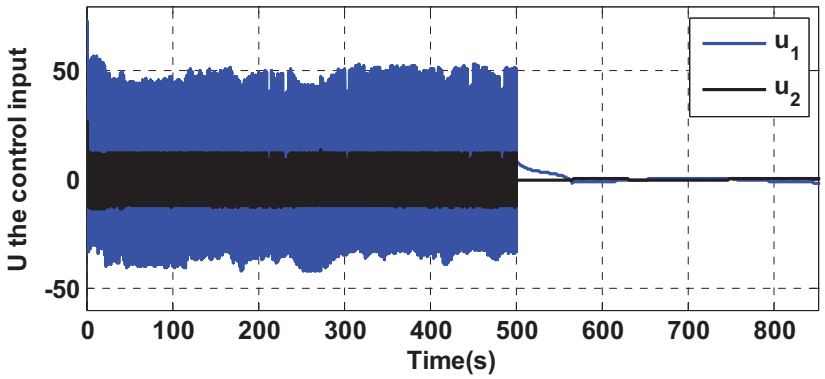

Figure 6. The actual control input to the system $\hat{U}^*$.

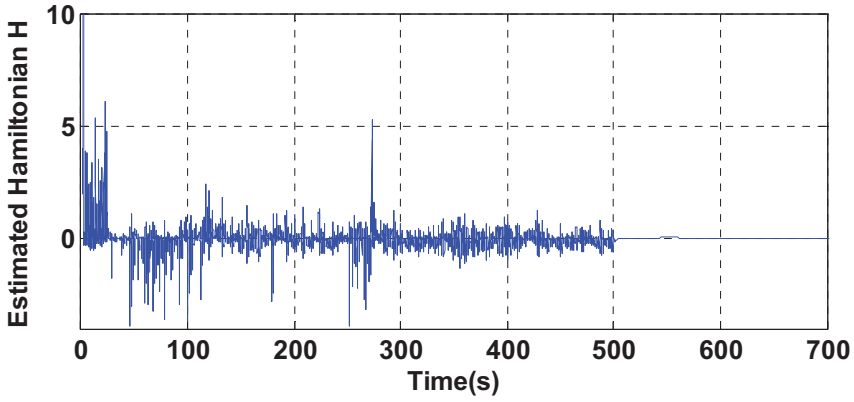

Figure 7. Approximation of the Hamiltonian.

Finally, in the case of Figs. 3 through 6, Fig. 7 demonstrates the estimated Hamiltonian in equation (29). To demonstrate the importance of the secondary term in the tuning law in (30), the online OLA design is attempted with $\Sigma(x,\hat{u})=0$. That is, the learning algorithm only seeks to minimize the auxiliary HJB residual (32) and does not consider system stability. Fig. 8 shows the results of not considering the nonlinear system's stability while learning the optimal HJB function. From this figure, it is clear that the system state quickly escape to infinity, and the SOLA-based controller fails to learn the HJB function. Thus, the importance of the secondary term in (30) which ensures the stability of the system is revealed.

B. Observer Based Online Optimal Control Output Feedback Control

Consider the following nonlinear system in the form of (1)-(2) respectively as

$$\dot{x}=-x\left(\frac{\pi}{2}+\tan^{-1}(5x)\right)-\frac{5x^2}{2(1+25x^2)}+4x+z \tag{48}$$

$$\dot{z}=2x^2-x+\left\{1+\frac{1}{2}\cos\left(x\right)\right\}u\,. \tag{49}$$

$$y=x\,, \tag{50}$$

which is in the form of system (35). Here, we repeat the experiment of the part (a) with the assumption that z is not measurable. Using the HJB cost function (3) with $Q(x)=E^TE$ and $R=1$, the basis vector for the SOLA-based scheme implementation was selected as $\varphi(E)=\left[x_e\;\; x_e^2\;\; x_e^3\;\; x_e^2\tan^{-1}(5x_e)\;\; z_e\; z_e^2\; z_e^2\tan^{-1}(5z_e)\; z_e^3\right]^T$ while the tuning parameters were selected as $\alpha_1=200$, $\alpha_2=0.01$, and $\lambda_{T\min}{}^{-1}=0.04$. Moreover, $x_e=x-x_d$, $z_e=\hat{z}-z_d$, and $A_o=0.1$. The initial conditions of the system states were taken as $\left[x\;\; z\right]^T=\left[2\;\; -2\right]^T$ while all NN weights were initialized to zero. That is, no

initial stabilizing control was utilized for implementation of this online design for the nonlinear system. Moreover, it is desired that the output track $x_d = \sin(t/50)$ as the desired trajectory.

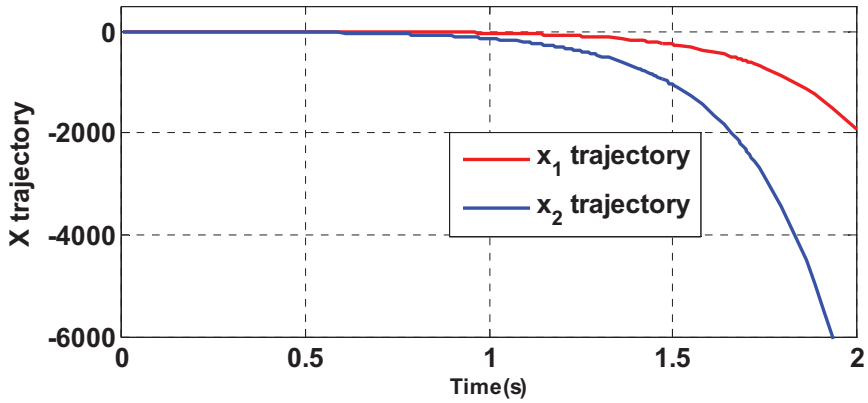

Figure 8. The system output without the OLA update.

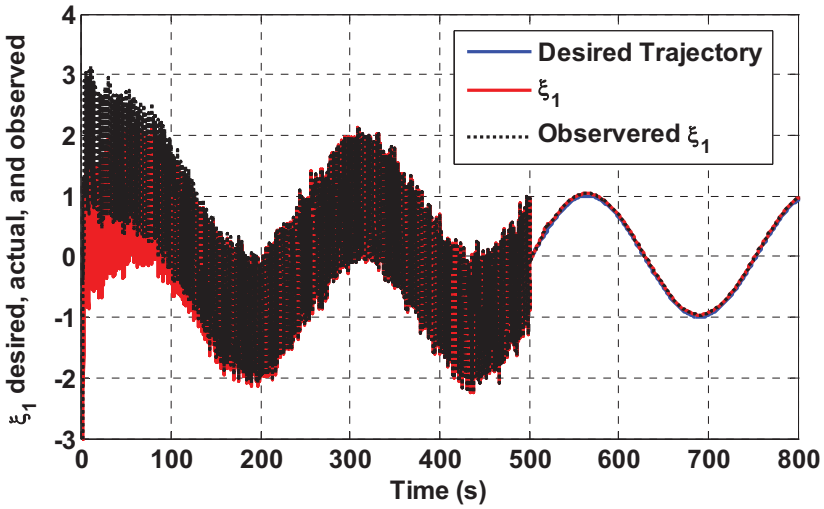

Figure 9. output ζ_1, the observed output $\hat{\zeta}_1$, and desired trajectory.

The simulation results are given in Figs. 9 and 10. In these figures the convergence of $\hat{\zeta}_1$ and $\hat{\zeta}_2$ to ζ_1 and ζ_2 are depicted. We can check from (A.12) and (A.14) that by properly choosing $\lambda_{T\min}$, the upper bound of $\|\tilde{\zeta}\|$ can be arbitrarily adjusted as small as desired. Therefore, after a transient response time (about 100 seconds), the observed state $\hat{\zeta}$ is equal to ζ and the online optimal controller can rely on

the observed value instead of the real value. Therefore, after removing the PE condition, the tracking error convergences to a uniformly ultimate bounded region close enough to the origin.

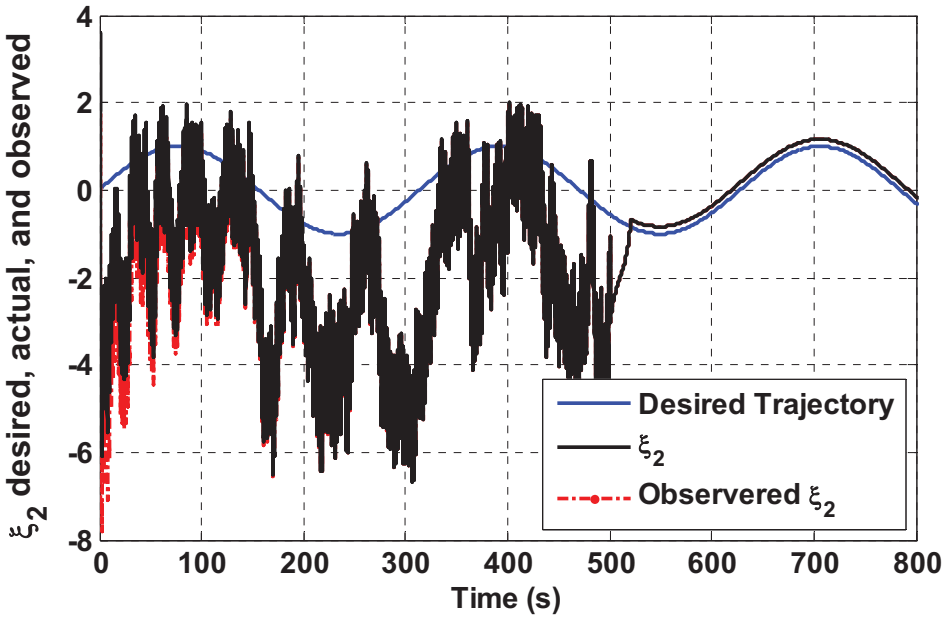

Figure 10. System output ζ_2, observed output $\hat{\zeta}_2$, and desired trajectory.

VI. CONCLUSIONS

This work proposed an optimal scheme for stabilizing nonlinear MIMO strict feedback systems using a single OLA to solve the Hamilton Jacobi-Bellman equation forward-in-time. In the presence of known dynamics, the regulation problem was undertaken. Then, by using a backstepping approach, the control input to the system is derived. Moreover, as a practical application, this scheme is developed to the optimal output feedback of SISO systems. A nonlinear observer is designed in order to estimate the unknown states in the output feedback case. UUB stability of the overall system is guaranteed in the presence of OLA approximation error. Simulation results were also

provided to verify the theoretical conjectures. Future work is to extend the results of the output feedback case of SISO to MIMO systems.

APPENDIX

Proof of Theorem 1. Consider the following positive definite Lyapunov candidate

$$J_{HJB} = \alpha_2 J_1(E) + \frac{1}{2}\tilde{\Theta}^T\tilde{\Theta} = \frac{1}{2}\left(\alpha_2 E^T E + \tilde{\Theta}^T\tilde{\Theta}\right) \tag{A.1}$$

whose first derivate with respect to time is given by

$$\dot{J}_{HJB} = \alpha_2 J_{1E}^T(E)\dot{E} + \tilde{\Theta}^T\dot{\tilde{\Theta}} = \alpha_2 E\dot{E} + \tilde{\Theta}^T\dot{\tilde{\Theta}} \tag{A.2}$$

where $J_1(E)$ is given in *Lemma 1* and $J_{1E}(E) = E$. One can easily find out that Equation (A.2), along the system trajectories (4), (6), and (8) is equal to $\dot{J}_{HJB}$ along the system (13) and (34). Therefore the optimal tracking problem of (4), (6), (8) is reduced to optimal stabilization of the system (13).

To begin the proof of the overall stability, observe that if $\|E\| = 0$, then $J_{HJB} = \tilde{\Theta}^T\tilde{\Theta}/2$ with $\dot{J}_{HJB} = 0$, and the parameter estimation error $\left\|\tilde{\Theta}\right\|$ remains constant and bounded [3]. On the other hand, to successfully accomplish the online learned objective, the states are required to satisfy $\|E\| > 0$. Therefore, the remainder of this proof considers the case of $\|E\| > 0$ (i.e. online learning is being performed). Then, substituting the nonlinear dynamics (13) with control input (28) applied along with the OLA estimation error dynamics (34) into (31) reveals

$$\dot{J}_{HJB} = \alpha_2 J_E^T(E)\left(F(E) - \frac{1}{2} G(X) R^{-1} G(X)^T \nabla_E^T \varphi(E) \hat{\Theta} \right)$$
$$- \frac{\alpha_1}{\rho^2}\left(\tilde{\Theta}^T \nabla_E \varphi(E)\left(\dot{E}^* + \frac{D\nabla_E \varepsilon}{2} \right)\right)^2 - \frac{\alpha_1}{8\rho^2}\left(\tilde{\Theta}^T \nabla_E \varphi(E) D \nabla_E^T \varphi(E) \tilde{\Theta} \right)^2$$
$$- \frac{3\alpha_1}{4\rho^2} \tilde{\Theta}^T \nabla_E \varphi(E)\left(\dot{E}^* + \frac{D\nabla_E \varepsilon}{2} \right) \tilde{\Theta}^T \nabla_E \varphi(E) D \nabla_E^T \varphi(E) \tilde{\Theta}$$
$$- \frac{\alpha_1}{\rho^2} \tilde{\Theta}^T \nabla_E \varphi(E)\left(\dot{E}^* + \frac{D\nabla_E \varepsilon}{2} \right) \varepsilon_{HJB}$$
$$- \frac{\alpha_1}{2\rho^2} \tilde{\Theta}^T \nabla_E \varphi(E) D \nabla_E^T \varphi(E) \tilde{\Theta} \varepsilon_{HJB}$$
$$- \Sigma(E, \hat{u}_1) \frac{\alpha_2}{2} \tilde{\Theta}^T \nabla_E \varphi(E) G(X) R^{-1} G(x)^T J_{1E}^T(x).$$

Next, completing the squares with respect to $\tilde{\Theta}^T \nabla_E \varphi(E) D \nabla_E^T \varphi(E) \tilde{\Theta}$ and $\tilde{\Theta}^T \nabla_x \varphi(x) \dot{E}^* + D\nabla_E \varepsilon / 2$ and taking the upper bound yields

$$\dot{J}_{HJB} \le \alpha_2 J_E^T(E)\left(F(E) - G(X) R^{-1} G(X)^T \nabla_E^T \varphi(E) \hat{\Theta} / 2 \right)$$
$$- \Sigma(x, \hat{U}) \frac{\alpha_2}{2} \tilde{\Theta}^T \nabla_E \varphi(E) G(x) R^{-1} G(x)^T J_E^T(x)$$

$$- \frac{\alpha_1}{32\rho^2} \left\| \tilde{\Theta}^T \nabla_E \varphi(E) \right\|^4 D_{\min}^2 + \frac{4\alpha_1}{\rho^2} \left\| \tilde{\Theta}^T \nabla_E \varphi(E) \right\|^2 \left\| \dot{E}^* + \frac{D\nabla_E \varepsilon}{2} \right\|^2 + \frac{3}{2\rho^2} \alpha_1 \varepsilon_{HJB}^2.$$

Now, completing the square with respect $\left\| \tilde{\Theta}^T \nabla_E \varphi(E) \right\|^2$ renders

$$\dot{J}_{HJB} \le \alpha_2 J_E^T(E)\left(F(E) - \frac{1}{2} G(X) R^{-1} G(X)^T \nabla_E^T \varphi(E) \hat{\Theta} \right)$$
$$- \Sigma(E, \hat{U}) \frac{\alpha_2}{2} \tilde{\Theta}^T \nabla_E \varphi(E) G(X) R^{-1} G(X)^T J_E^T(E)$$
$$- \frac{\alpha_1}{64\rho^2} \left\| \tilde{\Theta}^T \nabla_E \varphi(E) \right\|^4 D_{\min}^2 + \frac{\alpha_1 256}{\rho^2 D_{\min}^2} \left\| \dot{E}^* + \frac{D\nabla_E \varepsilon}{2} \right\|^4 + \frac{3\alpha_1}{2\rho^2} \varepsilon_{HJB}^2.$$

Next, observing the bound in (21) and applying the Cauchy-Schwarz inequality, $\dot{J}_{HJB}$ is upper bounded according to

$$\dot{J}_{HJB} \le \alpha_2 J_{1E}^T(E)\left(F(E) - \frac{1}{2} G(X) R^{-1} G(X)^T \nabla_E^T \varphi(E) \hat{\Theta} \right)$$

$$-\Sigma(E,\hat{U}) \frac{\alpha_2}{2} \tilde{\Theta}^T \nabla_E \varphi(E) G(X) R^{-1} G(X)^T J_{1E}^T(E)$$

$$-\frac{\alpha_1}{\rho^2} \left\| \tilde{\Theta} \right\|^4 \beta_1 + \frac{\alpha_1}{\rho^2} \eta(\varepsilon) + \frac{\alpha_1}{\rho^2} \beta_2 \delta^4(E) \qquad \text{(A.3)}$$

with $\beta_1 = \nabla \varphi_{\min}^4 D_{\min}^2 / 64$, $\beta_2 = 1024 / D_{\min}^2 + 3/2$, and $\eta(\varepsilon) = 64 D_{\max}^4 \varepsilon'^4_M / D_{\min}^2 + 3(\varepsilon'^4_M + \varepsilon'^4_M D_{\max}^2)/2$, and $0 < \nabla \varphi_{\min} \le \left\| \nabla \varphi(E) \right\|$ is ensured by $\|E\| > 0$ for a constant $\nabla \varphi_{\min}$. Now, the cases of $\Sigma(E,\hat{U}) = 0$ and $\Sigma(E,\hat{U}) = 1$ will be considered.

Case 1. For $\Sigma(E,\hat{U}) = 0$, the first term in (A.3) is less than zero by the definition of the operator in (31). Recalling $\delta(E) \equiv \sqrt[4]{K^* \left\| J_E \right\|}$ and observing $\left\| 1/\rho^2 \right\| \le 1$, (A.3) is rewritten as

$$\dot{J}_{HJB} \le -\left(\alpha_2 \dot{E}_{\min} - \alpha_1 \beta_2 K^* \right) \left\| J_E(E) \right\| - \frac{\alpha_1 \beta_1}{\rho^2} \left\| \tilde{\Theta} \right\|^4 + \alpha_1 \frac{\eta(\varepsilon)}{\rho^2} \qquad \text{(A.4)}$$

and (A.4) is less than zero provided $\alpha_2 / \alpha_1 > \beta_2 K^* / \dot{x}_{\min}$ and the following inequalities hold

$$\begin{aligned} &\left\| J_E(E) \right\| > \alpha_1 \eta(\varepsilon) / \left(\alpha_2 \dot{x}_{\min} - \alpha_1 \beta_2 K^* \right) \equiv b_{JE0} \text{ or} \\ &\left\| \tilde{\Theta} \right\| > \sqrt[4]{\eta(\varepsilon)/\beta_1} \equiv b_{\Theta 0}. \end{aligned} \qquad \text{(A.5)}$$

Note that $\|E\| > 0$ and the operator (26) ensure the existence of a constant $\dot{E}_{\min}$ satisfying $0 < \dot{E}_{\min} < \left\| \dot{E} \right\|$. According to standard Lyapunov extensions [3], the inequalities above guarantee that $\dot{J}_{HJB}$ is less than zero outside of a compact set. Thus,

$\left\|J_E(E)\right\|$ as well as the OLA parameter estimation error $\left\|\tilde{\Theta}\right\|$ remain bounded for the case $\Sigma(E,\hat{U})=0$. Recalling the Lyapunov candidate $J_E(E)$ is radially unbounded and continuously differentiable (*Lemma 1*), the boundedness of $\left\|J_E(E)\right\|$ implies the boundedness of the system states, $\|E\|$.

Case 2. Next, consider the case of $\Sigma(E,\hat{U})=1$ which implies the OLA based input (28) may not stabilizing. To begin, add and subtract $\alpha_2 J_{1E}^T(E)D(\nabla_E^T\varphi(E)\Theta+\nabla_E\varepsilon)/2$ to (A.3) to get

$$\begin{aligned}\dot{J}_{HJB} &\le \alpha_2 J_{1E}^T(E)\left(F(E)-\frac{1}{2}D(\nabla_E^T\varphi(E)\Theta+\nabla_E\varepsilon)\right)\\ &+\frac{\alpha_2}{2}J_{1E}^T(E)D\nabla_E\varepsilon\\ &-\frac{\alpha_1}{\rho^2}\left\|\tilde{\Theta}^T\right\|^4\beta_1+\frac{\alpha_1}{\rho^2}\eta(\varepsilon)+\frac{\alpha_1}{\rho^2}\beta_2\delta^4(E)\\ &=\alpha_2 J_{1E}^T(x)\left(F(E)+G(X)U^*\right)+\frac{\alpha_2}{2}J_{1E}^T(E)D\nabla_E\varepsilon\\ &-\frac{\alpha_1}{\rho^2}\left\|\tilde{\Theta}\right\|^4\beta_1+\frac{\alpha_1}{\rho^2}\eta(\varepsilon)+\frac{\alpha_1}{\rho^2}\beta_2K^*\left\|J_{1E}\right\|.\end{aligned}$$

Next, using *Lemma 2* and recalling the boundedness of D, $\dot{J}_{HJB}$ is rewritten as

$$\dot{J}_{HJB}\le-\alpha_2\bar{Q}_{\min}\left\|J_{1E}(E)\right\|^2+\alpha_2\left(D_{\max}\varepsilon_M'/2+\alpha_1\beta_2K^*/(\alpha_2\rho^2)\right)\left\|J_{1E}(E)\right\|-\frac{\alpha_1}{\rho^2}\left\|\tilde{\Theta}\right\|^4\beta_1+\frac{\alpha_1}{\rho^2}\eta(\varepsilon)$$

where $\bar{Q}_{Min}>0$ satisfies $\bar{Q}_{Min}\le\|Q(E)\|$ and is ensured by the condition $\|E\|>0$. As a final step, complete the square with respect to $\left\|J_{1E}(E)\right\|^2$ to reveal

$$\dot{J}_{HJB}\le-\frac{\alpha_2}{2}\bar{Q}_{\min}\left\|J_{1E}(E)\right\|^2-\frac{\alpha_1}{\rho^2}\left\|\tilde{\Theta}\right\|^4\beta_1+\frac{\alpha_1}{\rho^2}\eta(\varepsilon)$$

$$+\frac{\alpha_2}{4\bar{Q}_{\min}} D_{\max}^2 \varepsilon_M'^2 + \frac{\alpha_1^2}{\alpha_2 \rho^4 \bar{Q}_{\min}} \beta_2^2 K^{*2} \tag{A.6}$$

and $\dot{J}_{HJB} < 0$ provided the following inequalities hold

$$\left\| J_{1E}(x) \right\| > \sqrt{D_{\max}^2 \varepsilon_M'^2 / (2\bar{Q}_{\min}^2)} \equiv b_{JE}' \quad \text{and}$$

$$\left\| \tilde{\Theta} \right\| > \sqrt[4]{\eta(\varepsilon) / \beta_1 + \alpha_1 \beta_2^2 K^{*2} / (\beta_1 \alpha_2 \bar{Q}_{\min})} \equiv b_\Theta \tag{A.7}$$

According to standard Lyapunov extensions [3], the inequalities in (A.7) guarantee that $\dot{J}_{HJB}$ is less than zero outside of a compact set. Thus, $\left\| J_E(E) \right\|$ as well as the OLA parameter estimation error estimation errors, $\left\| \tilde{\Theta} \right\|$, remain bounded for the case $\Sigma(E, \hat{U}) = 1$. Recalling the Lyapunov candidate $J_E(E)$ is a radially unbounded and continuously differentiable (*Lemma 1*), the boundedness of $\left\| J_E(E) \right\|$ implies the boundedness of the system states, $\|E\|$.

The overall bounds for the cases $\Sigma(E, \hat{U}) = 0$ and $\Sigma(E, \hat{U}) = 1$ are then given by $\left\| J_E(E) \right\| \le b_{JE}$ and $\left\| \tilde{\Theta} \right\| \le b_\Theta$ for computable positive constants $b_{JE} = \max(b_{JE0}, b_{JE1})$ and $b_\Theta = \max(b_{\Theta 0}, b_{\Theta 1})$. Note that b_{Jx0} and $b_{\Theta 1}$ in (A.5) and (A.7), respectively, can be reduced through appropriate selection of α_1 and α_2. To complete the proof, subtract (22) , (27) and (28) from (23) to reveal

$$V^*(E) - \hat{V}(E) = \tilde{\Theta}^T \varphi(x) + \varepsilon(x)$$
$$U^* - \hat{U} = -\frac{1}{2} R^{-1} G(X)^T \nabla_E^T \varphi(X) \tilde{\Theta} - \frac{1}{2} R^{-1} G(X)^T \nabla_E \varepsilon(E)$$

.

Next, observing that the boundedness of the system states ensures the existence of positive constants φ_M and φ'_M such that $\|\varphi\| \le \varphi_M$ and $\|\nabla_E \varphi\| \le \varphi'_M$, respectively, and taking norm and the limit as $t \to \infty$ when $\Sigma(E,\hat{U}) = 0$ reveals

$$\left\|V^* - \hat{V}\right\| \le \left\|\tilde{\Theta}\right\| \|\varphi(E)\| + \varepsilon_M \le b_\Theta \varphi_M + \varepsilon_M \equiv \varepsilon_{r1}$$

$$\left\|U^*(x) - \hat{U}(x)\right\| \le \frac{\lambda_{\max}}{2}(R^{-1}) G_M b_\Theta \varphi'_M + \frac{\lambda_{\max}}{2}(R^{-1}) g_M \varepsilon'_M \equiv \varepsilon_{r2}.$$

■

Proof of Theorem 2. Consider the following positive definite Lyapunov candidate

$$J_{1HJB} = \gamma_2 J_1(\hat{E}) + \frac{1}{2}\tilde{\Theta}_1^T \tilde{\Theta}_1 + \frac{1}{2}\tilde{\zeta}^T P \tilde{\zeta} \tag{A.8}$$

whose first derivative with respect to time is given by

$$\begin{aligned} \dot{J}_{1HJB} = \; & \gamma_2 J_{1E}^T(\hat{E})\dot{\hat{E}} + \tilde{\Theta}_1^T \dot{\tilde{\Theta}}_1 + \dot{\tilde{\zeta}}^T P \tilde{\zeta} + \tilde{\zeta}^T P \dot{\tilde{\zeta}} \\ = \; & \gamma_2 \hat{E}\dot{\hat{E}} + \tilde{\Theta}_1^T \dot{\tilde{\Theta}}_1 + \tilde{\zeta}^T (A^T P + PA)\dot{\tilde{\zeta}} \end{aligned} \tag{A.9}$$

where $J_1(\hat{E})$ and $J_{1E}(\hat{E})$ are given in *Lemma 1*. With the same steps as of *Theorem 1* we get

$$\dot{J}_{1HJB} \le \gamma_2 J_{1E}^T(\hat{E})\left(\omega(e_1) - B(y)R_1^{-1}\mathrm{B}(y)^T \nabla_E^T \varphi_1(\hat{E})\hat{\Theta}_1 / 2\right)$$

$$-\Sigma_1(y,\hat{U})\frac{\gamma_2}{2}\tilde{\Theta}^T \nabla_{\hat{E}}\varphi_1(\hat{E})\mathrm{B}(y)R_1^{-1}\mathrm{B}(y)^T J_{1\hat{E}}^T(\hat{E}) - \frac{\gamma_1}{32\rho_1^2}\left\|\tilde{\Theta}_1^T \nabla_{\hat{E}}\varphi_1(\hat{E})\right\|^4 D_{\min}^2$$

$$+\frac{4\gamma_1}{\rho_1^2}\left\|\tilde{\Theta}_1^T \nabla_E \varphi_1(\hat{E})\right\|^2 \| \dot{\hat{E}}^* + \frac{D_1 \nabla_{\hat{E}}\varepsilon}{2} \|^2 + \frac{3}{2}\frac{\gamma_1}{\rho_1^2}\varepsilon_{HJB}^2 - \tilde{\zeta}^T T \dot{\tilde{\zeta}}.$$

Now, completing the square with respect $\| \tilde{\Theta}^T \nabla_E \varphi(\hat{E}) \|^2$ renders

$$\dot{J}_{1HJB} \le -\tilde{\zeta}^T T \dot{\tilde{\zeta}} + \gamma_2 J_{1E}^T(\hat{E})\left(\omega(e_1) - \frac{1}{2}B(y)R_1^{-1}B(y)^T \nabla_{\hat{E}}^T \varphi_1(\hat{E})\hat{\Theta}_1\right)$$

$$+\gamma_2 J_{1E}^T(\hat{E})\left(\omega(e_1)-\frac{1}{2}B(y)R_1^{-1}B(y)^T\nabla_{\hat{E}}^T\varphi_1(\hat{E})\hat{\Theta}_1\right)$$

$$-\Sigma_1(\hat{E},\hat{U})\frac{\gamma_2}{2}\tilde{\Theta}_1^T\nabla_{\hat{E}}\varphi_1(\hat{E})B(y)R_1^{-1}B(y)^T J_{1\hat{E}}^T(\hat{E})$$

$$-\frac{\gamma_1}{64\rho_1^2}\left\|\tilde{\Theta}_1^T\nabla_{\hat{E}}\varphi_1(\hat{E})\right\|^4 D_{1\min}^2+\frac{\gamma_1 256}{\rho_1^2 D_{1\min}^2}\left\|\dot{\hat{E}}^*+\frac{D_1\nabla_E\varepsilon}{2}\right\|^4+\frac{3\gamma_1}{2\rho_1^2}\varepsilon_{HJB}^2.$$

Next, observing the bound $\|\varphi_1(e_1)+\mathrm{B}(y)U_1^*\|\le\delta_1(\hat{E})$ which is similar to (21) and applying the Cauchy-Schwarz inequality, $\dot{J}_{HJB}$ is upper bounded according to

$$\dot{J}_{1HJB}\le\gamma_2 J_{1E}^T(\hat{E})\left(\omega(e_1)-\frac{1}{2}\mathrm{B}(y)R_1^{-1}\mathrm{B}(y)^T\nabla_{\hat{E}}^T\varphi_1(\hat{E})\hat{\Theta}_1\right)$$
$$-\Sigma_1(\hat{E},\hat{U}_1)\frac{\gamma_2}{2}\tilde{\Theta}_1^T\nabla_{\hat{E}}\varphi_1(\hat{E})B(y)R^{-1}\mathrm{B}(y)^T J_{1\hat{E}}^T(\hat{E})$$

$$-\frac{\gamma_1}{\rho_1^2}\left\|\tilde{\Theta}_1\right\|^4\tau_1+\frac{\gamma_1}{\rho_1^2}\eta_1(\varepsilon)+\frac{\gamma_1}{\rho_1^2}\tau_2\delta_1^4(\hat{E})-\tilde{\zeta}^T T\dot{\tilde{\zeta}}, \qquad \text{(A.10)}$$

with $\tau_1=\nabla\varphi_{1\min}^4 D_{1\min}^2/64$, $\tau_2=1024/D_{1\min}^2+3/2$, and $\eta_1(\varepsilon)=64D_{1\max}^4\varepsilon'^4_{1M}/D_{1\min}^2++3(\varepsilon'^4_{1M}+\varepsilon'^4_{1M}D_{1\max}^2)/2$, and $0<\nabla\varphi_{1\min}\le\|\nabla\varphi_1(\hat{E})\|$ is ensured by $\|\hat{E}\|>0$ for a constant $\nabla\varphi_{1\min}$. Now, the cases of $\Sigma_1(\hat{E},\hat{U}_1)=0$ and $\Sigma_1(\hat{E},\hat{U}_1)=1$ will be considered.

Case 1. For $\Sigma_1(\hat{E},\hat{U}_1)=0$, the first term in (A.10) is less than zero by the definition of the operator in (31). Recalling $\delta_1(\hat{E})\equiv\sqrt[4]{K_1^*\|J_{1\hat{E}}\|}$ and observing $\|1/\rho_1^2\|\le 1$, (A.10) is rewritten as

$$\dot{J}_{HJB}\le-\tilde{\zeta}^T T\dot{\tilde{\zeta}}-\left(\gamma_2\dot{\hat{E}}_{\min}-\gamma_1\tau_2K^*\right)\left\|J_E(\hat{E})\right\|-\frac{\gamma_1}{\rho_1^2}\left\|\tilde{\Theta}\right\|^4\tau_1+\frac{1}{\rho_1^2}\gamma_1\eta_1(\varepsilon), \qquad \text{(A.11)}$$

and (A.11) is less than zero provided $\gamma_2/\gamma_1 > \tau_2 K_1^*/\hat{\dot{E}}_{\min}$ and the following inequalities hold

$$\| J_{1\hat{E}}(\hat{E}) \| > \gamma_1 \eta_1(\varepsilon) / \left(\gamma_2 \hat{\dot{E}}_{\min} - \gamma_1 \tau_2 K_1^* \right) \equiv b_{1JE0}, \text{ or}$$

$$\| \tilde{\Theta}_1 \| > \sqrt[4]{\eta_1(\varepsilon)/\tau_1} \equiv b_{1\Theta 0}, \text{ or}$$

$$\left\| \tilde{\zeta} \right\| > \rho_1^{-1} \sqrt{\gamma_1 \eta_1(\varepsilon)/\lambda_{T\min}} \equiv b_{1\zeta} \tag{A.12}$$

Case 2. Next, consider the case of $\Sigma_1(\hat{E}, \hat{U}_1) = 1$ which implies the OLA based input $\hat{U}_1^* = -R_1^{-1} \mathrm{B}(y)^T \nabla_{\hat{E}}^T \varphi_1(\hat{E}) \hat{\Theta}_1 / 2$ may not stabilizing. To begin, add and subtract $\alpha_2 J_{1E}^T(\hat{E}) D(\nabla_E^T \varphi(\hat{E})\Theta + \nabla_E \varepsilon)/2$ to (A.10) to get

$$\dot{J}_{HJB} \le -\tilde{\zeta}^T T \dot{\tilde{\zeta}} + \alpha_2 J_{1\hat{E}}^T(\hat{E}) \left(\omega(e_1) - \frac{1}{2} D(\nabla_{\hat{E}}^T \varphi(\hat{E})\Theta + \nabla_{\hat{E}} \varepsilon) \right)$$

$$+ \frac{\alpha_2}{2} J_{1E}^T(\hat{E}) D_1 \nabla_{\hat{E}} \varepsilon - \frac{\gamma_1}{\rho_1^2} \| \tilde{\Theta}^T \|^4 \tau_1 + \frac{\gamma_1}{\rho_1^2} \eta_1(\varepsilon) + \frac{\gamma_1}{\rho_1^2} \delta_1^4(\hat{E})$$

$$= -\tilde{\zeta}^T T \dot{\tilde{\zeta}} + \gamma_2 J_{1\hat{E}}^T(x) \left(\omega(e_1) + \mathrm{B}(y) U_1^* \right) + \gamma_2 J_{1E}^T(\hat{E}) D_1 \nabla_{\hat{E}} \varepsilon / 2$$

$$- \frac{\gamma_1}{\rho_1^2} \left\| \tilde{\Theta}_1 \right\|^4 \tau_1 + \frac{\gamma_1}{\rho_1^2} \eta(\varepsilon) + \frac{\gamma_1}{\rho_1^2} \tau_2 K_1^* \left\| J_{1\hat{E}} \right\|.$$

Next, using *Lemma 2* and recalling the boundedness of D, $\dot{J}_{1HJB}$ is rewritten as

$$\dot{J}_{1HJB} \le -\tilde{\zeta}^T T \dot{\tilde{\zeta}} - \gamma_2 \bar{Q}_{\min} \| J_{1\hat{E}}(\hat{E}) \|^2$$
$$+ \gamma_2 \left(\frac{D_{1\max} \varepsilon_M'}{2} + \frac{\gamma_1 \tau_2 K_1^*}{\gamma_2 \rho_1^2} \right) \| J_{1\hat{E}}(\hat{E}) \| - \frac{\gamma_1}{\rho_1^2} \| \tilde{\Theta}_1 \|^4 \tau_1 + \frac{\gamma_1 \eta_1(\varepsilon)}{\rho_1^2} \text{ where } \bar{Q}_{1min} > 0$$

satisfies $\bar{Q}_{1min} \le \| Q_1(\hat{E}) \|$ and is ensured by the condition $\| \hat{E} \| > 0$. As a final step, complete the square with respect to $\| J_{1E}(\hat{E}) \|^2$ to reveal

$$\dot{J}_{HJB} \le -\tilde{\zeta}^T Q_1 \dot{\tilde{\zeta}} - \frac{\alpha_2}{2}\bar{Q}_{1\min} \| J_{1\hat{E}}(\hat{E}) \|^2 - \frac{\gamma_1}{\rho_1^2} \| \tilde{\Theta}_1 \|^4 \tau_1 + \frac{\gamma_1}{\rho_1^2}\eta_1(\varepsilon)$$

$$+ \frac{\gamma_2}{4\bar{Q}_{1\min}} D_{1\max}^2 \varepsilon_{1M}'^2 + \frac{\gamma_1^2}{\gamma_2 \rho_1^4 \bar{Q}_{1\min}} \tau_2^2 K^{*2} \tag{A.13}$$

and $\dot{J}_{HJB} < 0$ provided the following inequalities hold

$$\| J_{1E}(\hat{E}) \| > \sqrt{\frac{D_{1\max}^2 \varepsilon_{1M}'^2}{2\bar{Q}_{1\min}^2}} \equiv b_{1JE}' \quad \text{, or}$$

$$\| \tilde{\Theta}_1 \| > \sqrt[4]{\frac{\eta_1(\varepsilon)}{\tau_1} + \frac{\gamma_1}{\tau_1 \gamma_2 \bar{Q}_{1\min}} \tau_2^2 K^{*2}} \equiv b_{1\Theta}' \quad \text{, or}$$

$$\| \tilde{\zeta} \| > \frac{1}{\lambda_{T\min}} \sqrt[2]{\frac{\eta_1(\varepsilon)}{\tau_1} + \frac{\gamma_1}{\tau_1 \gamma_2 \bar{Q}_{1\min}} \tau_2^2 K^{*2}} \equiv b_{1\zeta}' \,. \tag{A.14}$$

According to standard Lyapunov extensions [3], the inequalities in (A.14) guarantee that $\dot{J}_{1HJB}$ is less than zero outside of a compact set. Thus, $\| J_{1E}(\hat{E}) \|$ as well as the OLA parameter estimation error estimation error $\| \tilde{\Theta}_1 \|$ remain bounded for the case $\Sigma_1(\hat{E}, \hat{U}_1) = 1$. Recalling the Lyapunov candidate $J_{1E}(\hat{E})$ is a radially unbounded and continuously differentiable (*Lemma 1*), the boundedness of $\| J_{1E}(\hat{E}) \|$ implies the boundedness of the states $\| \hat{E} \|$.

The overall bounds for the cases $\Sigma_1(\hat{E}, \hat{U}_1) = 0$ and $\Sigma_1(\hat{E}, \hat{U}_1) = 1$ are given by $\| J_{1E}(\hat{E}) \| \le b_{1JE}$ and $\| \tilde{\Theta}_1 \| \le b_{1\Theta}$ for computable positive constants $b_{1JE} = \max(b_{1JE0}, b_{1JE1})$ and $b_{1\Theta} = \max(b_{1\Theta 0}, b_{1\Theta 1})$. Note that b_{1Jx0} and $b_{1\Theta 1}$ in (A.12) and (A.14), respectively, can be reduced through appropriate selection of γ_1 and γ_2. To complete the proof, subtract (22) , (27) and (28) from (23) to get

$$V_1^*(\hat{E}) - \hat{V}_1(\hat{E}) = \tilde{\Theta}_1^T \varphi_1(\hat{E}) + \varepsilon_1(x)$$
$$U_1^* - \hat{U}_1 = -\frac{1}{2} R_1^{-1} \mathrm{B}(y)^T \nabla_{\hat{E}}^T \varphi_1(\hat{E}) \tilde{\Theta}_1 - \frac{1}{2} R_1^{-1} \mathrm{B}(y)^T \nabla_{\hat{E}} \varepsilon_1(\hat{E})$$

.

Next, observing that the boundedness of the system states ensures the existence of positive constants φ_{1M} and φ'_{1M} such that $\| \varphi_1 \| \leq \varphi_{1M}$ and $\| \nabla_{\hat{E}} \varphi_1 \| \leq \varphi'_{1M}$, respectively, and taking norm and the limit as $t \to \infty$ when $\Sigma_1(E, \hat{U}_1) = 0$ reveals

$$\| V_1^* - \hat{V}_1 \| \leq \| \tilde{\Theta}_1 \| \| \varphi_1(\hat{E}) \| + \varepsilon_M \leq b_{1\Theta} \varphi_{1M} + \varepsilon_{1M} \equiv \varepsilon_{r1}$$

$$\| U_1^*(x) - \hat{U}_1(x) \| \leq \frac{1}{2} \lambda_{\max}(R_1^{-1}) \mathrm{B}_{1M} b_{1\Theta} \varphi'_{1M} + \frac{\lambda_{\max}}{2} (R_1^{-1}) \mathrm{B}_M \varepsilon'_{1M} \equiv \varepsilon_{r2}.$$

■

REFERENCES

[1] Khalil H K. Nonlinear Systems (3rd Ed). Printice-Hall: 2002.

[2] Narendra K S, Annaswamy A M. Stable Adaptive Systems. Prentice-Hall: Englewood Cliffs, NJ, 1989.

[3] Lewis F L, Jagannathan S, Yesildirek A. Neural Network Control of Robot Manipulators and Nonlinear Systems. Taylor and Francis: Philadelphia, PA, 1999.

[4] Jagannathan S. Neural network control of nonlinear discrete-time systems, CRC Press, 2006.

[5] Kristic M, Kanellakopoulos I, Kokotovic P. Nonlinear and Adaptive Control Design. John Wiley and Sons: 1995.

[6] Dierks T, Jagannathan S. Optimal control of affine nonlinear continuous-time systems. in Proc. of the American Control Conference 2010; 1568-1573.

[7] Li Z H, Krstic M. Optimal design of adaptive tracking controllers for non-linear systems. Automatica 1997; 33(8): 1459-1473.

[8] Lewis F L, Syrmos V L. Optimal Control (2nd ed), Wiley: Hoboken, NJ, 1995.

[9] Beard R, Saridis G, Wen J. Improving the performance of stabilizing controls for nonlinear systems. IEEE Control Systems Magazine 1996; 16(5): 27-35.

[10] Si J, Barto A G, Powell W B, Wunsch D. Handbook of Learning and Approx. Dynamics Prog. Wiley: IEEE Press, 2004.

[11] Dierks T, Jagannathan S. Optimal control of affine nonlinear discrete-time systems, in Proc. of the Mediterranean Conference on Control and Automation 2009; 1390 – 1395.

[12] Vrabie D, Pastravanu O, Abu-Khalaf M, Lewis F L. Adaptive optimal control for continuous-time linear systems based on policy iteration. Automatica 2009, 45(2): 477-484.

[13] Vrabie D, Vamvoudakis K, Lewis F L. Adaptive optimal controllers based on generalized policy iteration in a continuous-time framework. Proc. of the IEEE Mediterranean Conf. on Control and Automation 2009; 1402-1409.

[14] Watkins C H. Learning from delayed rewards. University of Cambridge: PhD Dissertation, 1989.

[15] Wang D, Huang J. Neural network-based adaptive dynamic surface control for a class of uncertain nonlinear systems in strict-feedback form. Neural Networks, IEEE Transactions on 2005; 16(1): 195-202.

[16] Zhang T, Ge S S, Hang C C, Adaptive neural network control for strict-feedback nonlinear systems using backstepping design. Automatica 2000; 36: 1835–1846.

[17] Barron A R. Universal approximation bounds for superpositions of a sigmoidal function. Information Theory, IEEE Transactions on 1993; 39(3): 930-945.

[18] Lewis F L, Vamvourdakis K G. Reinforcement Learning for Partially Observable Dynamics Process: Adaptive Dynamic Programming Using Measured Output Data. IEEE Transaction System, Man, and Cybernetics, Part B 2011; 41(1): 14-251.

IV. NEURAL NETWORK-BASED OPTIMAL ADAPTIVE OUTPUT FEEDBACK CONTROL OF A HELICOPTER UAV

SUMMARY— Helicopter unmanned aerial vehicles (UAV) are widely used for both military and civilian operations. Because the helicopter UAVs are underactuated nonlinear mechanical systems, high-performance controller design for them presents a challenge. This paper introduces an optimal controller design via output feedback for trajectory tracking of a helicopter UAV using a neural network (NN). The output-feedback control system utilizes the backstepping methodology, employing kinematic and dynamic controllers and a NN observer. The online approximator-based dynamic controller learns the infinite-horizon Hamilton-Jacobi-Bellman (HJB) equation in continuous time and calculates the corresponding optimal control input by minimizing a cost function forward-in-time without using value and policy iterations. Optimal tracking is accomplished by using a single NN utilized for cost function approximation. The overall closed-loop system stability is demonstrated using Lyapunov analysis. Finally, simulation results are provided to demonstrate the effectiveness of the proposed control design for trajectory tracking.

Index Terms— Hamilton-Jacobi-Bellman equation, helicopter UAV, neural network (NN), nonlinear optimal control.

1. INTRODUCTION

Helicopter unmanned aerial vehicles (UAVs) are autonomous rotorcraft and due to their versatility and maneuverability, they are invaluable for applications where human intervention may be restricted. For unmanned helicopter control [1], it is essential to produce moments and forces on the UAV to position the helicopter such that the desired

regulated state is achieved, and to control the helicopter's velocity, position, and orientation such that it tracks a desired trajectory. The dynamics of the helicopter UAV are not only nonlinear, but are also coupled with each other and underactuated, which makes the control design challenging. Both inputs and dynamics are coupled on a helicopter, particularly as a result of the swashplate mechanical linkages and the torques created by drag against the rotors. In other words, a helicopter has six degrees of freedom (DOF) which must be controlled with only four control inputs in order to manipulate the thrust and the three rotational torques.

In order to develop the controllers for such helicopters, Koo and Sastry [1] have utilized an approximate linearization-based control [1] that transforms the system into linear form. Mettler et al. [2] have introduced a model for the helicopter independent of an accompanying control scheme [2]. Hovakimyan et al. [3] have implemented an output feedback control scheme with a neural network (NN)-based controller using feedback linearization [3]. Johnson and Kannan [4] have employed an inner and outer loop control using pseudo-control hedging [4], and Ahmed et al. [5] have introduced a backstepping-based controller for the helicopter [5]. Frazzoli [6] and Mahoney [7] have both generated control schemes for Lyapunov-based control of helicopter UAVs. However, none of these works [1]-[7] presents an optimal control scheme for the underactuated unmanned helicopter.

Optimal control of a linear system minimizing a quadratic cost function can be achieved by solving the Riccati equation [8]. In contrast, the optimal control of nonlinear systems often requires solving the nonlinear Hamilton-Jacobi-Bellman (HJB) equation, which does not have a closed-form solution.

Therefore, Enns and Si [9] have used neural network dynamic programming (NDP)-based optimal control of a helicopter UAV [9] using offline training and value and policy iterations. Lee et al. [10] introduced a robust command augmentation system using a NN, but inversion errors can lead to problems [10].

Since value and policy iteration-based optimal schemes are not suitable for hardware implementation, in the recent NDP literature, Dierks and Jagannathan [11] introduced an optimal regulation and tracking controller for nonlinear discrete-time systems in affine form. Here, the discrete-time HJB equation is solved online and forward-in-time. An online approximator (OLA) is tuned to learn the HJB equation, with a second OLA utilized to minimize the cost function [11]. Dierks and Jagannathan [12] have extended this NDP scheme to continuous-time systems in affine form by using a single online approximator (SOLA) [12]. However, such NDP-based optimal control schemes are not available to nonlinear systems in strict feedback form which use backstepping technique.

Therefore, a SOLA-based scheme for the optimal tracking control of a helicopter's nonlinear continuous-time feedback system is considered in this paper via a backstepping approach. A kinematic controller generates the desired velocities. The dynamic controller learns the continuous-time HJB equation and then calculates the corresponding optimal control input by minimizing a cost function forward-in-time by assuming known system dynamics. A single NN is used for approximating the cost function with the NN weights tuned online. A NN observer is employed to obtain the states from the outputs. By selecting suitable NN weight update laws, Lyapunov analysis is utilized to demonstrate the stability of the closed-loop system. It is shown that the

approximated control input approaches the optimal control input over time. Simulation results are included for both hovering and following a desired maneuver.

The main contribution of this paper includes the development of an optimal controller for tracking a trajectory of an unmanned underactuated helicopter, forward in time and without using value and policy iterations, where the helicopter system is expressed in strict feedback form appropriate for backstepping control. The controller tuning is independent of the trajectory. A NN-based OLA is utilized to approximate the cost function and the overall stability is guaranteed.

The optimal controller has been previously developed for affine nonlinear systems by using state feedback [11]-[12], but has not yet been developed for strict feedback system such as a rotary-wing aircraft. This optimal controller in [11]-[12] required that $f(V) = 0$ for $V = 0$ **.** However, since the dynamics are transformed into tracking error form, this condition is met. In addition, the proposed controller uses output feedback by using an observer [14] similar to that used for a quadrotor [14], but the present application to a helicopter is novel, and is not accompanied by the virtual and kinematic controllers employed in [14]. The proposed effort extends the work of [7], [12], and [14], from the fields of helicopter, optimal, and quadrotor control, respectively and adds a closed-loop stability proof that is involved yet demonstrating the convergence of the output feedback controller.

2. HELICOPTER DYNAMICS MODEL

Consider the helicopter shown in Figure 1 with six degrees of freedom (DOF) defined in the inertial coordinate frame Q^a, where its position coordinates are given by $\rho = \begin{bmatrix} x & y & z \end{bmatrix} \in Q^a$ and its orientation described as yaw, pitch, and roll, respectively, is

given by Euler angles $\Theta = \begin{bmatrix} \phi & \theta & \psi \end{bmatrix} \in Q^a$. The equations of motion are expressed in the body fixed frame Q^b which is associated with the helicopter's center of mass. The ${}^b x$-axis is defined parallel to the helicopter's direction of travel and the ${}^b y$-axis is defined perpendicular to the helicopter's direction of travel, while the ${}^b z$-axis is defined as projecting orthogonally downwards from the *xy*-plane of the helicopter. The dynamics of the helicopter is given by the Newton-Euler equation in the body fixed coordinate system and can be written as in [7] but in the form provided in [14] as

$$M \begin{bmatrix} \dot{v} \\ \dot{\omega} \end{bmatrix} = \bar{S}(\omega) + \begin{bmatrix} 0^{3\times 1} \\ N_2 \end{bmatrix} + \begin{bmatrix} G(R) \\ 0^{3\times 1} \end{bmatrix} + U + \tau_d \tag{1}$$

where the mass-inertia matrix M is defined as $M = diag\{mI \quad \mathcal{J}\} \in \mathbb{R}^{6\times 6}$, $m \in \mathbb{R}$ is a positive scalar denoting the mass of the helicopter, $I \in \mathbb{R}^{3\times 3}$ is the identity matrix, $\mathcal{J} \in \mathbb{R}^{3\times 3}$ is the positive-definite inertia matrix, $\bar{S}(\omega) = \begin{bmatrix} 0^{3\times 1} & -\omega \times \mathcal{J}\omega \end{bmatrix}^T$, $N_2 \in \mathbb{R}^{3\times 1}$ represents the nonlinear aerodynamic effects, $G(R) = m\bar{g}e_3 \in \mathbb{R}^{3\times 1}$ represents the gravity vector with $\bar{g}$ the gravitational acceleration, and $\tau_d = \begin{bmatrix} \tau_{d1}^T & \tau_{d2}^T \end{bmatrix}^T$ represents unknown bounded disturbances such that $\|\tau_d\| < \tau_M$ for all time t, with τ_M a known positive constant. Also, $v = \begin{bmatrix} v_x & v_y & v_z \end{bmatrix} \in \mathbb{R}^{3\times 1}$ and $\omega = \begin{bmatrix} \omega_x & \omega_y & \omega_z \end{bmatrix} \in \mathbb{R}^{3\times 1}$ represent the translational velocity and angular velocity vectors, respectively. The kinematics of the helicopter are given by

$$\dot{\rho} = Rv \tag{2}$$

and

$$\dot{\Theta} = T^{-1}\omega \tag{3}$$

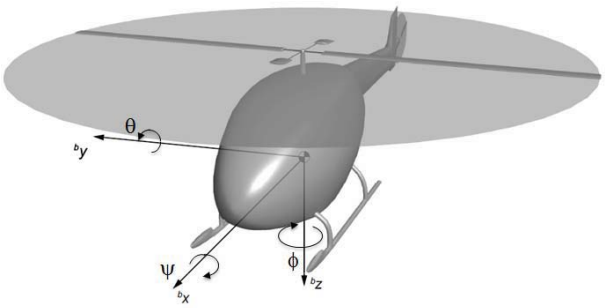

Figure 1. Helicopter orientation representation.

The translational rotation matrix used to relate a vector in body fixed frame to the inertial coordinate frame is defined as [14]

$$R(\Theta)=\begin{bmatrix} c_\theta c_\phi & s_\psi s_\theta c_\phi - c_\psi s_\phi & c_\psi s_\theta c_\phi + s_\psi s_\phi \\ c_\theta s_\phi & s_\psi s_\theta s_\phi + c_\psi c_\phi & c_\psi s_\theta s_\phi - s_\psi c_\phi \\ -s_\theta & s_\psi c_\theta & c_\psi c_\theta \end{bmatrix}$$

with $\|R\|_F < R_{\max}$ for a known constant $R_{\max}$ and $R^{-1}=R^T$, where $s_\bullet$ and $c_\bullet$ denote the $\sin(\bullet)$ and $\cos(\bullet)$ functions, respectively. The transformation matrix from the angular velocity to the derivative of the orientation is given by

$$T(\Theta)=\frac{1}{c_\theta}\begin{bmatrix} 0 & s_\psi & c_\psi \\ 0 & c_\theta c_\psi & -c_\theta s_\psi \\ c_\theta & s_\theta s_\psi & s_\theta c_\psi \end{bmatrix}$$

and is bounded according to $\|T\|_F < T_{\max}$ for a known constant $T_{\max}$, provided $-\pi/2<\psi<\pi/2$ and $-\pi/2<\theta<\pi/2$ such that the helicopter trajectory does not pass through any singularities [1], with $t_\bullet$ used to represent $\tan(\bullet)$. Throughout this work, $\|\bullet\|$ denotes a Euclidean norm and $\|\bullet\|_F$ denotes a Frobenius norm. Note also that $(\times)$ denotes the vector cross product. The nonlinear aerodynamic effects taken into

consideration for modeling of the helicopter are given by $N_2 = Q_M e_3 - Q_T e_2$, with Q_M and Q_T aerodynamic constants for which values are given in the simulation section, and originally found in [7]. Note that e_1, e_2, and e_3 are unit vectors directed along the *x*-, *y*-, and *z*-axes, respectively, in the inertial reference frame. The vector $U \in \mathbb{R}^{6\times 1}$ is given by

$$U = \begin{bmatrix} E_3^b & 0^{3\times 3} \\ 0^{3\times 1} & diag([p_{11}\ p_{22}\ p_{33}]) \end{bmatrix} \begin{bmatrix} u \\ w_1 \\ w_2 \\ w_3 \end{bmatrix},$$ where the control vector

$u_v = \begin{bmatrix} u & w_1 & w_2 & w_3 \end{bmatrix}$, with u providing the thrust in the z-direction, w_1, w_2 and w_3 providing the rotational torques in the *x*-, *y*-, and *z*-directions, respectively, p_{ii} positive definite constants that make up a gain array, and $E_3^b = [0\,0\,1]^T$. Defining the new augmented variables $X = [\rho^T\ \Theta^T]^T \in \mathbb{R}^{6\times 1}$ and $V = [v^T\ \omega^T]^T \in \mathbb{R}^{6\times 1}$, (1) can be rewritten in a form suitable for backstepping as

$$\dot{X} = AV + \xi \tag{4}$$

$$\dot{V} = f(V) + M^{-1}U \tag{5}$$

where $f(V) = M^{-1}(\bar{S}(\omega) + [0^{3\times 1}\ \mathcal{N}_2]^T) + \bar{G}$ with $\bar{G} = M^{-1}[G(R)\quad 0^{3\times 1}] \in \mathbb{R}^{6\times 1}$, and $\xi \in \mathbb{R}^{6\times 1}$ is the bounded sensor measurement noise such that $\| \xi \| \leq \xi_M$ for a known constant ξ_M. Equation (5) is in the body reference frame, while equation (4) is in the earth reference frame. Note that these last two equations take the form

$$\dot{x}_1 = f_1(x_1) + g_1(x_1)x_2 + \xi$$
$$\dot{x}_2 = f_2(x_2) + g_2(x_2)u$$

with $f_1(x_1) = 0$. This system is a candidate for backstepping control [13]. The dynamic controller operates in the body reference frame, with equation (4) necessary to bring these results back to the earth reference frame. Also, $A = diag\left(\begin{bmatrix} R & T^{-1} \end{bmatrix}\right) \in \mathbb{R}^{6\times 6}$. Writing explicitly, $f(V)$ yields

$$f(V) = \begin{bmatrix} 1/m & 0 & 0 & 0 & 0 & 0 \\ 0 & 1/m & 0 & 0 & 0 & 0 \\ 0 & 0 & 1/m & 0 & 0 & 0 \\ 0 & 0 & 0 & 1/\mathcal{J}_x & 0 & 0 \\ 0 & 0 & 0 & 0 & 1/\mathcal{J}_y & 0 \\ 0 & 0 & 0 & 0 & 0 & 1/\mathcal{J}_z \end{bmatrix} \left(\begin{bmatrix} 0 \\ 0 \\ 0 \\ -\omega_x \times \mathcal{J}_x \omega_x \\ -\omega_y \times \mathcal{J}_y \omega_y \\ -\omega_z \times \mathcal{J}_z \omega_z \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ -Q_T \\ Q_M \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ m\overline{g} \\ 0 \\ 0 \\ 0 \end{bmatrix} \right)$$

In this section, the dynamic model of the helicopter with six degrees of freedom has been presented. The control methodology is addressed next.

3. MOETHODOLOGY

The overall control objective for the unmanned helicopter is to track a desired trajectory $X_d(t)$ and a desired heading (yaw) while maintaining stable flight. Full knowledge of the helicopter states is required to achieve the control objective which is in practice not possible. Therefore, a NN observer is designed to estimate the states from the outputs. This output feedback control scheme consists of a kinematic controller to generate the desired velocity for the dynamic controller, a virtual controller and an optimal controller. First the kinematic controller is introduced and subsequently, the observer design is given.

3.1.Kinematic Controller

To design the kinematic controller for the unmanned helicopter, define the position tracking error as

$$\delta_1 = \rho_d - \rho \tag{6}$$

The observer's velocity estimate $\hat{v}$ from Section III.3.2 may be used to obtain the desired velocity, v_d as in [7]

$$v_d = \hat{v} - \frac{1}{m}\delta_1$$

Note that the notation $(\hat{\bullet})$ is used to denote an estimate. In addition, it is important to note that there exist desired trajectories which may reach unstable operating regions as the orientation about the *x*- and *y*- axes approaches $\pm\pi/2$. It is possible to avoid these singularities by redefining Euler angles or with an alternative approach employing quaternions, but there are still physical constraints to be considered. In other words, if the main rotor blades move into a plane perpendicular to the ground, the helicopter becomes unstable. This is a consequence of the physical limitations of helicopters. Therefore, trajectories requiring that these orientations be maintained should not be assigned to the helicopter.

3.2.Observer Design

The following section extends the work in [14] by Dierks and Jagannathan to a helicopter system. An observer is used to estimate the system states based on the system model and outputs. The helicopter states to be estimated are given by V with the observer's estimate of these states given by $\hat{V}$ and the state estimation error given by $\tilde{V} = V - \hat{V}$. The output is X, with the integrated observer's estimate of the output given by $\hat{X}$, and the error between the actual output and the integrated observer's estimate of the output given by $\tilde{X}$, with $\tilde{X} = X - \hat{X}$.

The observer is NN-based, and functions by estimating the output and comparing the estimate to the actual output. Referring back to (4) and (5), if A and ξ are known, then X may be easily obtained by integrating $\dot{X}$, and rearranging and solving yields V. But since X is known, A may be accurately obtained, meaning that ξ and the NN reconstruction error are the only sources of error in determining V.

To begin, a NN basis vector x_o is selected such that $\hat{x}_o = \begin{bmatrix} 1 & \hat{X} & \hat{V}^T & \tilde{X}^T \end{bmatrix}^T$, with the NN estimate of $f_{o1} = W_o^T \sigma\left(V_o^T x_o\right) + \varepsilon_o$ given by $\hat{f}_{o1} = \hat{W}_o^T \sigma\left(V_o^T \hat{x}_o\right)$. The NN reconstruction error ε_o is bounded such that $\|\varepsilon_o\| \le \varepsilon_{Mo}$, with ε_{Mo} a known constant. For this neural network estimate, $\sigma(\bullet)$ represents the activation function and W_o and V_o are

the weights, with $\hat{W}_o$ an estimate of W_o, which has an upper bound $\|W_o\|_F \leq W_{Mo}$. Similarly to the estimates of the states and outputs and their errors, the observer weight error is defined such that $\tilde{W}_o = W_o - \hat{W}_o$.

The NN estimate is then used to calculate the observer's estimate of the states as

$$\dot{\hat{Z}} = \hat{f}_{o1} + K_{o2} A^{-1} \tilde{X} \tag{7}$$

which is promptly used in addition to the output error to calculate the estimate of the output by using the dynamic equation

$$\dot{\hat{X}} = A\hat{Z} + K_{o1}\tilde{X} \tag{8}$$

At this point, the states may be estimated as

$$\hat{V} = \hat{Z} + K_{o3} A^{-1} \tilde{X} \tag{9}$$

The weight update law is then used to update the weights as below

$$\dot{\hat{W}}_o = F_o \sigma\left(V_o^T x_o\right) \tilde{X}^T - \kappa_{o1} F_o \hat{W}_o \tag{10}$$

with $F_o = F_o^T > 0$ and $\kappa_{o1} > 0$ tunable gains. The observer NN weights are randomly initialized. The observer error dynamics are

$$\begin{aligned} \dot{\tilde{X}} &= A\tilde{V} - (K_{o1} - K_{o3})\tilde{X} + \xi \\ \dot{\tilde{Z}} &= (f_o + (A^T - K_{o3}\dot{A}^{-1})\tilde{X}) - \hat{f}_{o1} - K_{o2} A^{-1} \tilde{X} \\ &\quad -(A^T - K_{o3}\dot{A}^{-1})\tilde{X} \end{aligned} \tag{11}$$

and the observer estimation error dynamics are given by

$$\dot{\tilde{V}} = -K_{o3}\tilde{V} + \tilde{f}_{o1} - A^{-1}(K_{o2} - K_{o3}(K_{o1} - K_{o3}))\tilde{X} - A^T \tilde{X} + \xi_1 \tag{12}$$

with ξ_1 a vector of positive constants containing a number of error terms. In (11), note that $\dot{A}^{-1}$ denotes the derivative of A^{-1} rather than the inverse of $\dot{A}$. In addition, positive gains K_{o1}, K_{o2}, K_{o3} are selected such that $K_{o1} > K_{o3}$, $K_{o3} > 2N_o/\kappa_{o1}$, and $K_{o2} = K_{o3}\left(K_{o1} - K_{o3}\right)$, where N_o is the number of hidden layer neurons. These equations are useful for proving Theorem 1, which is now introduced.

Theorem 1 [14] (Boundedness of observer estimation errors). Given the observer defined in (7), (8), and (9), with NN weight update law as given in (10), then there are positive gains K_{o1}, K_{o2}, K_{o3} for which $K_{o1} > K_{o3}$, $K_{o3} > 2N_o/\kappa_{o1}$, $K_{o2} = K_{o3}\left(K_{o1} - K_{o3}\right)$,

where N_o is the number of hidden layer neurons, such that the observer estimation error $\tilde{X}$, as well as $\tilde{V}$ and $\tilde{W}_o$ are UUB, with bounds

$$\left\|\tilde{X}\right\| > \sqrt{\frac{2\eta_o}{K_{o1}-K_{o3}}} \text{ or } \left\|\tilde{V}\right\| > \sqrt{\eta_o \Big/ \left(\frac{K_{o3}}{2} - \frac{N_o}{\kappa_{o1}}\right)} \text{ or } \left\|\tilde{W}_o\right\|_F > \sqrt{2\eta_o/\kappa_{o1}}$$

In addition, selecting the values of K_{o1}, K_{o2}, K_{o3} and κ_{o1} allows the bound on the errors to be made arbitrarily small.

3.3.Virtual Controller

The next step is to design the virtual controller, which is used to obtain the virtual control ouput or desired input $u_d = [\zeta\ w_{1d}\ w_{2d}\ w_{3d}]^T$. The steps given here follow the approach taken in [7]. This process is performed by first defining a set of error terms. The first, $\delta_1 = \rho_d - \rho$, was introduced with the kinematic controller. The second error term to be minimized is $\hat{\delta}_2 = m(\hat{v} - v_d)$, with $\hat{\delta}_2$ a velocity tracking error that incorporates the helicopter's mass. The third and fourth errors to be considered are $\epsilon_3 = \phi_d - \phi$ and $\epsilon_4 = \dot{\phi} - \dot{\phi}_d$, with ϕ_d the desired heading, which consider the error in the helicopter's heading and the rate at which this error is changing. A fifth error term considers the error in the thrust and may be expressed as

$$\hat{\delta}_3 = m\bar{g}e_3 - m\dot{v}_d + \hat{\delta}_2 + \frac{1}{m}\delta_1 - \zeta R(\Theta)e_3 \tag{13}$$

with all of the variables in (13) as previously defined. For convenience, a term

$$Y_d = \hat{\delta}_2 + \hat{\delta}_3 + \frac{d}{dt}(m\bar{g}e_3 - m\dot{v}_d + \hat{\delta}_2 + \frac{1}{m}\delta_1)$$

is introduced prior to the final error term necessary for this development, allowing this final error term to be written as

$$\hat{\delta}_4 = Y_d - (\dot{\zeta}R(\Theta)e_3 + \zeta R(\Theta)skew(\hat{\omega})e_3)$$

The choice of these particular error terms is analyzed in further detail in [7]. Selecting

$$\begin{aligned}&\tilde{\zeta} R(\Theta) e_3 - \zeta R(\Theta) skew(e_3)\tilde{w}_d \\ &= \hat{\delta}_3 + \hat{\delta}_4 + \dot{Y}_d - 2\dot{\zeta} R(\Theta) skew(\hat{\omega}) e_3\end{aligned} \tag{14}$$

to be solved for control of the main rotor thrust, pitch, and roll, and the following equation

$$\ddot{\phi} = \ddot{\phi}_d - \epsilon_3 - \epsilon_4 \tag{15}$$

to be solved for control of the yaw [7], a solution for both equations is given by

$$\begin{aligned}&\begin{bmatrix} \tilde{w}_{1d} \\ \tilde{w}_{2d} \\ \tilde{\zeta} \end{bmatrix} = \begin{bmatrix} 0 & \zeta & 0 \\ -\zeta & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}^{-1} R(\Theta)^T \\ &(\dot{Y}_d - 2\dot{\zeta} R(\Theta) skew(\hat{\omega}) e_3 + \hat{\delta}_3 + \hat{\delta}_4)\end{aligned} \tag{16}$$

from which $\tilde{w}_{1d}$, $\tilde{w}_{2d}$, and $\tilde{\zeta}$ may be obtained (with $\tilde{\zeta}$ obtained recursively). This solution was obtained by making use of the property $skew(e_3)\tilde{w}_d = e_3 \times \tilde{w}_d = -skew(\tilde{w}_d)e_3$, and rearranging and rewriting (14). Defining the relationship between the angular velocity and the orientation (from Section 2) as

$$\dot{\Theta} = \frac{1}{\cos(\theta)} \begin{pmatrix} 0 & s_\psi & c_\psi \\ 0 & c_\theta c_\psi & -c_\theta s_\psi \\ c_\theta & s_\theta s_\psi & s_\theta c_\psi \end{pmatrix} \hat{\omega} = T^{-1}\hat{\omega} \tag{17}$$

it is now possible to rearrange (5) in terms of $\dot{\hat{\omega}}$ and set $\dot{\hat{\omega}} = \tilde{w}_d$, while considering only the virtual control inputs. Doing this yields

$$\tilde{w}_d = -\mathcal{J}^{-1}\hat{\omega} \times \mathcal{J}\hat{\omega} + \left|Q_M\right| \mathcal{J}^{-1} e_3 - \left|Q_T\right| \mathcal{J}^{-1} e_2 + \mathcal{J}^{-1} P w_d .$$

Taking the derivative of (17), rearranging, and considering only the yaw (first element in orientation vector) results in

$$\ddot{\phi} = -e_1^T T^{-1} \dot{T} T^{-1} \hat{\omega} + \frac{1}{c_\theta}(s_\psi \tilde{w}_{2d} + c_\psi \tilde{w}_{3d}) \tag{18}$$

Then, employing both (15) and (18) and rearranging allows $\tilde{w}_{3d}$ to be obtained as

$$\tilde{w}_{3d} = \frac{c_\theta}{c_\psi}(\ddot{\phi}_d - \epsilon_4 - \epsilon_3 + e_1^T T^{-1} \dot{T} T^{-1} \hat{\omega} - \frac{s_\psi}{c_\theta} \tilde{w}_{2d}) \tag{19}$$

Now the real inputs are obtained. To do this, first restate a portion of the dynamics to obtain w_d from (5) as

$w_d = P^{-1}(\mathcal{J}\tilde{w}_d + \hat{\omega} \times \mathcal{J}\hat{\omega} - Q_M e_3 + Q_T e_2)$ with $P = diag([p_{11}\ p_{22}\ p_{33}]^T)$ a set of gains, and then obtain ζ by double-integrating from $\ddot{\zeta} = \tilde{\zeta}$ by using the value that has just been obtained for $\tilde{\zeta}$. Combining the preceding results allows one to obtain the feedforward portion of the control input as

$$u_d = [\zeta\ w_{1d}\ w_{2d}\ w_{3d}]^T \tag{20}$$

from the values that have just been obtained for ζ, w_{1d}, w_{2d}, and w_{3d}. Proof that the inputs generated by these equations assures convergence is provided in the Appendix.

3.4.Hamilton-Jacobi-Bellman Equation

In this section, based on the information provided by the kinematic controller, the optimal control input is designed to ensure that the unmanned helicopter system in (1) tracks a desired trajectory $X_d(t)$ in an optimal manner. This work extends that of [12] to output feedback control. For optimal tracking, the desired dynamics are defined as

$$\dot{V}_d = f(V_d) + g u_v^* \tag{21}$$

where $f(V_d) \in \mathbb{R}^{6\times1}$ is the internal dynamics of the helicopter system rewritten in terms of the desired state $V_d \in \mathbb{R}^{6\times1}$, g is bounded satisfying $g_{min} \le \| g \|_F \le g_{max}$, and $u_v^* \in \mathbb{R}^{6\times1}$ is the desired control input corresponding to the desired states. For reference, g is provided here explicitly as $g = M^{-1}\begin{bmatrix} E_3^b & 0^{3\times3} \\ 0^{3\times1} & diag([p_{11}\ p_{22}\ p_{33}]) \end{bmatrix}$.

Under these conditions, the optimal control input for the unmanned helicopter system given in (21) can be determined [8]. Next, the state tracking error is defined as

$$e = \hat{V} - V_d \tag{22}$$

Now, taking the derivative of (22), considering the estimated dynamics $\dot{\hat{V}} = f(\hat{V}) + gu_v$, and including (21), the tracking error dynamics in (22) can be written as

$$\dot{e} = f(\hat{V}) + gu_v - \dot{V}_d = f_e\left(e\right) + gu_e \tag{23}$$

where $f_e(e) = f(\hat{V}) - f(V_d)$ and $u_e = u_v - u_v^*$. The dynamics $f_e(e)$ and g are assumed to be known throughout this paper; however, this assumption may be relaxed if the uncertainties are estimated online using NNs. It is important to note that for the tracking error dynamics (23) with $e = 0$, there exists a unique equilibrium point (solution) to $f_e(e) = 0$ on the compact set $\Upsilon \subset \mathbb{R}^{6\times1}$ with $f_e\left(e = 0\right) = 0$ [12]. In this section, based on the information provided by the kinematic controller, the optimal control input is designed to ensure that the unmanned helicopter system in (1) tracks a desired trajectory $X_d(t)$ in an optimal manner. This work extends that of [12] to output feedback control. For optimal tracking, the desired dynamics are defined with $f_e\left(e = 0\right) = 0$ [12]. In other words, when the system dynamics are converted into the

tracking error form given by (23), the origin becomes the unique equilibrium point. $f_e(e)=0$

In order to control (23) in an optimal manner, the control policy should be sel $W_T(e(t))=\int_t^\infty r(e(\tau),u_e(\tau))d\tau\ u_e$ ected such that it minimizes the cost function given by

$$H_T(e,u_e)=r(e,u_e)+W_{Te}^T(e)(f_e(e)+gu_e) \tag{24}$$

where $r(e(\tau),u_e(\tau))=Q(e)+u_e^TBu_e$ and is the penalty on the states, with $B\in\mathbb{R}^{6\times6}$ a positive semi-definite matrix. After this, the Hamiltonian for the HJB tracking problem is defined in terms of the cost function as

$$H_T(e,u_e)=r(e,u_e)+W_{Te}^T(e)(f_e(e)+gu_e) \tag{25}$$

where $W_{Te}(e)$ is the gradient of $W_T(e)$ with respect to . Now, applying stationarity condition $\partial H(e,u_e)/\partial u_e=0$, the optimal control input is found to be

$$u_e^*(e)=-B^{-1}g^TW_{Te}^*(e)/2 \tag{26}$$

with $u_e^*(e)\in\mathbb{R}^4$. Substituting the optimal control input from (26) into the Hamiltonian (25) generates the HJB equation for the tracking problem as

$$0=Q_e(e)+W_{Te}^{*T}(e)f_e(e)-W_{Te}^{*T}(e)gB^{-1}g^TW_{Te}^*(e)/4 \tag{27}$$

with $W_T^*(0)$. The control input must be selected such that the cost function in (24) is finite, and it is assumed that there is an admissible controller [12]. At this point, Lemma 1 is introduced.

Lemma 1 (Boundedness of system state errors) [12]. Given the unmanned helicopter system with cost function (24) and optimal control input (26), let $J_1(e)$ be a continuously differentiable, radially unbounded Lyapunov candidate function such that

$\dot{J}_1(e) = J_{1e}^T(e)\dot{e} = J_{1e}^T(e)(f_e(e) + gu_e^*) < 0$ with $J_{1e}(e)$ the partial derivative of $J_1(e)$. In addition, let $\bar{Q}(e) \in \mathbb{R}^{6\times 6}$ be a positive definite matrix satisfying $\| \bar{Q}(e) \| = 0$ only if $\| e \| = 0$ and $\bar{Q}_{min} \le \| \bar{Q}(e) \| \le \bar{Q}_{max}$ for $e_{min} \le \| e \| \le e_{max}$ for positive constants $\bar{Q}_{min}$, $\bar{Q}_{max}$, e_{min}, and e_{max}. Also, let $\bar{Q}(e)$ satisfy $\lim_{e\to\infty} \bar{Q}(e) = \infty$ as well as

$$W_e^{*T} \bar{Q}(e) J_{1e} = r(e, u_e^*) = Q(e) + u_e^{*T} B u_e^* \tag{28}$$

then the following relation is true

$$J_{1e}^T (f_e(e) + gu_e^*) = -J_{1e}^T \bar{Q}(e) J_{1e} \tag{29}$$

Proof for Lemma 1 is provided in the Appendix.

Next, it is apparent that an expression including the optimally augmented control input in (26) can be written as

$$u_V = u_d - B^{-1} g^T W_{Te}^*(e) / 2 \tag{30}$$

with the desired feedforward control input u_d obtained from the virtual controller (20) in the previous section. Next, the SOLA is introduced.

3.5.Single Online Approximator (SOLA)-Based Optimal Control of Helicopter

Usually, in adaptive-critic based techniques, two OLAs [12] are used for optimal control, with one used to approximate the cost function while the other is used to generate the control action. In this paper, the adaptive critic for optimal control of a helicopter is realized online using a single OLA. For the SOLA to learn the cost function, the cost function is rewritten using the OLA representation as

$$W(e) = \Gamma^T \Phi(e) + \varepsilon \tag{31}$$

where Γ is the constant target OLA vector, $\Phi(e)$ is a linearly independent basis vector that satisfies $\Phi(e) = 0$, and ε is the OLA reconstruction error. The basis vector

used in this case is the same as in the previous section. The target OLA vector and reconstruction errors are assumed to be upper bounded according to $\| \Gamma \| \leq \Gamma_M$ and $\| \varepsilon \| \leq \varepsilon_M$, respectively [14]. The gradient of the OLA cost function in (31) is written as

$$\partial W(e) / \partial e = W_e(e) = \nabla_e^T \Phi(e) \Gamma + \nabla_e \varepsilon \tag{32}$$

Using (32), the optimal control input in (26) and the HJB equation in (27) can be written as

$$\begin{aligned} u_e^* &= -B^{-1} g^T \nabla_e^T \Phi(e) \Gamma / 2 - B^{-1} g^T \nabla_e \varepsilon / 2 \\ H^*(e, \Gamma) &= Q(e) + \Gamma^T \nabla_e \Phi(e) f_e(e) \\ -\Gamma^T \nabla_e \Phi(e) C \nabla_e^T \Phi(e) \Gamma / 4 &+ \varepsilon_{HJB} = 0 \end{aligned} \tag{33}$$

where $C = g B^{-1} g^T > 0$ is bounded such that $C_{min} \leq \| C \| \leq C_{max}$ for known constants C_{min} and C_{max} and

$$\begin{aligned} \varepsilon_{HJB} &= \nabla_e \varepsilon^T (f_e(e) - \frac{1}{2} C (\nabla_e^T \Phi(e) \Gamma + \nabla_e \varepsilon)) + \frac{1}{4} \nabla_e \varepsilon^T C \nabla_e \varepsilon \\ &= \nabla_e \varepsilon^T (f_e(e) + g u_e^*) + \frac{1}{4} \nabla_e \varepsilon^T C \nabla_e \varepsilon \end{aligned}$$

is the OLA reconstruction error. The OLA estimate of (31) is

$$\hat{W}(e) = \hat{\Gamma}^T \Phi(e) \tag{34}$$

with $\hat{\Gamma}$ the OLA estimate of the target vector Γ. In the same way, the estimate for the optimal control input based on (33) in terms of $\hat{\Gamma}$ can be expressed as

$$\hat{u}_e^* = -B^{-1} g^T \nabla_e^T \Phi(e) \hat{\Gamma} / 2 \tag{35}$$

The overall control input

$$\hat{u}_V = u_d + \hat{u}_e^* \tag{36}$$

is therefore now based on the NN estimate.

Lyapunov analysis performed in the appendix shows that the estimated control inputs approach the optimal control inputs with a bounded error. Employing (33) and (34) , the approximate Hamiltonian may now be written as

$$\hat{H}^*(e,\hat{\Gamma}) = Q(e) + \hat{\Gamma}^T \nabla_e \Phi(e) f_e(e) - \hat{\Gamma}^T \nabla_e \Phi(e) C \nabla_e^T \Phi(e) \hat{\Gamma} / 4 \tag{37}$$

Considering the definition of the OLA approximation of the cost function (34) and the Hamiltonian function (37), it is clear that both converge to zero when $\| e \| = 0$. Consequently, once the system state errors have converged to zero, the cost function approximation is no longer updated [14]. Recollecting the HJB equation in (25), the OLA estimate $\hat{\Gamma}$ should be tuned to minimize $\hat{H}^*(e,\hat{\Gamma})$. However, merely tuning $\hat{\Gamma}$ to minimize $\hat{H}^*(e,\hat{\Gamma})$ does not ensure the stability of the nonlinear helicopter system during the OLA learning process.

Therefore, the OLA tuning algorithm [12] is designed to minimize (37) while considering the system stability and is given below

$$\begin{aligned} \dot{\hat{\Gamma}} = & -\alpha_1 \frac{\hat{\beta}}{(\hat{\beta}^T \hat{\beta} + 1)^2} (Q(e) + \hat{\Gamma}^T \nabla_e \Phi(e) f_e(e) \\ & - \hat{\Gamma}^T \nabla_e \Phi(e) C \nabla_e^T \Phi(e) \hat{\Gamma} / 4) + \Sigma(e, \hat{u}_e^*) 0.5 \alpha_2 \nabla_e \Phi(e) C J_{1e}(e) \end{aligned} \tag{38}$$

where $\hat{\beta} = \nabla_e \Phi(e) f_e(e) - \nabla_e \Phi(e) C \nabla_e^T \Phi(e) \hat{\Gamma} / 2$, $\alpha_1 > 0$ and $\alpha_2 > 0$ are design constants, $J_{1e}(e)$ is defined in *Lemma 1*, and the operator $\Sigma(e, \hat{u}_e^*)$ is given by

$$\Sigma(e, \hat{u}_e^*) = \begin{cases} 0 & \text{if } J_{1e}^T(e)\dot{e} = J_{1e}^T(e) \\ & (f_e(e) - gB^{-1}g^T \nabla_e^T \Phi(e) \hat{\Gamma} / 2) < 0 \\ 1 & \text{otherwise} \end{cases} \tag{39}$$

Note that the weight update law is different than that in [12] as it is based on the observer's estimate of the states, rather than on the actual states themselves. The first

term in (38) is the portion of the tuning law which minimizes (37) and is derived using a normalized gradient descent scheme with the auxiliary HJB error defined as below

$$E_{HJB} = (\hat{H}^*(e,\hat{\Gamma}))^2 / 2 \tag{40}$$

The second term in the OLA tuning law in (38) ensures that the system states remain bounded while the SOLA scheme learns the optimal cost function.

The dynamics of the OLA parameter estimation error is considered as $\tilde{\Gamma} = \Gamma - \hat{\Gamma}$. Since this yields

$Q(e) = -\Gamma^T \nabla_e \Phi(e) f_e(e) + \Gamma^T \nabla_e \Phi(e) C \nabla_e^T \Phi(e) \Gamma / 4 - \varepsilon_{HJB}$ from (33), the approximate HJB equation in (37) can be expressed in terms of $\tilde{\Gamma}$ as

$$\begin{aligned} \hat{H}(e,\hat{\Gamma}) &= -\tilde{\Gamma}^T \nabla_e \Phi(e) f_e(e) + \frac{1}{2} \tilde{\Gamma}^T \nabla_e \Phi(e) C \nabla_e^T \Phi(e) \Gamma \\ &- \frac{1}{4} \tilde{\Gamma}^T \nabla_e \Phi(e) C \nabla_e^T \Phi(e) \tilde{\Gamma} - \varepsilon_{HJB} \end{aligned} \tag{41}$$

Then, since $\dot{\tilde{\Gamma}} = \dot{\hat{\Gamma}}$ and $\hat{\beta} = \nabla_e \Phi(e)(\dot{e}^* + C \nabla_e \varepsilon / 2) + \nabla_e \Phi(e) C \nabla_e^T \Phi(e) \tilde{\Gamma} / 2$, where $\dot{e} = f_e(e) + g u_e$, the error dynamics of (38) are

$$\begin{aligned} \dot{\tilde{\Gamma}} = \frac{\alpha_1}{\rho_1^2} &\left(\nabla_e \Phi(e) \left(\dot{e}_1^* + \frac{C \nabla_e \varepsilon}{2} \right) + \frac{\nabla_e \Phi(e) C \nabla_e^T \Phi(e) \tilde{\Gamma}}{2} \right) \\ &\left(\tilde{\Gamma}^T \nabla_e \Phi(e) \left(\dot{e}_1^* + \frac{C \nabla_e \varepsilon}{2} \right) + \frac{\tilde{\Gamma}^T \nabla_e \Phi(e) C \nabla_e^T \Phi(e) \tilde{\Gamma}}{2} + \varepsilon_{HJB} \right) \\ &- \Sigma(e, \hat{u}_e^*) \frac{\alpha_2}{2} \nabla_e \Phi(e) g B^{-1} g^T J_{1e}(e) \end{aligned} \tag{42}$$

where $\rho_1 = (\hat{\beta}^T \hat{\beta} + 1)$. Next, it is necessary to examine the stability of the SOLA-based adaptive scheme for optimal control along with the stability of the helicopter system.

3.6.Stability Analysis

The proofs to be introduced shortly are built on the basis of [7] and [12]. It is found that the control input consists of a predetermined feedforward term and an optimal feedback term that is a function of the gradient of the optimal cost function. In order to implement the optimal control in (24), the SOLA-based control law is used to learn the optimal feedback tracking control after necessary modifications, such that the OLA tuning algorithm is able to minimize the Hamiltonian while maintaining the stability of the helicopter system.

Lemma 1 has been introduced already and gives the boundedness of $\| J_{1e} \|$ and therefore the system state errors,. First, however, a definition is needed.

Definition: An equilibrium point e_e is said to be uniformly ultimately bounded (UUB) if there exists a compact set $S \subset \mathbb{R}^n$ such that for every $e_0 \in S$ there exists a bound D and time $T(D, e_0)$ such that $\| e(t) - e_e \| \leq D$ for all $t \geq t_0 + T$.

This definition will be used for Theorem 2, which will be provided shortly. Lemma 2 is now provided because it provides a stability condition needed for the proof for Theorem 2. Theorem 2 establishes the feedforward term stability and the stability of the entire resulting system.

Lemma 2 [12] *(Stability condition)*. Consider the affine system given by (4) and (5) under no disturbances and with known system dynamics, and with the smooth cost function given[12] in (24). If the applied control input is optimal, then the closed-loop system is asymptotically stable.

Theorem 2 (Overall system stability). Given the unmanned helicopter system with target HJB equation (27), let the tuning law for the SOLA be given by (38), and let

the feedforward control input be as in (20). Then there exist constants b_{Je} and b_{Γ} such that the OLA approximation error $\tilde{\Gamma}$ and $\| J_{1e}(e) \|$ are UUB for all $t \geq t_0 + T$ with ultimate bounds given by $\| J_{1e}(e) \| \leq b_{Je}$ and $\| \tilde{\Gamma} \| \leq b_{\Gamma}$. Further, OLA weight estimation error satisfy $\| W^* - \hat{W} \| \leq \varepsilon_{r1}$ and corresponding control input is bounded $\| u_v^* - \hat{u}_v^* \| \leq \varepsilon_{r2}$ for small constants ε_{r1} and ε_{r2}.

Note that a logical extension of Theorem 2 is that because $\| u_v^* - \hat{u}_v^* \| \leq \varepsilon_{r2}$, it is also the case that $\| V_d - V \| \leq \varepsilon_{r3}$, for a positive constant ε_{r3}. This is true because the system has known dynamics, with the optimal control input u_v^* generating the desired states V_d, and the neural-network-based estimate of the optimal control input $\hat{u}_v^*$ generating the actual states V. Proof is provided in the appendix.

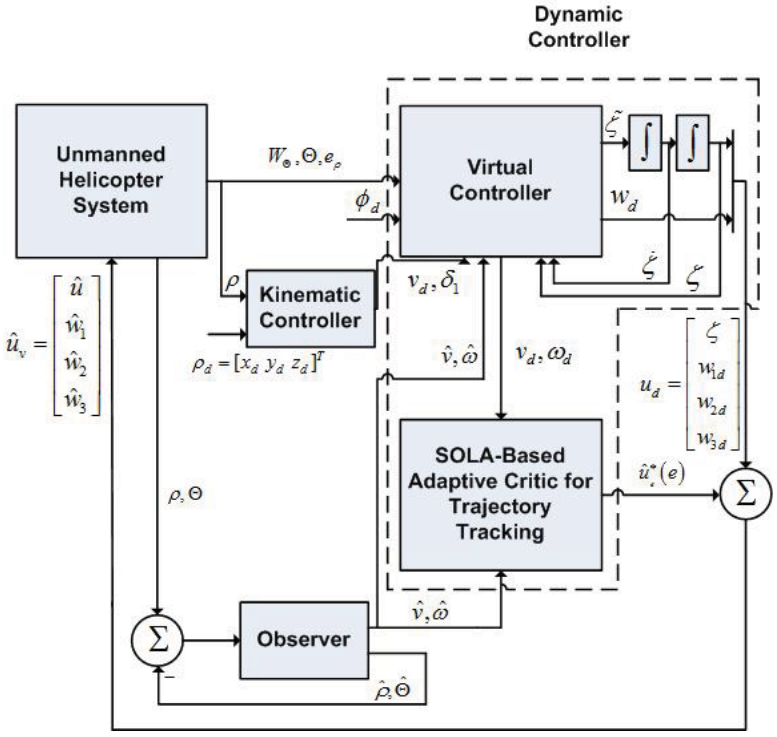

Figure 2. Output feedback control scheme.

In Figure 2, the entire NN-based output feedback control scheme for optimal tracking of the desired trajectory by the helicopter is illustrated. Note that the dynamic controller is comprised of the items within the dashed boundary. This output feedback control scheme consists of a kinematic controller to generate the desired velocity for the dynamic controller, a virtual controller to provide a feedforward term u_d, an optimal controller to generate the NN-based optimal feedback term $\hat{u}_e^*$, and an observer to estimate the states. Summing the control terms from the virtual and optimal (SOLA-based) controllers yields the NN-based control input for the helicopter dynamics $\hat{u}_v$, which is an estimate of the desired control input u_v.

4. SIMULATION RESULTS

All simulations are performed in Simulink and demonstrate the performance of the proposed control scheme when the helicopter is hovering, landing, and tracking trajectories. The simulations take into account the aerodynamic features presented as part of the helicopter model earlier in this paper.

The constants used for simulation are $\overline{g} = 9.8m/s^2$, $m = 9.6kg$, $p = diag([1.1 \quad 1.1 \quad 1.1]^T)$ $\mathcal{J} = diag([0.4\ 0.56\ 0.29]^T)\ kg{\cdot}m^2$, $\alpha_1 = 100$, $\alpha_2 = 1$, $l_t = 1.2m$ (x-axis dimension from the helicopter's center of gravity to the tail rotor), $l_m = 0.27m$, $Q_M = 0.002$, and $Q_T = 0.0002$. The optimal controller used seven hidden layer neurons for all simulations in this section, with gains $B = diag([0.1\ 0.1\ 0.1\ 0.001]^T)$. The basis function is $\Phi(e) = [1 \quad e_i \quad e_i^2 \quad e_i^3 \quad \sin(e_i) \quad \sin(2e_i) \quad \tanh(e_i) \quad \tanh(2e_i)]$ for i = 1to.

The basis function was augmented with a '1' for $\Phi_{1,1}(e)$ $\Phi_{1,1}(e)$ to aid the convergence rate. All NN weights are initialized to zero except for the observer's weights, which are randomly initialized. The observer NN uses five hidden layer neurons, with gains $K_{o1} = 22$, $K_{o2} = 121$, and $K_{o3} = 11$, $F_o = 10$ and $\kappa_{o1} = 1$, and a basis function as previously given. A disturbance is added between 39 and 40 seconds. The disturbance is a fast ramp to simulate a gust of wind, is applied to the position in three dimensions, and is expressed mathematically as $p = p + 1.6(t - 39)$.

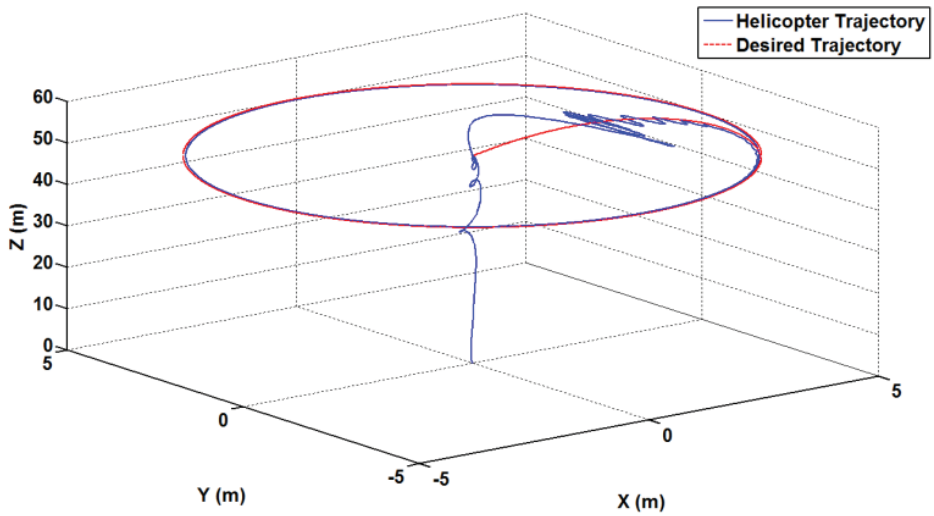

Figure 3. 3-D perspective of position during a take-off and circular maneuver.

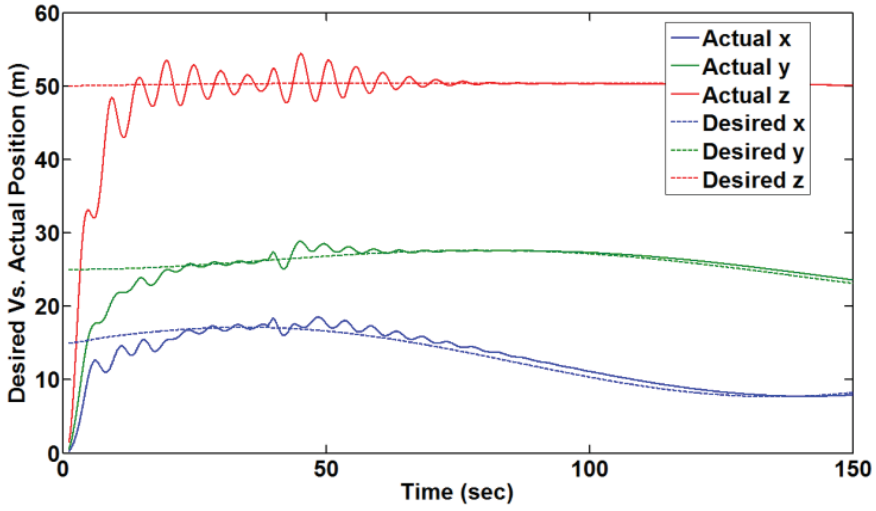

Figure 4. Helicopter position vs. time for the case of hovering.

Figure 3 demonstrates the helicopter's ability to follow a trajectory in two dimensions while hovering. The desired trajectory is defined as $(x_d, y_d, z_d) = (5(1-e^{-0.01t})\sin(0.025t), 5(1-e^{-0.01t})\cos(0.025t), 50)$ in meters. The figure shows that the helicopter can take off and follow the desired circular trajectory after a transient response. Figure 4 shows the actual and desired trajectories with respect to time during hovering. As expected, they track the target values despite the bounded disturbance.

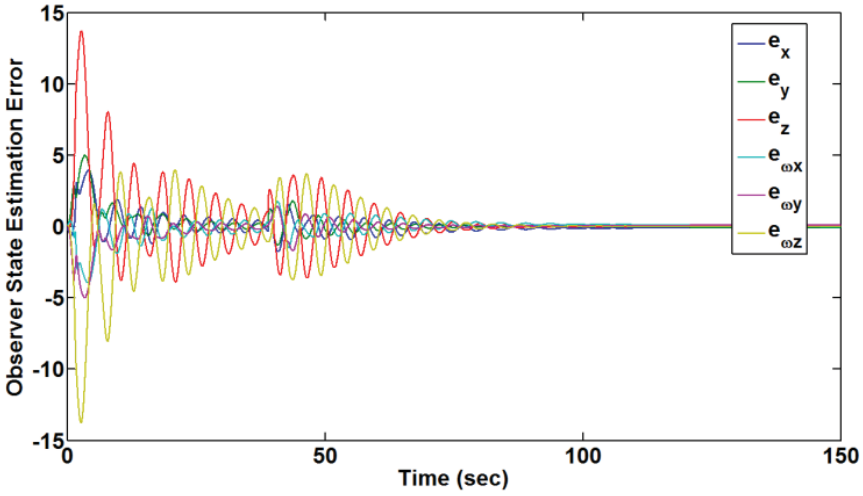

Figure 5. Observer state estimation errors during take-off and hover operation.

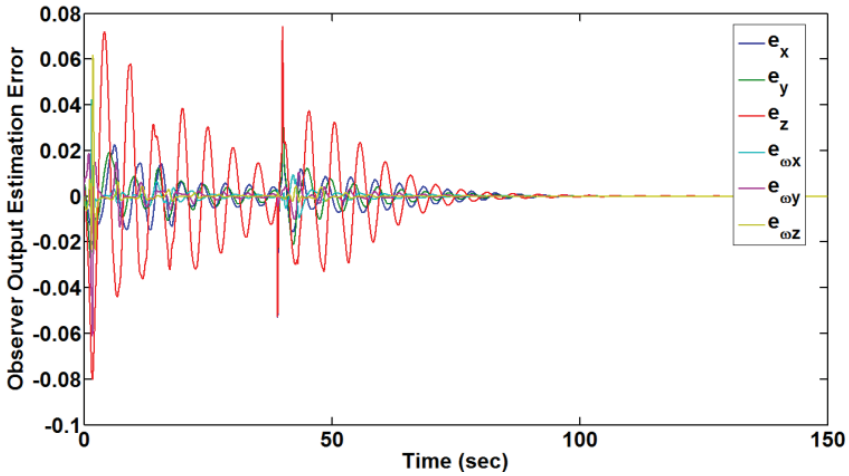

Figure 6. Observer output estimation error during take-off and hover maneuver.

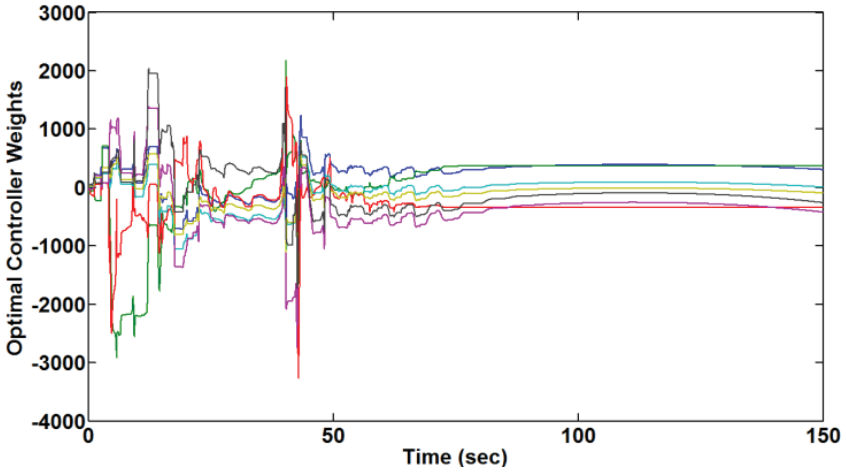

Figure 7. Cost function weights with respect to take-off and hover maneuver.

Figures 5 and 6 depict the observer performance for tracking the system states and system outputs respectively. These figures show that the observer errors converge near zero. Figures 7 and 8 depict the optimal controller weights and the control inputs respectively while Figure 9 shows the cumulative cost.

In other words, Figures 7 through 9 show the boundedness of the cost function weights, control input, and the cost function, as claimed in the Theorem. Finally, Figure 10 illustrates the Hamiltonian computed with respect to (37). The figure shows that the cost function is successfully estimated as the Hamiltonian converges to zero. This numerically proves that the designed controller has an optimal behavior after 20 seconds of starting the maneuver.

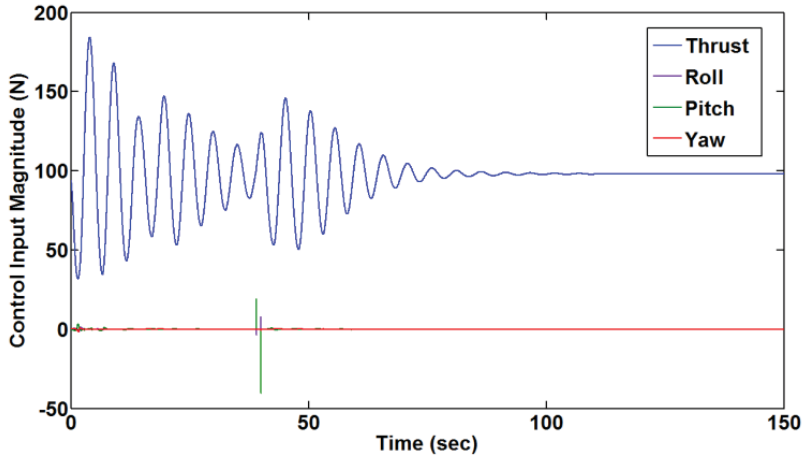

Figure 8. Control inputs applied to the helicopter with respect to Figure 3.

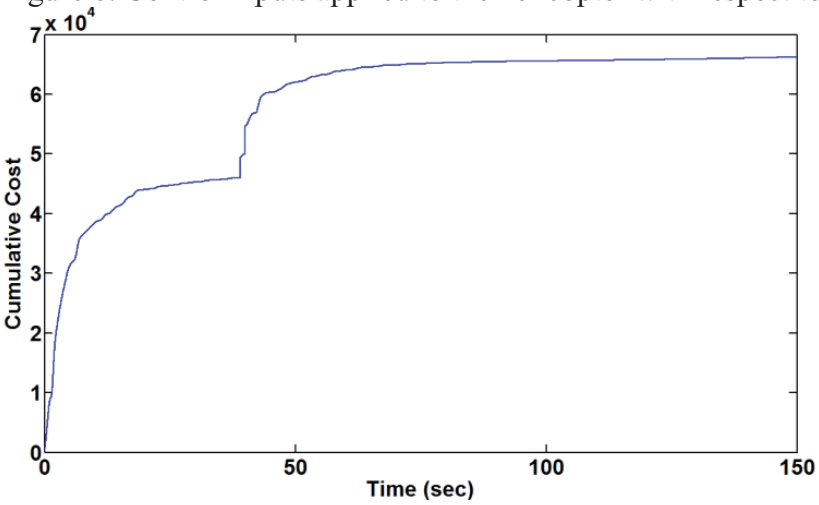

Figure 9. Cumulative cost to the maneuver of Figure 3.

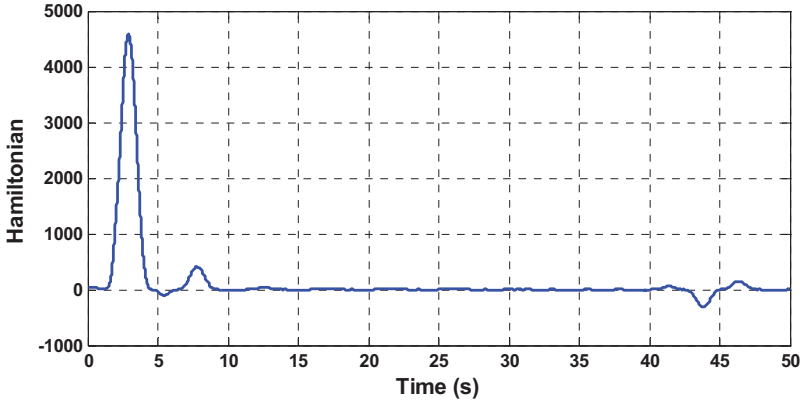

Figure 10. the Hamiltonian with respect to Figure 3 computed using (37).

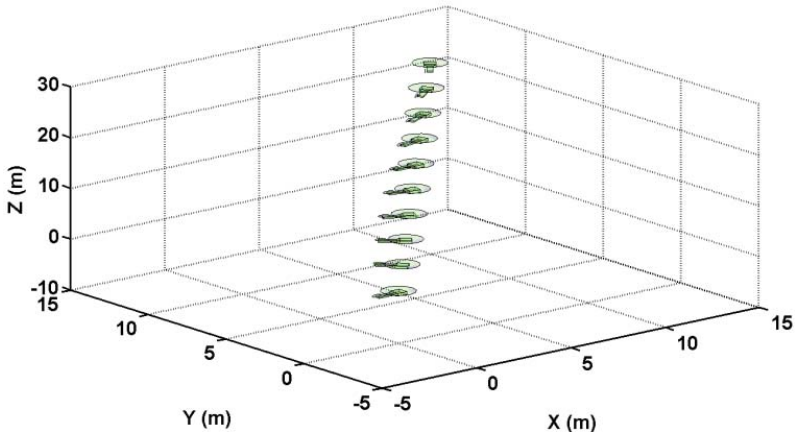

Figure 11. 3-D perspective of position and orientation during landing.

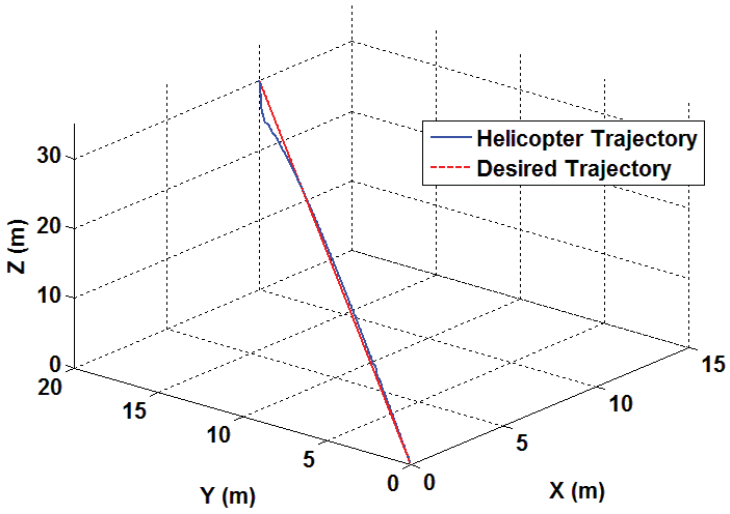

Figure 12. 3-D perspective of position during landing maneuver.

Figure 11 provides a 3-D view of the helicopter landing, showing both the position and the orientation throughout the maneuver. In Figure 12, a 3-D view of this landing maneuver is provided where the desired trajectory is plotted versus the helicopter position. It is well known that the process of landing has its own difficulties for any aircraft. This is why a trajectory is prescribed for landing instead of directly having a set point on the origin. This assures that the helicopter will smoothly land without any risk of crash.

5. CONCLUSIONS

A NN-based optimal control law has been proposed which uses a single online approximator for optimal regulation and tracking control of an unmanned helicopter with known dynamics having a dynamic model in strict-feedback form. The SOLA-based adaptive approach is designed to learn the infinite horizon continuous-time HJB equation, and the corresponding optimal control input that minimizes the HJB equation is calculated forward-in-time.

A feedforward controller has been introduced to compensate for the helicopter's weight and requirement for rotor thrust when in hover, and to permit trajectory tracking. Furthermore, it has been shown that the estimated control input approaches the target optimal control input with a small bounded error. A kinematic control structure has been used to obtain the desired velocity such that the desired position is achieved. A NN-based observer has been employed for obtaining the states from the outputs. The stability of the system has been analyzed, and simulation results confirm that the unmanned helicopter is capable of regulation and trajectory tracking.

APPENDIX

Proof of Lemma 1: Applying the optimal control input to an affine nonlinear system, the cost function becomes

$$\dot{W}^*(e) = W_e^{*T}(e)\dot{e} = W_e^{*T}(e)(f_e(e) + gu_e^*) = -Q_e(e) - u_e^{*T}Bu_e^*$$

Since $W_e^{*T}\bar{Q}_e(e)J_{1e} = r(e,u_e^*) = Q_e(e) + u_e^{*T}Bu_e^*$

one may obtain

$$(f_e(e) + gu_e^*) = -(W_e^*W_e^{*T})^{-1}W_e^*(Q_e(e) + u_e^{*T}Bu_e^*) =$$
$$-(W_e^*W_e^{*T})^{-1}W_e^*W_e^{*T}\bar{Q}_e(e)J_{1e} = -\bar{Q}_e(e)J_{1e}$$

from which one then has

$$J_{1e}^T(f_e(e)+gu_e^*)=-J_{1e}^T\bar{Q}_e(e)J_{1e}$$

concluding the proof for Lemma 1. □

Proof for Lemma 2: see reference [12].

Proof of Theorem 2: First, begin with the positive definite Lyapunov function candidate

$$\begin{aligned}J&=\alpha_2J_1(e)+\tilde{\Gamma}^T\tilde{\Gamma}/2+\delta_1^T\delta_1/2+\hat{\delta}_2^T\hat{\delta}_2/2+\hat{\delta}_3^T\hat{\delta}_3/2\\&+\epsilon_3^T\epsilon_3/2+\hat{\delta}_4^T\hat{\delta}_4/2+\epsilon_4^T\epsilon_4/2+\tilde{X}^T\tilde{X}/2+\tilde{V}^T\tilde{V}/2\\&+0.5tr\left\{\tilde{W}_o^TF_o^{-1}\tilde{W}_o\right\}\end{aligned}$$

The proof may then be divided into steps, with the first part of the Lyapunov function candidate considered first.

Step 1: Consider the optimal control Lyapunov function candidate $J_{HJB}=\alpha_2J_1(e)+\tilde{\Gamma}^T\tilde{\Gamma}/2$. Differentiating, one obtains $\dot{J}_{HJB}=\alpha_2J_{1e}^T(e)\dot{e}+\tilde{\Gamma}^T\dot{\tilde{\Gamma}}$. With $J_1(e)$ and $J_{1e}(e)$ as previously given. If $\|e\|=0$, $J_{HJB}(e)=\tilde{\Gamma}^T\tilde{\Gamma}/2$, $\dot{J}_{HJB}(e)=0$, and $\|\tilde{\Gamma}\|$ remains a bounded constant. For online learning, however, it is the case that $\|e\|>0$. For convenience, define $\dot{e}_1^*=f_e(e)+gu_e^*$. Then, using the affine nonlinear system, the optimal control input, and the tuning law's error dynamics along with the derivative of the Lyapunov candidate function J_{HJB}, one arrives at

$$\begin{aligned}\dot{J}_{HJB}&=\alpha_2J_{1e}^T(e)(f_e(e)-0.5gB^{-1}g^T\nabla_e^T\Phi(e)\hat{\Gamma})\\&-\left(\alpha_1/\rho_1^2\right)(\tilde{\Gamma}^T\nabla_e\Phi(e)(\dot{e}_1^*+0.5C\nabla_e\varepsilon))^2\\&-\left(\alpha_1/8\rho_1^2\right)(\tilde{\Gamma}^T\nabla_e\Phi(e)C\nabla_e^T\Phi(e)\tilde{\Gamma})^2\end{aligned}$$

$$
\begin{aligned}
&-\frac{3\alpha_1}{4\rho_1^2}\tilde{\Gamma}^T\nabla_e\Phi(e)(\dot{e}_1^*+\frac{C\nabla_e\varepsilon}{2})\tilde{\Gamma}^T\nabla_e\Phi(e)C\nabla_e^T\Phi(e)\tilde{\Gamma}\\
&-\left(\alpha_1/\rho_1^2\right)\tilde{\Gamma}^T\nabla_e\Phi(e)(\dot{e}_1^*+0.5C\nabla_e\varepsilon)\varepsilon_{HJB}\\
&-\left(\alpha_1/2\rho_1^2\right)\tilde{\Gamma}^T\nabla_e\Phi(e)C\nabla_e^T\Phi(e)\tilde{\Gamma}\varepsilon_{HJB}\\
&-\Sigma(e,\hat{u}_e)0.5\alpha_2\tilde{\Gamma}^T\nabla_e\Phi(e)gB^{-1}g^TJ_{1e}^T(e)
\end{aligned}
\tag{43}
$$

For convenience, all terms excluding the first and last in (43) are considered first, with this portion of J_{HJB} given by J_2:

$$
\begin{aligned}
\dot{J}_2=&-\left(\alpha_1/\rho_1^2\right)\tilde{\Gamma}^T(\nabla_e\Phi(e)(\dot{e}_1^*+0.5C\nabla_e\varepsilon)\\
&+0.5\nabla_e\Phi(e)C\nabla_e^T\Phi(e))(\tilde{\Gamma}^T\nabla_e\Phi(e)(\dot{e}_1^*+0.5C\nabla_e\varepsilon)\\
&+0.25\tilde{\Gamma}^T\nabla_e\Phi(e)C\nabla_e^T\Phi(e)\tilde{\Gamma}+\varepsilon_{HJB})
\end{aligned}
\tag{44}
$$

Multiplying through the $\tilde{\Gamma}^T$ term in (44) and expanding yields

$$
\begin{aligned}
\dot{J}_2=&-\left(\alpha_1/\rho_1^2\right)(\tilde{\Gamma}^T\nabla_e\Phi(e)(\dot{e}_1^*+0.5C\nabla_e\varepsilon))^2\\
&-\left(\alpha_1/8\rho_1^2\right)(\tilde{\Gamma}^T\nabla_e\Phi(e)C\nabla_e^T\Phi(e)\tilde{\Gamma})^2\\
&-\left(3\alpha_1/4\rho_1^2\right)(\tilde{\Gamma}^T\nabla_e\Phi(e)(\dot{e}_1^*+0.5C\nabla_e\varepsilon))\\
&(\tilde{\Gamma}^T\nabla_e\Phi(e)C\nabla_e^T\Phi(e)\tilde{\Gamma})\\
&-\left(\alpha_1/\rho_1^2\right)(\tilde{\Gamma}^T\nabla_e\Phi(e)(\dot{e}_1^*+0.5C\nabla_e\varepsilon))\varepsilon_{HJB}\\
&-\left(\alpha_1/2\rho_1^2\right)(\tilde{\Gamma}^T\nabla_e\Phi(e)C\nabla_e^T\Phi(e)\tilde{\Gamma})\varepsilon_{HJB}
\end{aligned}
\tag{45}
$$

Completing the squares with respect to $\tilde{\Gamma}^T\nabla_e\Phi(e)(\dot{e}_1^*+0.5C\nabla_e\varepsilon)$ and $\tilde{\Gamma}^T\nabla_e\Phi(e)C\nabla_e^T\Phi(e)\tilde{\Gamma}$ allows (45) to be rewritten as

$$
\begin{aligned}
\dot{J}_2=&-\left(\alpha_1/2\rho_1^2\right)(\tilde{\Gamma}^T\nabla_e\Phi(e)(\dot{e}_1^*+0.5C\nabla_e\varepsilon))^2\\
&-\left(\alpha_1/16\rho_1^2\right)(\tilde{\Gamma}^T\nabla_e\Phi(e)C\nabla_e^T\Phi(e)\tilde{\Gamma})^2+\left(\alpha_1/\rho_1^2\right)\varepsilon_{HJB}^2\\
&-\left(3\alpha_1/4\rho_1^2\right)(\tilde{\Gamma}^T\nabla_e\Phi(e)(\dot{e}_1^*+0.5C\nabla_e\varepsilon))\\
&(\tilde{\Gamma}^T\nabla_e\Phi(e)C\nabla_e^T\Phi(e)\tilde{\Gamma})+\left(\alpha_1/2\rho_1^2\right)\varepsilon_{HJB}^2\\
&-\left(\alpha_1/2\rho_1^2\right)(\tilde{\Gamma}^T\nabla_e\Phi(e)(\dot{e}_1^*+0.5C\nabla_e\varepsilon)+\varepsilon_{HJB})^2\\
&-\left(\alpha_1/16\rho_1^2\right)(\tilde{\Gamma}^T\nabla_e\Phi(e)C\nabla_e^T\Phi(e)\tilde{\Gamma}+4\varepsilon_{HJB})^2
\end{aligned}
\tag{46}
$$

Now, because the terms $-\left(\alpha_1/2\rho_1^2\right)(\tilde{\Gamma}^T\nabla_e\Phi(e)(\dot{e}_1^*+0.5C\nabla_e\varepsilon)+\varepsilon_{HJB})^2$ and $-\left(\alpha_1/16\rho_1^2\right)(\tilde{\Gamma}^T\nabla_e\Phi(e)C\nabla_e^T\Phi(e)\tilde{\Gamma}+4\varepsilon_{HJB})^2$ are negative definite, they cannot cause instability and will therefore be neglected from the remainder of the analysis. Rewriting (46) without these two terms and completing the square with respect to $\tilde{\Gamma}^T\nabla_e\Phi(e)C\nabla_e^T\Phi(e)\tilde{\Gamma}$ results in

$$\dot{J}_2 \le -\left(\alpha_1/2\rho_1^2\right)(\tilde{\Gamma}^T\nabla_e\Phi(e)(\dot{e}_1^*+0.5C\nabla_e\varepsilon))^2-\left(\alpha_1/32\rho_1^2\right)(\tilde{\Gamma}^T\nabla_e\Phi(e)C\nabla_e^T\Phi(e)\tilde{\Gamma})^2$$
$$-\left(\alpha_1/8\rho_1^2\right)(0.5\tilde{\Gamma}^T\nabla_e\Phi(e)C\nabla_e^T\Phi(e)\tilde{\Gamma}+6\tilde{\Gamma}^T\nabla_e\Phi(e)(\dot{e}_1^*+0.5C\nabla_e\varepsilon))^2$$
$$+\left(9\alpha_1/2\rho_1^2\right)(\tilde{\Gamma}^T\nabla_e\Phi(e)(\dot{e}_1^*+0.5C\nabla_e\varepsilon))^2+\left(3\alpha_1/2\rho_1^2\right)\varepsilon_{HJB}^2 \qquad (47)$$

Because the third term in (47) is negative semi-definite, it cannot cause instability and will therefore be neglected from the remainder of the analysis. In addition, the first and fourth terms in (47) are summed before rewriting (47) as

$$\begin{aligned}\dot{J}_2 &\le -\left(\alpha_1/32\rho_1^2\right)(\tilde{\Gamma}^T\nabla_e\Phi(e)C\nabla_e^T\Phi(e)\tilde{\Gamma})^2\\&+\left(4\alpha_1/\rho_1^2\right)(\tilde{\Gamma}^T\nabla_e\Phi(e)(\dot{e}_1^*+0.5C\nabla_e\varepsilon))^2+\left(3\alpha_1/2\rho_1^2\right)\varepsilon_{HJB}^2\end{aligned} \qquad (48)$$

Taking bounds on (48) and completing the square with respect to $\left\|\tilde{\Gamma}^T\nabla_e\Phi(e)\right\|^2$ results in

$$\begin{aligned}\dot{J}_2 \le{}& \left(3\alpha_1/2\rho_1^2\right)\varepsilon_{HJB}^2-\left(\alpha_1/64\rho_1^2\right)\left\|\tilde{\Gamma}^T\nabla_e\Phi(e)\right\|^4 C_{\min}^2\\&-\frac{\alpha_1 C_{\min}^2}{\rho_1^2}\left(\frac{\left\|\tilde{\Gamma}^T\nabla_e\Phi(e)\right\|^2}{8}-\frac{16}{C_{\min}^2}\left\|\dot{e}_1^*+0.5C\nabla_e\varepsilon\right\|^2\right)^2\\&+\left(256\alpha_1/\rho_1^2C_{\min}^2\right)\left\|\dot{e}_1^*+0.5C\nabla_e\varepsilon\right\|^4\end{aligned} \qquad (49)$$

Because the third term in (49) is negative semi-definite, it cannot cause instability and will therefore be neglected from the remainder of the analysis. Rewriting (49)

$$\begin{aligned}\dot{J}_2 &\le \left(3\alpha_1/2\rho_1^2\right)\varepsilon_{HJB}^2 - \left(\alpha_1/64\rho_1^2\right)\left\|\tilde{\Gamma}^T\nabla_e\Phi(e)\right\|^4 C_{\min}^2 \\ &+\left(256\alpha_1/\rho_1^2 C_{\min}^2\right)\left\|\dot{e}_1^* + 0.5C\nabla_e\varepsilon\right\|^4\end{aligned} \tag{50}$$

Applying the Pythagorean Theorem, noting that $\varepsilon_M^{'}$ is an upper bound such that $\left\|\nabla_e\varepsilon\right\| \le \varepsilon_M^{'}$, and employing the relationship $\left|\varepsilon_{HJB}\right| \le \varepsilon_M^{'}\delta(e) + \varepsilon_M^{'2}C_{\max}$ with $\delta(e) = \sqrt[4]{K^*\left\|J_{1e}(e)\right\|}$ and with K^* a constant allows (50) to be rewritten as

$$\begin{aligned}\dot{J}_2 &\le \left(3\alpha_1/2\rho_1^2\right)(\varepsilon_M^{'4} + \delta^4(e) + \varepsilon_M^{'4}C_{\max}^2) \\ &-\left(\alpha_1/64\rho_1^2\right)\left\|\tilde{\Gamma}^T\nabla_e\Phi(e)\right\|^4 C_{\min}^2 \\ &+\left(256\alpha_1/\rho_1^2 C_{\min}^2\right)\left(2\left\|\dot{e}_1^*\right\|^2 + 2\left\|0.5C\nabla_e\varepsilon\right\|^2\right)^2\end{aligned} \tag{51}$$

The last term in (51) may be rewritten as a result of the property that

$$\begin{aligned}\left(2\|x\|^2 + 2\|y\|^2\right)^2 &= 4\left(\|x\|^4 + \|y\|^4 + 2\|x\|^2\|y\|^2\right) \\ &\le 4\left(\|x\|^4 + \|y\|^4 + \|x\|^4 + \|y\|^4\right) = 8\left(\|x\|^4 + \|y\|^4\right)\end{aligned}$$

as

$$\begin{aligned}\dot{J}_2 &\le \left(3\alpha_1/2\rho_1^2\right)(\varepsilon_M^{'4} + \delta^4(e) + \varepsilon_M^{'4}C_{\max}^2) \\ &-\left(\alpha_1/64\rho_1^2\right)\left\|\tilde{\Gamma}^T\nabla_e\Phi(e)\right\|^4 C_{\min}^2 \\ &+\left(2048\alpha_1/\rho_1^2 C_{\min}^2\right)\left(\left\|\dot{e}_1^*\right\|^4 + \left\|0.5C\nabla_e\varepsilon\right\|^4\right)\end{aligned} \tag{52}$$

Making use of the fact that $\delta(e) = \left\|\dot{e}_1^*\right\|$ and with $\left\|\nabla_e\varepsilon\right\| \le \varepsilon_M^{'}$ an upper bound on the OLA reconstruction error, (52) may be written as

$$\dot{J}_2 \le \alpha_1\eta(\varepsilon)/\rho_1^2 - \alpha_1\beta_1\left\|\tilde{\Gamma}\right\|^4/\rho_1^2 + \alpha_1\beta_2\delta^4(e)/\rho_1^2 \tag{53}$$

With $\beta_1 = \Phi_{\min}^4 C_{\min}^2 / 64$, $\beta_2 = \left(2048/C_{\min}^2\right) + 3/2$, and $\eta(\varepsilon) = 128\varepsilon_M'^4 C_{\max}^4 / C_{\min}^2 + (3/2)\left(\varepsilon_M'^4 + \varepsilon_M'^4 C_{\max}^2\right)$. Now, looking back at (43), it is necessary to consider the case $J_{1e}(e)\dot{e} < 0$ and $\Sigma(e, \hat{u}_e) = 0$:

$$\dot{J}_{HJB} \le -\alpha_2 \| J_{1e}(e)\dot{e} \| + \frac{\alpha_1 \eta(\varepsilon)}{\rho_1^2} - \frac{\alpha_1 \beta_1}{\rho_1^2} \left\| \tilde{\Gamma} \right\|^4 + \frac{\alpha_1 \beta_2}{\rho_1^2} \delta^4(e)$$

This result may be rewritten taking a bound on $\dot{e}$ as

$$\dot{J}_{HJB} \le -\alpha_2 \| J_{1e}(e) \| \dot{e}_{\min} + \left(\alpha_1 \eta(\varepsilon)/\rho_1^2\right)$$
$$-\left(\alpha_1 \beta_1 / \rho_1^2\right) \left\| \tilde{\Gamma} \right\|^4 + \left(\alpha_1 \beta_2 / \rho_1^2\right) K^* \left\| J_{1e}(e) \right\|$$

Combining terms results in

$$\dot{J}_{HJB} \le -\| J_{1e}(e) \| (\alpha_2 \dot{e}_{\min} - \left(\alpha_1 \beta_2 / \rho_1^2\right) K^*)$$
$$+\left(\alpha_1 \eta(\varepsilon)/\rho_1^2\right) - \left(\alpha_1 \beta_1 / \rho_1^2\right) \left\| \tilde{\Gamma} \right\|^4$$

This is negative definite provided that $\alpha_2/\alpha_1 > \beta_2 K^* / \dot{e}_{min}$ and $\| J_{1e}(e) \| > \alpha_1 \eta(\varepsilon) / (\alpha_2 \rho_1^2 \dot{e}_{min} - \alpha_1 \beta_2 K^*) \equiv b_{Je0}$, or $\| \tilde{\Gamma} \| > \sqrt[4]{\eta(\varepsilon)/\beta_1} \equiv b_{\Gamma 0}$.

Next, it is necessary to consider the case $J_{1e}(e)\dot{e} > 0$ and $\Sigma(e, \hat{u}_e) = 1$:

$$\dot{J}_{HJB} \le \alpha_2 J_{1e}^T(e)(f_e(e) - 0.5C\nabla_e^T \Phi(e)\hat{\Gamma})$$
$$+\left(\alpha_1 \eta(\varepsilon)/\rho_1^2\right) - \left(\alpha_1 \beta_1 / \rho_1^2\right) \left\| \tilde{\Gamma} \right\|^4 + \left(\alpha_1 \beta_2 / \rho_1^2\right) \delta^4(e) \qquad (54)$$
$$-0.5\alpha_2 \tilde{\Gamma}^T \nabla_e \Phi(e) C J_{1e}^T(e)$$

The term $0.5\alpha_2 J_{1e}^T(e) C(\nabla_e^T \Phi(e)\Gamma + \nabla_e \varepsilon)$ is added and subtracted from (54) to obtain

$$\begin{aligned}
&\dot{J}_{HJB} \le \alpha_2 J_{1e}^T(e)(f_e(e) - 0.5C\nabla_e^T\Phi(e)\hat{\Gamma}) + \left(\alpha_1\eta(\varepsilon)/\rho_1^2\right) \\
&-0.5\alpha_2\tilde{\Gamma}^T\nabla_e\Phi(e)CJ_{1e}^T(e) - \left(\alpha_1\beta_1/\rho_1^2\right)\left\|\tilde{\Gamma}\right\|^4 \\
&+0.5\alpha_2 J_{1e}^T(e)C(\nabla_e^T\Phi(e)\Gamma + \nabla_e\varepsilon) \\
&-0.5\alpha_2 J_{1e}^T(e)C(\nabla_e^T\Phi(e)\Gamma + \nabla_e\varepsilon) + \left(\alpha_1\beta_2/\rho_1^2\right)\delta^4(e)
\end{aligned} \tag{55}$$

Now, using the relationship for $\bar{Q}(e)$ given in (29), taking the bounds on $\bar{Q}(e)$, C, and $\nabla_e\varepsilon$ and the norm on $J_{1e}^T(e)$ allows (55) to be rewritten as

$$\begin{aligned}
&\dot{J}_{HJB} \le \alpha_2\bar{Q}_{\min}\left\|J_{1e}^T(e)\right\|^2 + \left(\alpha_1\eta(\varepsilon)/\rho_1^2\right) - \left(\alpha_1\beta_1/\rho_1^2\right)\left\|\tilde{\Gamma}\right\|^4 \\
&+\left(\alpha_1\beta_2/\rho_1^2\right)K^*\left\|J_{1e}(e)\right\| + 0.5\alpha_2\left\|J_{1e}^T(e)\right\|C_{\max}\varepsilon_M^{'}
\end{aligned}$$

This last equation may be rewritten as below

$$\begin{aligned}
&\dot{J}_{HJB} \le \frac{\alpha_1\eta(\varepsilon)}{\rho_1^2} - \frac{\alpha_1\beta_1}{\rho_1^2}\left\|\tilde{\Gamma}\right\|^4 - \alpha_2\frac{\bar{Q}_{\min}}{2}\left\|J_{1e}(e)\right\|^2 \\
&-\frac{\alpha_2\bar{Q}_{\min}}{2}\left(\left\|J_{1e}(e)\right\| - \left(\frac{C_{\max}\varepsilon_M^{'}}{2\bar{Q}_{\min}} + \frac{\alpha_1\beta_2K^*}{\rho_1^2\bar{Q}_{\min}\alpha_2}\right)\right)^2 \\
&+\frac{\alpha_2\bar{Q}_{\min}}{2}\left(\frac{C_{\max}\varepsilon_M^{'}}{2\bar{Q}_{\min}} + \frac{\alpha_1\beta_2K^*}{\rho_1^2\bar{Q}_{\min}\alpha_2}\right)^2
\end{aligned}$$

Lemma 2 yields

$$\begin{aligned}
&\dot{J}_{HJB} \le -0.5\alpha_2\bar{Q}_{min}\,\|J_{1e}(e)\|^2 - \alpha_1\,\|\tilde{\Gamma}\|^4\,\beta_1/\rho_1^2 \\
&+\alpha_1\eta(\varepsilon)/\rho_1^2 + \alpha_2C_{max}^2\varepsilon_M^{'2}/(4\bar{Q}_{min}) + \alpha_1^2\beta_2^2K^{*2}/(\alpha_2\rho_1^4\bar{Q}_{min})
\end{aligned}$$

with $0 < \bar{Q}_{min} \le \|Q(e)\|$. This is negative definite provided that $\|J_{1e}(e)\| > \sqrt{C_{\max}^2\varepsilon_M^{'2}/2\bar{Q}_{\min}^2} \equiv b_{Je1}$ and $\|\tilde{\Gamma}\| > \sqrt[4]{\left(\eta(\varepsilon)/\beta_1\right) + \left(\alpha_1\beta_2^2K^{*2}/\beta_1\alpha_2\bar{Q}_{\min}\right)} \equiv b_{\Gamma 1}$.

Therefore, $\left\|J_{1e}(e)\right\|$, $\|\tilde{\Gamma}\|$, and $\|e\|$ are UUB. In addition, defining the bounds $\Phi_{\min} \le \left\|\Phi(e)\right\| \le \Phi_{\max}$ and $\left\|\nabla_e\Phi\right\| \le \Phi_{\max}^{'}$, $\|W^* - \hat{W}\| \le \left\|\tilde{\Gamma}\right\|\left\|\Phi(e)\right\| + \varepsilon_M \le b_\Gamma\Phi_{\max} + \varepsilon_M \equiv \varepsilon_{r1}$

and $\| u_e^* - \hat{u}_e \| \leq (1/2)\lambda_{\max}(B^{-1}) g_M b_\Gamma \Phi'_{\max} + \lambda_{\max}(B^{-1}) g_M \varepsilon'_M \equiv \varepsilon_{r2}$, with $\lambda_{\max}(B^{-1})$ denoting the maximum eigenvalues of B^{-1}. The second part of the Lyapunov candidate function is considered next.

Step 2: Consider the feedforward control Lyapunov function candidate $J_{feedforward} = S_1 + S_2 + S_3 + S_4$ with $S_1 = 0.5\delta_1^T \delta_1$, $S_2 = 0.5\hat{\delta}_2^T \hat{\delta}_2$, $S_3 = 0.5\hat{\delta}_3^T \hat{\delta}_3 + 0.5\epsilon_3^T \epsilon_3$, and $S_4 = 0.5\hat{\delta}_4^T \hat{\delta}_4 + 0.5\epsilon_4^T \epsilon_4$. It has been shown that this selection of Lyapunov candidate will guarantee stability in [7]. Applying elements integral to (20) gives the derivative of the Lyapunov function

$$\dot{J}_{feedforward} = \dot{S}_1 + \dot{S}_2 + \dot{S}_3 + \dot{S}_4 =$$
$$-(1/m)\delta_1^T \delta_1 - \hat{\delta}_2^T \hat{\delta}_2 - \hat{\delta}_3^T \hat{\delta}_3 - \hat{\delta}_4^T \hat{\delta}_4 - \epsilon_3^T \epsilon_3 - \epsilon_4^T \epsilon_4$$

so $\dot{J}_{feedforward} < 0$, which is an asymptotically stable result. To quickly review the elements in this stability analysis, δ_1 is used for the error in the tracking along with ϵ_3, $\hat{\delta}_2$ regulates the translational velocity, $\hat{\delta}_3$ and $\hat{\delta}_4$ take roll and pitch angles into consideration, and ϵ_3 and ϵ_4 are used for the error in the orientation (yaw) and corresponding angular velocity.

Step 3: Consider the stability of the entire system. Combining

$$\dot{J}_{HJB} + \dot{J}_{feedforward} + \dot{J}_o = -\frac{\alpha_2 \bar{Q}_{e,min} \| J_{1e}(e) \|^2}{2} - \frac{\alpha_1 \| \tilde{\Gamma} \|^4 \beta_1}{\rho_1^2}$$
$$+ \frac{\alpha_1 \eta(\varepsilon)}{\rho_1^2} + \frac{\alpha_2 C_{max}^2 \varepsilon_M'^2}{(4\bar{Q}_{e,min})} + \frac{\alpha_1^2 \beta_2^2 K^{*2}}{(\alpha_2 \rho^4 \bar{Q}_{e,min})}$$
$$-(1/m)\delta_1^T \delta_1 - \hat{\delta}_2^T \hat{\delta}_2 - \hat{\delta}_3^T \hat{\delta}_3 - \hat{\delta}_4^T \hat{\delta}_4 - \epsilon_3^T \epsilon_3 - \epsilon_4^T \epsilon_4$$
$$+0.5\tilde{X}^T \dot{\tilde{X}} + 0.5\tilde{V}^T \dot{\tilde{V}} + 0.5tr\left\{ \tilde{W}_o^T F_o^{-1} \dot{\tilde{W}}_o \right\}$$

Lemma 1 and Lemma 2 will then ensure $\dot{J}_{HJB} < 0$ given that

$$|| J_{1e}(e) ||> \sqrt{C_{max}^2 \varepsilon_M^{'2} / (2\bar{Q}_{e,min}^2)} \equiv b_{Je1'} \tag{56}$$

and

$$|| \tilde{\Gamma} ||> \sqrt[4]{\eta(\varepsilon) / \beta_1 + \alpha_1 \beta_2^2 K^{*2} / (\beta_1 \alpha_2 \bar{Q}_{e,min})} \equiv b_{\Gamma 1} \tag{57}$$

if $\Sigma(e, \hat{u}_e) = 1$, or

$$\alpha_2 / \alpha_1 > \beta_2 K^* / \dot{e}_{min} \tag{58}$$

and

$$|| J_{1e}(e) ||> \alpha_1 \eta(\varepsilon) / (\alpha_2 \rho_1^2 \dot{e}_{min} - \alpha_1 \beta_2 K^*) \equiv b_{Je0} \tag{59}$$

or

$$|| \tilde{\Gamma} ||> \sqrt[4]{\eta(\varepsilon) / \beta_1} \equiv b_{\Gamma 0} \tag{60}$$

if $\Sigma(e, \hat{u}_e) = 0$. In addition, it is also necessary that

$$\left\| \tilde{X} \right\| > \sqrt{2\eta_o / (K_{o1} - K_{o3})} \tag{61}$$

or

$$\left\| \tilde{V} \right\| > \sqrt{\eta_o / (0.5 K_{o3} - (N_o / \kappa_{o1}))} \tag{62}$$

or

$$\left\| \tilde{W}_o \right\|_F > \sqrt{2\eta_o / \kappa_{o1}} \tag{63}$$

which allows the conclusion that $|| W^*(e) - \hat{W}(e) ||\leq || \tilde{\Gamma} ||| \Phi(e) || + \varepsilon_M \leq b_\Gamma \Phi_M + \varepsilon_M \equiv \varepsilon_{r1}$ and

$$\begin{aligned} &|| u_e^*(e) - \hat{u}_e(e) ||\leq \lambda_{max}(B^{-1}) g_M b_\Gamma \Phi_M^{'} / 2 \\ &+ \lambda_{max}(B^{-1}) g_M \varepsilon_M^{'} / 2 \equiv \varepsilon_{r2} \end{aligned}$$

Then $\dot{J}_{HJB} + \dot{J}_{feedforward} + \dot{J}_o < 0$ provided that the conditions in (56) - (63) hold. Because the feedforward term is asymptotically stable, This result may be extended with the same bounds to the final control input such that

$$\| u_v^* - \hat{u}_v^* \| \leq \lambda_{max}(B^{-1}) g_M b_\Gamma \Phi_M^{'} / 2 + \lambda_{max}(B^{-1}) g_M \varepsilon_M^{'} / 2 \equiv \varepsilon_{r2}$$

In other words, the overall system is UUB with the bounds from (56) and (57), completing the proof.

At this point, bounds have been provided showing the convergence of the state estimate to the true state, and the convergence of the true state to the desired state, which collectively result in the convergence of the estimated state to the desired state.

The dynamics presented at the beginning of the paper provide $\overline{S}(\omega) = \begin{bmatrix} 0^{3\times1} & -\omega \times \mathcal{J}\omega \end{bmatrix}^T$. However, this is a simplification of the real dynamics, which include an additional term such that $\overline{S}(\omega) = [R(\Theta) K w_d \quad -\omega \times \mathcal{J}\omega \,]^T$, with $K = \frac{1}{l_M} \begin{bmatrix} 0 & 1 & 0 \\ -1 & 0 & 1/l_t \\ 0 & 0 & 0 \end{bmatrix}$. This coupling term is relatively small, but the robustness against neglecting the term has been demonstrated using a nonlinear controller and is available for the interested reader in [7] for the case of state feedback. The case for output feedback is given below, following the approach given in [7]. Define $\alpha_1 = \left(\delta_1, \hat{\delta}_2, \hat{\delta}_3, \hat{\delta}_4, \epsilon_3, \epsilon_4 \right) \in \mathbb{R}^{14}$ as a vector of errors that have previously been introduced. The error dynamics for the complete system model remain as in [7]. The norms of the errors are described as $\gamma = \left(|\delta_1|, |\hat{\delta}_2|, |\hat{\delta}_3|, |\hat{\delta}_4|, |\epsilon_3|, |\epsilon_4| \right) \in \mathbb{R}^6$. Differentiating the feedforward Lyapunov function, $\dot{J}_{feedforward} = -\gamma^T \Lambda \gamma$, with Λ a positive gain matrix. A set of variables is defined for

convenience and simplified notation, such that $\pi_0 = \left(0,1,(m+1)/m,\left(2m^2+m+1\right)/m^2,0,0\right)$, $\pi_1 = \left(0,0,0,0,\sqrt{2},\sqrt{2}\right)^T$, $\pi_2 = \left(0,0,1,1,0,0\right)^T$, $\pi_3 = \left(\frac{2m+1}{m^3},\frac{2m^3+m+1}{m^3},\frac{2(m+1)}{m},\frac{2m+1}{m},0,0\right)^T$, $\pi_4 = \pi_1$, $\pi_5 = \left((m+1)/m^2,1/m,(2m+1)/m,1,0,0\right)$, and $\pi_6 = \left(1/m,1,1,0,0,0\right)^T$, as well as $\tau_1 = \left(0,0,0,0,\sqrt{2}\right)^T$, $\tau_3 = \left(0,0,0,\sqrt{2},0\right)^T$, $\tau_4 = \left(0,m,0,0,0\right)$, $\tau_5 = \left(m,0,0,0,0\right)^T$, and $c_0 = \left\|I_{3\times 3}\right\|$, $c_2 = \left(2m^2+2m+1\right)/m$, $c_3 = (m+1)/m^2$, $c_4 = 2c_3$, $c_5 = c_2$, $c_6 = c_5 + 2k_1c_3 + d_1k_1c_4$, with $d_1 = 4 + c_0$.

Bounding the small body forces results in

$$\dot{J}_{feedforward} \le -\gamma^T\Lambda\gamma + \left|\hat{\delta}_2\right|\left|R(\Theta)Kw_d\right| + \frac{m+1}{m}\left|\hat{\delta}_3\right|\left|R(\Theta)Kw_d\right| + \left(\left(2m^2+m+1\right)/m\right)\left|\hat{\delta}_4\right|\left|R(\Theta)Kw_d\right| = -\gamma^T\Lambda\gamma + \sigma\left|w_d\right|\pi_0^T\gamma$$

where σ corresponds to the offset between the main rotor shaft and the helicopter's center of gravity. Next, it is necessary to determine the bound resulting in tracking with UUB stability. Defining $\left|w_0\right| = \left|Q_M\right| + \left|Q_T\right|$, one may rewrite $Pw_d = \mathcal{J}\tilde{w}_d - \hat{\omega}\times\mathcal{J}\hat{\omega} - \left|Q_M\right|e_3 + \left|Q_T\right|e_2$ as $\left|w_d\right| \le c_0\hat{\omega}^2 + c_0\left|\tilde{w}_d\right| + \left|w_0\right|$. Bounding (19) with the orientation constraints from Section 3.1 and the trajectory $\chi = \left(\left|\ddot{\rho}_d\right|,\left|\rho_d^{(3)}\right|,\left|\rho_d^{(4)}\right|,\left|\dot{\Theta}_d\right|,\left|\ddot{\Theta}_d\right|\right)$ and employing (16) yields an upper bound such that

$$\left|\tilde{w}_d\right| \le \left|\tilde{w}_d^{(1,2)}\right| + \left|\tilde{w}_d^3\right| \le \left|\tilde{w}_d^{(1,2)}\right| + \left\langle\tau_1,\chi\right\rangle + \left\langle\pi_1,\gamma\right\rangle + 4\hat{\omega}^2 + \sqrt{2}\left|\tilde{w}_d^2\right| \le \left(\left(1+\sqrt{2}\right)/\left|\zeta\right|\right)\left|\zeta\right|\left|\tilde{w}_d^{(1,2)}\right| + \left\langle\tau_1,\chi\right\rangle + \left\langle\pi_1,\gamma\right\rangle + 4\hat{\omega}^2$$

It is also possible to upper-bound $\dot{Y}_d$, $\hat{\omega}$, the main rotor thrust control input ζ, $\dot{\zeta}$, and the control input torques w_d.

Setting bounds k_0, k_1, and k_2 such that $\left|J_{feedforward}(0)\right| \le k_0^2$, $\left|\hat{\omega}(0)\right| \le k_1$, and $\left|w_0\right| = \left(\left|Q_M\right| + \left|Q_T\right|\right) \le k_2$, with p_3^0, p_3^1, q_3^0, and q_3^1 positive constants and defining two bounds for the trajectory, B_1 and B_2 such that

$$B_1(\Delta) = \frac{k_0\left[\left(mg-\Delta\right)\left(1-\left(\sigma k_2 / k_0\right)\left|\pi_0\right|\right) + \Delta\sigma p_3^1\left|\pi_0\right| - \sigma\left(c_6 + p_3^0\left|\pi_0\right|\right)\right]}{\sigma\left(p_3^0 - \Delta p_3^1\right)\left(q_3^0 - \Delta q_3^0\right)}$$

and $B_2(\Delta) = \left(\Delta - \left|\pi_6\right| k_0\right) / \left|\tau_5\right|$ with

$$\Delta_* = \left(\begin{array}{c} \arg\sup \\ \Delta \ge mg - \left(k_0(c_4 + 2\left|\pi_0\right|\left|\pi_5\right|)\right) / \left(k_1\left|\pi_0\right|\right) \end{array} \left[\min\left\{B_1(\Delta), B_2(\Delta)\right\}\right] \right)_{\text{and}}$$

$$B_* = \begin{array}{c} \sup \\ \Delta \ge mg - \left(k_0(c_4 + 2\left|\pi_0\right|\left|\pi_5\right|)\right) / \left(k_1\left|\pi_0\right|\right) \end{array} \left[\min\left\{B_1(\Delta), B_2(\Delta)\right\}\right] \text{then if } \left|\chi\right| \le B_*,$$

the closed-loop system is locally UUB.

Expressing part of this result mathematically, $\left|J_{feedforward}(t)\right| \le k_0^2$, $\zeta > mg - \Delta_*$, and $\hat{\omega}(t) < k_1 + k_0\left|\pi_0\right|\left|\pi_4\right|(mg + \Delta_*) + \left|\pi_0\right|\left|q_1(\zeta)\right| B_*$, with $q_1(\zeta) = 2\tau_4 + \left|\zeta\right|\tau_3$. This locally UUB result may be obtained by guaranteeing that trajectory bounds B_1 and B_2 are positive, control input ζ is lower bounded, and the angular velocity $\hat{\omega}(t)$ is upper bounded, which may be demonstrated by following an approach similar to that taken in [7]. The result is that these three requirements are satisfied for all time and the closed-loop system is locally UUB. □

REFERENCES

[1] T. J. Koo and S. Sastry, "Output tracking control design of a helicopter model based on approximate linearization," in *Proc. 37th IEEE Conf. on Decision and Control*, Tampa, FL, 1998, pp. 3635-3640.

[2] B. F. Mettler, M. B. Tischlerand, and T. Kanade, "System Identification modelling of a small-scale rotorcraft for flight control design," *Int. Journal of Robotics Research*, Vol. 20, pp. 795-807, 2000.

[3] N. Hovakimyan, N. Kim, A. J. Calise, and J. V. R. Prasad, "Adaptive output feedback for high-bandwidth control of an unmanned helicopter," in *Proc. of AIAA Guidance, Navigation, and Control Conf.*, Montreal, Canada, 2001, pp. 1-11.

[4] E. N. Johnson and S. K. Kannan, "Adaptive trajectory control for autonomous helicopters," *Journal of Guidance, Control and Dynamics*, Vol. 28, pp. 524-538, 2005.

[5] B. Ahmed, H. R. Pota and M. Garratt, "Flight control of a rotary wing UAV using backstepping," *Int. Journal of Robust and Nonlinear Control*, Vol. 20, pp. 639-658, 2010.

[6] E. Frazzoli, M. A. Dahleh and E. Feron, "Trajectory tracking control design for autonomous helicopters using a backstepping algorithm," in *Proc. of American Control Conference*, Chicago, 2000, pp. 4102-4107.

[7] R. Mahoney and T. Hamel, "Robust trajectory tracking for a scale model autonomous helicopter," *Int. Journal of Robust and Nonlinear Control*, Vol. 14, pp. 1035-1059, 2004.

[8] F. L. Lewis and V. L. Syrmos, *Optimal Control*, 2nd ed. Hoboken, NJ: Wiley, 1995.

[9] R. Enns and J. Si, "Helicopter trimming and tracking control using direct neural dynamic programming," *IEEE Trans. on Neural Networks*, Vol. 14, pp. 929-939, 2003.

[10] S. Lee, C. Ha and B. S. Kim, "Adaptive nonlinear control system design for helicopter robust command augmentation," *Aerospace Science and Technology*, Vol. 9, pp. 241-251, 2005.

[11] T. Dierks and S. Jagannathan, "Optimal control of affine nonlinear discrete-time systems," in *Proc. of the Mediterranean Conf. on Control and Automation*, Thessaloniki, Greece, 2009, pp. 1390-1395.

[12] T. Dierks and S. Jagannathan, "Optimal control of affine nonlinear continuous-time systems," in *Proc. of American Control Conf.*, Baltimore, Maryland, 2010, pp. 1568-1573.

[13] H.K. Khalil, *Nonlinear Systems*, 3rd ed. Upper Saddle River, NJ: Prentice-Hall, 2002.

[14] T. Dierks and S. Jagannathan, "Output feedback control of a quadrotor UAV using neural networks," *IEEE Trans. on Neural Networks*, Vol. 21, pp. 50-66, 2010.

[15] F. L. Lewis, S. Jagannathan, and A. Yesilderek, *Neural Network Control of Robot Manipulators and Nonlinear Systems*. London: Taylor & Francis, 1999.

V. OPTIMAL ADAPTIVE CONTROL OF NONLINEAR CONTINUOUS-TIME SYSTEMS IN STRICTFEEDBACK FORM WITH UNKNOWN INTERNAL DYNAMICS

SUMMARY— This paper focuses on optimal adaptive control of nonlinear continuous-time systems in strict feedback form with uncertain internal dynamics. First, it is shown that the optimal tracking problem of strict feedback systems can be reduced to an optimal regulation problem by designing both feedforward and optimal adaptive feedback controllers which stabilize the tracking error dynamics of the affine nonlinear continuous-time system. Then, an optimal adaptive feedback scheme is introduced to estimate the infinite horizon cost or value function for affine nonlinear continuous-time systems with unknown internal dynamics. The optimal adaptive control input is then obtained by using the cost or value function estimation which is shown to minimize the Hamilton-Jacobi-Bellman (HJB) estimation error in a forward-in-time manner without using any value or policy iterations by assuming states are measurable. Simultaneously the unknown internal dynamics are also estimated separately. Finally, the optimal adaptive control scheme is revisited by using the output feedback. Simulation examples are provided to validate the theoretical results.

Keywords- Online Nonlinear Optimal Control; Adaptive Control; Strict Feedback Systems; Output Feedback Control

1. INTRODUCTION

Stabilization of nonlinear systems is now an established field [1]-[4]. Many control techniques such as feedback linearization [1][5], sliding mode [1], backstepping [5], adaptive control [2][5], and online approximators (OLA's)-based methods [2]-[3] are

developed for stabilization of both continuous and discrete-time nonlinear systems. However, a controller should not only stabilize a nonlinear system but also minimize a prescribed performance index [6],[7],[8]. This problem becomes more challenging when the nonlinear system dynamics become partially uncertain [9][10].

It is well known that the optimal control of linear systems can be obtained by solving the Riccati equation [8]. In contrast, the optimal control of nonlinear continuous or discrete-time systems is a much more challenging task that usually requires the solution of the Hamilton-Jacobi-Bellman (HJB) equation which does not have a closed-form solution. For the case of infinite horizon optimal control, linear systems require the algebraic Reccati (ARE) to be solve for once, whereas the nonlinear systems require the solution of the HJB equation in real-time [9]. These offline optimal schemes are usually impractical when the system dynamics are uncertain. Thus, online adaptive approximation-based optimal controller designs referred to as adaptive critic designs (ACD) [10]-[13] are introduced recently in the literature.

The ACD techniques [11] tend to solve the optimal control forward-in-time by finding the solution to the HJB equation in an iterative manner via value or policy iterations instead of the offline methodology. Recently, state dependent Riccati equation (SDRE) [16] is also proposed to address optimal control in an iterative and numerical way by assuming the nonlinear system has a linear state dependent representation. Q-learning is an alternate way to solve the optimal control [14][18] for discrete-state and continuous-time systems by using reinforcement learning without using system dynamics. In policy iteration based schemes, an initial admissible controller is normally needed.

In [6], a single online approximator-based ACD technique is introduced for continuous-time nonlinear system in affine form since traditional ACD schemes require two approximators such as neural networks (NNs). Lyapunov stability is included and policy and value iterations are not needed while an initial admissible controller is not required. Instead, value function and policy are updated at each sampling interval for implementation. However, full knowledge of the system dynamics is needed [6].

Several other online optimal methods are introduced to solve the optimal control of continuous and discrete time when the system dynamics are not fully known [10][11][14][13]. Although [12] does not require a full knowledge of the system dynamics, it is applicable to unknown linear systems. Continuing the work of [14], the work in [9] proposed an approach to control partially unknown nonlinear systems using iterative solution of HJB equation via policy iteration by requiring an initial admissible controller. While the recent iterative methods tend to offer closed loop stability, they require significant number of iterations for convergence [13] and consequently they are unsuitable for hardware implementation. Besides, available ACD techniques [6][11] are available for nonlinear systems in affine form.

In the recent literature, the optimal control of nonlinear strict feedback continuous-time systems [19] is introduced without using policy iterations when the system dynamics are known. Continuing the work of [19], the proposed current work tackles the optimal control of nonlinear continuous-time systems with unknown internal dynamics. The optimal control law and cost function are approximated by online parametric structures in a forward-in-time manner where the internal dynamics are also

identified using an adaptive scheme. Moreover, an initial admissible controller is not required and policy iterations are not utilized.

In the nonlinear adaptive control literature, strict feedback nonlinear systems are represented in a variety of forms [5],[15]-[22] and their stability is studied using the standard backstepping scheme [1][5] without any optimality. In addition, in a few papers [4],[15]-[22] the control of such unknown strict feedback systems using neural network (NN)-based adaptive schemes is given. More recently, the inverse optimal control of strict feedback systems is introduced in [7] when the dynamics are assumed to be known. However, in the inverse optimal control problem, first the control law is designed, and then the associated cost function is identified for that control law in contrast with traditional optimal control schemes where a control law is designed based on a given cost function.

In this paper, it is demonstrated that by using a feedforward controller, the optimal tracking problem of the partially unknown strict feedback systems is equivalent to optimally stabilizing an affine system expressed in the tracking error form. Subsequently, the optimal adaptive scheme is developed for affine nonlinear continuous-time systems without needing the internal dynamics and policy or value iterations. It is shown that the proposed approach can estimate the optimal value or cost function which in turn becomes the solution to the HJB equation. The internal dynamics is also being estimated separately first by using linear in the unknown parameters (LIP) assumption. To relax the LIP, NN approximators are utilized subsequently. Lyapunov theory is utilized to demonstrate the convergence of the adaptive optimal control scheme for the overall nonlinear system while explicitly considering the approximation errors resulting

from the use of the online approximator (OLA) in the backstepping approach. Finally, the optimal adaptive control of such systems by using output feedback is also introduced.

In the proposed method, an initial stabilizing control is not required in contrast to [11] and the proposed scheme is developed forward-in-time whereas some original works have to have an offline solution to the system dynamics or the HJB equation [9]. In addition, this scheme is developed without using value and/or policy iterations which are commonly used in the available ACD techniques [12][13]. It is shown that the approximated control input approaches the optimal value over time when a linear in the unknown parameter adaptive control is utilized. When a NN is utilized for value function estimation, it is shown that the convergence of the closed system will be only a function of its reconstruction error which converges to zero when a large number of neurons are selected. This is also an advantage over [22] that also uses a NN based adaptive approach without policy iteration, where the bounds do not converge to zero even if the reconstruction errors becomes zero.

The paper is organized as follows. Section II demonstrates that the optimal control of a class of strict feedback nonlinear continuous-time systems is equivalent to optimally controlling error dynamics which is in affine form if a proper feedforward term is chosen. Section III introduces an online optimal stabilization scheme for affine nonlinear continuous-time systems with uncertain internal dynamics. Next, Section IV combines the results of the previous sections to a state feedback based optimal adaptive tracking control of strict-feedback systems. Section V presents the optimal adaptive tracking control of a class of strict-feedback nonlinear continuous-time systems in the

absence of state measurements. Finally, Section VI evaluates the theoretical results by some numerical examples.

2. THE TRACKING PROBLEM FOR STRICT FEEDBACK SYSTEMS

Consider the multi-input multi-output (MIMO) nonlinear continuous-time system in the absence of disturbances described by

$$\dot{x}_i = f_i\left(x_1,...,x_i\right) + g_i\left(x_1,...,x_i\right)x_{i+1} \text{ for } \quad 1 \le i \le N-1 \text{ and } \quad N \ge 2 \tag{1}$$

$$\dot{x}_N = f_N\left(x_1,...,x_N\right) + g_N\left(x_1,...,x_N\right)u\,, \tag{2}$$

$$y = x_1 \tag{3}$$

where each $x_i \in \Re^m$ denotes a state vector, $u \in \Re^m$ represents the input vector with $f_i\left(x_1,...,x_i\right) \in \Re^m$, and $g_i\left(x_1,...,x_i\right) \in \Re^{m\times m}$ being nonlinear smooth functions. It is assumed that systems (1)-(2) is reachable while its internal dynamics, $f_i\left(x_1,...,x_i\right)$, can be represented as LIP as

$$f_i\left(x_1,...,x_i\right) = \Psi(x)\Lambda_i \tag{4}$$

where $\Psi(x) \in \Re^{m\times m}$ and $\Lambda_i \in \Re^{m\times 1}$ with Λ_i being unknown. Without loss of generality, it is assumed that $\Psi^T(x) = \Psi(x)$. Under the LIP assumption, the internal dynamics are estimated as $\hat{f}_i\left(x_1,...,x_i\right) = \Psi(x)\hat{\Lambda}_i$ where $\hat{\Lambda}_i$ is the estimate of the target parameter vector Λ_i, and the estimation error is then given by

$$f_i\left(x_1,...,x_i\right) - \hat{f}_i\left(x_1,...,x_i\right) = \tilde{f}_i\left(x_1,...,x_i\right) = \Psi(x)(\Lambda_i - \hat{\Lambda}_i) = \Psi(x)\tilde{\Lambda}_i\,. \tag{5}$$

The update law for $\hat{\Lambda}_i$ will be designed based on Lyapunov stability analysis. Here for(1), the system state x_{i+1} is treated as the virtual control input. Nonetheless, the

overall system (1)-(3) is being controlled through the control input u . The following assumption is needed before we proceed.

Assumption 1. It is assumed that $\left\| g_i\left(x_1,...,x_i\right)\right\| \neq 0$ $(1 \le i \le N)$ belongs to $\Omega \in \Re^n$, and it is bounded satisfying $g^i_{\min} \le \| g_i\left(x_1,...,x_i\right)\|_F \le g^i_{\max}$ where ||.|| is the Frobenius norm and $g^i_{\min}$ and $g^i_{\max}$ are positive constants.

Under the Assumption 1, the optimal control input for the nonlinear system (1)-(2) can be obtained [8] by using a backstepping approach. In other words, the objective of our scheme is to design an adaptive controller u in order to have the output y to track a desired trajectory x_{1d} in an *optimal manner* even when the internal dynamics, $f_i(.)$, are unknown. To this end, by applying the backstepping approach [5], the system given by (1)-(2) tracks a predesigned trajectory $\left(x_{2d},...,x_{Nd}\right)$. Now, we follow the steps in the standard backstepping scheme to design the optimal adaptive scheme for strict-feedback systems.

To stabilize the tracking error, $e_1 = x_1 - x_{1d}$, the backstepping approach will use N steps [1] which are presented next.

Step.1: It is desired that x_1 to follow a smooth desired trajectory x_{1d} . Therefore, define the tracking error $x_1 - x_{1d} = e_1$. The system dynamics in (1) can be rewritten as

$$\dot{x}_1 - \dot{x}_{1d} = \dot{e}_1 = -\dot{x}_{1d} + f_1\left(x_1\right) + g_1\left(x_1\right)x_{2d} + g_1\left(x_1\right)\left(x_2 - x_{2d}\right) \ , \qquad (6)$$

With the assumption that $\hat{f}_1\left(x_1\right)$ is the estimate of $f_1\left(x_1\right)$, define the estimation error as $f_1\left(x_1\right) - \hat{f}_1\left(x_1\right) = \tilde{f}_1\left(x_1\right)$. Then

$$\dot{x}_1 - \dot{x}_{1d} = \dot{e}_1$$
$$= -\dot{x}_{1d} + \hat{f}_1(x_1) - \hat{f}_1(x_{1d}) + \hat{f}_1(x_{1d}) + \tilde{f}_1(x_1) + g_1(x_1)(x_{2d}^* + x_{2d}^a)$$
$$+ g_1(x_1)(x_2 - x_{2d}) - g_1(x_1)(x_2 - x_{2d})$$

$$= \hat{f}_1(e_1) + \tilde{f}_1(x_1) + g_1(x_1)x_{2d}^* + g_1(x_1)e_2, \quad (7)$$

where virtual control input x_{2d} is chosen such that $x_{2d} = x_{2d}^* + x_{2d}^a$ with x_{2d}^* being the optimal feedback control input and x_{2d}^a the feedforward virtual control input. The input x_{2d}^a is selected by solving

$$-\dot{x}_{1d} + \hat{f}_1(x_{1d}) + g_1(x_1)x_{2d}^a - g_1(x_1)(x_2 - x_{2d}) = 0. \quad (8)$$

Here, $\hat{f}_1(x_1) - \hat{f}_1(x_{1d})$ is denoted as $\hat{f}_1(e_1)$ for simplicity. In the right hand side (RHS) of (4), the effect of the third and the last terms are cancelled during Lyapunov stability proof. This can be accomplished by choosing a proper desired trajectory (and the corresponding virtual controller) in the next step. Section III is devoted to present the existence of the optimal feedback control input x_{2d}^* and its design. Inevitably, e_2 cannot be zero due to dynamics of the second system of (1) and the desired output x_1 trajectory. Since the second to the ($N-1$) steps are quite similar, we skip to the i^{th} step.

Step. i: In this step, we need an optimal controller for the system (1)-(3) such that $e_i \to 0$. To this end, the system i in (1) can be rewritten as

$$\dot{x}_i - \dot{x}_{id} = \dot{e}_i = -\dot{x}_{id} + f_i(x_1,...,x_i) + g_i(x_1,...,x_i)x_{(i+1)d} + g_i(x_1,...,x_i)(x_{i+1} - x_{i+1d})$$

$$= \hat{f}_i(e_1,...,e_i) + \tilde{f}_i(x_1,...,x_i) + g_i(x_1,...,x_i)x_{(i+1)d}^* + g_i(x_1,...,x_i)e_{i+1} - g_{i-1}^T(x_1,...,x_{i-1})e_{i-1}, \quad (9)$$

where x_{id} is chosen such that $x_{(i+1)d} = x_{(i+1)d}^* + x_{(i+1)d}^a$, with the virtual control input $x_{(i+1)d}^a$ satisfying (similar to the step (7))

$$-\dot{x}_{id}+\hat{f}_i\left(x_1,...,x_i\right)+g_i\left(x_1,...,x_i\right)x_{(i+1)di}^{a}=\hat{f}_i\left(x_1,...,x_i\right)-\hat{f}_i\left(x_{1d},...,x_{id}\right)-g_{i-1}^{T}\left(x_1,...,x_{i-1}\right)e_{i-1}$$

$$=\hat{f}_i\left(e_1,...,e_i\right)-g_{i-1}^{T}\left(x_1,...,x_{i-1}\right)e_{i-1}. \tag{10}$$

As mentioned in the previous step, there exists an optimal solution for the virtual input $x_{(i+1)d}^{*}$ which will be designed in the next section. Moreover, the third term of (6) inevitably shows up due to the design procedure, while the fourth term is deliberately added due to stability considerations.

Step. N: In this step, similar to the previous steps, the system input will be designed. To this end, the system (2) can be rewritten as

$$\dot{x}_N-\dot{x}_{Nd}=\dot{e}_n=-\dot{x}_{Nd}+\hat{f}_N\left(x_1,...,x_N\right)+g_N\left(x_1,...,x_i\right)u$$

$$=\hat{f}_N\left(e_1,...,e_N\right)+\tilde{f}_N\left(x_1,...,x_N\right)+g_N\left(x_1,...,x_N\right)u^{*}-g_{N-1}^{T}\left(x_1,...,x_{N-1}\right)e_N, \tag{11}$$

where x_{id} is chosen such that $u=u^{*}+u^{a}$, with the feedforward control input u^{a} is selected from

$$-\dot{x}_{Nd}+\hat{f}_N\left(x_1,...,x_N\right)+g_N\left(x_1,...,x_N\right)u^{a}=\hat{f}_N\left(e_1,...,e_N\right)-g_{N-1}^{T}\left(x_1,...,x_{N-1}\right)e_{N-1}, \tag{12}$$

The optimal feedback control input, u^{*}, exists and will be designed in Section III.

It is obvious that u^{a} is required only in the case that the knowledge of f_i (.) i.e. Λ_i is given. Therefore, $\hat{\Lambda}_i$ should be estimated directly or indirectly [2] by using an adaptive scheme. Here, we use an indirect approach that estimates the internal dynamics by using a state estimator. Now, consider the following state estimator for the strict feedback system described by

$$\dot{\hat{x}}_i=\Psi(x)\hat{\Lambda}_i+g_i\left(x_1,...,x_i\right)x_{i+1}+\Upsilon_i\tilde{x}_i+\overline{\Upsilon}_i\tilde{x}_i\left(\tilde{x}_i^{T}\overline{\Upsilon}_i\tilde{x}_i\right)$$

$$\text{for}\quad 1\le i\le N-1\ \text{ and }\quad N\ge 2 \tag{13}$$

$$\dot{\hat{x}}_N = \Psi(x)\hat{\Lambda}_N + g_N\left(x_1,...,x_N\right)u + \Upsilon_N\tilde{x}_N + \overline{\Upsilon}_N\tilde{x}_N\left(\tilde{x}_N^T\overline{\Upsilon}_N\tilde{x}_N\right), \tag{14}$$

where $\Upsilon_i, \overline{\Upsilon}_i$ for $1 \le i \le N$ are chosen as positive definite constant design matrices. Define the state estimation error as $\tilde{x}_i = x_i - \hat{x}_i$. Now, by subtracting the dynamics (13)-(14) from (1)-(2) yields the state estimation error dynamics as

$$\dot{\tilde{x}}_i = \Psi(x)\tilde{\Lambda}_i - \Upsilon_i\tilde{x}_i - \overline{\Upsilon}_i\tilde{x}_i\left(\tilde{x}_i^T\overline{\Upsilon}_i\tilde{x}_i\right) \text{ for } \quad 1 \le i \le N \text{ and } N \ge 2 \tag{15}$$

The following lemma is stated in order to convert the strict-feedback system into an affine system.

Lemma 1. Consider the tracking dynamics defined in (4), (6), and (8). Assume that the virtual and real control input vector $U = \begin{bmatrix} x_{2d} & \cdots & x_{Nd} & u \end{bmatrix}$ is designed such that $U = U^a + U^*$ where $U^a = \begin{bmatrix} x_{2d}^a & \cdots & x_{Nd}^a & u^a \end{bmatrix}$ is the feedforward control input designed in (5), (7), (9) and $U^* = \begin{bmatrix} x_{2d}^* & \cdots & x_{Nd}^* & u^* \end{bmatrix}$ represent the feedback control input which optimally stabilizes the system

$$\begin{bmatrix} \dot{e}_1 \\ \vdots \\ \dot{e}_N \end{bmatrix} = \begin{bmatrix} \hat{f}_1(e_1) \\ \vdots \\ \hat{f}_N(e_1,...,e_N) \end{bmatrix} + \begin{bmatrix} g_1\left(x_1\right) & & 0 \\ & \ddots & \\ 0 & & g_N\left(x_1,...,x_N\right) \end{bmatrix} U^*. \tag{16}$$

In this case, optimal control of (1) and (2) is equivalent to the optimal controller design for (10) with $\hat{\Lambda}_i$ is being updated using

$$\dot{\hat{\Lambda}}_i = \Psi^T(x)e_i + \Psi^T(x)\tilde{x}_i, \tag{17}$$

where $\alpha_\Lambda > 0$ is a design parameter. In the other words, by applying $U = U^a + U^*$ to the system (1) and (2), the system dynamics (1) and (2) are transformed into the error system given by (10).

Proof. By choosing $J_1 = \left(E^T E + \tilde{\Lambda}^T \tilde{\Lambda} + \tilde{X}^T \tilde{X}\right)/2$ with $E^T = \begin{bmatrix} e_1^T & \cdots & e_N^T \end{bmatrix}$, $\tilde{X}^T = \begin{bmatrix} \tilde{x}_1^T & \cdots & \tilde{x}_N^T \end{bmatrix}$ and $\Lambda^T = \begin{bmatrix} \Lambda_1^T & \cdots & \Lambda_N^T \end{bmatrix}$ as the Lyapunov candidate. Taking the derivative and evaluating the system dynamics (4), (6), (8) along the desired trajectory we have

$$\dot{J}_1 = E^T \dot{E} + \tilde{\Lambda}^T \dot{\tilde{\Lambda}} + \tilde{X}^T \dot{\tilde{X}} = E^T \dot{E} - tr(\tilde{\Lambda}^T \dot{\hat{\Lambda}})$$

$$= \sum_{i=1}^{N-1} e_i^T \left(\hat{f}_i(e_1,...,e_i) + g_i(x_1,...,x_i) x_{(i+1)d}^* \right) + e_N^T \left(\hat{f}_N(e_1,...,e_N) + g_N(x_1,...,x_N) u^* \right)$$

$$\sum_{i=1}^{N} e_i^T \tilde{f}_i(x_1,...,x_i) + \sum_{i=1}^{N-1} e_i^T g_i(x_1,...,x_i) e_{i+1} - \sum_{i=2}^{N} e_i^T g_{i-1}^T(x_1,...,x_{i-1}) e_{i-1}$$

$$-\sum_{i=1}^{N} \tilde{\Lambda}_i^T \Psi^T(x) e_i - \sum_{i=1}^{N} \tilde{\Lambda}_i^T \Psi^T(x) \tilde{x}_i + \sum_{i=1}^{N} \tilde{x}_i^T \Psi(x) \tilde{\Lambda}_i - \sum_{i=1}^{N} \tilde{x}_i^T \Upsilon_i \tilde{x}_i$$

$$= \left(\sum_{i=1}^{N} e_i^T \hat{f}_i(e_1,...,e_i) + E^T \begin{bmatrix} g_1(x_1) & & 0 \\ & \ddots & \\ 0 & & g_N(x_1,...,x_N) \end{bmatrix} U^* \right)$$

$$-\sum_{i=1}^{N} \tilde{x}_i^T \Upsilon_i \tilde{x}_i - \sum_{i=1}^{N} \left(\tilde{x}_i^T \overline{\Upsilon}_i \tilde{x}_i \right)^2 \tag{18}$$

From (11) one can easily recognize that if the optimal controller U^* is stabilizing then the first term in the RHS of equation (17) becomes negative and therefore $\dot{J}_1$ becomes negative semidefinite which implies that the closed-loop signals are bounded. This in turn guarantees the tracking error E and the state estimation error $\tilde{X}$ go to zero as time approaches infinity using Barbalat's Lemma [4]. Moreover, convergence of $\tilde{X}$ implies that $\Psi(x)\tilde{\Lambda}_i$ will converge to zero based on (15). Therefore, if the input is persistently exciting, $\Psi(x)$ will not be zero while $\tilde{\Lambda}_i$ will converge to zero over time

which implies that $\hat{f}_i$ will converge to f_i provided U^* optimally stabilizes the affine system. ■

***Remark 1*:** In fact, this lemma shows that by properly selecting the feedforward term for strict feedback systems, the optimal tracking control problem reduces to optimally stabilizing the nonlinear continuous-time systems in affine form described by(10). However, the unknown parameters, Λ, requires a state estimator (4) (or observer) though the states are measurable.

***Remark 2*:** It is possible to choose $\dot{\hat{\Lambda}}_i = \Psi^T(x)e_i$ instead of (15) to relax the state estimator (13)-(14). Then, by choosing $J_1' = \left(E^T E + \tilde{\Lambda}^T \tilde{\Lambda}\right)/2$, it can be shown that the first derivative of the Lyapunov function candidate becomes

$$\dot{J}_1' = \sum_{i=1}^{N} e_i^T \hat{f}_i\left(e_1,...,e_i\right) + E^T \begin{bmatrix} g_1\left(x_1\right) & & 0 \\ & \ddots & \\ 0 & & g_N\left(x_1,...,x_N\right) \end{bmatrix} U^*. \tag{19}$$

By selecting the feedback control input optimally, the affine system can be stabilized and the error E approaches to zero asymptotically with the parameter estimation error bounded. However, it is not possible to show that $\hat{f}_i$ will converge to f_i. Therefore, direct method of estimating Λ may not be suitable.

The next step is to design U^* in equation (11) that stabilizes the system(10) in an optimal manner. Since (10) is a nonlinear continuous-time system in affine form, the next section will focus on designing an optimal adaptive controller that stabilizes a generic affine nonlinear continuous-time system. This will provide the necessary optimal stabilizing term U^* in (10) and (11) that makes $\dot{J}_1$ negative semidefinite.

3. OPTIMAL ADAPTIVE CONTROL OF AFFINE SYSTEMS WITH UNKNOWN INTERNAL DYNAMICS

Consider the nonlinear continuous-time system in affine form in the absence of disturbances described by

$$\dot{\chi} = \bar{f}(\chi) + \bar{g}(\chi)\upsilon \ , \tag{20}$$

where $\chi \in \Re^n$ represent the system states, $\bar{f}(\chi) \in \Re^n$, $\bar{g}(\chi) \in \Re^{nxm}$ are nonlinear smooth functions with $\bar{g}(\chi)$ satisfying $g_{min} \le ||\bar{g}(\chi)||_F \le g_{max}$, and $\upsilon \in \Re^m$ is the control input. Without loss of generality, assume that the system is controllable, $x = 0$ a unique equilibrium point on $\Omega \in \Re^n$ with $\bar{f}(0) = 0$. Under these conditions, the optimal control input for the nonlinear system (1) can be calculated [8]. Additionally, the internal dynamics $\bar{f}(\chi)$ is considered unknown whereas expressed as LIP i.e. it can be represented as

$$\bar{f}(\chi) = \mu(\chi)\lambda \, , \tag{21}$$

with $\mu(x) \in \Re^{n \times t}$ being a smooth regression function and $\lambda \in \Re^{t \times 1}$ as an unknown parameter vector. Here the control coefficient matrix, $\bar{g}(\chi)$, is a known function throughout the development of this paper.

Remark 3: It is important to mention that (20) is a more generic affine representation of (10). As we mentioned in Section II, our aim is design an optimal adaptive controller to stabilize (10). Therefore, this section will focus of optimal stabilization of affine continuous-time system (20). Next section will use these results to the particular case of (10).

The infinite horizon cost function for (20) is given by

$$V(\chi(t)) = \int_t^\infty r(\chi(\tau), \upsilon(\tau))d\tau \ , \tag{22}$$

where $r(\chi(\tau),\upsilon(\tau)) = Q(\chi) + \upsilon^T R\upsilon$, $Q(\chi) \geq 0$ is the positive semi-definite penalty on the states, and $R \in \Re^{mxm}$ is a positive definite matrix. Selecting the state penalty $Q(\chi)$ to be positive definite ensures that variations in any direction of the state x affects the cost $V(\chi(t))$ (Lewis and Syrmos, 1995). Moving on, the control input υ is required to be selected such that the cost function (22) is finite; or υ must be admissible [11].

Next, we define the Hamiltonian for the cost function (22) with an associated admissible control input υ to be

$$H(\chi,\upsilon) = r(\chi,\upsilon) + V_x^T(\chi)(\bar{f}(\chi) + \bar{g}(\chi)\upsilon) \ , \tag{23}$$

where $V_\chi(\chi)$ is the gradient of the $V(\chi)$ with respect to χ. It is well known that the optimal control input $\upsilon^*(\chi)$ that minimizes the cost function (22) also minimizes the Hamiltonian (23); therefore, the optimal control is found by using the stationary condition $\partial H(\chi,\upsilon)/\partial\upsilon = 0$ and revealed to be

$$\upsilon^*(\chi) = -\frac{1}{2}R^{-1}\bar{g}(\chi)^T V_\chi^*(\chi). \tag{24}$$

Substituting the optimal control (24) into the Hamiltonian (23) while observing $H(\chi,\upsilon^*,V_\chi^*) = 0$ reveals the HJB equation and the necessary and sufficient condition for optimal control to be

$$0 = Q(\chi) + V_\chi^{*T}(\chi)\bar{f}(\chi) - V_\chi^{*T}(\chi)\bar{g}(\chi)R^{-1}V^*(\chi)/4 \tag{25}$$

with $V^*(0) = 0$. It is clear from (19) that the optimal control can be generated by solving the cost or value function through the HJB equation (20). For linear systems, equation (25) yields the standard algebraic Riccati equation (ARE) [8].

Lemma 2 [6]. Given the nonlinear system (20) with associated cost function (22) and optimal control (24), let $\bar{J}(\chi)$ be a continuously differentiable, radially unbounded Lyapunov candidate such that $\dot{\bar{J}}(\chi) = J_\chi^T(\chi)\dot{\chi} = \bar{J}_\chi^T(\chi)\left(\bar{f}(\chi) + \bar{g}(\chi)\upsilon^*\right) < 0$ where $\bar{J}_\chi(\chi)$ is the partial derivative of the radially unbounded $\bar{J}(\chi)$ with respect to χ. Moreover, let $\bar{Q}(\chi)$ be a positive definite matrix satisfying $\left\|Q(\chi)\right\| = 0$ only if $\|\chi\| = 0$ and $Q_{\min} \le \left\|\bar{Q}(\chi)\right\| \le Q_{\max}$ for $\chi_{\min} \le \|\chi\| \le \chi_{\max}$ for positive constants $Q_{\min}$, $Q_{\max}$, $\chi_{\min}$ and $\chi_{\min}$. In addition, let $Q(\chi)$ satisfy $\lim_{\chi\to\infty} Q(\chi) = \infty$ as well as

$$V_\chi^{*T}\bar{Q}(\chi)J_\chi = r(\chi,\upsilon^*) = Q(\chi) + \upsilon^{*T}R\upsilon^* . \tag{26}$$

Then, the following relation holds

$$J_\chi^T(\bar{f}(\chi) + \bar{g}(\chi)\upsilon^*) = -\bar{J}_\chi^T Q(\chi)\bar{J}_\chi . \tag{27}$$

Proof: is referred to [6].

In [14], the closed loop dynamics $\bar{f}(\chi) + \bar{g}(\chi)\upsilon^*$ are required to satisfy a Lipschitz condition such that $\left\|\bar{f}(\chi) + \bar{g}(\chi)\upsilon^*\right\| \le K$ for a constant K. In contrast in this work, the optimal closed loop dynamics are assumed to be upper bounded by a function of the system states such that

$$\left\|\bar{f}(\chi) + \bar{g}(\chi)\upsilon^*\right\| \le \delta(\chi) . \tag{28}$$

The generalized bound $\delta(\chi)$ in this work is taken as $\delta(\chi) \equiv \sqrt[4]{k^* Q(\chi)}$ to satisfy general bounds with k^* is a constant. The assumption of a time-varying upper bound in (28) is a less stringent assumption than the constant upper bound required in [14].

To begin the development, we rewrite the cost function (22) using an approximator representation as

$$V(\chi) = \vartheta^T \varphi(\chi) + \varepsilon(\chi) \ , \tag{29}$$

where $\vartheta \in \Re^L$ is the constant target vector, $\varphi(\chi): \Re^n \to \Re^L$ is a linearly independent basis vector which satisfies $\phi(0) = 0$, and $\varepsilon(\chi)$ is the reconstruction error. The target vector and reconstruction errors are assumed to be upper bounded according to $\|\vartheta\| \le \vartheta_M$ and $\|\varepsilon(\chi)\| \le \varepsilon_M$, respectively [3]. In addition, it will be assumed that the gradient of the reconstruction error with respect to χ is upper bounded according to $\|\partial\varepsilon(\chi)/\partial\chi\| = \|\nabla_\chi \varepsilon(\chi)\| \le \varepsilon'_M$ [22]. The gradient of the cost function (29) is written as

$$\partial V(\chi)/\partial\chi = V_\chi(\chi) = \nabla_\chi^T \varphi(\chi)\vartheta + \nabla_\chi \varepsilon(\chi) \ . \tag{30}$$

Similar to standard adaptive control literature [2], it will be assumed that the reconstruction error $\varepsilon(\chi)$ is negligible, although Corollary 1 will show that with a bounded $\varepsilon(\chi)$ the closed loop system will still stay stable but bounded. Now, using (30), the optimal control (24) and HJB function (25) are rewritten as

$$\upsilon^*(\chi) = -\frac{1}{2} R^{-1} \overline{g}(\chi)^T \nabla_\chi^T \phi(\chi)\vartheta \tag{31}$$

and

$$H^*(\chi, \vartheta) = Q(\chi) + \vartheta^T \nabla_\chi \phi(\chi) \overline{f}(\chi) - \frac{1}{4} \vartheta^T \nabla_\chi \phi(\chi) \Pi \nabla_\chi^T \phi(\chi) \vartheta = 0 \tag{32}$$

where $\Pi = g(\chi) R^{-1} g(\chi)^T > 0$ is bounded such that $\Pi_{\min} \le \|\Pi\| \le \Pi_{\max}$ for known constants $\Pi_{\min}$ and $\Pi_{\max}$. Moving on, the estimate of (18) is now written as

$$\hat{V}(\chi) = \hat{\vartheta}^T \varphi(\chi) \tag{33}$$

where $\hat{\vartheta}$ is the OLA estimate of the target parameter vector ϑ. Similarly, the estimate of the optimal control (20) is written in terms of $\hat{\vartheta}$ as

$$\hat{\upsilon}(\chi) = -\frac{1}{2} R^{-1} \bar{g}(\chi)^T \nabla_\chi^T \varphi(\chi) \hat{\vartheta} . \tag{34}$$

In the development of this work, it will be shown that an initial stabilizing control is not required to implement the proposed optimal adaptive scheme in contrast to [11][14] which require initial control policies to be stabilizing. Moreover, Lyapunov theory will show that the estimated optimal control input (34) approaches the real optimal control input (24) with the evolution of time. The proposed optimal adaptive parameter tuning law described next ensures that the system states remain bounded and that (34) will become admissible.

Now, using (33) and (34), the approximate Hamiltonian can be written as

$$\hat{H}(\chi, \hat{\vartheta}) = Q(\chi) + \hat{\vartheta}^T \nabla_\chi \phi(\chi) \hat{\bar{f}}(\chi) - \frac{1}{4} \hat{\vartheta}^T \nabla_\chi \phi(\chi) \Pi \nabla_\chi^T \phi(\chi) \hat{\vartheta} \tag{35}$$

Remark 3: Observing the definition of the approximation of the cost function (33) and the estimation of Hamiltonian function (35), it is evident that both become zero when $\|\chi\| = 0$. Thus, once the system states have converged to zero, the cost function approximation can no longer be updated. This can be viewed as a persistency of excitation (PE) requirement for the inputs to the cost function OLA. That is, the system states must be persistently exiting long enough for the OLA to learn the optimal cost function.

Recalling the HJB equation shown in (25), the estimate $\hat{\vartheta}$ should be tuned to minimize $\hat{H}(\chi, \hat{\vartheta})$. However, tuning to minimize $\hat{H}(\chi, \hat{\vartheta})$ alone does not ensure the stability of the nonlinear system (20) during adaptation. Therefore, the proposed tuning

algorithm is designed to minimize (35) while considering the stability of (20) and written as

$$\dot{\hat{\vartheta}} = -\alpha_1 \frac{\hat{\sigma}}{(\hat{\sigma}^T\hat{\sigma}+1)^2}\left(Q(\chi)+\hat{\vartheta}^T\nabla_\chi\varphi(\chi)\hat{\bar{f}}(\chi)-\frac{1}{4}\hat{\vartheta}^T\nabla_\chi\varphi(\chi)\Pi\nabla_\chi^T\varphi(\chi)\hat{\vartheta}\right)$$

$$+\Sigma(\chi,\hat{\upsilon})\frac{\alpha_2}{2}\nabla_\chi\varphi(\chi)\bar{g}(\chi)R^{-1}\bar{g}(\chi)^T J_{1\chi}(\chi) \tag{36}$$

where $\hat{\bar{f}}(\chi)=\mu(\chi)\hat{\lambda}$ is the estimate of the internal dynamics with $\hat{\lambda}$ being the estimate of λ. Moreover, $\hat{\sigma}=\nabla_\chi\varphi\hat{\bar{f}}(\chi)-\nabla_\chi\varphi(\chi)\Pi\nabla_\chi^T\varphi(\chi)\hat{\vartheta}/2$, $\alpha_1>0$ and $\alpha_2>0$ are design constants. The indicator function $\Sigma(\chi,\hat{u}_1)$ used to determine the stability condition of the closed system and it is defined as follows

$$\Sigma(x,\hat{\upsilon})=\begin{cases}0 & \text{if } J_{1\chi}^T\dot{\chi}=J_{1\chi}^T(f(\chi)-\frac{1}{2}g(\chi)R^{-1}g(\chi)^T\nabla_\chi^T\varphi(\chi)\hat{\vartheta})<0\\ 1 & \text{otherwise}\end{cases}. \tag{37}$$

Here, $J_1(\chi)$ is a positive definite radially unbounded function of χ. The first term in (36) is the portion of the tuning law which seeks to minimize (35) and was derived using a normalized gradient descent scheme with the auxiliary HJB error defined as

$$E_{HJB}=\frac{1}{2}\hat{H}(\chi,\hat{\vartheta})^2 \tag{38}$$

Meanwhile, the second term in the tuning law (36) is included as it is required for the process of stability proof. The first portion of the tuning law $\dot{\hat{\vartheta}}$ in (36) utilizes $(\hat{\sigma}^T\hat{\sigma}+1)^2$ instead of the traditional $(\hat{\sigma}^T\hat{\sigma}+1)$ used for normalization. This modification was also utilized in [6][14] for the critic update. However, the update is different from the critic update proposed in [14] since online approximators are utilized in [14] whereas only a single network is used in this work to generate the optimal

controller. Moreover, the work in [6] uses a single NN for estimating the optimal input whereas the update law is in switching form and internal dynamics $\bar{f}(x)$ is assumed to be known.

Moving on, the dynamics of the parameter estimation error is given by $\tilde{\vartheta} = \vartheta - \hat{\vartheta}$. From (32), it can be observed that

$$Q(\chi) = -\vartheta^T \nabla_\chi \varphi(\chi) \bar{f}(\chi) + \vartheta^T \nabla_\chi \varphi(\chi) \Pi \nabla_\chi^T \varphi(\chi) \vartheta / 4 \tag{39}$$

The approximate HJB equation (35) can be rewritten in terms of $\tilde{\vartheta}$ as

$$\hat{H}(\chi, \hat{\vartheta}) = -\tilde{\vartheta}^T \nabla_\chi \varphi(\chi) \hat{\bar{f}}(\chi) + \frac{1}{2} \tilde{\vartheta}^T \nabla_\chi \varphi(\chi) \Pi \nabla_\chi^T \varphi(\chi) \vartheta - \frac{1}{4} \tilde{\vartheta}^T \nabla_\chi \varphi(\chi) \Pi \nabla_\chi^T \varphi(\chi) \tilde{\vartheta} . \tag{40}$$

Observing $\dot{\tilde{\vartheta}} = -\dot{\hat{\vartheta}}$ and $\hat{\sigma} = \nabla_\chi \varphi(\chi)(\dot{\chi}^* - \hat{\bar{f}}(\chi)) + \nabla_\chi \varphi(\chi) \Pi \nabla_\chi^T \varphi(\chi) \tilde{\vartheta} / 2$ where $\dot{\chi}^* = \bar{f}(\chi) + \bar{g}(\chi) \upsilon^*$ the error dynamics of (36) can be written as:

$$\dot{\tilde{\vartheta}} = \alpha_1 \frac{1}{\rho^2} \left(\nabla_\chi \varphi \left(\dot{\chi}^* - \tilde{\bar{f}}(\chi) \right) + \frac{1}{2} \nabla_\chi \varphi(\chi) \Pi \nabla_\chi^T \varphi(\chi) \tilde{\vartheta} \right)$$

$$\times \left(-\tilde{\vartheta} \nabla_\chi \varphi(\chi) \left(\dot{\chi}^* - \tilde{\bar{f}}(\chi) \right) - \vartheta^T \nabla_\chi \varphi(\chi) \tilde{\bar{f}}(\chi) - \frac{1}{4} \tilde{\vartheta}^T \nabla_\chi \varphi(\chi) \Pi \nabla_\chi^T \varphi(\chi) \tilde{\vartheta} \right)$$

$$- \frac{\alpha_2}{2} \Sigma(\chi, \hat{\upsilon}) \nabla_\chi \varphi(\chi) \bar{g}(\chi) R^{-1} \bar{g}(\chi)^T J_{1\chi}(\chi) , \tag{41}$$

where $\rho = (\hat{\sigma}^T \hat{\sigma} + 1)$ and $\tilde{\bar{f}} \equiv \bar{f} - \hat{\bar{f}}$ and (39) is used to substitute the value of $Q(\chi)$. The update law for estimation of the internal dynamics is chosen as

$$\dot{\hat{\lambda}} = \mu(\chi) \tilde{\chi} \tag{42}$$

which appears to be similar to the update law (17) with a state estimator given by

$$\dot{\hat{\chi}} = \mu(\chi) \hat{\lambda} + \bar{g}(\chi) \upsilon + p \tilde{\chi} + q \tilde{\chi} (\tilde{\chi}^T q \tilde{\chi}) \tag{43}$$

where the state estimation error is defined as $\tilde{\chi} = \chi - \hat{\chi}$ and $p, q \in \Re^n$ as a positive definite matrix. Moreover, $\sigma_M(\blacktriangle)$ and $\sigma_m(\blacktriangle)$ will be denoted as the maximum and the minimum singular values of $\blacktriangle$ respectively. And therefore, the state estimation error dynamics are given by

$$\dot{\tilde{\chi}} = \mu(\chi)\tilde{\lambda} - p\tilde{\chi} - q\tilde{\chi}\left(\tilde{\chi}^T q \tilde{\chi}\right). \tag{44}$$

Next, the stability of the optimal adaptive scheme is examined along with the stability of the nonlinear system (20).

Theorem 1: (Overall Stability Proof of the Optimal Adaptive Scheme). Consider the nonlinear system (20) with the cost function and internal dynamics are provided by a LIP adaptive system. Let the cost function parameter estimation update law be given by (36) and the adaptation gains/parameters are chosen such that

$$\alpha_2 / \alpha_2 > \frac{4096\alpha_1}{\Pi_{\min}^4} k^* / \dot{x}_{\min}$$

$$\sigma_m(p)\sigma_M^{-1}(p) > \frac{5}{2}\alpha_1 \left\|\nabla_\chi \varphi(\chi)\right\|_{\max}^2 \left\|\vartheta\right\|_{\max}^2$$

$$\sigma_m^2(q)\sigma_M^{-2}(q) > \frac{4096}{\Pi_{\min}^4}\alpha_1 \tag{45}$$

hold. Then, the state vector χ and the state estimation error $\tilde{\chi}$ uniformly converge to zero in the large while the cost function and state estimation parameters are bounded. Moreover, the cost function parameter errors $\tilde{\vartheta}$ and the internal dynamics parameter errors $\tilde{\lambda}$ converge to zero provided the input is persistently exciting which also means that $\hat{V} \to V^*$ and $\hat{\upsilon} \to \upsilon^*$.

Proof: See the appendix.

Next when a NN is utilized to estimate the cost function instead of a standard LIP adaptive system, the cost function reconstruction error $\varepsilon(\chi)$ is not equal to zero. The following corollary provides the convergence bounds of the closed-loop system.

Definition 2 (Lewis et al., 1999)*:* An equilibrium point χ_e is said to be *uniformly ultimately bounded (UUB)* if there exists a compact set $S \subset \Re^n$ so that for all $\chi_0 \in S$ there exists a bound B and a time in terval $T(B, \chi_o)$ such that $\|\chi(t) - \chi_e\| \leq B$ for all $t \geq t_0 + T$.

Corollary 1: (Overall Stability Proof of NN-based Optimal Adaptive Scheme). Given the nonlinear system (20) with the target HJB equation (25), let the cost function parameter estimation update law be given by (36) and the update law parameters are chosen such that for any $\alpha_2 > 0$, $p > 0$, and $q > 0$ (45) holds. Then, there exists positive constants, b_x , b_ϑ , and b_f such that the cost function parameter error $\tilde{\vartheta}$, the state χ, and the internal dynamics parameter error $\tilde{\lambda}$ are *UUB* for all $t \geq t_0 + T$ with ultimate bounds given $\|\chi\| \leq b_\chi$,and $\|\tilde{\vartheta}\| \leq b_\vartheta$. Further, as ε_{HJB} gets smaller ($\varepsilon_{HJB} \to 0$) by choosing more appropriate NN basis function and making the number of neurons larger, the convergence bounds gets smaller i.e. $V^* \to \hat{V}$ and $\hat{\upsilon} \to \upsilon^*$ consequently.

Proof: See appendix.

It is worth mentioning that compared with [22], Theorem 1 and Corollary 1 represent a more powerful result since the convergence bounds approach to zero with $\varepsilon_{HJB} \to 0$ while they do not in [22]. In the following section, we extend the above design scheme to optimal adaptive control of strict feedback systems with unknown internal dynamics.

4. OPTIMAL TRAJECTORY AND CONTROL INPUT DESIGN

Due to *Lemma 1*, the objective of this section is to optimally make the tracking error E in (10) to converge to zero. It is desired to design the optimal control vector defined by $\left[x_{2d}^*,...,x_{Nd}^*,u^*\right]$ such that the tracking error $(e_1,...,e_N)$ is stable while minimizing the cost function

$$V=\int_t^\infty r(E(\tau),U^*(\tau))d\tau\,, \tag{46}$$

where $E=\left[e_1,...,e_N\right]^T$, $U^*=\left[x_{2d}^*,...,x_{Nd}^*,u^*\right]^T$, and $\left[x_1,...,x_N\right]=X$. In (12), $r(E,U^*)=Q(E)+U^{*T}RU^*$, $Q(E)\geq 0$ is the positive semi-definite penalty on the states, and $R>0\in\Re^{M\times M}$ is a positive definite matrix with $M=mN$ .since the size of X is m times that of x_i.

Next, consider the optimal stabilization problem from for an affine system (17) in the error domain

$$\dot{E}=\hat{F}(E)+G(X)U^*\,, \tag{47}$$

where $\left[\hat{f}_1^T(e_1)\quad\cdots\quad\hat{f}_N^T(e_1,...,e_N)\right]^T=\hat{F}(E)$ and $G(X)=diag\left[g_1(x_1),...,g_N(x_1,...,x_N)\right]$. It is desired that E converges to zero while the cost function (12) is minimized.

Moving on, the control input U^* needs to be designed such that the cost function (12) will be finite. We define the Hamiltonian for the cost function (11) with an associated admissible control input U to be [8]

$$H(E,U)=r(E,U)+V_E^T(E)\left(F(E)+G(X)U\right), \tag{48}$$

where $V_E(E)$ is the gradient of the $V(E)$ with respect to E. *In the sequel, we will use the same terminology for denoting gradient of functions i.e. for any function* $\Omega(\psi)$, $\Omega_\psi(\psi)$ *means gradient of* $\Omega(\psi)$ *with respect to* ψ. It is well-known that the optimal trajectory U^* that minimizes the cost function (12) also minimizes the Hamiltonian (14); therefore, the optimal control is found by using the stationarity condition $\partial H(E,U)/\partial U = 0$ and revealed to be [8]

$$U^*(E) = -\frac{1}{2}R^{-1}G(X)^T V_E^*(E). \tag{49}$$

By substituting the optimal control (15) into the Hamiltonian (14) while observing $H(E,U) = 0$ reveals the HJB equation and the necessary and sufficient condition for optimal control to be [8]

$$Q(E) + V_E^T(E)F(X) - \frac{1}{4}V_E^T(E)G(X)R^{-1}G(X)^T V_E(E) = 0 \quad , \tag{50}$$

with $V^*(0) = 0$. For linear systems, equation (16) yields the standard algebraic Riccati equation (ARE) [8]. Before proceeding, the following technical lemma is required.

Lemma 2 [6]. Given the nonlinear system (13) with associated cost function (12) and optimal control (15), let $J(E)$ be a continuously differentiable, radially unbounded Lyapunov candidate such that $\dot{J}(E) = J_E^T(E)\dot{E} = J_E^T(E)\left(F(E) + G(X)U\right) < 0$ where $J_E^T(E)$ is the radially unbounded partial derivative of $J^T(E)$. Moreover, let $\bar{Q}(E)$ be a positive definite matrix satisfying $\left\|\bar{Q}(E)\right\| = 0$ only if $\|E\| = 0$ and $\bar{Q}_{\min} \le \left\|\bar{Q}(E)\right\| \le \bar{Q}_{\max}$ for $\varpi_{\min} \le \|E\| \le \varpi_{\max}$ for positive constants $\bar{Q}_{\min}$, $\bar{Q}_{\max}$, $\varpi_{\min}$ and $\varpi_{\max}$. In addition, let $\bar{Q}(E)$ satisfy $\lim_{E\to\infty}\bar{Q}(E) = \infty$ as well as

$$V_E^{*T}\bar{Q}(E)J_E = r(E,u^*) = Q(E) + U^{*T}RU^*. \tag{51}$$

Then, the following relation holds

$$J_E^T(F(E) + G(E)U^*) = -J_E^T\bar{Q}(E)J_E. \tag{52}$$

Proof: When the optimal control (15) is applied to the nonlinear system(13), the cost function (12) becomes a Lyapunov function rendering

$$\dot{V}^*(E) = V_E^{*T}(E)\dot{E} = V_E^{*T}\left(E\right)\left(F(E) + G(x)U^*\right) = -Q(E) - U^{*T}RU^* \tag{53}$$

From (15), after manipulation and substitution of (17), equation (19) is rewritten as

$$\begin{aligned} F(E) + G(X)U^* &= -(V_E^*V_E^{*T})^{-1}V_E^*(Q(E) + U^{*T}RU^*) \\ &= -(V_E^*V_E^{*T})^{-1}V_E^*V_E^{*T}\bar{Q}(E)J_E = -\bar{Q}(E)J_E \end{aligned} \tag{54}$$

Now, multiply both sides of (20) by J_E^T yields the desired relationship in (18). ■

The generalized bound $\delta(E)$ is taken as $\delta(E) < \sqrt[4]{K^*Q(E)}$ in this work where $\|J_E\|$ can be selected to satisfy general bounds and K^* is a constant. For example, if $\delta(E) = K_1\|E\|$ for a constant K_1, then it can be shown that selecting $J(E) = (E^TE)^{(5/2)}/5$ with $J_E(E) = (E^TE)^{(3/2)}E^T$ satisfies the bound. The assumption of a time-varying upper bound in (13) is a less stringent assumption than the constant upper bound required in [13]. The next section develops an approach for optimally stabilize the affine system which is required for optimal tracking of original strict feedback systems.

Moving on, we rewrite the cost function (11) using an OLA representation as

$$V(E) = \Theta^T\varphi(E) + \varepsilon(E), \tag{55}$$

where $\Theta \in \Re^L$ is the constant target OLA vector, $\varphi(E): \Re^n \to \Re^L$ is a linearly independent basis vector which satisfies $\varphi(0) = 0$. It has been shown [13] that by increasing the dimension of the regression vector $\varphi(E)$, the OLA reconstruction error decreases $\varepsilon(E)$. It is assumed that the cost function can be represented by (22) with $\varepsilon(E) = 0$ i.e. it is LIP in Θ. The target OLA vector is assumed to be bounded above according to $\|\Theta\| \le \Theta_M$ [3]. The gradient of the OLA cost function (22) is written as

$$\partial V(E) / \partial E = V_E(E) = \nabla_E^T \varphi(E)\Theta . \tag{56}$$

Now, using (23), the optimal control (14) and HJB equation (16) are rewritten as

$$U^*(E) = -\frac{1}{2} R^{-1} G(E)^T \nabla_E^T \varphi(E)\Theta \tag{57}$$

and

$$H^*(E,\Theta) = Q(E) + \Theta^T \nabla_E \varphi(E) F(E) - \frac{1}{4}\Theta^T \nabla_E \varphi(E) D \nabla_E^T \varphi(E)\Theta = 0 \ , \tag{58}$$

where $D = G(E)R^{-1}G(E)^T > 0$ is bounded such that $D_{\min} \le \|D\| \le D_{\max}$ for known constants $D_{\min}$ and $D_{\max}$. Moving on, the estimate of (11) is now written as

$$\hat{V}(E) = \hat{\Theta}^T \varphi(E) \quad , \tag{59}$$

where $\hat{\Theta}$ is the estimate of the target parameter vector Θ. Similarly, the estimate of the optimal control (14) is written in terms of $\hat{\Theta}$ as

$$\hat{U}^* = -\frac{1}{2} R^{-1} G(X)^T \nabla_E^T \varphi(E)\hat{\Theta} . \tag{60}$$

It is shown [6] that an initial stabilizing control is not required to implement the proposed scheme in contrast to [11] and [13], which require initial control policies to be

stabilizing. In fact, the proposed OLA parameter tuning law described next ensures that the system states remain bounded and that (28) will become admissible.

Now, using (22), the approximate Hamiltonian can be written as

$$\hat{H}(E,\hat{\Theta}) = Q(E) + \hat{\Theta}^T \nabla_E \varphi(E) \hat{F}(E) - \frac{1}{4} \hat{\Theta}^T \nabla_E \varphi(E) D \nabla_E^T \varphi(E) \hat{\Theta} . \tag{61}$$

Observing the definition of cost function estimation (27) and the Hamiltonian function (29), it is evident that both become zero when $\|E\| = 0$. Thus, once the system states have converged to zero, the cost function estimation can no longer be updated. This can be viewed as a persistency of excitation (PE) requirement for the inputs to the cost function estimator [11], [13]. That is, the system states must be exiting long enough for the OLA to learn the optimal cost function.

Recalling the HJB equation in (16), the OLA estimate $\hat{\Theta}$ should be tuned to minimize $\hat{H}(E,\hat{\Theta})$. However, tuning to minimize $\hat{H}(E,\hat{\Theta})$ alone does not ensure the stability of the nonlinear system (13) during the adaptation process. Therefore, the proposed tuning algorithm is designed to minimize (29) while considering the stability of (13) and written as

$$\dot{\hat{\Theta}} = -\beta_1 \frac{\hat{\sigma}}{(\hat{\sigma}^T \hat{\sigma} + 1)^2} \left(Q(E) + \hat{\Theta}^T \nabla_E \varphi(E) \hat{F}(E) - \frac{1}{4} \hat{\Theta}^T \nabla_E \varphi(E) D \nabla_E^T \varphi(E) \hat{\Theta} \right)$$

$$+ \frac{\beta_2}{2} \Sigma(E,\hat{U}) \nabla_E \varphi(E) D \nabla_E^T \varphi(E) J_{1E}(E) \tag{62}$$

where $\hat{\sigma} = \nabla_E \varphi F(E) - \nabla_E \varphi(E) D \nabla_E^T \varphi(E) \hat{\Theta} / 2$, $\beta_1 > 0$ and $\beta_2 > 0$ are design constants, $J_1(E)$ is a Lyapunov function described in *Lemma 2*. Th term $\Sigma(E,\hat{U})$ used to determine the stability condition of the closed system and it is defined as

$$\Sigma(E,\hat{U}) = \begin{cases} 0 & \text{if} \quad J_{1E}^T \dot{E} = J_{1E}^T (F(E) - \frac{1}{2} G(E) R^{-1} G(E)^T \nabla_E^T \varphi(E) \hat{\Theta}) < 0 \\ 1 & \text{otherwise} \end{cases}. \tag{63}$$

where $J_1(E)$ is a Lyapunov function whose derivative with respect to E is denoted by J_{1E}. The first term in (30) is the portion of the tuning law which seeks to minimize (29) and was derived using a normalized gradient descent scheme with the auxiliary HJB error defined as

$$E_{HJB} = \hat{H}(E,\hat{\Theta})^2 / 2 \ . \tag{64}$$

Meanwhile, the second term in the parameter tuning law (30) is included as it is required in the process of Lyapunov-based stability proof of the overall closed loop system. Moving on, we now form the dynamics of the cost function parameter estimation error $\tilde{\Theta} = \Theta - \hat{\Theta}$. Observing $Q(E) = -\Theta^T \nabla_E \varphi(E) F(E) + \Theta^T \nabla_E \varphi(E) D \nabla_E^T \varphi(E) \Theta / 4$ from (24), the approximate HJB equation (29) can be rewritten in terms of $\tilde{\Theta}$ as

$$\hat{H}(E,\hat{\Theta}) = -\tilde{\Theta}^T \nabla_E \varphi(E) F(E) + \frac{1}{2} \tilde{\Theta}^T \nabla_E \varphi(E) D \nabla_E^T \varphi(E) \Theta$$

$$-\frac{1}{4} \tilde{\Theta}^T \nabla_E \varphi(E) D \nabla_E^T \varphi(E) \tilde{\Theta}. \tag{65}$$

Next, observing $\dot{\tilde{\Theta}} = -\dot{\hat{\Theta}}$ and $\hat{\sigma} = \nabla_E \varphi(E) \dot{E}^* + \nabla_E \varphi(E) D \nabla_E^T \varphi(E) \tilde{\Theta} / 2$ where $\dot{E}^* = F(E) + G(X) U^*$, the error dynamics of (20) are written as

$$\dot{\tilde{\Theta}} = \frac{\beta_1}{\rho^2} \left(\nabla_E \varphi(E) \dot{E}^* + \frac{\nabla_E \varphi(E) D \nabla_E^T \varphi(E) \tilde{\Theta}}{2} \right) \times$$

$$\left(\tilde{\Theta}^T \nabla_E \varphi(E) \dot{E}^* + \frac{1}{4} \tilde{\Theta}^T \nabla_E \varphi(E) D \nabla_E^T \varphi(E) \tilde{\Theta} \right)$$

$$-\frac{\beta_2}{2} \Sigma(E,\hat{U}) \nabla_E \varphi(E) D \nabla_E^T \varphi(E) J_{1E}(E) \tag{66}$$

where $\rho = (\hat{\sigma}^T \hat{\sigma} + 1)$. Next, the stability of the optimal adaptive control scheme is examined along with the stability of the nonlinear system (13). As the next step an identifier is also required and represented in section II equations (13)-(14).

Theorem 2: (Stability of Optimal Adaptive Control Scheme with Partially Unknown Dynamics). Given the nonlinear system (4), (6), and (8) with the target HJB equation (16), and let the tuning law for the internal dynamics and the cost function estimation be given by (17) and (30) respectively. Then, when the design parameter is selected as

$$\beta_2 / \beta_1 > \frac{4096\beta_1}{D_{\min}^4} K^* / \dot{E}_{\min}$$

$$\sigma_m(\Upsilon)\sigma_M^{-1}(\Upsilon) > \frac{5}{2}\beta_1 \left\|\nabla_E \varphi(E)\right\|_{\max}^2 \left\|\Theta\right\|_{\max}^2$$

$$\sigma_m^2(\bar{\Upsilon})\sigma_M^{-2}(\bar{\Upsilon}) > \frac{4096}{D_{\min}^4}\beta_1, \tag{67}$$

Then the closed loop system is globally uniformly stable such that the tracking error E, the internal dynamics parameter error $\tilde{\Lambda}$, and cost function parameter error $\tilde{\Theta}$ converge to zero which also implies that $\hat{V} \to V^*$ and $\hat{U} \to U^*$.

Proof. See the appendix. ■

The block diagram of the proposed state feedback-based optimal adaptive control scheme is shown in Figure 1 where value or policy iterations are not utilized. Only the value function and control inputs are updated once per sampling interval.

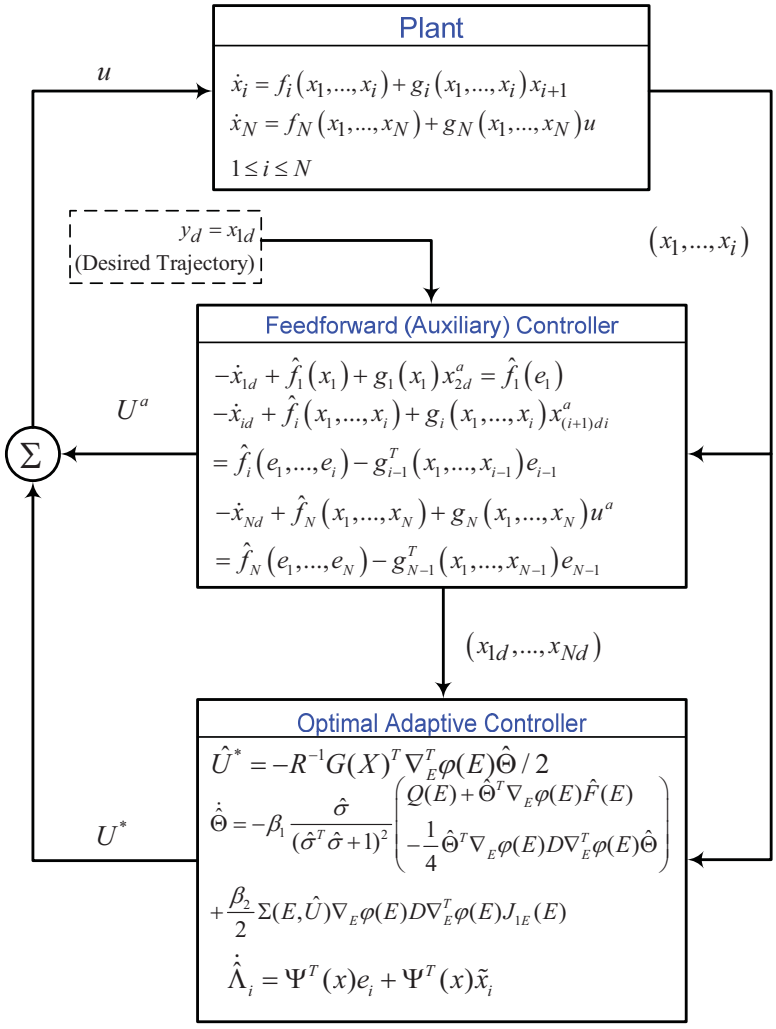

Figure 1. The block diagram of the proposed optimal adaptive with a state feedback approach.

5. OBSERVER BASED OUTPUT FEEDBACK CONTROL

Practically, the states are not measurable in a vast class of nonlinear systems. In this section, we consider the control problem of strict feedback control of the system (2)-(3) where $f_i(.)$ is unknown in parameters and $g_i(.)$ is known, whereas the state vector is not measured and only the output $y = h(x)$ is given. The multi-input multi-output (MIMO) feedback control of strict feedback systems will have to mitigate several challenges and will be relegated for a future publication. For example, selecting different

outputs can change the relative degree of the system which in turn can complicate the process of the controller design. Therefore, we consider the system (1)-(3) to a single-input and single-output (SISO) case. This problem is still difficult as no known output feedback-based optimal control scheme is available in the forward-in-time manner for nonlinear systems, although recently for linear systems some results are achieved [17]. Now, assume that (1)-(3) is represented in a SISO representation i.e. $x_i \in \Re$ and $u \in \Re$. It is shown in [5] that, in this case, there exists a mapping $\zeta = (\zeta_1,...,\zeta_N) = \aleph(x_1,...,x_N)$ that transforms the system (1)-(3) into a new state space representation as

$$\begin{aligned}
\dot{\zeta}_1 &= \zeta_2 + \omega_1(y) \\
\dot{\zeta}_2 &= \zeta_3 + \omega_2(y) \\
&\vdots \\
\dot{\zeta}_{N-1} &= \zeta_N + \omega_{N-1}(y) \\
\dot{\zeta}_N &= \omega_N(y) + b\beta(y)u \\
y &= \zeta_1 = h(x)
\end{aligned} \tag{68}$$

where $\omega_i(y) \in \Re$ are unknown nonlinear functions of the output that can be represented as LIP functions in the following form

$$\omega_i(y) = \Psi(y)\Xi_i . \tag{69}$$

with $\Psi(y)$ is a given regression vector and Ξ_i being the unknown parameters with respect to $\omega_i(y)$. The transformation $\aleph$ exists only when the relative degree of (1)-(3) is equal to N (in SISO case). Assume that $\hat{\Xi}_i$ is an estimation to Ξ_i ,

$\hat{\omega}(y_1) \equiv \Psi(y_1)\hat{\Xi}_i$, and $\tilde{\omega}(y_1) \equiv \Psi(y_1)\left\{\Xi_i - \hat{\Xi}_i\right\} \equiv \Psi(y_1)\tilde{\Xi}_i$ with $\hat{\Xi}_i$ to be the estimation of Ξ_i. To overcome the need for state availability, define the observer dynamics as

$$\dot{\hat{\zeta}} = A\hat{\zeta} + k(y - \hat{y}) + \hat{\omega}(y) + b\beta(y)u + A^T\hat{E} + \lambda b^T c(y - \hat{y})^3$$
$$\hat{y} = c^T\hat{\zeta} \tag{70}$$

where

$$A = \begin{bmatrix} 0 & & I \\ \vdots & \ddots & \\ 0 & \cdots & 0 \end{bmatrix}, \quad b = \begin{bmatrix} 0 \\ \vdots \\ 0 \\ 1 \end{bmatrix}, \quad c = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \quad \omega(y) = \begin{bmatrix} \omega_1(y) \\ \vdots \\ \omega_n(y) \end{bmatrix},$$

with $A_o = A - kc^T$ being Hurwitz, $\lambda > 0$ as a scalar constant , $\hat{E}$ is the estimated tracking error. By defining $\tilde{\zeta} = \zeta - \hat{\zeta}$, the observer error dynamics takes the following form

$$\dot{\tilde{\zeta}} = A_o\tilde{\zeta} + \tilde{\omega}(y_1) - A^T\hat{E}, \tag{71}$$

with Σ being the solution of the Lyapunov function $A_0^T\Sigma + \Sigma A_0 = -\mathrm{M}$ where $\mathrm{M} > 0$. Now, the same backstepping approach of Section II can be applied with the assumption that ζ_i for $i = 2,...,N$ are not measured but estimated using the observer (36). By following the Steps 1 through N, we get

$$\begin{bmatrix} \dot{\hat{e}}_1 \\ \vdots \\ \dot{\hat{e}}_N \end{bmatrix} = \begin{bmatrix} \hat{\omega}_1(\hat{e}_1) \\ \vdots \\ \hat{\omega}_N(\hat{e}_1) \end{bmatrix} + \begin{bmatrix} \tilde{\omega}_1(y_1) \\ \vdots \\ \tilde{\omega}_N(y_1) \end{bmatrix} + \begin{bmatrix} 1 & \cdots & & 0 \\ 0 & \ddots & & \vdots \\ \vdots & & 1 & 0 \\ 0 & \cdots & 0 & \beta(y) \end{bmatrix} U^* + A\tilde{\zeta}.$$

$$\equiv \hat{\omega}(\hat{e}_1) + \tilde{\omega}(y_1) + \mathrm{B}(y)U^* + A\tilde{\zeta} + \lambda b^T c(y - \hat{y})^3, \tag{72}$$

with $\hat{e}_i = \hat{\zeta}_i - \zeta_{id}$ that implies $e_1 = \hat{e}_1 = y - y_d$ since $\hat{\zeta}_1 = \zeta_1 = y$. With the assumption that $\hat{\Xi}_i$ is an estimation to Ξ_i, we have $\hat{\omega}(\hat{e}_1) \equiv \{\Psi(y_1) - \Psi(y_{1d})\}\hat{\Xi}_i$ and $\tilde{\omega}(y_1) \equiv \Psi(y_1)\{\Xi_i - \hat{\Xi}_i\}$. The desired trajectory for the feedforward controller is designed as

$$
\begin{aligned}
&-\dot{y}_{1d} + \hat{\omega}_1(y) + \hat{\zeta}^a_{2d} = \hat{\omega}_1(\hat{y}_1) - \hat{\omega}_1(\hat{y}_{1d}) \equiv \hat{\omega}_1(\hat{e}_1) \\
&\vdots \\
&-\dot{\zeta}_{id} + \hat{\omega}_i(y) + \hat{\zeta}^a_{(i+1)di} = \hat{\omega}_i(\hat{e}_i) - \hat{e}_{i-1} \\
&\vdots \\
&-\dot{\zeta}_{Nd} + \hat{\omega}_N(y) + b\beta(y)u^a = \hat{\omega}_N(\hat{e}_N) - \hat{e}_{N-1}.
\end{aligned} \tag{73}
$$

Equation (39) is identical to the equations derived for the Steps 1 to N in Section II by using the system output y and the estimated states $\hat{\zeta}$ (by the observer) instead of the real value of ζ. As a consequence, an estimation error term $A\tilde{\zeta}$ appears in error dynamics (38). Additionally, in (38) $U_1^* = \begin{bmatrix} \zeta_{2d}^* & \cdots & \zeta_{Nd}^* & u^* \end{bmatrix}$. Moreover, the update law for unknown internal function parameters Π_i is given as

$$
\dot{\hat{\Xi}}_i = \Psi(y_1)(2\Sigma + I)\hat{E} + 2\Psi(y_1)\Sigma\zeta_d. \tag{74}
$$

Theorem 2 will show that the state estimation error $\tilde{\zeta} = \zeta - \hat{\zeta}$ is guaranteed to be stable which is necessary for the overall stability of the closed loop system. Using $r_1(\hat{E}, U^*) = Q_1(\hat{E}) + U_1^{*T} R_1 U_1^*$, where $\hat{E} = \begin{bmatrix} \hat{e}_1, ..., \hat{e}_N \end{bmatrix}^T$ the target HJB equation takes the following form

$$
Q_1(\hat{E}) + V_{1E}^T(\hat{E})\omega(e_1) - \frac{1}{4} V_{1\hat{E}}^T(\hat{E}) B(y) R_1^{-1} B(y)^T V_{1\hat{E}}(\hat{E}) = 0. \tag{75}
$$

with $\hat{E}=\begin{bmatrix}\hat{e}_1 & \cdots & \hat{e}_N\end{bmatrix}^T$, $V_1=\int_t^\infty r_1(\hat{E}(\tau),U^*(\tau))d\tau$ as the cost function, $R_1\in\Re^{N\times N}$ with $R_1>0$ since $m=1$, and $Q_1(\hat{E})$ is a positive semi-definite function of $\hat{E}$. Therefore, the optimal controller for this case can be represented as

$$U_1^*(\hat{E})=-\frac{1}{2}R_1^{-1}\mathrm{B}\left(y\right)^T V_{1\hat{E}}^*(\hat{E}). \tag{76}$$

Now, consider an adaptive representation as

$$V_1(\hat{E})=\Theta_1^T\varphi_1(\hat{E}), \tag{77}$$

and update law as

$$\dot{\hat{\Theta}}_1=-\gamma_1\frac{\hat{\sigma}_1}{\rho_1^2}\left(Q_1(\hat{E})+\hat{\Theta}_1^T\nabla_{\hat{E}}\varphi_1(\hat{E})\hat{\omega}(y)-\frac{1}{4}\hat{\Theta}^T\nabla_{\hat{E}}\varphi_1(\hat{E})D_1\nabla_{\hat{E}}^T\varphi_1(\hat{E})\hat{\Theta}_1\right)$$

$$+\frac{\gamma_2}{2}\Sigma(\hat{E},\hat{U})\nabla_{\hat{E}}\varphi(\hat{E})D_1\nabla_{\hat{E}}^T\varphi(\hat{E})J_{1\hat{E}}(\hat{E}) \tag{78}$$

The indicator function $\Sigma(\hat{E},\hat{U})$ is defined as follows

$$\Sigma(\hat{E},\hat{U})=\begin{cases}0 & \text{if}\quad J_{1\hat{E}}^T\dot{\hat{E}}=J_{1\hat{E}}^T(F(\hat{E})-\frac{1}{2}D_1\nabla_{\hat{E}}^T\varphi(\hat{E})\hat{\Theta}_1)<0\\ 1 & \text{otherwise}\end{cases}. \tag{79}$$

Here, $J_1(\hat{E})$ is a positive definite radially unbounded function of $\hat{E}$, $\gamma_1,\gamma_2>0$ are real design parameters, Θ_1 is the target parameter and $\varphi_1(\hat{E})$ the basis function for the estimation of $V_1(\hat{E})$. Moreover,

$$\hat{\sigma}_1=\nabla_{\hat{E}}\varphi_1\omega\left(\hat{e}_1\right)-\nabla_{\hat{E}}\varphi_1(\hat{E})D_1\nabla_{\hat{E}}^T\varphi_1(\hat{E})\hat{\Theta}_1/2, \tag{80}$$

with $D_1=B(y)R_1^{-1}B(y)^T>0$ where $D_{1\min}<\|D_1\|<D_{1\max}$. It is finally assumed that $\left\|\omega\left(\hat{e}_1\right)+B(y)U_1^*\right\|\le\delta_1(E)\equiv\sqrt[4]{K_1^*Q_1(\hat{E})}$ with $K_1^*>0$ and also $\rho\equiv(\hat{\sigma}^T\hat{\sigma}+1)$.

Lemma 2. Consider the tracking dynamics defined in (4), (6), and (8). Assume that the virtual and real control input vector $U_1 = \begin{bmatrix} \zeta_{2d} & \cdots & \zeta_{Nd} & u \end{bmatrix}$ is designed such that $U = U_1^a + U_1^*$ where $U_1^a = \begin{bmatrix} \zeta_{2d}^a & \cdots & \zeta_{Nd}^a & u^a \end{bmatrix}$ being the feedforward control input is designed in (39) and $U_1^* = \begin{bmatrix} x_{2d}^* & \cdots & x_{Nd}^* & u^* \end{bmatrix}$ represents the optimal feedback control input which stabilizes the following system

$$\begin{bmatrix} \dot{\hat{e}}_1 \\ \vdots \\ \dot{\hat{e}}_N \end{bmatrix} = \begin{bmatrix} \hat{\omega}_1(\hat{e}_1) \\ \vdots \\ \hat{\omega}_N(\hat{e}_1) \end{bmatrix} + \begin{bmatrix} 1 & \cdots & & 0 \\ 0 & \ddots & & \vdots \\ \vdots & & 1 & 0 \\ 0 & \cdots & 0 & \beta(y) \end{bmatrix} U_1^* . \tag{81}$$

In this case, optimal control of (35) is equivalent to the optimal controller design (81) under the condition that $\hat{\Xi}_i$ is being updated using the update (74). In the other words, by applying $U = U^a + U^*$ to the system (35), the system dynamics (35) is transformed into the error dynamic system given by (81).

Proof. By choosing $J_2 = \left(\hat{E}^T \hat{E} + tr(\tilde{\Xi}^T \tilde{\Xi}) + \tilde{\zeta}^T \tilde{\zeta} \right) / 2$ with as the Lyapunov candidate, taking derivative and evaluating the system dynamics (37),(38), and (74), we have

$$\begin{aligned} \dot{J}_1 &= \hat{E}^T \dot{\hat{E}} + tr(\tilde{\Xi}^T \dot{\tilde{\Xi}}) + 2\tilde{\zeta}^T \Sigma \dot{\tilde{\zeta}} = \hat{E}^T \dot{\hat{E}} - tr(\tilde{\Xi}^T \dot{\hat{\Xi}}) - 2\tilde{\zeta}^T \Sigma \dot{\hat{\zeta}} \\ &= \hat{E}^T \hat{\omega}(\hat{e}_1) + \hat{E}^T \tilde{\omega}(y_1) + \hat{E}^T \mathrm{B}(y) U^* + \hat{E}^T A \tilde{\zeta} - 2\tilde{\Xi}^T \Psi(y_1) \Sigma \hat{E} - \tilde{\Xi}^T \Psi(y_1) \Sigma \zeta_d \\ &\quad - \tilde{\zeta}^T M \tilde{\zeta} + 2\tilde{\zeta}^T \Sigma \tilde{\omega}(y_1) - 2\tilde{\zeta}^T \Sigma A^T \hat{E} \\ &= \hat{E}^T \hat{\omega}(\hat{e}_1) + \hat{E}^T \tilde{\omega}(y_1) + \hat{E}^T \mathrm{B}(y) U^* + \hat{E}^T A_0 \tilde{\zeta} - \tilde{\Xi}^T \Psi(y_1)(2\Sigma + I)\hat{E} - 2\tilde{\Xi}^T \Psi(y_1) \Sigma \zeta_d \\ &\quad - \tilde{\zeta}^T M \tilde{\zeta} + 2(\hat{E} + \zeta_d)^T \Sigma \tilde{\omega}(y_1) - \tilde{\zeta}^T A^T \hat{E} \\ &= \hat{E}^T \left(\hat{\omega}(\hat{e}_1) + \mathrm{B}(y) U_1^* \right) - \tilde{\zeta}^T M \tilde{\zeta} - \lambda \left(\tilde{\zeta}^T b^T c \tilde{\zeta} \right)^2 \end{aligned} \tag{82}$$

One may easily recognize that from (82), the existence of an optimal controller to make the first term of the RHS of the equation is sufficient to stabilize (81) and send $\tilde{\zeta}$ to zero. This implies that by optimally stabilizing (81), the first term of the RHS of (82) gets negative and $\tilde{\zeta}$ also converge to zero, therefore the stability of (35) is equivalent that of (81) and convergence of $\tilde{\zeta}$. ■

We can now introduce *Theorem 2* under the case where the states are not measured while the output is only available.

Theorem 3: (Output Feedback Optimal Adaptive Control Scheme). Assume that the states of the nonlinear system (1) through (3) are not measurable while the output is only available with $m = 1$. Assume also that x_i are transformed using $\aleph[x_1,...,x_N]$ to ζ which converts the system dynamics into (35). Consider the nonlinear system (35), the observer (36), the target HJB equation (40), the unknown parameter update law (74), and the tuning law for cost function estimation (43) with

$$\gamma_2 / \gamma_1 > \frac{4096\gamma_1}{D_{1\min}^4} K_1^* / \dot{\hat{E}}_{\min}$$

$$\sigma_m(M)\sigma_M^{-1}(M) > \frac{5}{2}\gamma_1 \left\| \nabla_{\hat{E}} \varphi_1(\hat{E}) \right\|_{\max}^2 \left\| \Theta_1 \right\|_{\max}^2$$

$$\lambda > \frac{4096}{D_{1\min}^4}\gamma_1 . \tag{83}$$

The closed system tracking error $\hat{E}$ and the observer error $\tilde{\zeta}$ are asymptotically stable and the cost function parameter estimation error $\tilde{\Theta}_1$ converges to zero ($\hat{V} \to V^*$ and $\hat{U} \to U^*$) if the input is persistently exciting.

Proof. Refer to the appendix ■

The block diagram of the proposed output feedback-based optimal control scheme is shown in Figure 2 where value and policy iterations are not utilized. Only the value function and control input are updated at the sampling interval. The interesting point of this approach is now revealed in this figure where the observer is providing the parameters ζ when it is applied to the real system. This means that by guaranteeing the existence of $\aleph(X)$, the user does not need the system representation (35).

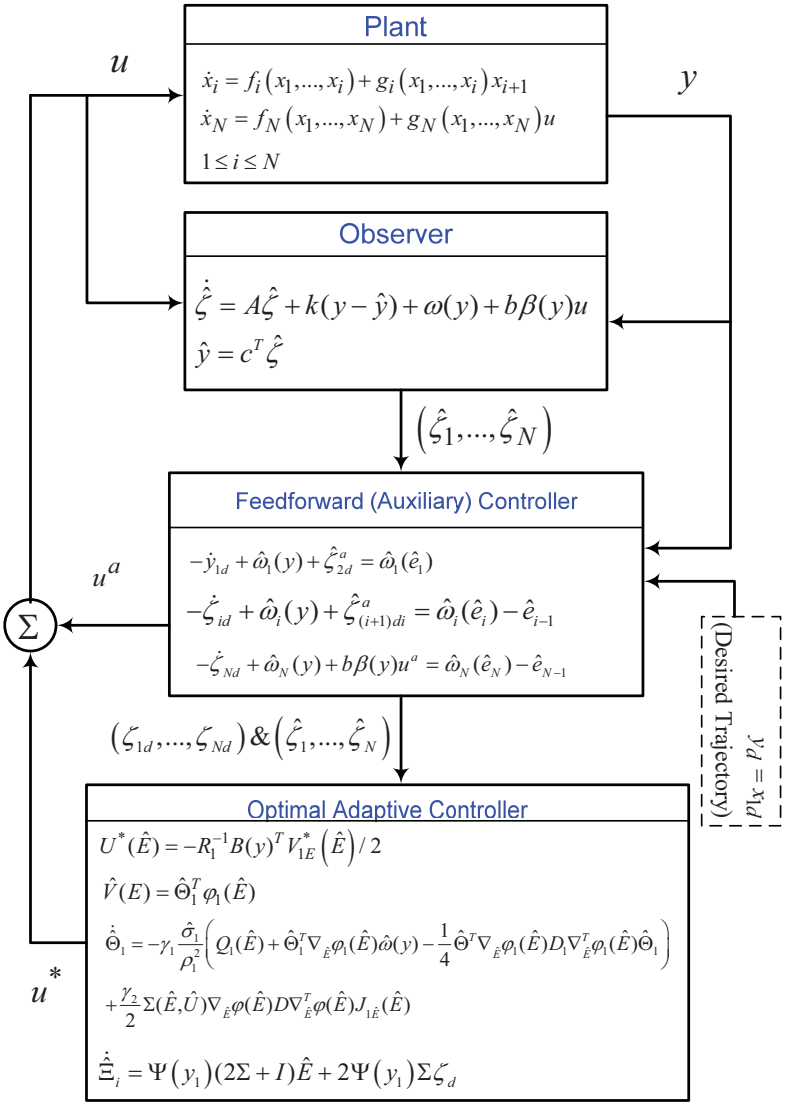

Figure 2. The block diagram of the proposed optimal adaptive with an output feedback approach.

6. NUMERICAL RESULTS

In this section, first a MIMO system is considered and a state feedback optimal approach is designed and verified in simulation. Subsequently, the output feedback-based optimal scheme is evaluated in another example.

6.1. Optimal Adaptive Control of a MIMO Affine System with Unknown Internal Dynamics

Consider an affine system represented as follows

$$\begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \end{bmatrix} = f(x) + g(x)\begin{pmatrix} u_1 \\ u_2 \end{pmatrix} \quad \text{with}$$

$$f(x) = \begin{bmatrix} x_2 \\ -x_1\left(\frac{\pi}{2} + \tan^{-1}(5x_1)\right) - \frac{5x_1^2}{2(1+25x_1^2)} + 4x_2 \end{bmatrix} \text{and}$$

$$g(x) = \begin{bmatrix} 1 & 0 \\ 0 & 3 \end{bmatrix} \tag{84}$$

where $(x_1, x_2) \in \Re^2$ are the states of the system, $f(x)$ is assumed to be an unknown function, and $g(x)$ is given to the control scheme. The basis function for the estimation of $f(x)$ is chosen as $\psi(x) = [x_1, x_2, x_1^2, x_2^2, x_1x_2, x_1^3, x_2^3, x_1^2x_2, x_1x_2^2]$ and therefore $\lambda \in \Re^{9\times 2}$. The cost function estimation basis function is chosen as $\varphi(x) = [x_1^2, x_2^2, x_1x_2, x_1^3, x_2^3, x_1^2x_2, x_1x_2^2, x_1^4, x_2^4, x_1^2x_2^2, x_1^2x_2^2]$ and therefore, $\vartheta \in \Re^{11\times 1}$. It is assumed that the system is initiated on $(x_1, x_2)\big|_{t=0} = (10,10)$ and it is going to be optimally stabilized to the zero while the internal dynamics is unknown. Moreover, it is obvious that finding k^* is difficult, therefore, in order to make (45) hold, we choose $\alpha_1 \ll \alpha_2$, $\alpha_1 \ll \sigma_m(p)\sigma_M^{-1}(p)$. In this case $\alpha_1 = 0.007$, $\alpha_2 = 0.05$, and $p = diag(0.1, 0.1)$ work properly.

The control scheme of section III is applied to the system (84) using MATLAB. It is obvious that the plant (84) is unstable and therefore, not only the controller should optimally stabilize the system, it should also be able to learn the unknown internal dynamics $f(x)$. Figure 3 shows the convergence of the states of the system and Figure 4 illustrates the applied control inputs with respect to the state trajectory of the Figure 1. Figures 5 and 6 illustrate the convergence of the weights of the cost function $\vartheta(t)$ and the internal dynamics $\lambda(t)$. Finally, Figure 7 show that the Hamiltonian $H(x,u)$ converges in a short time and therefore, after about 0.5 seconds, the applied control signal is optimal.

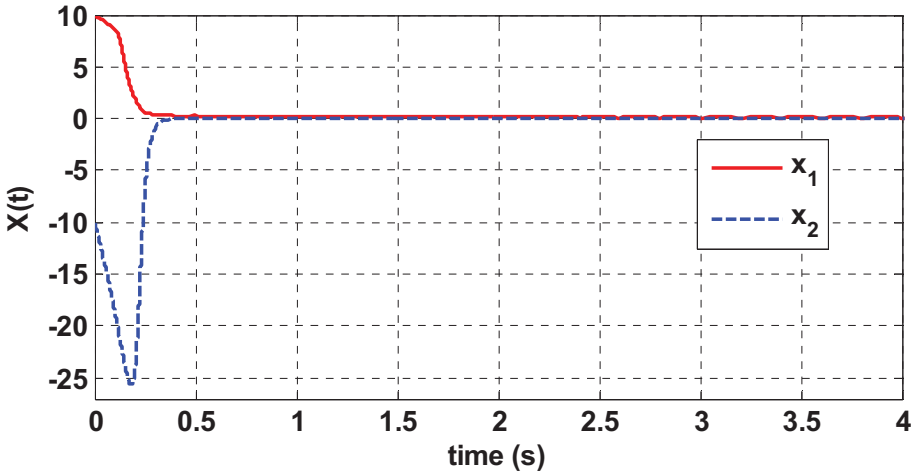

Figure 3. Convergence of the states with the optimal adaptive scheme.

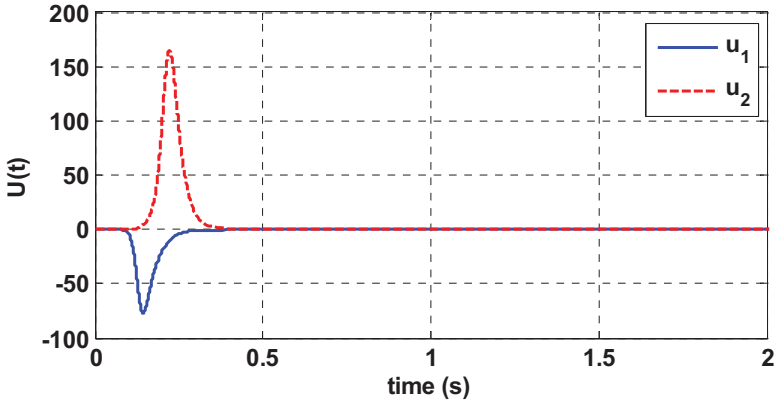

Figure 4. The applied control input.

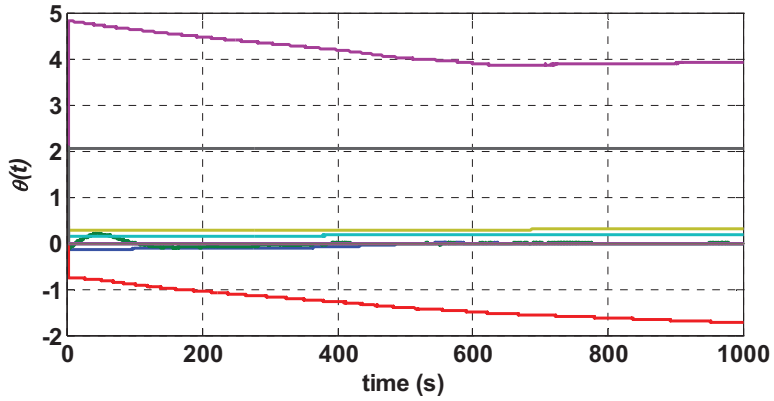

Figure 5. Convergence of the cost function weights $\vartheta(t)$.

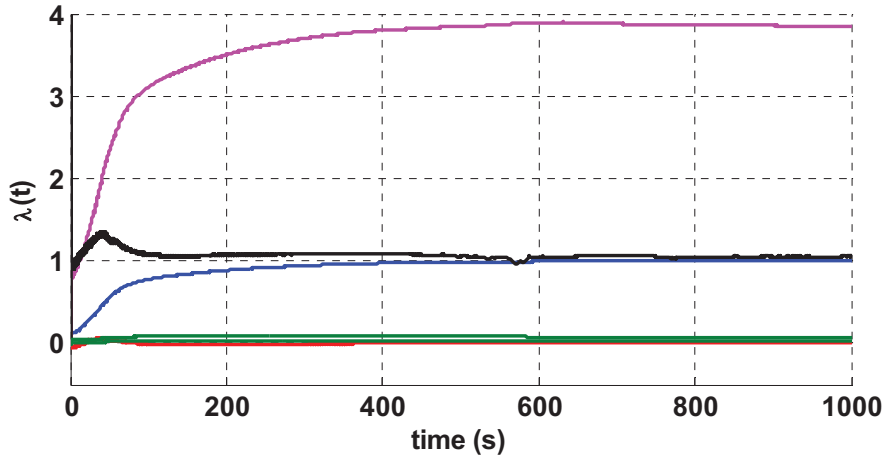

Figure 6. Convergence of the internal dynamics estimation weights $\lambda(t)$.

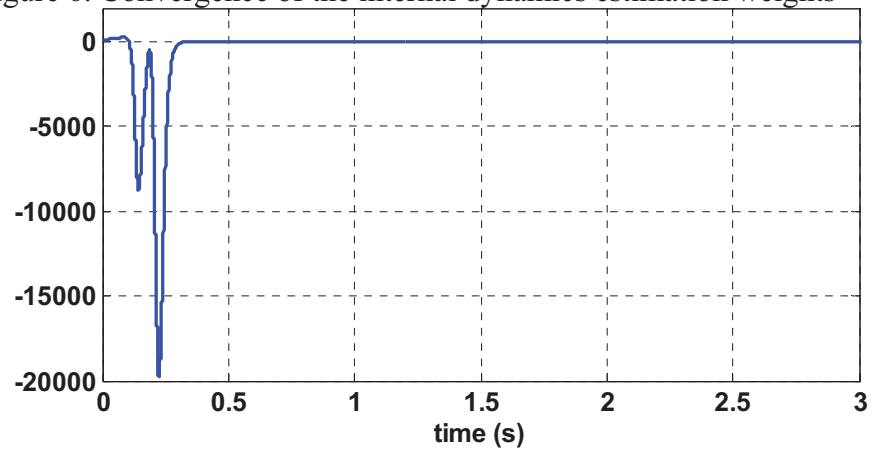

Figure 7. Hamiltonian convergence.

6.2. *Optimal Adaptive Control of a MIMO Strict Feedback System with Unknown Internal Dynamics*

Consider the following nonlinear system in the form of (1)-(2) respectively as

$$\begin{bmatrix}\dot{x}_1\\ \dot{x}_2\end{bmatrix} = f_1(x) + g_1(x)\begin{pmatrix}z_1\\ z_2\end{pmatrix} \quad \text{with } f_1(x) = \begin{bmatrix} x_2 \\ -x_1\left(\frac{\pi}{2} + \tan^{-1}(5x_1)\right) - \frac{5x_1^2}{2(1+25x_1^2)} + 4x_2 \end{bmatrix}$$

$$\text{and } g_1(x) = \begin{bmatrix}1 & 0\\ 0 & 3\end{bmatrix} \tag{85}$$

$$\begin{bmatrix}\dot{z}_1\\ \dot{z}_2\end{bmatrix} = f_2(x,z) + g_2(x,z)u \text{ with } f_2(x,z) = \begin{bmatrix}-4z_1 + x_2^2 - 2x_1\\ -3z_2 + 2x_2^2 - x_1\end{bmatrix} \text{ and}$$

$$g_2(x,z) = \begin{bmatrix}\left(1+z_1^2\right)^{-1} & 0\\ 0 & 1+\frac{1}{2}\cos\left(z_1 + x_1\right)\end{bmatrix} \tag{86}$$

Using the cost function (12) with $Q(E) = E^T E$ and $R = 1$, the basis vector for the SOLA-based scheme implementation is selected as

$\varphi(E) = \begin{bmatrix} x_{e1} & x_{e2} & x_{e1}x_{e2} & x_{e1}^2 & x_{e2}^2 & x_{e1}^2\tan^{-1}(5x_{e1}) & x_{e1}^3 \\ z_{e1} & z_{e2} & z_{e1}z_{e2} & z_{e1}^2 & z_{e2}^2 & z_{e1}^2\tan^{-1}(5z_{e1}) & z_{e1}^3 \end{bmatrix}^T$ where, $x_{e1} = x_1 - x_{1d}$, $x_{e2} = x_2 - x_{2d}$, $z_{e1} = z_1 - z_{1d}$, and $z_{e2} = z_2 - z_{2d}$. Moreover, it is obvious that finding k^* is difficult, therefore, in order to make (67) hold, we choose $\alpha_1 \ll \alpha_2$, $\alpha_1 \ll 1$, $\alpha_1 < \alpha_\Lambda$. In this case $\alpha_1 = 0.007$, $\alpha_2 = 0.05$, and $\alpha_\Lambda = 0.1$ work properly. The initial conditions of the system states were taken as $\begin{bmatrix} x_1 & x_2 & z_1 & z_2 \end{bmatrix}^T = [10 \quad -5 \quad 2 \quad 2]^T$ while all NN weights were initialized to zero. That is, no initial stabilizing control was utilized for implementation of this online design for the nonlinear system. Furthermore, it is desired that the output track $X_d = \begin{bmatrix}\sin(t/50) & \sin(t/40)\end{bmatrix}^T$ as the desired trajectory.

Figures 8 and 9 illustrate the closed loop behavior while tracking the desired trajectory. It is apparent from the figures that the optimal controller takes around 100

seconds to make the tracking error asymptotically zero. Reminding that the system (85) is unstable and the system tends to diverge if enough energy is not applied, having a longer period of time to make tracking error go to zero becomes inevitable where the optimal controller tries to spend less energy. In Figure 10, the desired trajectory is designed as a virtual controller to make the output track the desired output. Convergence of the unknown weights of the cost function $\hat{\Theta}$ is illustrated in the Figure 11. Figures 12 and 13 represent the convergence of the unknown parameters of estimated functions $\hat{f}_1(x)$ and $\hat{f}_2(x,z)$ respectively. Observing the Figure 14, one can realize that the Hamiltonian in (14) converges to zero after 50 seconds that implies that the cost function is well estimated and the applied control input is optimal. Finally, Figure 15 shows the control input applied to the system that can be represented as $\hat{U}^* + \hat{U}^a$.

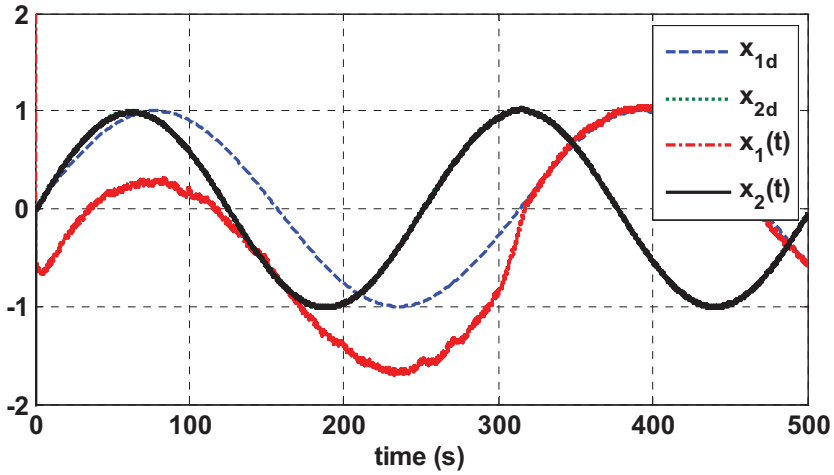

Figure 8. Performance of the output feedback optimal adaptive controller with a desired trajectory $X_d = [\sin(t/50), \sin(t/40)]^T$.

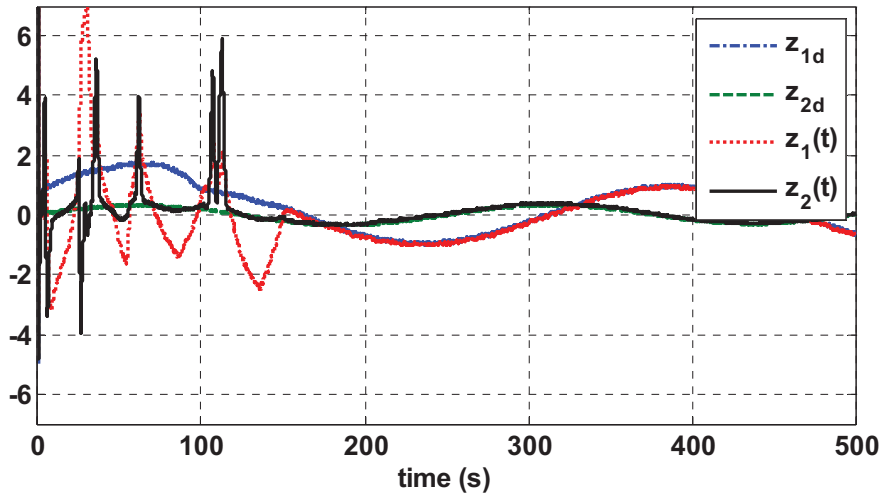

Figure 9. Tracking performance of the virtual controller z .

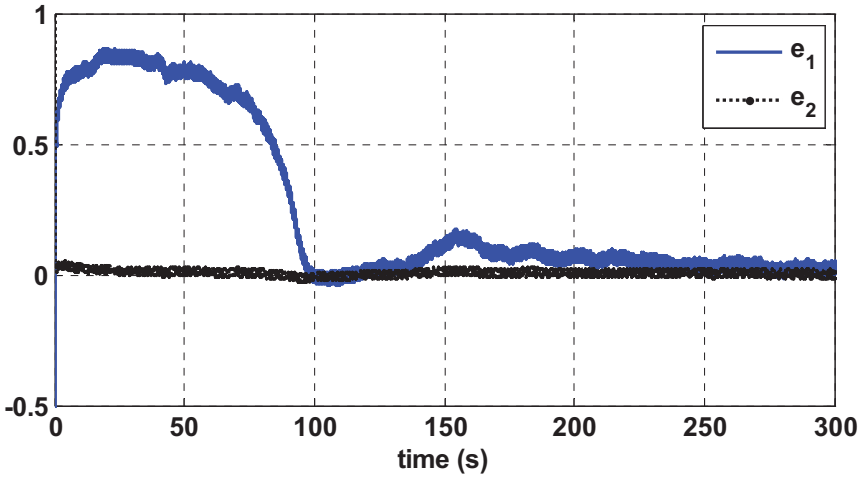

Figure 10. Tracking error.

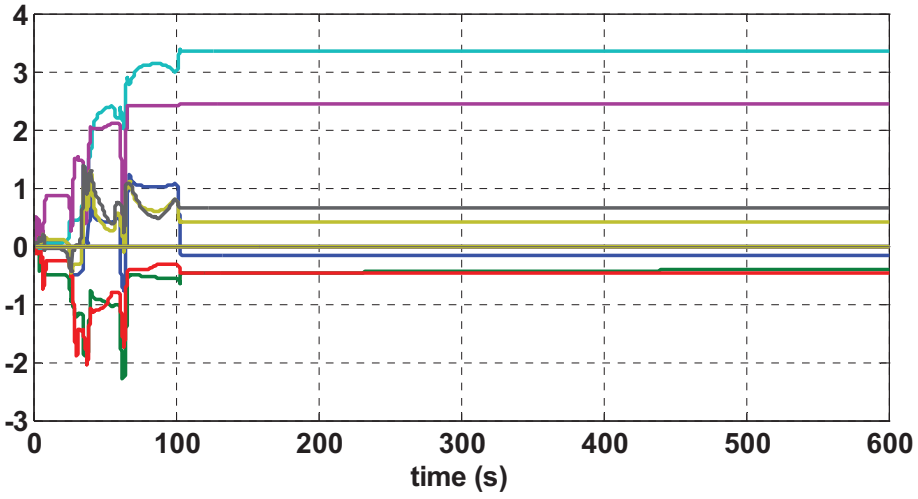

Figure 11. The cost function parameter $\hat{\Theta}$ convergence.

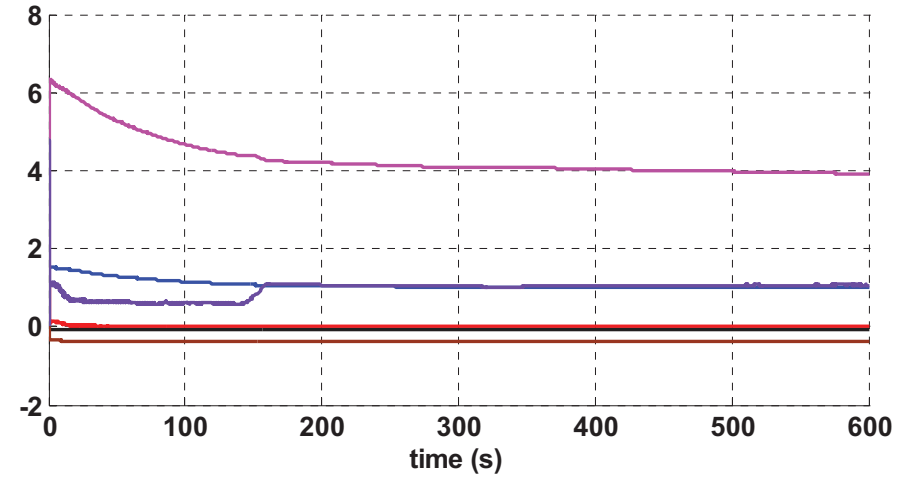

Figure 12. Parameter convergence $\hat{\Lambda}_1$

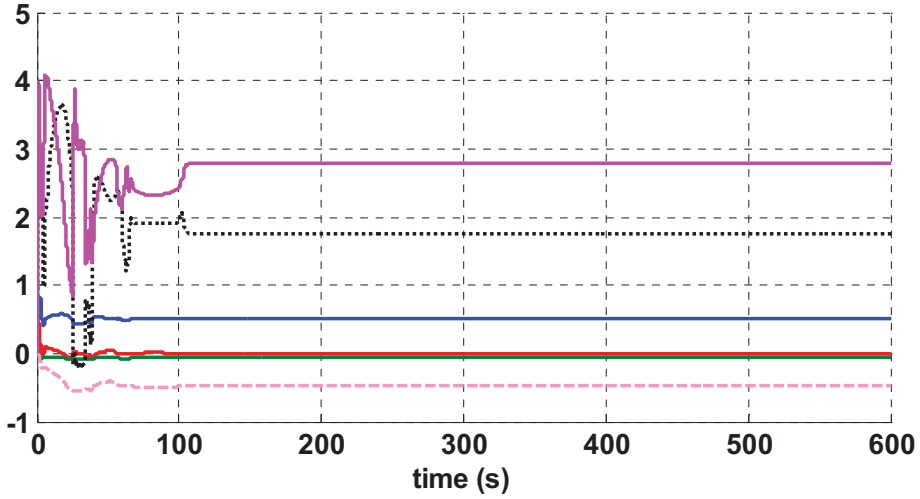

Figure 13. Parameter convergence $\hat{\Lambda}_2$

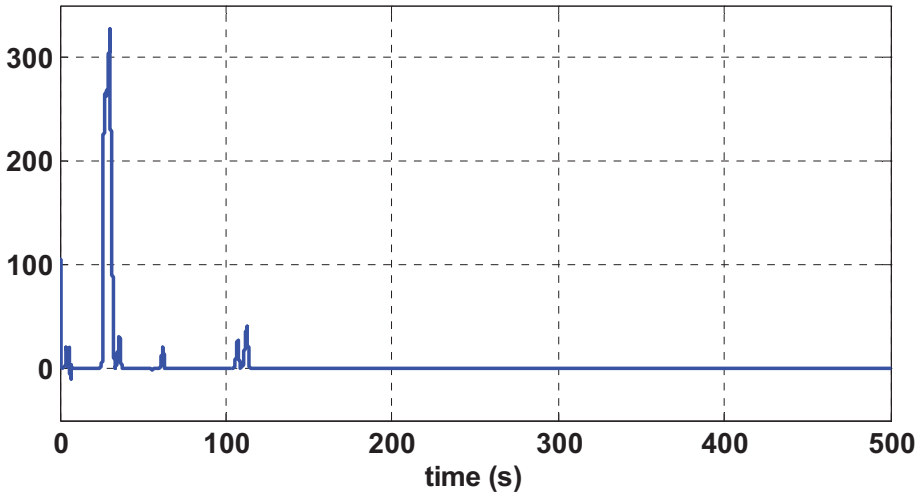

Figure 14. Hamiltonian Convergence $H(E,U)$.

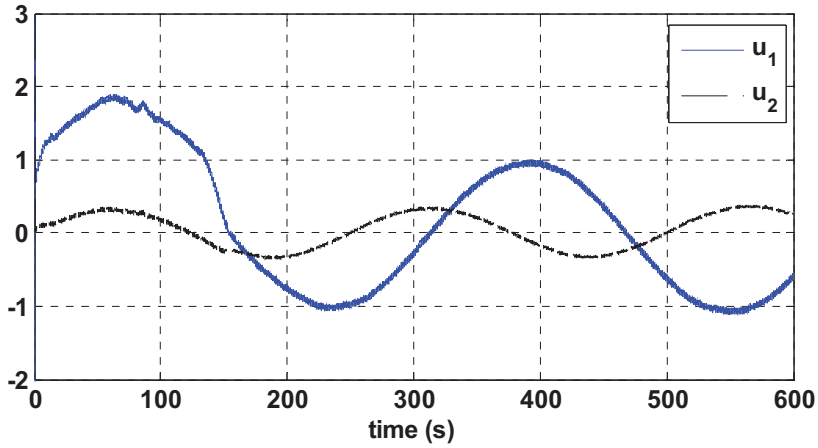

Figure 15. The control input with $\hat{U}^* + \hat{U}^a$.

6.3.Observer Based Online Optimal Control Output Feedback Control

Consider the following nonlinear system in the form of (1)-(3) respectively as

$$\dot{x} = -x\left(\frac{\pi}{2} + \tan^{-1}(5x)\right) - \frac{5x^2}{2(1+25x^2)} + 4x + z \tag{87}$$

$$\dot{z} = 2x^2 - x + \left\{1 + \frac{1}{2}\cos(x)\right\}u\,. \tag{88}$$

$$y = x\,, \tag{89}$$

which is in the form of system (35). Here, we repeat the experiment of the part (b) with the difference that z is not measurable. Using the HJB cost function (40) with $Q(x) = E^T E$ and $R = 1$, the basis vector for the cost function estimation was selected as $\varphi(E) = \left[x_e \;\; x_e^2 \;\; x_e^3 \;\; x_e^2 \tan^{-1}(5x_e) \;\; z_e \; z_e^2 \; z_e^2 \tan^{-1}(5z_e) \; z_e^3\right]^T$ while the tuning parameters were selected as $\alpha_1 = 200$ and $\alpha_2 = 0.01$. Moreover, $x_e = x - x_d$, $z_e = \hat{z} - z_d$, and $A_o = 0.1$. The initial conditions of the system states were taken as $\left[x \;\; z\right]^T = \left[2 \;\; -2\right]^T$ while all NN weights were initialized to zero. That is, no initial stabilizing control was utilized for implementation of this online design for the nonlinear system. Moreover, it is desired that the output track $x_d = \sin(t/50)$ as the desired trajectory.

Figures 16 and 17 illustrate the state convergence trajectory comparing with their desired and observed values. The results show that the tracking errors converge to zero after a transient behavior. Figure 16 illustrates the convergence of the cost function estimation parameters $\hat{\Theta}_1$. The internal dynamics estimated parameters $\hat{\Xi}_1$ and $\hat{\Xi}_2$ are shown in Figure 19 both together. The Hamiltonian estimation error in (40) is depicted in Figure 20 that shows the optimal scheme is able to optimally navigate the tracking system after 20 seconds of convergence. Finally the applied control input is given in the Figure 21.

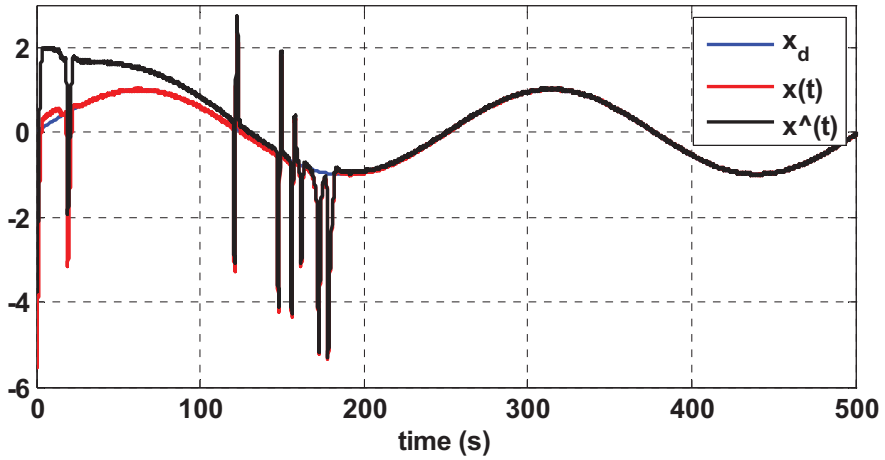

Figure 16. Trajectory $x(t)$ along with its desired and observed values.

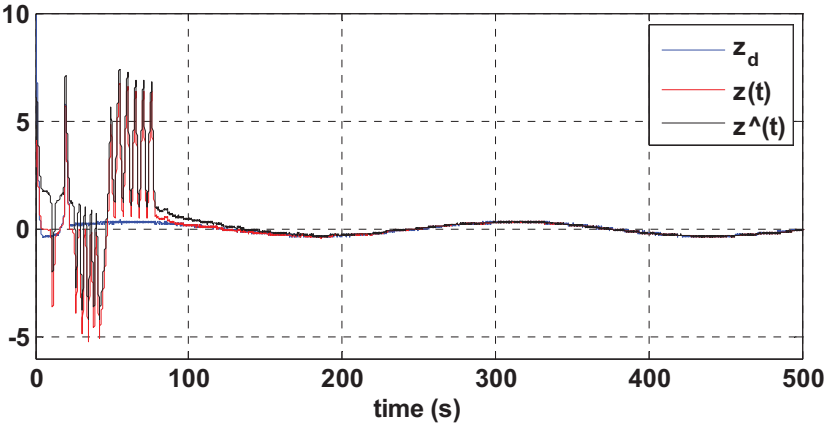

Figure 17. Trajectory $z(t)$ along with its desired and observed values.

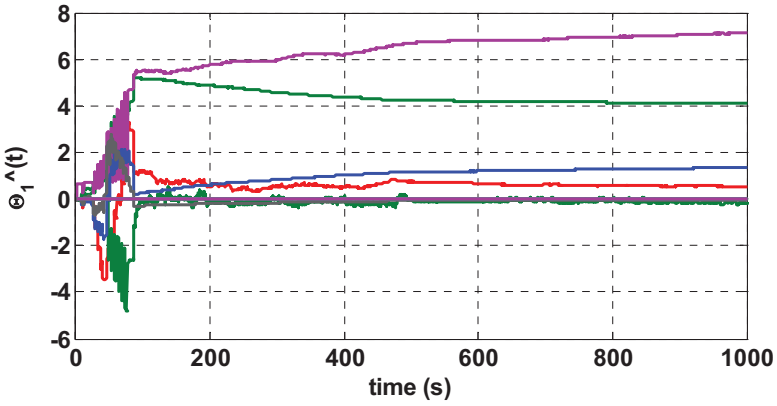

Figure 18. The cost function parameter estimation $\hat{\Theta}_1$.

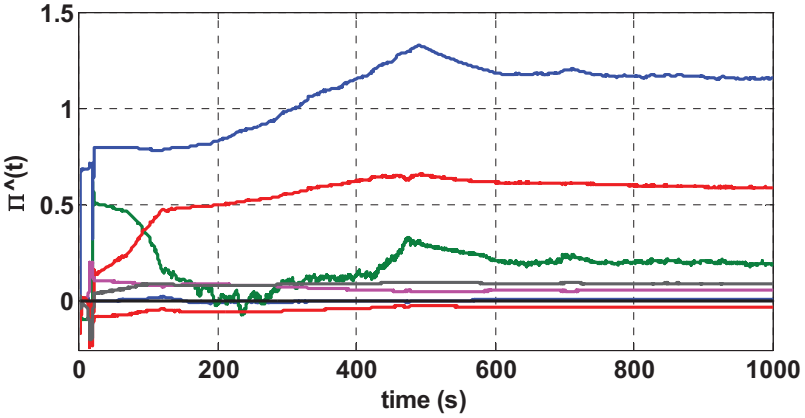

Figure 19. The internal dynamics parameter estimation $\hat{\Pi}(t) = (\hat{\Pi}_1, \hat{\Pi}_2)$.

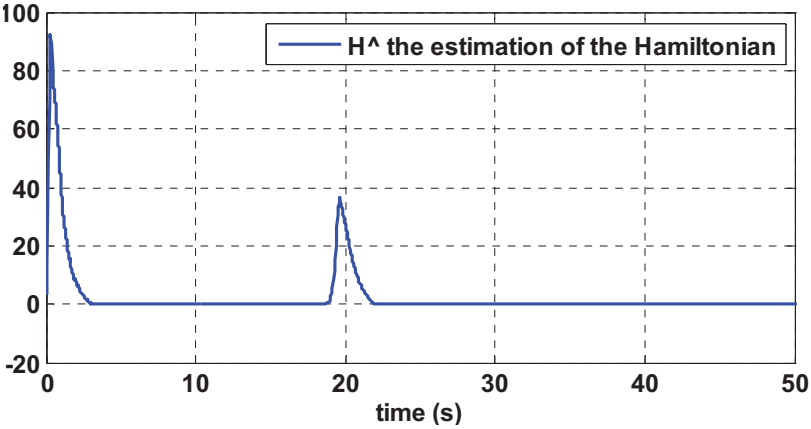

Figure 20. The Hamiltonian estimation error convergence.

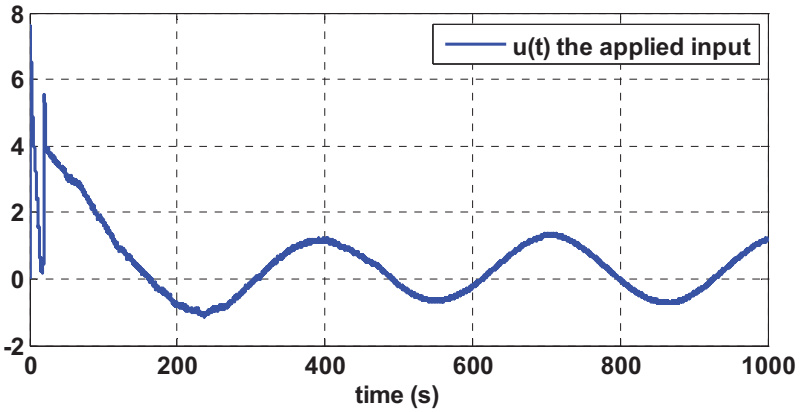

Figure 21. The applied control input $u(t)$.

7. CONCLUSIONS

This work proposes an adaptive optimal scheme for stabilizing nonlinear MIMO systems for affine and strict feedback nonlinear continuous-time systems with unknown internal dynamics. An adaptive approximator is proposed to solve the Hamilton Jacobi-Bellman equation forward-in-time while the other unknown dynamics/states of the system are estimated by state estimator with adaptive elements. Using Lyapunov theorem, this work shows that the problem of online optimal tracking of affine/strict feedback systems can be divided to two steps: a) finding a feedforward tracking controller that is required for all of the tracking problems; and b) designing an optimal controller that optimally stabilizes the tracking error dynamics. Furthermore, it is shown that the problem is solvable while internal dynamics and the states of system are unknown. Numerical results demonstrate the approach to unstable plants whose optimal stabilization/tracking is more challenging.

APPENDIX

Proof of Theorem 1. The Lyapunov function chosen for the stability proof is given as the following:

$$J_{HJB} = \alpha_2 J_1(\chi) + \frac{1}{2}\tilde{\vartheta}^T\tilde{\vartheta} + \frac{1}{2}tr(\tilde{\lambda}^T\tilde{\lambda}) + \frac{1}{2}\tilde{\chi}^T\tilde{\chi} \tag{A.1}$$

By taking the derivative we have:

$$\dot{J}_{HJB} = \alpha_2 J_{1\chi}^T(\chi)\dot{\chi} + \tilde{\vartheta}^T\dot{\tilde{\vartheta}} + tr(\tilde{\lambda}^T\dot{\tilde{\lambda}}) + \tilde{\chi}^T\dot{\tilde{\chi}}\,. \tag{A.2}$$

Now, by applying the system and error dynamics of (20), (41), and (42) we have:

$$\begin{aligned}
\dot{J}_{HJB} &= \alpha_2 J_{1\chi}^T(\chi)\left(\bar{f}(\chi) - \frac{1}{2}\bar{g}(\chi)R^{-1}\bar{g}(\chi)^T\nabla_\chi^T\varphi(\chi)\hat{\vartheta}\right) \\
&+\alpha_1\frac{1}{\rho^2}\left(\nabla_\chi\varphi\left(\dot{\chi}^* - \bar{\bar{f}}(\chi)\right) + \frac{1}{2}\nabla_\chi\varphi(\chi)\Pi\nabla_\chi^T\varphi(\chi)\tilde{\vartheta}\right) \\
&\times\left(-\tilde{\vartheta}\nabla_\chi\varphi(\chi)\left(\dot{\chi}^* - \bar{\bar{f}}(\chi)\right) - \vartheta^T\nabla_\chi\varphi(\chi)\bar{\bar{f}}(\chi) - \frac{1}{4}\tilde{\vartheta}^T\nabla_\chi\varphi(\chi)\Pi\nabla_\chi^T\varphi(\chi)\tilde{\vartheta}\right) \\
&-\frac{\alpha_2}{2}\Sigma(\chi,\hat{\upsilon})\tilde{\vartheta}^T\nabla_\chi\varphi(\chi)\bar{g}(\chi)R^{-1}\bar{g}(\chi)^T J_{1\chi}(\chi) - \tilde{\chi}^T p\tilde{\chi} \\
&= \alpha_2 J_{1\chi}^T(\chi)\left(\bar{f}(\chi) - \frac{1}{2}\bar{g}(\chi)R^{-1}\bar{g}(\chi)^T\nabla_\chi^T\varphi(\chi)\hat{\vartheta}\right) \\
&-\alpha_1\frac{1}{\rho^2}\left(\nabla_\chi\varphi\left(\dot{\chi}^* - \bar{\bar{f}}(\chi)\right) + \frac{1}{2}\nabla_\chi\varphi(\chi)\Pi\nabla_\chi^T\varphi(\chi)\tilde{\vartheta}\right)^2 \\
&-\alpha_1\frac{1}{\rho^2}\left(-\tilde{\vartheta}^T\nabla_x\varphi\left(\dot{\chi}^* - \bar{\bar{f}}(\chi)\right) - \frac{1}{2}\tilde{\vartheta}^T\nabla_\chi\varphi(\chi)\Pi\nabla_\chi^T\varphi(\chi)\tilde{\vartheta}\right) \\
&\times\left(-\vartheta^T\nabla_\chi\varphi(\chi)\bar{\bar{f}}(\chi) + \frac{1}{4}\tilde{\vartheta}^T\nabla_\chi\varphi(\chi)\Pi\nabla_\chi^T\varphi(\chi)\tilde{\vartheta}\right) \\
&-\frac{\alpha_2}{2}\Sigma(\chi,\hat{\upsilon})\tilde{\vartheta}^T\nabla_\chi\varphi(\chi)\bar{g}(\chi)R^{-1}\bar{g}(\chi)^T J_{1\chi}(\chi) - \tilde{\chi}^T p\tilde{\chi} - \left(\tilde{\chi}^T q\tilde{\chi}\right)^2 \\
&= \alpha_2 J_{1\chi}^T(\chi)\left(\bar{f}(\chi) - \frac{1}{2}\bar{g}(\chi)R^{-1}\bar{g}(\chi)^T\nabla_\chi^T\varphi(\chi)\hat{\vartheta}\right) \\
&-\alpha_1\frac{1}{\rho^2}\left(\tilde{\vartheta}^T\nabla_x\varphi\bar{\bar{f}}(\chi) - \frac{1}{2}\tilde{\vartheta}^T\nabla_\chi\varphi(\chi)\Pi\nabla_\chi^T\varphi(\chi)\tilde{\vartheta} - \tilde{\vartheta}^T\nabla_\chi\varphi\dot{\chi}^*\right)^2 \\
&-\frac{\alpha_1}{8}\frac{1}{\rho^2}\left(\tilde{\vartheta}^T\nabla_\chi\varphi(\chi)\Pi\nabla_\chi^T\varphi(\chi)\tilde{\vartheta}\right)^2 \\
&-\alpha_1\frac{1}{\rho^2}\begin{pmatrix}
-\tilde{\vartheta}^T\nabla_\chi\varphi\left(\bar{\bar{f}}(\chi) - \dot{\chi}^*\right)\bar{\bar{f}}(\chi)^T\nabla_\chi^T\varphi(\chi)\vartheta \\
+\frac{1}{4}\tilde{\vartheta}^T\nabla_\chi\varphi\left(\bar{\bar{f}}(\chi) - \dot{\chi}^*\right)\tilde{\vartheta}^T\nabla_\chi\varphi(\chi)\Pi\nabla_\chi^T\varphi(\chi)\tilde{\vartheta} \\
+\frac{1}{2}\tilde{\vartheta}^T\nabla_\chi\varphi(\chi)\Pi\nabla_\chi^T\varphi(\chi)\tilde{\vartheta}\vartheta^T\nabla_\chi\varphi(\chi)\bar{\bar{f}}(\chi)
\end{pmatrix}
\end{aligned}$$

$$-\frac{\alpha_2}{2}\Sigma(\chi,\hat{\upsilon})\tilde{\vartheta}^T\nabla_\chi\varphi(\chi)\bar{g}(\chi)R^{-1}\bar{g}(\chi)^T J_{1\chi}(\chi)-\tilde{\chi}^T p\tilde{\chi}-\left(\tilde{\chi}^T q\tilde{\chi}\right)^2 \qquad \text{(A.3)}$$

Due to the linear system $\dot{\tilde{\chi}}=-p\tilde{\chi}+\bar{\bar{f}}$ given by equation (44), one can easily conclude that $\|\tilde{\chi}\|\geq\sigma_M^{-2}(p)\left\|\bar{\bar{f}}\right\|$. Moreover we know that $\tilde{\chi}^T p\tilde{\chi}\geq\|\tilde{\chi}\|^2\sigma_m(p)$, where $\sigma_m(p)$ and $\sigma_M(p)$ are the minimum and maximum singular value of the matrix p. Thus,

$$-\tilde{\chi}^T p\tilde{\chi}\leq-\sigma_m(p)\|\tilde{\chi}\|^2\leq-\sigma_m(p)\sigma_M^{-1}(p)\left\|\bar{\bar{f}}\right\|^2 \qquad \text{(A.4)}$$

Therefore, by also applying the Cauchy Schwarz inequality yields

$$\begin{aligned}
\dot{J}_{HJB}\leq{}&\alpha_2 J_{1\chi}^T(x)\left(\bar{f}(\chi)-\frac{1}{2}\bar{g}(\chi)R^{-1}\bar{g}(\chi)^T\nabla_\chi^T\varphi(\chi)\hat{\vartheta}\right)\\
&-\alpha_1\frac{1}{2\rho^2}\left(\tilde{\vartheta}^T\nabla_\chi\varphi\left(\bar{\bar{f}}(\chi)-\dot{\chi}^*\right)-\frac{1}{2}\tilde{\vartheta}^T\nabla_\chi\varphi(\chi)\Pi\nabla_\chi^T\varphi(\chi)\tilde{\vartheta}\right)^2\\
&-\frac{\alpha_1}{8}\frac{1}{\rho^2}\left(\tilde{\vartheta}^T\nabla_\chi\varphi(\chi)\Pi\nabla_\chi^T\varphi(\chi)\tilde{\vartheta}\right)^2-\sigma_m(p)\sigma_M^{-1}(p)\left\|\bar{\bar{f}}\right\|^2-\sigma_m^2(q)\sigma_M^{-2}(q)\left\|\bar{\bar{f}}\right\|^4\\
&-\alpha_1\frac{1}{\rho^2}\begin{pmatrix}-\tilde{\vartheta}^T\nabla_\chi\varphi\left(\bar{\bar{f}}(\chi)-\dot{x}^*\right)\bar{\bar{f}}^T(\chi)\nabla_\chi^T\varphi(\chi)\vartheta\\+\frac{1}{4}\tilde{\vartheta}^T\nabla_\chi\varphi\left(\bar{\bar{f}}(\chi)-\dot{\chi}^*\right)\tilde{\vartheta}^T\nabla_\chi\varphi(\chi)\Pi\nabla_\chi^T\varphi(\chi)\tilde{\vartheta}\\+\frac{1}{2}\tilde{\vartheta}^T\nabla_\chi\varphi(\chi)\Pi\nabla_\chi^T\varphi(\chi)\tilde{\vartheta}\bar{\bar{f}}^T(\chi)\nabla_\chi\varphi(\chi)\vartheta\end{pmatrix}\\
&-\frac{\alpha_2}{2}\Sigma(\chi,\hat{\upsilon})\tilde{\vartheta}^T\nabla_\chi\varphi(\chi)\bar{g}(\chi)R^{-1}\bar{g}(\chi)^T J_{1\chi}(\chi)\\
={}&\alpha_2 J_{1\chi}^T(x)\left(\bar{f}(\chi)-\frac{1}{2}\bar{g}(\chi)R^{-1}\bar{g}(\chi)^T\nabla_\chi^T\varphi(\chi)\hat{\vartheta}\right)\\
&-\alpha_1\frac{1}{2\rho^2}\left(\tilde{\vartheta}^T\nabla_\chi\varphi\left(\bar{\bar{f}}(\chi)-\dot{\chi}^*\right)-\frac{1}{2}\tilde{\vartheta}^T\nabla_\chi\varphi(\chi)\Pi\nabla_\chi^T\varphi(\chi)\tilde{\vartheta}\right)^2\\
&-\frac{\alpha_1}{16}\frac{1}{\rho^2}\left(\tilde{\vartheta}^T\nabla_\chi\varphi(\chi)\Pi\nabla_\chi^T\varphi(\chi)\tilde{\vartheta}\right)^2-\sigma_m(p)\sigma_M^{-1}(p)\left\|\bar{\bar{f}}\right\|^2-\sigma_m^2(q)\sigma_M^{-2}(q)\left\|\bar{\bar{f}}\right\|^4\\
&+\alpha_1\frac{1}{\rho^2}\tilde{\vartheta}^T\nabla_\chi\varphi\left(\bar{\bar{f}}(\chi)-\dot{x}^*\right)\bar{\bar{f}}^T(\chi)\nabla_\chi^T\varphi(\chi)\vartheta\\
&-\frac{\alpha_1}{32}\frac{1}{\rho^2}\left(\tilde{\vartheta}^T\nabla_\chi\varphi(\chi)\Pi\nabla_\chi^T\varphi(\chi)\tilde{\vartheta}\right)^2-\alpha_1\frac{1}{4\rho^2}\tilde{\vartheta}^T\nabla_\chi\varphi\left(\bar{\bar{f}}(\chi)-\dot{\chi}^*\right)\tilde{\vartheta}^T\nabla_\chi\varphi(\chi)\Pi\nabla_\chi^T\varphi(\chi)\tilde{\vartheta}
\end{aligned}$$

$$-\frac{\alpha_1}{32}\frac{1}{\rho^2}\left(\tilde{\vartheta}^T\nabla_\chi\varphi(\chi)\Pi\nabla_\chi^T\varphi(\chi)\tilde{\vartheta}\right)^2-\alpha_1\frac{1}{2\rho^2}\tilde{\vartheta}^T\nabla_\chi\varphi(\chi)\Pi\nabla_\chi^T\varphi(\chi)\tilde{\vartheta}\tilde{\bar{f}}^T(\chi)\nabla_\chi\varphi(\chi)\vartheta$$

$$-\frac{\alpha_2}{2}\Sigma(\chi,\hat{\upsilon})\tilde{\vartheta}^T\nabla_\chi\varphi(\chi)\bar{g}(\chi)R^{-1}\bar{g}(\chi)^TJ_{1\chi}(\chi)$$

$$=\alpha_2J_{1\chi}^T(x)\left(\bar{f}(\chi)-\frac{1}{2}\bar{g}(\chi)R^{-1}\bar{g}(\chi)^T\nabla_\chi^T\varphi(\chi)\hat{\vartheta}\right)$$

$$-\alpha_1\frac{1}{2\rho^2}\left(\tilde{\vartheta}^T\nabla_\chi\varphi\left(\tilde{\bar{f}}(\chi)-\dot{\chi}^*\right)-\frac{1}{2}\tilde{\vartheta}^T\nabla_\chi\varphi(\chi)\Pi\nabla_\chi^T\varphi(\chi)\tilde{\vartheta}\right)^2$$

$$-\frac{\alpha_1}{16}\frac{1}{\rho^2}\left(\tilde{\vartheta}^T\nabla_\chi\varphi(\chi)\Pi\nabla_\chi^T\varphi(\chi)\tilde{\vartheta}\right)^2-\sigma_m(p)\sigma_M^{-1}(p)\left\|\tilde{\bar{f}}\right\|^2-\sigma_m^2(q)\sigma_M^{-2}(q)\left\|\tilde{\bar{f}}\right\|^4$$

$$+\alpha_1\frac{1}{\rho^2}\tilde{\vartheta}^T\nabla_\chi\varphi\left(\tilde{\bar{f}}(\chi)-\dot{x}^*\right)\tilde{\bar{f}}^T(\chi)\nabla_\chi^T\varphi(\chi)\vartheta$$

$$-\frac{\alpha_1}{32}\frac{1}{\rho^2}\left(\tilde{\vartheta}^T\nabla_\chi\varphi(\chi)\Pi\nabla_\chi^T\varphi(\chi)\tilde{\vartheta}+4\tilde{\vartheta}^T\nabla_\chi\varphi\left(\tilde{\bar{f}}(\chi)-\dot{\chi}^*\right)\right)^2$$

$$+\frac{\alpha_1}{2}\frac{1}{\rho^2}\left(\tilde{\vartheta}^T\nabla_\chi\varphi\left(\tilde{\bar{f}}(\chi)-\dot{\chi}^*\right)\right)^2$$

$$-\frac{\alpha_1}{32}\frac{1}{\rho^2}\left(\tilde{\vartheta}^T\nabla_\chi\varphi(\chi)\Pi\nabla_\chi^T\varphi(\chi)\tilde{\vartheta}+8\tilde{\bar{f}}^T(\chi)\nabla_\chi\varphi(\chi)\vartheta\right)+2\alpha_1\frac{1}{\rho^2}\left(\tilde{\bar{f}}^T(\chi)\nabla_\chi\varphi(\chi)\vartheta\right)^2$$

$$-\frac{\alpha_2}{2}\Sigma(\chi,\hat{\upsilon})\tilde{\vartheta}^T\nabla_\chi\varphi(\chi)\bar{g}(\chi)R^{-1}\bar{g}(\chi)^TJ_{1\chi}(\chi)\,. \tag{A.5}$$

Now, again applying the Cauchy Schwarz inequality to obtain

$$\dot{J}_{HJB}\le\alpha_2J_{1\chi}^T(x)\left(\bar{f}(\chi)-\frac{1}{2}\bar{g}(\chi)R^{-1}\bar{g}(\chi)^T\nabla_\chi^T\varphi(\chi)\hat{\vartheta}\right)$$

$$-\alpha_1\frac{1}{2\rho^2}\left(\tilde{\vartheta}^T\nabla_\chi\varphi\left(\tilde{\bar{f}}(\chi)-\dot{\chi}^*\right)-\frac{1}{2}\tilde{\vartheta}^T\nabla_\chi\varphi(\chi)\Pi\nabla_\chi^T\varphi(\chi)\tilde{\vartheta}\right)^2$$

$$-\frac{\alpha_1}{16}\frac{1}{\rho^2}\left(\tilde{\vartheta}^T\nabla_\chi\varphi(\chi)\Pi\nabla_\chi^T\varphi(\chi)\tilde{\vartheta}\right)^2-\sigma_m(p)\sigma_M^{-1}(p)\left\|\tilde{\bar{f}}\right\|^2-\sigma_m^2(q)\sigma_M^{-2}(q)\left\|\tilde{\bar{f}}\right\|^4$$

$$+\alpha_1\frac{1}{\rho^2}\left(\tilde{\vartheta}^T\nabla_\chi\varphi\left(\tilde{\bar{f}}(\chi)-\dot{\chi}^*\right)\right)^2+\frac{5}{2}\alpha_1\frac{1}{\rho^2}\left(\tilde{\bar{f}}^T(\chi)\nabla_\chi\varphi(\chi)\vartheta\right)^2$$

$$-\frac{\alpha_2}{2}\Sigma(\chi,\hat{\upsilon})\tilde{\vartheta}^T\nabla_\chi\varphi(\chi)\bar{g}(\chi)R^{-1}\bar{g}(\chi)^TJ_{1\chi}(\chi) \tag{A.6}$$

Now, by applying Cauchy Schwartz again

$$\dot{J}_{HJB} \le \alpha_2 J_{1\chi}^T(x)\left(\overline{f}(\chi)-\frac{1}{2}\overline{g}(\chi)R^{-1}\overline{g}(\chi)^T\nabla_\chi^T\varphi(\chi)\hat{\vartheta}\right)$$
$$-\alpha_1\frac{1}{2\rho^2}\left(\tilde{\vartheta}^T\nabla_\chi\varphi\left(\tilde{\overline{f}}(\chi)-\dot{\chi}^*\right)-\frac{1}{2}\tilde{\vartheta}^T\nabla_\chi\varphi(\chi)\Pi\nabla_\chi^T\varphi(\chi)\tilde{\vartheta}\right)^2$$
$$-\frac{\alpha_1}{32}\frac{1}{\rho^2}\left(\tilde{\vartheta}^T\nabla_\chi\varphi(\chi)\Pi\nabla_\chi^T\varphi(\chi)\tilde{\vartheta}\right)^2-\left(\sigma_m(p)\sigma_M^{-1}(p)-\frac{5}{2}\alpha_1\frac{1}{\rho^2}\left\|\nabla_\chi\varphi(\chi)\vartheta\right\|^2\right)\left\|\tilde{\overline{f}}\right\|^2$$
$$-\sigma_m^2(q)\sigma_M^{-2}(q)\left\|\tilde{\overline{f}}\right\|^4$$
$$-\frac{\alpha_1}{64}\frac{1}{\rho^2}\left\|\tilde{\vartheta}^T\nabla_\chi\varphi(\chi)\right\|^4\Pi_{\min}^2+2\alpha_1\frac{1}{\rho^2}\left\|\tilde{\vartheta}^T\nabla_\chi\varphi(\chi)\right\|^2\left\|\tilde{\overline{f}}(\chi)\right\|^2$$
$$-\frac{\alpha_1}{64}\frac{1}{\rho^2}\left\|\tilde{\vartheta}^T\nabla_\chi\varphi(\chi)\right\|^4\Pi_{\min}^2+2\alpha_1\frac{1}{\rho^2}\left\|\tilde{\vartheta}^T\nabla_\chi\varphi(\chi)\right\|^2\left\|\dot{\chi}^*\right\|^2$$
$$-\frac{\alpha_2}{2}\Sigma(\chi,\hat{\upsilon})\tilde{\vartheta}^T\nabla_\chi\varphi(\chi)\overline{g}(\chi)R^{-1}\overline{g}(\chi)^TJ_{1\chi}(\chi)$$
$$=\alpha_2 J_{1\chi}^T(x)\left(\overline{f}(\chi)-\frac{1}{2}\overline{g}(\chi)R^{-1}\overline{g}(\chi)^T\nabla_\chi^T\varphi(\chi)\hat{\vartheta}\right)+\alpha_1\frac{4096\alpha_1}{\Pi_{\min}^4}\frac{1}{\rho^2}\left\|\dot{\chi}^*\right\|^4$$
$$-\alpha_1\frac{1}{2\rho^2}\left(\tilde{\vartheta}^T\nabla_\chi\varphi\left(\tilde{\overline{f}}(\chi)-\dot{\chi}^*\right)-\frac{1}{2}\tilde{\vartheta}^T\nabla_\chi\varphi(\chi)\Pi\nabla_\chi^T\varphi(\chi)\tilde{\vartheta}\right)^2$$
$$-\frac{\alpha_1}{32}\frac{1}{\rho^2}\left(\tilde{\vartheta}^T\nabla_\chi\varphi(\chi)\Pi\nabla_\chi^T\varphi(\chi)\tilde{\vartheta}\right)^2-\left(\sigma_m(p)\sigma_M^{-1}(p)-\frac{5}{2}\alpha_1\frac{1}{\rho^2}\left\|\nabla_\chi\varphi(\chi)\vartheta\right\|^2\right)\left\|\tilde{\overline{f}}\right\|^2$$
$$-\left(\sigma_m^2(q)\sigma_M^{-2}(q)-\frac{4096\alpha_1}{\Pi_{\min}^4}\frac{1}{\rho^2}\right)\left\|\tilde{\overline{f}}\right\|^4$$
$$-\frac{\alpha_1\Pi_{\min}^2}{64}\frac{1}{\rho^2}\left(\left\|\tilde{\vartheta}^T\nabla_\chi\varphi(\chi)\right\|^2+\frac{64}{\Pi_{\min}^2}\left\|\tilde{\overline{f}}(\chi)\right\|^2\right)^2-\frac{\alpha_1\Pi_{\min}^2}{64}\frac{1}{\rho^2}\left(\left\|\tilde{\vartheta}^T\nabla_\chi\varphi(\chi)\right\|^2+\frac{64}{\Pi_{\min}^2}\left\|\dot{\chi}^*\right\|^2\right)^2$$
$$-\frac{\alpha_2}{2}\Sigma(\chi,\hat{\upsilon})\tilde{\vartheta}^T\nabla_\chi\varphi(\chi)\overline{g}(\chi)R^{-1}\overline{g}(\chi)^TJ_{1\chi}(\chi) \tag{A.7}$$

Now we can write:

$$\dot{J}_{HJB} \le \alpha_2 J_{1\chi}^T(x)\left(\overline{f}(\chi)-\frac{1}{2}\overline{g}(\chi)R^{-1}\overline{g}(\chi)^T\nabla_\chi^T\varphi(\chi)\hat{\vartheta}\right)+\alpha_1\frac{4096}{\Pi_{\min}^4}\frac{1}{\rho^2}\left\|\dot{\chi}^*\right\|^4$$
$$-\frac{\alpha_1}{32}\frac{1}{\rho^2}\left(\tilde{\vartheta}^T\nabla_\chi\varphi(\chi)\Pi\nabla_\chi^T\varphi(\chi)\tilde{\vartheta}\right)^2$$
$$-\left(\sigma_m(p)\sigma_M^{-1}(p)-\frac{5}{2}\alpha_1\frac{1}{\rho^2}\left\|\nabla_\chi\varphi(\chi)\vartheta\right\|^2\right)\left\|\tilde{\overline{f}}\right\|^2-\left(\sigma_m^2(q)\sigma_M^{-2}(q)-\frac{4096\alpha_1}{\Pi_{\min}^4}\frac{1}{\rho^2}\right)\left\|\tilde{\overline{f}}\right\|^4$$
$$-\frac{\alpha_2}{2}\Sigma(\chi,\hat{\upsilon})\tilde{\vartheta}^T\nabla_\chi\varphi(\chi)\overline{g}(\chi)R^{-1}\overline{g}(\chi)^TJ_{1\chi}(\chi) \tag{A.8}$$

Now, we consider the two cases where $\Sigma(\chi,\hat{\upsilon})=0$ and $\Sigma(\chi,\hat{\upsilon})=1$. For the case $\Sigma(\chi,\hat{\upsilon})=0$ we have:

$$\dot{J}_{HJB}\le\alpha_2 J_{1\chi}^{T}(x)\left(\overline{f}(\chi)-\frac{1}{2}\overline{g}(\chi)R^{-1}\overline{g}(\chi)^{T}\nabla_{\chi}^{T}\varphi(\chi)\hat{\vartheta}\right)+\alpha_1\frac{4096}{\Pi_{\min}^{4}}\frac{1}{\rho^{2}}\delta^{4}(x)$$
$$-\frac{\alpha_1}{32}\frac{1}{\rho^{2}}\left(\tilde{\vartheta}^{T}\nabla_{\chi}\varphi(\chi)\Pi\nabla_{\chi}^{T}\varphi(\chi)\tilde{\vartheta}\right)^{2}$$
$$-\left(\sigma_m(p)\sigma_M^{-1}(p)-\frac{5}{2}\alpha_1\frac{1}{\rho^{2}}\left\|\nabla_{\chi}\varphi(\chi)\vartheta\right\|^{2}\right)\left\|\tilde{\overline{f}}\right\|^{2}-\left(\sigma_m^{2}(q)\sigma_M^{-2}(q)-\frac{4096\alpha_1}{\Pi_{\min}^{4}}\frac{1}{\rho^{2}}\right)\left\|\tilde{\overline{f}}\right\|^{4}\qquad(A.9)$$

That implies $\dot{J}_{HJB}\le 0$ under the conditions

$$\alpha_2/\alpha_1>\frac{4096\alpha_1}{\Pi_{\min}^{4}}k^{*}/\dot{x}_{\min}$$
$$\sigma_m(p)\sigma_M^{-1}(p)>\frac{5}{2}\alpha_1\left\|\nabla_{\chi}\varphi(\chi)\right\|_{\max}^{2}\left\|\vartheta\right\|_{\max}^{2}$$
$$\sigma_m^{2}(q)\sigma_M^{-2}(q)>\frac{4096}{\Pi_{\min}^{4}}\alpha_1.\qquad(A.10)$$

For the case that $\Sigma(\chi,\hat{\upsilon})=1$ we have:

$$\dot{J}_{HJB}\le\alpha_2 J_{1\chi}^{T}(x)\left(\overline{f}(\chi)+\overline{g}(\chi)\upsilon^{*}\right)+\alpha_1\frac{4096}{\Pi_{\min}^{4}}\frac{1}{\rho^{2}}\left\|J_{1\chi}(x)\right\|$$
$$-\frac{\alpha_1}{32}\frac{1}{\rho^{2}}\left(\tilde{\vartheta}^{T}\nabla_{\chi}\varphi(\chi)\Pi\nabla_{\chi}^{T}\varphi(\chi)\tilde{\vartheta}\right)^{2}$$
$$-\left(\sigma_m(p)\sigma_M^{-1}(p)-\frac{5}{2}\alpha_1\frac{1}{\rho^{2}}\left\|\nabla_{\chi}\varphi(\chi)\vartheta\right\|^{2}\right)\left\|\tilde{\overline{f}}\right\|^{2}-\left(\sigma_m^{2}(q)\sigma_M^{-2}(q)-\frac{4096\alpha_1}{\Pi_{\min}^{4}}\frac{1}{\rho^{2}}\right)\left\|\tilde{\overline{f}}\right\|^{4}\qquad(A.11)$$

Therefore we have [6]

$$\dot{J}_{HJB}\le-\alpha_2 Q_{\min}\left\|J_{1\chi}(x)\right\|^{2}+\alpha_1\frac{4096}{\Pi_{\min}^{4}}\frac{1}{\rho^{2}}\left\|J_{1\chi}(x)\right\|-\frac{\alpha_1}{32}\frac{1}{\rho^{2}}\left(\tilde{\vartheta}^{T}\nabla_{\chi}\varphi(\chi)\Pi\nabla_{\chi}^{T}\varphi(\chi)\tilde{\vartheta}\right)^{2}$$
$$-\left(\sigma_m(p)\sigma_M^{-1}(p)-\frac{5}{2}\alpha_1\frac{1}{\rho^{2}}\left\|\nabla_{\chi}\varphi(\chi)\vartheta\right\|^{2}\right)\left\|\mu(\chi)\tilde{\lambda}\right\|^{2}$$
$$-\left(\sigma_m^{2}(q)\sigma_M^{-2}(q)-\frac{4096\alpha_1}{\Pi_{\min}^{4}}\frac{1}{\rho^{2}}\right)\left\|\mu(\chi)\tilde{\lambda}\right\|^{4}\qquad(A.12)$$

That implies $\left\|J_{1\chi}(x)\right\|$ converges in a bound such that

$$\left\|J_{1\chi}(\chi)\right\| \le \frac{\alpha_1}{\alpha_2 Q_{\min}} \frac{4096\alpha_1}{\Pi_{\min}^4} \frac{1}{\rho^2} \le \frac{\alpha_1}{\alpha_2 Q_{\min}} \frac{4096\alpha_1}{\Pi_{\min}^4}$$

$$\left\|\tilde{\lambda}\right\| \le \max\left(\frac{1}{\mu_{\min}} \sqrt[2]{\frac{1}{4\alpha_2 Q_{\min}} \left(\frac{4096\alpha_1}{\Pi_{\min}^4}\right)^2 \left(\sigma_m(p)\sigma_M^{-1}(p) - \frac{5}{2}\alpha_1 \left\|\nabla_\chi \varphi(\chi)\vartheta\right\|_{\max}^2\right)^{-1}}, \frac{1}{\mu_{\min}} \sqrt[4]{\frac{1}{4\alpha_2 Q_{\min}} \left(\frac{4096\alpha_1}{\Pi_{\min}^4}\right)^2 \left(\sigma_m^2(q)\sigma_M^{-2}(q) - \frac{4096\alpha_1}{\Pi_{\min}^4}\right)^{-1}} \right)$$

$$\left\|\tilde{\vartheta}\right\| \le \sqrt[4]{\frac{1}{4\alpha_2 Q_{\min} \Pi_{\max}^4 \left\|\nabla_\chi \varphi(\chi)\right\|_{\max}^4} \left(\frac{4096\alpha_1}{\Pi_{\min}^4}\right)^2}. \tag{A.13}$$

The results for the case $\Sigma(\chi,\hat{\upsilon}) = 1$ imply that the closed loop system converges to a compact set given by the bounds in (A.13). Moreover, these bounds can be arbitrary made small by choosing proper design parameters α_1, α_2, p, and q. Therefore if $\Sigma(\chi,\hat{\upsilon}) = 1$ occurs, the closed loop system converges to an arbitrary small bound where $\Sigma(\chi,\hat{\upsilon}) = 0$ holds and therefore $\tilde{\vartheta}$ and $\tilde{\lambda}$ will converge to zero by under the case PE condition. When that occurs, the state χ optimally converges to zero. ■

Proof of Corollary 1. The proof is similar to the proof of theorem 1 except it is assumed that $\varepsilon_{HJB} \neq 0$. It can be easily show that by the same Lyapunov function chosen as (A.1) we have:

$$\begin{aligned}
\dot{J}_{HJB} \le{}& \alpha_2 J_{1\chi}^T(x)\left(\overline{f}(\chi) - \frac{1}{2}\overline{g}(\chi)R^{-1}\overline{g}(\chi)^T \nabla_\chi^T \varphi(\chi)\hat{\vartheta}\right) + \alpha_1 \frac{4096}{\Pi_{\min}^4}\frac{1}{\rho^2}\left\|\dot{\chi}^*\right\|^4 \\
&- \frac{\alpha_1}{32}\frac{1}{\rho^2}\left(\tilde{\vartheta}^T \nabla_\chi \varphi(\chi)\Pi \nabla_\chi^T \varphi(\chi)\tilde{\vartheta}\right)^2 \\
&- \left(\sigma_m(p)\sigma_M^{-1}(p) - \frac{5}{2}\alpha_1 \frac{1}{\rho^2}\left\|\nabla_\chi \varphi(\chi)\vartheta\right\|^2\right)\left\|\tilde{\overline{f}}\right\|^2 - \left(\sigma_m^2(q)\sigma_M^{-2}(q) - \frac{4096\alpha_1}{\Pi_{\min}^4}\frac{1}{\rho^2}\right)\left\|\tilde{\overline{f}}\right\|^4 \\
&- \frac{\alpha_2}{2}\Sigma(\chi,\hat{\upsilon})\tilde{\vartheta}^T \nabla_\chi \varphi(\chi)\overline{g}(\chi)R^{-1}\overline{g}(\chi)^T J_{1\chi}(\chi) + \alpha_1 \frac{1}{\rho^2}\eta(\varepsilon)
\end{aligned} \tag{A.14}$$

with $\beta_1 = \nabla\varphi_{\min}^4 \Pi_{\min}^2 / 64$, $\beta_2 = 1024/\Pi_{\min}^2 + 3/2$, and $\eta(\varepsilon) = 64 D_{\max}^4 \varepsilon_M'^4 / D_{\min}^2 + 3(\varepsilon_M'^4 + \varepsilon_M'^4 \Pi_{\max}^2)/2$, and $0 < \nabla\varphi_{\min} \le \left\|\nabla\varphi(x)\right\|$ is ensured by $\left\|x\right\| > 0$ for a constant $\nabla\varphi_{\min}$. Moreover, it is assumed that $\left\|\partial\varepsilon(x)/\partial x\right\| = \left\|\nabla_x \varepsilon(x)\right\| \le \varepsilon_M'$

[6]. The inequality (A.14) implies that under the conditions of (A.10), and the case $\Sigma(\chi,\hat{\upsilon})=0$ the closed loop system converges to the following bounds

$$\left\|J_{1x}\right\| \le \frac{\alpha_1 \eta(\varepsilon)}{\left|\alpha_2 \dot{\chi}_{\min} - \alpha_1 \frac{4096}{\Pi_{\min}^4} k^*\right|} \equiv b_{Jx} \tag{A.15}$$

$$\left\|\tilde{\vartheta}^T\right\| \le \sqrt[4]{\frac{32\eta(\varepsilon)}{\alpha_1 \nabla_x \varphi_{\min}^4 \Pi_{\min}^2}} \equiv b_\vartheta \tag{A.16}$$

$$\left\|\mu(x)\tilde{\lambda}\right\| \le \sqrt{\frac{-\left(\sigma_m(p)\sigma_M^{-1}(p) - \frac{5}{2}\alpha_1 \frac{1}{\rho^2}\left\|\nabla_\chi \varphi(\chi)\vartheta\right\|^2\right) + \sqrt{\left(\sigma_m(p)\sigma_M^{-1}(p) - \frac{5}{2}\alpha_1 \frac{1}{\rho^2}\left\|\nabla_\chi \varphi(\chi)\vartheta\right\|^2\right)^2 + 4\alpha_1\eta(\varepsilon)\left(\sigma_m^2(q)\sigma_M^{-2}(q) - 2\frac{4096\alpha_1}{\Pi_{\min}^4}\right)}}{2\sigma_m^2(q)\sigma_M^{-2}(q) - 2\frac{4096\alpha_1}{\Pi_{\min}^4}}} \equiv b_\lambda \tag{A.17}$$

For the case of $\Sigma(\chi,\hat{\upsilon})=1$, we have

$$\dot{J}_{HJB} \le -\alpha_2 Q_{\min}\left\|J_{1\chi}(x)\right\|^2 + \alpha_1 \frac{4096}{\Pi_{\min}^4}\frac{1}{\rho^2}\left\|J_{1\chi}(x)\right\| - \frac{\alpha_1}{32}\frac{1}{\rho^2}\left(\tilde{\vartheta}^T \nabla_\chi \varphi(\chi)\Pi \nabla_\chi^T \varphi(\chi)\tilde{\vartheta}\right)^2$$

$$-\left(\sigma_m(p)\sigma_M^{-1}(p) - \frac{5}{2}\alpha_1 \frac{1}{\rho^2}\left\|\nabla_\chi \varphi(\chi)\vartheta\right\|^2\right)\left\|\mu(\chi)\tilde{\lambda}\right\|^2 \tag{A.18}$$

That implies the closed loop will converge to the following bounds

$$\left\|J_{1\chi}(\chi)\right\| \le \sqrt{\frac{-\alpha_1 \frac{4096}{\Pi_{\min}^4} + \sqrt{\left(\alpha_1 \frac{4096}{\Pi_{\min}^4}\right)^2 + 4\alpha_2 Q_{\min}\alpha_1\eta(\varepsilon)}}{2\alpha_2 Q_{\min}}} \equiv b_{J\chi}^{'} \tag{A.19}$$

$$\left\|\tilde{\vartheta}^T\right\| \le \sqrt[4]{\frac{32\eta(\varepsilon)}{\alpha_1 \nabla_x \varphi_{\min}^4 \Pi_{\min}^2} + \frac{32}{\alpha_2}\frac{4096}{\Pi_{\min}^4}k^*} \equiv b_\vartheta^{'}$$

$$-\left(\sigma_m^2(q)\sigma_M^{-2}(q) - \frac{4096\alpha_1}{\Pi_{\min}^4}\frac{1}{\rho^2}\right)\left\|\mu(\chi)\tilde{\lambda}\right\|^4 + \alpha_1 \frac{1}{\rho^2}\eta(\varepsilon) \tag{A.20}$$

$$\left\|\mu(x)\tilde{\lambda}\right\| \le \sqrt{\frac{-\left(\sigma_m(p)\sigma_M^{-1}(p)-\frac{5}{2}\alpha_1\frac{1}{\rho^2}\left\|\nabla_\chi\varphi(\chi)\vartheta\right\|^2\right)+\sqrt{\left(\sigma_m(p)\sigma_M^{-1}(p)-\frac{5}{2}\alpha_1\frac{1}{\rho^2}\left\|\nabla_\chi\varphi(\chi)\vartheta\right\|^2\right)^2+4\alpha_1\left(\eta(\varepsilon)+\alpha_1\frac{4096}{\Pi_{\min}^4}k^*\right)\left(\sigma_m^2(q)\sigma_M^{-2}(q)-2\frac{4096\alpha_1}{\Pi_{\min}^4}\right)}}{2\sigma_m^2(q)\sigma_M^{-2}(q)-2\frac{4096\alpha_1}{\Pi_{\min}^4}}} \equiv b_\lambda^{'} \quad \text{(A.21)}$$

Therefore, The overall bounds for the cases $\Sigma(x,\hat{u}_1)=0$ and $\Sigma(x,\hat{u}_1)=1$ are then given by $\left\|J_{1x}(x)\right\| \le \max(b_{J\chi}, b_{J\chi}^{'})$, $\left\|\tilde{\Theta}\right\| \le \max(b_\vartheta, b_\vartheta^{'})$, and $\left\|\mu(x)\tilde{\lambda}\right\| \le \max(b_\lambda, b_\lambda^{'})$. As it is mentioned in the hypothesis, by ε_{HJB} becoming small, the convergence bounds become smaller for the case of $\Sigma(x,\hat{u}_1)=0$, although for the case of $\Sigma(x,\hat{u}_1)=1$, the controller makes the system to converge in the bound where $\Sigma(x,\hat{u}_1)=0$ and therefore by $\varepsilon_{HJB}\to 0$ the cost function estimation and the estimated optimal controller will converge to their ideal values. ■

Proof of Theorem 2. The Lyapunov function chosen for the stability proof is chosen as the following:

$$J_{HJB} = \alpha_2 J_1(E) + \frac{1}{2}\tilde{\Theta}^T\tilde{\Theta} + \frac{1}{2}tr(\tilde{\Lambda}^T\tilde{\Lambda}) + \frac{1}{2}\tilde{X}^T\tilde{X} \quad \text{(A.22)}$$

where, same as Lemma 2, $J_1(E)$ is chosen to be $J_1(E)=E^TE$ for the sake of simplicity. First taking the derivative with respect to time to get

$$\dot{J}_{HJB} = \alpha_2 J_{1E}^T(E)\dot{E} + \tilde{\Theta}^T\dot{\tilde{\Theta}} + tr(\tilde{\Lambda}^T\dot{\tilde{\Lambda}}) + \tilde{X}^T\dot{\tilde{X}} \quad \text{(A.23)}$$

One can easily find out that equation (A.2), along the system trajectories (4), (6), and (8) is equal to $\dot{J}_{HJB}$ along the system (13) and (34). Therefore, the optimal tracking problem of (4), (6), (8) is reduced to optimal stabilization of the system (13).

To begin the proof of the overall stability, observe that if $\|E\| = 0$, then $J_{HJB} = \tilde{\Theta}^T \tilde{\Theta} / 2$ with $\dot{J}_{HJB} = 0$, and the parameter estimation error $\|\tilde{\Theta}\|$ remains constant and bounded [3]. On the other hand, to successfully accomplish the online learned objective, the states are required to satisfy $\|E\| > 0$. Therefore, the remainder of this proof considers the case of $\|E\| > 0$ (i.e. online learning is being performed). Then, substituting the nonlinear dynamics (13) with control input (28) applied along with the weight estimation error dynamics (34) into (31) reveals (by following the similar steps of the theorem 1 proof):

$$\begin{aligned}
\dot{J}_{HJB} &= \beta_2 J_{1E}^T(E)\left(F(E) - \frac{1}{2} G(X) R^{-1} G(X)^T \nabla_E^T \varphi(E) \hat{\Theta} \right) \\
&-\beta_1 \frac{1}{\rho^2}\left(\tilde{\Theta}^T \nabla_E \varphi \tilde{F}(E) - \frac{1}{2} \tilde{\Theta}^T \nabla_E \varphi(E) D \nabla_E^T \varphi(E) \tilde{\Theta} - \tilde{\Theta}^T \nabla_E \varphi \dot{E}^* \right)^2 \\
&-\frac{\beta_1}{8} \frac{1}{\rho^2} \left(\tilde{\Theta}^T \nabla_E \varphi(E) D \nabla_E^T \varphi(E) \tilde{\Theta} \right)^2 \\
&-\beta_1 \frac{1}{\rho^2} \begin{pmatrix} -\tilde{\Theta}^T \nabla_E \varphi\left(\tilde{F}(E) - \dot{E}^* \right) \tilde{F}(E)^T \nabla_E^T \varphi(E) \Theta \\ +\frac{1}{4} \tilde{\Theta}^T \nabla_E \varphi\left(\tilde{F}(E) - \dot{E}^* \right) \tilde{\Theta}^T \nabla_E \varphi(E) D \nabla_E^T \varphi(E) \tilde{\Theta} \\ +\frac{1}{2} \tilde{\Theta}^T \nabla_E \varphi(E) D \nabla_E^T \varphi(E) \tilde{\Theta} \Theta^T \nabla_E \varphi(E) \tilde{F}(E) \end{pmatrix} \\
&-\frac{\beta_2}{2} \Sigma(E, \hat{U}) \tilde{\Theta}^T \nabla_E \varphi(E) G(E) R^{-1} G(E)^T J_{1E}(E) - \tilde{X}^T \Upsilon \tilde{X} - \left(\tilde{X}^T \overline{\Upsilon} \tilde{X} \right)^2
\end{aligned} \tag{A.24}$$

By following the similar steps of the theorem 1 proof, for the case $\Sigma(E, \hat{U}) = 0$ we have:

$$\begin{aligned}
\dot{J}_{HJB} &\le \beta_2 J_{1E}^T(E)\left(F(E) - \frac{1}{2} G(X) R^{-1} G(X)^T \nabla_E^T \varphi(E) \hat{\Theta} \right) + \beta_1 \frac{4096}{D_{\min}^4} \frac{1}{\rho^2} \delta^4(E) \\
&-\frac{\beta_1}{32} \frac{1}{\rho^2} \left(\tilde{\Theta}^T \nabla_E \varphi(E) D \nabla_E^T \varphi(E) \tilde{\Theta} \right)^2 \\
&-\left(\sigma_m(\Upsilon) \sigma_M^{-1}(\Upsilon) - \frac{5}{2} \beta_1 \frac{1}{\rho^2} \left\| \nabla_E \varphi(E) \Theta \right\|^2 \right) \left\| \tilde{F} \right\|^2 - \left(\sigma_m^2(\overline{\Upsilon}) \sigma_M^{-2}(\overline{\Upsilon}) - \frac{4096 \beta_1}{D_{\min}^4} \frac{1}{\rho^2} \right) \left\| \tilde{F} \right\|^4
\end{aligned} \tag{A.25}$$

That implies $\dot{J}_{HJB} \le 0$ under the conditions

$$\beta_2 / \beta_1 > \frac{4096\beta_1}{D_{\min}^4} K^* / \dot{E}_{\min}$$

$$\sigma_m(\Upsilon)\sigma_M^{-1}(\Upsilon) > \frac{5}{2}\beta_1 \left\|\nabla_E \varphi(E)\right\|_{\max}^2 \left\|\Theta\right\|_{\max}^2$$

$$\sigma_m^2(\bar{\Upsilon})\sigma_M^{-2}(\bar{\Upsilon}) > \frac{4096}{D_{\min}^4}\beta_1 . \tag{A.26}$$

For the case that $\Sigma(\chi, \hat{\upsilon}) = 1$ we have:

$$\dot{J}_{HJB} \le \beta_2 J_{1E}^T(E)\left(F(E) + G(X)U^*\right) + \beta_1 \frac{4096}{D_{\min}^4}\frac{1}{\rho^2}\left\|J_{1E}(E)\right\|$$

$$-\frac{\beta_1}{32}\frac{1}{\rho^2}\left(\tilde{\Theta}^T \nabla_E \varphi(E) D \nabla_E^T \varphi(E)\tilde{\Theta}\right)^2$$

$$-\left(\sigma_m(\Upsilon)\sigma_M^{-1}(\Upsilon) - \frac{5}{2}\beta_1\frac{1}{\rho^2}\left\|\nabla_E\varphi(E)\Theta\right\|^2\right)\left\|\tilde{F}\right\|^2 - \left(\sigma_m^2(\bar{\Upsilon})\sigma_M^{-2}(\bar{\Upsilon}) - \frac{4096\alpha_1}{D_{\min}^4}\frac{1}{\rho^2}\right)\left\|\tilde{F}\right\|^4 \tag{A.27}$$

Therefore we have [6]

$$\dot{J}_{HJB} \le -\beta_2 Q_{\min}\left\|J_{1E}(E)\right\|^2 + \beta_1\frac{4096}{D_{\min}^4}\frac{1}{\rho^2}\left\|J_{1E}(E)\right\| - \frac{\beta_1}{32}\frac{1}{\rho^2}\left(\tilde{\Theta}^T\nabla_E\varphi(E)D\nabla_E^T\varphi(E)\tilde{\Theta}\right)^2$$

$$-\left(\sigma_m(\Upsilon)\sigma_M^{-1}(\Upsilon) - \frac{5}{2}\beta_1\frac{1}{\rho^2}\left\|\nabla_E\varphi(E)\Theta\right\|^2\right)\left\|\Psi(E)\tilde{\Lambda}\right\|^2$$

$$-\left(\sigma_m^2(\bar{\Upsilon})\sigma_M^{-2}(\bar{\Upsilon}) - \frac{4096\beta_1}{D_{\min}^4}\frac{1}{\rho^2}\right)\left\|\Psi(E)\tilde{\Lambda}\right\|^4 \tag{A.28}$$

That implies $\left\|J_{1E}(E)\right\|$ converges in a bound such that

$$\left\|J_{1E}(E)\right\| \le \frac{\beta_1}{\beta_2 Q_{\min}}\frac{4096\beta_1}{D_{\min}^4}\frac{1}{\rho^2} \le \frac{\beta_1}{\beta_2 Q_{\min}}\frac{4096\beta_1}{D_{\min}^4}$$

$$\left\|\tilde{\Theta}\right\| \le \sqrt[4]{\frac{1}{4\beta_2 Q_{\min} D_{\max}^4 \left\|\nabla_E\varphi(E)\right\|_{\max}^4}\left(\frac{4096\beta_1}{D_{\min}^4}\right)^2}$$

$$\left\|\tilde{\Lambda}\right\| \le \max\left(\begin{array}{l} \dfrac{1}{\Psi_{\min}}\sqrt{\dfrac{1}{4\beta_2 Q_{\min}}\left(\dfrac{4096\beta_1}{D_{\min}^4}\right)^2\left(\sigma_m(\Upsilon)\sigma_M^{-1}(\Upsilon)-\dfrac{5}{2}\beta_1\dfrac{1}{\rho^2}\left\|\nabla_E\varphi(E)\Theta\right\|_{\max}^2\right)} \\ ,\dfrac{1}{\Psi_{\min}}\sqrt[4]{\dfrac{1}{4\beta_2 Q_{\min}}\left(\dfrac{4096\beta_1}{D_{\min}^4}\right)^2\left(\sigma_m(\bar{\Upsilon})\sigma_M^{-1}(\bar{\Upsilon})-\dfrac{4096\beta_1}{D_{\min}^4}\right)^{-1}} \end{array}\right). \quad \text{(A.29)}$$

The results for the case $\Sigma(E,\hat{U})=1$ imply that the closed loop system converges a compact set bounded by the bounds given in (A.29). Moreover, these bounds can be arbitrary made small by choosing proper design parameters β_1, β_2, Υ, and $\bar{\Upsilon}$. Therefore if $\Sigma(E,\hat{U})=1$ occurs, the closed loop system converges to an arbitrary small bound where $\Sigma(E,\hat{U})=0$ holds and therefore $\tilde{\Theta}$ and $\tilde{\Lambda}$ will converge to zero provided the input satisfied the PE condition. When that occurs, the state E optimally converges to zero. ■

Proof of Theorem 3. Consider the following positive definite Lyapunov candidate

$$J_{HJB}=\alpha_2 J_1(E)+\frac{1}{2}\tilde{\Theta}_1^T\tilde{\Theta}_1+\frac{1}{2}tr(\tilde{\Xi}^T\tilde{\Xi})+\frac{1}{2}tr(\tilde{\zeta}^T\Sigma\tilde{\zeta}) \quad \text{(A.30)}$$

where $J_1(E)=V(E)$ similar to Theorem 1. Using the result of *Lemma 2* and Theorem 1 (A.3) we have:

$$\begin{aligned}
\dot{J}_{HJB} &= \gamma_2 J_{1\hat{E}}^T(\hat{E})\left(\omega(y)-\frac{1}{2}D_1\nabla_{\hat{E}}^T\varphi_1(\hat{E})\hat{\Theta}_1\right) \\
&-\gamma_1\frac{1}{\rho_1^2}\left(\tilde{\Theta}_1^T\nabla_{\hat{E}}\varphi_1(\hat{E})\tilde{\omega}(y)-\frac{1}{2}\tilde{\Theta}_1^T\nabla_{\hat{E}}\varphi_1(\hat{E})D_1\nabla_{\hat{E}}^T\varphi_1(\hat{E})\tilde{\Theta}_1-\tilde{\Theta}_1^T\nabla_E\varphi_1\dot{\hat{E}}^*\right)^2 \\
&-\frac{\gamma_1}{8}\frac{1}{\rho_1^2}\left(\tilde{\Theta}_1^T\nabla_{\hat{E}}\varphi_1(\hat{E})D_1\nabla_{\hat{E}}^T\varphi_1(\hat{E})\tilde{\Theta}_1\right)^2 \\
&-\gamma_1\frac{1}{\rho_1^2}\left(\begin{array}{l} -\tilde{\Theta}_1^T\nabla_{\hat{E}}\varphi_1\left(\tilde{\omega}(y)-\dot{\hat{E}}^*\right)\tilde{\omega}(y)^T\nabla_{\hat{E}}^T\varphi_1(\hat{E})\Theta_1 \\ +\dfrac{1}{4}\tilde{\Theta}_1^T\nabla_{\hat{E}}\varphi\left(\tilde{\omega}(y)-\dot{\hat{E}}^*\right)\tilde{\Theta}_1^T\nabla_{\hat{E}}\varphi_1(\hat{E})D\nabla_{\hat{E}}^T\varphi_1(\hat{E})\tilde{\Theta}_1 \\ +\dfrac{1}{2}\tilde{\Theta}_1^T\nabla_{\hat{E}}\varphi_1(\hat{E})D_1\nabla_{\hat{E}}^T\varphi_1(\hat{E})\tilde{\Theta}_1\Theta_1^T\nabla_{\hat{E}}\varphi_1(\hat{E})\tilde{\omega}(y) \end{array}\right) \\
&-\frac{\gamma_2}{2}\Sigma(\hat{E},\hat{U}_1)\tilde{\Theta}_1^T\nabla_{\hat{E}}\varphi_1(\hat{E})D_1J_{1\hat{E}}(\hat{E})-\tilde{\zeta}^T M\tilde{\zeta}-\lambda\left(\tilde{\zeta}^T b^T c\tilde{\zeta}\right)^2
\end{aligned} \quad \text{(A.31)}$$

By following the similar steps of the Theorem 1 proof, for the case $\Sigma(E,\hat{U})=0$ we have

$$\dot{J}_{HJB} \le \gamma_2 J_{1\hat{E}}^T(\hat{E})\left(\omega(y)-\frac{1}{2}D_1\nabla_{\hat{E}}^T\varphi_1(\hat{E})\hat{\Theta}_1\right)+\gamma_1\frac{4096}{D_{1\min}^4}\frac{1}{\rho_1^2}\delta_1^4(\hat{E})$$
$$-\frac{\gamma_1}{32}\frac{1}{\rho_1^2}\left(\tilde{\Theta}^T\nabla_{\hat{E}}\varphi(\hat{E})D_1\nabla_{\hat{E}}^T\varphi(\hat{E})\tilde{\Theta}\right)^2$$
$$-\left(\sigma_m(M)\sigma_M^{-1}(M)-\frac{5}{2}\gamma_1\frac{1}{\rho^2}\left\|\nabla_{\hat{E}}\varphi(\hat{E})\Theta_1\right\|^2\right)\left\|\tilde{\omega}(y)\right\|^2-\left(\hat{\lambda}-\frac{4096\gamma_1}{D_{1\min}^4}\frac{1}{\rho^2}\right)\left\|\tilde{\omega}(y)\right\|^4 \quad \text{(A.32)}$$

That implies $\dot{J}_{HJB}\le 0$ under the conditions

$$\gamma_2/\gamma_1 > \frac{4096\gamma_1}{D_{1\min}^4}K_1^*/\dot{\tilde{E}}_{\min}$$
$$\sigma_m(M)\sigma_M^{-1}(M) > \frac{5}{2}\gamma_1\left\|\nabla_{\hat{E}}\varphi_1(\hat{E})\right\|_{\max}^2\left\|\Theta_1\right\|_{\max}^2$$
$$\hat{\lambda} > \frac{4096}{D_{1\min}^4}\gamma_1. \quad \text{(A.33)}$$

For the case that $\Sigma(\chi,\hat{\upsilon})=1$ we have:

$$\dot{J}_{HJB} \le \gamma_2 J_{1\hat{E}}^T(\hat{E})\left(\hat{\omega}(\hat{e}_1)+B(y)U_1^*\right)+\beta_1\frac{4096}{D_{1\min}^4}\frac{1}{\rho_1^2}\left\|J_{1\hat{E}}(\hat{E})\right\|$$
$$-\frac{\beta_1}{32}\frac{1}{\rho_1^2}\left(\tilde{\Theta}_1^T\nabla_{\hat{E}}\varphi(\hat{E})D_1\nabla_{\hat{E}}^T\varphi_1(\hat{E})\tilde{\Theta}_1\right)^2$$
$$-\left(\sigma_m(M)\sigma_M^{-1}(M)-\frac{5}{2}\gamma_1\frac{1}{\rho_1^2}\left\|\nabla_{\hat{E}}\varphi_1(\hat{E})\Theta_1\right\|^2\right)\left\|\tilde{\omega}(y)\right\|^2-\left(\hat{\lambda}-\frac{4096\gamma_1}{D_{1\min}^4}\frac{1}{\rho_1^2}\right)\left\|\tilde{\omega}(y)\right\|^4 \quad \text{(A.34)}$$

Therefore we have [6]:

$$\dot{J}_{HJB} \le -\gamma_2 Q_{1\min}\left\|J_{1\hat{E}}(\hat{E})\right\|^2+\gamma_1\frac{4096}{D_{1\min}^4}\frac{1}{\rho_1^2}\left\|J_{1\hat{E}}(\hat{E})\right\|-\frac{\gamma_1}{32}\frac{1}{\rho_1^2}\left(\tilde{\Theta}_1^T\nabla_{\hat{E}}\varphi(\hat{E})D_1\nabla_{\hat{E}}^T\varphi(\hat{E})\tilde{\Theta}_1\right)^2$$
$$-\left(\sigma_m(M)\sigma_M^{-1}(M)-\frac{5}{2}\gamma_1\frac{1}{\rho_1^2}\left\|\nabla_{\hat{E}}\varphi_1(\hat{E})\Theta_1\right\|^2\right)\left\|\Psi(E)\tilde{\Xi}\right\|^2-\left(\hat{\lambda}-\frac{4096\gamma_1}{D_{1\min}^4}\frac{1}{\rho_1^2}\right)\left\|\Psi(E)\tilde{\Xi}\right\|^4 \quad \text{(A.35)}$$

That implies $\left\|J_{1E}(E)\right\|$ converges in a bound such that

$$\left\|J_{1\hat{E}}(\hat{E})\right\| \le \frac{\gamma_1}{\gamma_2 Q_{1\min}}\frac{4096\gamma_1}{D_{1\min}^4}\frac{1}{\rho_1^2} \le \frac{\gamma_1}{\gamma_2 Q_{1\min}}\frac{4096\gamma_1}{D_{1\min}^4}$$

$$\left\|\tilde{\Theta}_1\right\| \le \sqrt[4]{\frac{1}{4\gamma_2 Q_{1\min} D_{1\max}^4 \left\|\nabla_{\hat{E}}\varphi(\hat{E})\right\|_{\max}^4}\left(\frac{4096\gamma_1}{D_{1\min}^4}\right)^2}$$

$$\left\|\tilde{\Xi}\right\| \le \max\left(\begin{array}{l}\dfrac{1}{\Psi_{\min}}\sqrt{\dfrac{1}{4\gamma_2 Q_{1\min}}\left(\dfrac{4096\gamma_1}{D_{1\min}^4}\right)^2\left(\sigma_m(M)\sigma_M^{-1}(M)-\dfrac{5}{2}\gamma_1\dfrac{1}{\rho_1^2}\left\|\nabla_E\varphi(E)\Theta_1\right\|_{\max}^2\right)} \\ ,\dfrac{1}{\Psi_{\min}}\sqrt[4]{\dfrac{1}{4\beta_2 Q_{1\min}}\left(\dfrac{4096\gamma_1}{D_{1\min}^4}\right)^2\left(\lambda-\dfrac{4096\gamma_1}{D_{1\min}^4}\right)^{-1}}\end{array}\right). \quad \text{(A.36)}$$

The results for the case $\Sigma(\hat{E},\hat{U}_1)=1$ imply that the closed loop system converges a compact set bounded by the bounds given in (A.36). Moreover, these bounds can be arbitrary made small by choosing proper design parameters γ_1, γ_2, M, and λ. Therefore if $\Sigma(\hat{E},\hat{U}_1)=1$ occurs, the closed loop system converges to an arbitrary small bound where $\Sigma(\hat{E},\hat{U}_1)=0$ holds and therefore $\tilde{\Theta}_1$ and $\tilde{\Xi}$ will converge provided the input satisfies the PE condition holds. When that occurs, the state $\hat{E}$ optimally converges to zero. Since the estimation parameter error $\tilde{\Xi}$ also converges to zero, the tracking error $y-y_d$ will converge to zero. ■

REFERENCES

[1] H. K. Khalil, *Nonlinear Systems* (3rd Ed). Printice-Hall: 2002.

[2] S. K. Narendra, A. M. Annaswamy, *Stable Adaptive Systems*, Prentice-Hall: Englewood Cliffs, NJ, 1989.

[3] F. L. Lewis, S. Jagannathan, A. Yesildirek, *Neural Network Control of Robot Manipulators and Nonlinear Systems.* Taylor and Francis: Philadelphia, PA, 1999.

[4] S. Jagannathan, Neural Network Control of Nonlinear Discrete-time Systems, CRC Press, 1998.

[5] M. Kristic, I. Kanellakopoulos, and P. Kokotovic, *Nonlinear and Adaptive Control Design*. John Wiley and Sons: 1995.

[6] T. Dierks, S. Jagannathan, "Optimal control of affine nonlinear continuous-time systems," *Proc. of the American Control Conference* 2010, pp. 1568-1573.

[7] Z. H. Li, M. Krstic, "Optimal design of adaptive tracking controllers for non-linear systems," *Automatica* 1997; vol. 33, no.8, pp.1459-1473.

[8] F. L. Lewis and V. L. Syrmos, *Optimal Control* (2^{nd} ed), Wiley: Hoboken, NJ, 1995.

[9] D. Vrabie and F. L. Lewis, "Neural network approach to continuous-time direct adaptive optimal control for partially unknown nonlinear systems," *Neural Networks*, no. 22, pp. 237-246, 2009.

[10] H. Zargarzadeh, S. Jagannathan, J. Drallmeier, "Online near optimal control of unknown nonaffine systems with application to HCCI engines," *Adaptive Dynamic Programming And Reinforcement Learning (ADPRL), IEEE Symposium on*, *2011*.

[11] R. Beard, G. Saridis, and J. Wen, "Improving the performance of stabilizing controls for nonlinear systems," *IEEE Control Systems Magazine*, vol. 16, no. 5, pp. 27-35, 1996.

[12] J. Si, A. G. Barto, W. B. Powell, and D. Wunsch, *Handbook of Learning and Approx. Dynamics Prog.* Wiley: IEEE Press, 2004.

[13] T. Dierks and S. Jagannathan, "Optimal control of affine nonlinear discrete-time systems," *in Proc. of the Mediterranean Conference on Control and Automation*, pp. 1390 – 1395, 2009.

[14] D. Vrabie, O. Pastravanu, M. Abu-Khalaf, and F. L. Lewis, "Adaptive optimal control for continuous-time linear systems based on policy iteration," *Automatica*, vol. 45, no. 2, pp. 477-484, 2009.

[15] D. Vrabie, K. Vamvoudakis, and F. L. Lewis, "Adaptive optimal controllers based on generalized policy iteration in a continuous-time framework". *Proc. of the IEEE Mediterranean Conf. on Control and Automation*, pp. 1402-1409, 2009.

[16] J. Shamma and J. Cloutier, "Existence of SDRE stabilizing feedback," *IEEE Trans. Automat. Contr.*, vol. 48, pp. 513-517, 2003.

[17] C. H. Watkins, *Learning from delayed rewards*. University of Cambridge: PhD Dissertation, 1989.

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

[19] H. Zargarzadeh, T. Dierks, and S. Jagannathan, "State and Output Feedback-based Adaptive Optimal Control of Nonlinear Continuous-time Systems in Strict Feedback Form," to be in *Proc. of the American Control Conference* 2012.

[20] D. Wang, and J. Huang, "Neural network-based adaptive dynamic surface control for a class of uncertain nonlinear systems in strict-feedback form," *Neural Networks, IEEE Transactions on*, vol. 16, no. 1, pp.195-202, 2005.

[21] T. Zhang, S. S. Ge, and C. C. Hang, "Adaptive neural network control for strict-feedback nonlinear systems using backstepping design," *Automatica*, vol. 36, pp. 1835–1846, 2000.

[22] K. G. Vamvoudakis, and F. L. Lewis, "Online actor–critic algorithm to solve the continuous-time infinite horizon optimal control problem" *Automatica*, vol. 46, no. 5, pp. 878–888, 2010.

[23] F. L. Lewis, and K. G. Vamvoudakis, "Reinforcement Learning for Partially Observable Dynamics Process: Adaptive Dynamic Programming Using Measured Output Data," *IEEE Transaction System, Man, and Cybernetics, Part B*, vol. 41, no. 1, pp. 14-251, 2011.

SECTION
2. CONCLUSIONS AND FUTURE WORK

In this dissertation, the adaptive dynamics programming (ADP) based suite of optimal adaptive control schemes are developed for a class of nonlinear discrete and continuous time systems when the plant dynamics are partially or completely unknown. Moreover, in the case of discrete-time nonlinear systems, an extremum seeking method is developed to drive a closed loop system toward a setpoint that provides the best performance. This dissertation establishes the fact that it is possible to simultaneously stabilize an unknown nonlinear system while the control law is adaptively tuned to satisfy the Hamilton Jacobi Bellman (HJB) equation. The main advantage of the proposed work is that it is iteration free since proposed adaptive update laws are only updated once a sampling interval without needing a faster inner loop. The second advantage being that the system dynamics are not needed while in some cases an initial stabilizing controller is not needed.

2.1. CONCLUSIONS

For the case of unknown nonlinear discrete systems, Paper 1 provides a NN-based optimal adaptive control scheme for a nonaffine nonlinear discrete-time system in input-output while the system dynamics are fully unknown. The control law relaxed the policy or value iteration. The nonaffine representation in the input-output form allows the new identifier to identify the system dynamics by using a single NN while separating the internal dynamics, the input gain matrix, and the higher order residual terms. The NN optimal controller is able to stabilize the affine part of the identified system when the higher order terms are bounded. The auxiliary control law derived using singular

perturbation theory indeed helps in the stability of the overall closed-loop system by mitigating the higher order terms. The closed loop Lyapunov based stability proof guarantees that the overall system is uniformly ultimately bounded (UUB) and therefore the proposed controller is optimal with a bounded error. Finally, the scheme is successfully validated on a HCCI engine model which is a realistic physical example of unknown nonlinear nonaffine systems.

The second paper develops a new extremum seeking method for nonlinear discrete-time systems with unknown performance output function. The proposed extremum seeking scheme plays the role of an outer loop that seeks for unique extremum (optimum) operating set point. The stability analysis is presented in two steps: first via averaging analysis and then singularly perturbed systems analysis, showing UUB stability with an arbitrarily small bound provided the plant is stabilized in an UUB manner. The proposed method is applied to the HCCI engine model that is stabilized using the controller proposed in paper I . The simulation results show that the proposed method not only maximizes the performance but also is able to satisfy a constraint such as peak pressure rise rate for HCCI engines.

The third paper is based on optimal tracking online control of MIMO continuous time strict feedback systems in the form of state and output feedback control. The novel proposed feedforward control scheme reduces the problem to optimally controlling the closed loop tracking error dynamics in affine form. In fact it is shown that a suitable backstepping controller is able to provide an error dynamics which has an affine representation. For the affine representation of the error dynamics, the single online approximator (SOLA) based optimal online scheme guarantees the solution to the HJB

equation if the system is persistently excited and therefore the corresponding controller converges to the optimal one. No admissible initial controller is required in the control scheme due to the novel update law and the control approach is relaxed from iterative based solutions. The proposed optimal output feedback control scheme for strict feedback systems is shown to provide guaranteed stability while rendering optimality. The overall Lyapunov based stability proof shows the closed loop system will remain bounded with a bound which gets arbitrarily small if the dimension of the basis function vector utilized for approximating the cost function is high enough.

The fourth paper is an application of the second paper to one of the most challenging problem of underactuated mechanical system. Unmanned aerial vehicle (UAV) helicopters are underactuated systems whose dynamics represent a strict feedback form system. The designed NN based output feedback based optimal tracking controller is able to achieve aggressive maneuvers which is a significant contribution compared to the available results in the literature. In the absence of state vector, a NN observer generates the states while an OLA based online controller learns the solution to the HJB equation. The backstepping feedforward controller is able to compensate the helicopter's weight and the rotor thrust requirement for hovering. The Lyapunov stability analysis and simulation results show the unmanned helicopter is capable of tracking a desired trajectory in an optimal manner with bounded error.

The fifth paper continues the work of Paper III by proposing an optimal adaptive controlling scheme for affine/strict feedback systems whose internal dynamics are unknown. The feedforward backstepping controller scheme converts the strict feedback system to the problem of optimally stabilizing the tracking error dynamics in affine form.

The closed-loop stability is demonstrated even when the internal dynamics are unknown. Next, the observer-based output feedback control scheme generates optimal control input while minimizing a cost function and relaxing the admissible control input. It is demonstrated that the proposed adaptive update laws can force the tracking errors to zero. Simulation results concur the theoretical results developed in the paper.

2.2. FUTURE WORK

As part of future work, the optimal adaptive controller for the nonaffine nonlinear discrete-time system can be redesigned by using multiple models without using value and policy iterations. Multiple model approach allows optimal adaptive controllers for plants whose dynamics change significantly with the state or input such as the case of a change in fuel-type for the HCCI engine. Optimal control of such systems can be challenging due to switching behavior of the dynamics.

Furthermore, there are several potential areas that can be pursued. In the field of optimal adaptive control of unknown systems, the problem of finite horizon optimal control is still a completely new ground. Finite horizon optimal control is based on minimizing the index function in a finite interval which renders a more complicated problem in terms of the solution to the cost function in online manner restricted with the terminal conditions. A simple extension to the finite horizon problem is when the control inputs have constraints as every actuator has physical limits.

Optimal control of strict feedback systems can be attacked by a variety of cost functions. The cost function chosen in the current dissertation penalizes the tracking error derived by proposing a spatial type of feedforward controller. The problem becomes more difficult if the tracking error and the control input are solely penalized. It is also an interesting problem to assume the input gain matrix is also unknown. Moreover, optimal

control of strict feedback system with input and state constraints is not yet addressed. Optimal tracking control of UAV vehicles is very important due to the limitation of control input with mobility. Therefore, the hardware implementation of the proposed optimal controller on UAVs will also be an interesting work for the future.

VITA

Hassan Zargarzadeh was born in Dezfool, Khozestan, Iran. He earned the Bachelor of Science and Master of Science degrees in Electrical Engineering from Tehran Polytechnic and Iran University of Science and Technology in 2000 and 2009 respectively. He had been working with industry from 1999 to 2005 as an instrumentation engineer in Kaveh Glass Co. and as production manager in Panam Azma. He joined Missouri university of Science and Technology in January 2010 and received the degree of Doctor of Philosophy in Electrical Engineering in December 2012.